[PhD_A1_published_data]
**“Browne_et_al,2006.pdf”**
**“Browne_et_al,2006_supplement.doc”**

[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.1_Ph_otd1]
[PhD_A2.1.1_Ph_otd1_sequences]
**“Ph_otd1_general_read_me.txt”**
“**Ph_otd1_cDNA_sequences.txt**”
[PhD_A2.1.2_Ph_otd1_alignments]
**“alignment_Ph_otd1_all_aa_sequences.gif”**
**“alignment_Ph_otd1_all_bp_sequences.gif”**
**“difference_alignment_Ph_otd1_all_aa_sequences.gif”**
**“difference_alignment_Ph_otd1_all_bp_sequences.gif”**

[PhD_A2.2_Ph_otd2]
**“Ph_otd2_read_me.txt”**

[PhD_A2.3_Ph_hbn]
[PhD_A2.3.1_Ph_hbn_sequences[
**“Ph_hbn_general_read_me.txt”**
**“Ph_hbn_cDNA_sequences.txt”**
[PhD_A2.3.2_Ph_hbn_alignments]
**“alignment_Ph_hbn_all_aa_sequences.gif”**
**“alignment_Ph_hbn_all_bp_sequences.gif”**
**“difference_alignment_Ph_hbn_all_aa_sequences.gif”**
**“difference_alignment_Ph_hbn_all_bp_sequences.gif”**

[PhD_A2.4_Ph_al1]
[PhD_A2.4.1_Ph_al1_sequences]
**“Ph_al1_general_read_me.txt”**
**“Ph_al1_cDNA_sequences.txt“**
[PhD_A2.4.2_Ph_al1_alignments]
**“alignment_Ph_al1_all_aa_sequences.gif”**
**“alignment_Ph_al1_all_bp_sequences.gif”**
**“difference_alignment_Ph_al1_all_aa_sequences.gif”**
**“difference_alignment_Ph_al1_all_bp_sequences.gif”**

[PhD_A2.5_Ph_al2]
[PhD_A2.5.1_Ph_al2_sequences]
**“Ph_al2_general_read_me.txt”**
**“Ph_al2_cDNA_sequences.txt”**

  - [PhD_A2.5.2_Ph_al2_alignments]
    - **“alignment_Ph_al2_all_aa_sequences.gif”**
    - **“alignment_Ph_al2_all_bp_sequences.gif”**
    - **“difference_alignment_Ph_al2_all_aa_sequences.gif”**
    - **“difference_alignment_Ph_al2_all_bp_sequences.gif”**

- [PhD_A2.6_Ph_pby1_pby2]
  - [PhD_A2.6.1_Ph_pby1_pby2_sequences]
    - “**Ph_pby1_general_read_me.txt”**
    - **“Ph_pby1_cDNA_sequences.txt”**
    - **“Ph_pby2_vsdPCR.txt”**
  - [PhD_A2.6.2_Ph_pby1_pby2_alignments]
    - **“alignment_Ph_pby1_pby2_vsdPCR_aa_sequences.gif”**
    - **“alignment_Ph_pby1_pby2_vsdPCR_sequences.gif”**
    - **“alignment_Ph_pby1A_all_aa_sequences.gif”**
    - **“alignment_Ph_pby1A_all_bp_sequences.gif”**
    - **“alignment_Ph_pby1A_pby1B_pbyshared_aa.gif”**
    - **“alignment_Ph_pby1B_all_aa_sequences.gif”**
    - **“alignment_Ph_pby1B_all_bp_sequences.gif”**
    - **“difference_alignment_Ph_pby1_pby2_vsdPCR_aa_sequences.gif”**
    - **“difference_alignment_Ph_pby1_pby2_vsdPCR_sequences.gif”**
    - **“difference_alignment_Ph_pby1A_all_aa_sequences.gif”**
    - **“difference_alignment_Ph_pby1A_all_bp_sequences.gif”**
    - **“difference_alignment_Ph_pby1B_all_aa_sequences.gif”**
    - **“difference_alignment_Ph_pby1B_all_bp_sequences.gif”**

- [PhD_A2.7_Ph_awh]
  - [PhD_A2.7.1_Ph_awh_sequences]
    - **“Ph_awh_general_read_me.txt”**
    - **“Ph_awh_cDNA_sequences.txt”**
  - [PhD_A2.7.2_Ph_awh_alignments]
    - **“alignment_Ph_awh_all_aa_sequences.gif”**
    - **“alignment_Ph_awh_all_bp_sequences.gif”**
    - **“difference_alignment_Ph_awh_all_aa_sequences.gif”**
    - **“difference_alignment_Ph_awh_all_bp_sequences.gif”**

- [PhD_A2.8_Ph_six3]
  - [PhD_A2.8.1_Ph_six3_sequences]
    - **“Ph_six3_general_read_me.txt”**
    - **“Ph_six3_cDNA_sequences.txt”**
  - [PhD_A2.8.2_Ph_six3_alignments]
    - **“alignment_Ph_six3_all_aa_sequences.gif”**

**“alignment_Ph_six3_all_bp_sequences.gif”**
**“difference_alignment_Ph_six3_all_aa_sequences.gif”**
**“difference_alignment_Ph_six3_all_bp_sequences.gif”**

[PhD_A2.9_Ph_six4]
    [PhD_A2.9.1_Ph_six4_sequences]
        **“Ph_six4_general_read_me.txt”**
        **“Ph_six4_cDNA_sequences.txt”**
    [PhD_A2.9.2_Ph_six4_alignments]
        **“alignment_Ph_six4_all_aa_sequences.gif”**
        **“alignment_Ph_six4_all_bp_sequences.gif”**
        **“difference_alignment_Ph_six4_all_aa_sequences.gif”**
        **“difference_alignment_Ph_six4_all_bp_sequences.gif”**

[PhD_A2.10_Ph_gbx]
    [PhD_A2.10.1_Ph_gbx_sequences]
        **“Ph_gbx_general_read_me.txt”**
        **“Ph_gbx_cDNA_sequences.txt”**
    [PhD_A2.10.2_Ph_gbx_alignments]
        **“alignment_Ph_gbx_all_aa_sequences.gif”**
        **“alignment_Ph_gbx_all_bp_sequences.gif”**
        **“difference_alignment_Ph_gbx_all_aa_sequences.gif”**
        **“difference_alignment_Ph_gbx_all_bp_sequences.gif”**

[PhD_A2.11_Ph_kni1]
    [PhD_A2.11.1_Ph_kni1_sequences]
        **“Ph_kni1_general_read_me.txt”**
        **“Ph_kni1_cDNA_sequences.txt”**
    [PhD_A2.11.2_Ph_kni1_alignments]
        **“alignment_Ph_kni1_all_aa_sequences.gif”**
        **“alignment_Ph_kni1_all_bp_sequences.gif”**
        **“difference_alignment_Ph_kni1_all_aa_sequences.gif”**
        **“difference_alignment_Ph_kni1_all_bp_sequences.gif”**

[PhD_A2.12_Ph_kni2]
    [PhD_A2.12.1_Ph_kni2_sequences]
        **“Ph_kni2_general_read_me.txt”**
        **“Ph_kni2_cDNA_sequences.txt”**
    [PhD_A2.12.2_Ph_kni2_alignment]
        **“alignment_Ph_kni2_all_aa_sequences.gif”**
        **“alignment_Ph_kni2_all_bp_sequences.gif”**
        **“difference_alignment_Ph_kni2_all_aa_sequences.gif”**

**“difference_alignment_Ph_kni2_all_bp_sequences.gif”**

[PhD_A3_list_of_sequence_clones]
**“PhD_A3_List_of_Parhyale_goi_sequence_clones.pdf”**

[PhD_A4_oligonucleotides]
**“PhD_A4.1_degenerate_oligonucleotides.pdf”**
**“PhD_A4.2_specific_oligonucleotides.pdf”**
**“PhD_A4.3_siRNA_molecules.pdf”**

[PhD_A5_plasmid_sequences]
**“pSL_shuttle.txt”**
**“alignment_pSLshuttle_pSLfa1180fa_Ph-hsp70_SV40.gif”**
**“pMi[3xP3_DsRed; hsp70_hp-otd1].txt”**
**“pMi[3xP3_EGFP; hsp70_hp-DsRed].txt”**
**“pimpMiII.txt”**
**“alignment_pimpMiII_pMiII_9700.gif”**
**“pimpMiII[mse_EG-kni2intron-FP].txt”**

[PhD_A6_phylogeny_database]
**“PhD_A6.1_List_of_species.pdf”**
**“PhD_A6.2_List_of_non-Parhyale_sequences.pdf”**

[PhD_A6.3_phylogeny_ClustalW2_alignments]
**“alignment_aristaless-group.txt”**
**“alignment_Arrowhead-Lhx68_proteins.txt”**
**“alignment_Arrowhead-LIM-HD-group.txt”**
**“alignment_knirps-family.txt”**
**“alignment_paired-Pax-group.txt”**
**“alignment_sine-oculis-Six-family.txt”**
**“alignment_Unplugged-Gbx-proteins.txt”**
**“alignment_Unplugged-Gbx-similar_proteins.txt”**

[PhD_A7_abbreviations]
**“PhD_A7_Abbreviations.pdf”**

[PhD_A1_published_data]

**“Browne_et_al,2006.pdf”**

Dev Genes Evol (2006) 216: 581–595
DOI 10.1007/s00427-006-0074-7

ORIGINAL ARTICLE

**William E. Browne · Bernhard G. M. Schmid ·
Ernst A. Wimmer · Mark Q. Martindale**

# Expression of *otd* orthologs in the amphipod crustacean, *Parhyale hawaiensis*

Received: 11 January 2006 / Accepted: 20 March 2006 / Published online: 7 July 2006

**Abstract** The arthropod head is a complex metameric structure. In insects, *orthodenticle* (*otd*) functions as a ‘head gap gene’ and plays a significant role in patterning and development of the anterior head ectoderm, the protocerebrum, and the ventral midline. In this study, we characterize the structure and developmental deployment of two *otd* paralogs in the amphipod crustacean, *Parhyale hawaiensis*. *Photd1* is initially expressed at gastrulation through germband stages in a bilaterally symmetric, restricted region of the anterior head ectoderm and also in a single column of cells along the ventral midline. Late in embryogenesis, *Photd1* is expressed within the developing anterior brain and the expression along the embryonic midline has become restricted to a stereotypic group of segmentally reiterated cells. The second ortholog *Photd2*, however, has a unique temporal–spatial expression pattern and is not detected until after the head lobes have been organized in the developing ectoderm of the germband during late germband stages. Anteriorly, *Photd2* is coincident with the *Photd1* head expression domain; however, *Photd2* is not detected along the ventral midline during formation of the germband and only appears in the ventral midline late in embryonic development in a restricted group of cells distinct from those expressing *Photd1*. The early expression of *Photd1* in the anterior head ectoderm is consistent with a role as a head gap gene. The more posterior expression of *Photd1* is suggestive of a role in patterning the embryonic ventral midline. *Photd2* expression appears too late to play a role in early head patterning but may contribute to latter patterning in restricted regions of both the head and the ventral midline. The comparative analysis of *otd* reveals the divergence of gene expression and gene function associated with duplication of this important developmental gene.

**Keywords** Crustacean · Amphipod · *Parhyale* · *Orthodenticle* · Head development · Brain development · CNS · Midline · Neurogenesis

## Introduction

Genes that control the anterior head and nervous system of the fly, *Drosophila melanogaster*, have guided ideas regarding the development and evolution of the arthropod head. In *Drosophila*, a small number of transcription factor genes play a major role in the specification of the anterior head ectoderm and the anterior-most neuromeres that comprise the supraesophageal ganglion (Younossi-Hartenstein et al. 1997). These ‘head gap genes’ are deployed very early in fly development in overlapping functional domains (Cohen and Jürgens 1990; Schmidt-Ott et al. 1994) and their initially broad anterior domains of expression are progressively restricted and sharpened during the processes of both cellularization and gastrulation (Dalton et al. 1989; Finkelstein and Perrimon, 1990; Walldorf and Gehring 1992; Wimmer et al. 1993; Grossniklaus et al. 1994; Wimmer et al. 1997). The resulting anterior expression domains demarcate specific regions of the developing *Drosophila* brain (Hirth et al. 1995). Among these genes, orthologs of *orthodenticle* (*otd*) have been demonstrated to play a central role in anterior head and brain development in disparate bilaterians (e.g., Acampora et al. 1998; reviewed in Arendt and

---

Communicated by S. Roth

**Electronic Supplementary Material** Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00427-006-0074-7

W. E. Browne (✉) · M. Q. Martindale
Kewalo Marine Lab, Pacific Biosciences Research Center,
University of Hawaii, 41 Ahui St.,
Honolulu, HI 96813, USA
e-mail: wbrowne@hawaii.edu
Tel.: +1-808-5397326
Fax: +1-808-5994817

B. G. M. Schmid · E. A. Wimmer
Department of Developmental Biology, Johann-Friedrich-Blumenbach-Institute of Zoology and Anthropology,
Georg-August-University Göttingen, GZMB,
Justus-von-Liebig-Weg 11,
37077 Göttingen, Germany

Nubler-Jung 1996; Sharman and Brand 1998; Lichtneckert and Reichert 2005).

In *Drosophila* and other insects, the supraesophageal ganglia can be subdivided along the neuroaxis, from anterior to posterior, into three neuromere subunits the protocerebrum (PC), deutocerebrum (DC), and tritocerebrum (TC) (Reichert and Boyan 1997). In *Drosophila*, *otd* is initially expressed in a wide anterior circumferential stripe in the blastoderm embryo (Finkelstein and Perrimon 1990; Cohen and Jürgens 1991). As development progresses, the *otd* expression boundary becomes restricted to the embryonic brain and is detected throughout the developing PC neuromere and in the anterior DC neuromere. Mutant *otd* alleles delete or severely reduce the size of the PC neuromere and the PC-associated preoral medial commissure (Hirth et al. 1995). *otd* appears to, in part, functionally act anteriorly in the specification of cell identity via regulation of proneural gene expression (Younossi-Hartenstein et al. 1997), the intrasegmental regulation of the segment polarity genes *engrailed* (*en*) and *wingless* (*wg*) in embryos (Gallitano-Mendel and Finkelstein 1998), and later in larval and pupal development of medial structures including the ocelli of the adult fly (Wieschaus et al. 1992; Royet and Finkelstein 1995; Hirth et al. 1995). In addition, reciprocal negative regulation between the gene *unplugged* (*unpg*) and *otd* suggests that *otd* plays an important role in specification of the anterior boundary between the DC and TC neuromeres in *Drosophila* (Hirth et al. 2003; Lichtneckert and Reichert 2005). Expression of *otd* is also prominent in the embryonic ventral midline of *Drosophila* and plays an important role in the patterning of the segmentally reiterated ventral nerve cord (VNC) commissures, particularly the posterior commissure (Finkelstein et al. 1990; Klämbt et al. 1991).

The embryonic development of both the head and ventral midline in *Drosophila* is distinctive from most arthropods. Unique to the dipteran flies, the anterior head segments involute in late stages of development and the ventral midline is generated from a distinct group of mesectodermal cells. Given the highly divergent mode of embryonic head development in flies, the expression patterns of *orthodenticle* orthologs have been examined in an insect with less-derived embryonic development, the beetle *Tribolium castaneum*. In contrast to *Drosophila*, *Tribolium* possesses two orthologs of *otd*, *Tcotd-1*, and *Tcotd-2* (Li et al. 1996). The expression of the two *Tribolium otd* paralogs are divergent with respect to each other as well as to their putative ortholog in *Drosophila* (Li et al. 1996). Knockdown of *Tcotd-1* function via parental RNAi produces a headless embryonic phenotype that is more severe than the *Drosophila* gap gene phenotype, which indicates an important anterior determinant role of *Tcotd-1* in *Tribolium* (Schröder 2003). It is notable that the effects of knockdown restricted to zygotic expression of *Tcotd-1* were generally more moderate and did not eliminate the entire head but resemble the *Drosophila* gap gene phenotype (Schröder 2003). Thus, while *Tcotd-1* retains an early anterior gap gene patterning function analogous to its ortholog *otd* in *Drosophila*, *Tcotd-1* has an additional functional role in specifying the entire head region in *Tribolium*. This observation in *Tribolium* suggests the possibility of *otd* serving as an anterior determinant in other arthropods.

In the absence of equivalent information from additional arthropod taxa, these data can only apply to the most recent common ancestor of holometabolous beetles and flies. Thus, it is problematic to generalize details regarding the mode of ancestral arthropod head development exclusively from data produced in flies and other insects. Among arthropod taxa more distantly related to insects, only limited information is currently available. In the chelicerate *Archegozetes longisetosus*, expression of a single *orthodenticle* ortholog has been reported and cursory examination of expression during late embryogenesis reveals *Alotd* in an anterior head domain as well as along the ventral midline similar to *Tcotd-1* in *Tribolium* (Telford and Thomas 1998). Recent work suggests that the crustaceans are sister taxa to the insects (Friedrich and Tautz 1995; Dohle 2001; Giribet et al. 2001, 2005; Hwang et al. 2001; Richter 2002; Cook et al. 2005; Regier et al. 2005). Comparative neuroanatomy of crustaceans (e.g., Hanstrom 1928; Sandeman et al. 1992; Gerberding 1997; Strausfeld 1998; Harzsch 2003) suggests that a similar and presumably homologous supraesophageal neuromere ground plan is shared with the insects (Sandeman et al. 1992). Additional comparative studies between crustaceans and insects have suggested both strong similarities and notable differences in neuronal morphology (Whitington et al. 1993; Whitington 1996). These observations have been extended to suggest homologies among neuronal identities by correlating neuronal morphology with the expression of molecular markers (Duman-Scheel and Patel 1999; Browne et al. 2005). Using these studies, our understanding of insect neurogenesis, in combination with the embryonic staging and cell lineage data available for *Parhyale hawaiensis* (Gerberding et al. 2002; Browne et al. 2005), provides a framework from which to make a detailed analysis of the expression of head patterning orthologs in a model crustacean. At the level of segment morphology and segment specification, the heads of crustaceans, such as *Parhyale*, are quite different from the heads of insects, such as *Drosophila*. For example, insects do not possess paired appendages on the intercalary segment whereas crustaceans possess antennae on the homologous segment (An2). In addition, many aspects of the olfactory systems of insects and crustaceans appear to be considerably divergent (Strausfeld 1998; Strausfeld and Hildebrand 1999). Thus, the a priori expectation would be for gene orthologs controlling the morphology of insects to have divergent temporal and spatial expression in crustaceans.

In this paper, we report detailed expression patterns associated with orthologs of *otd* in the model amphipod crustacean, *P. hawaiensis*. To further characterize anterior head and ventral midline development in *Parhyale*, we cloned *otd* orthologs and carried out a detailed study of their expression during embryogenesis via in situ mRNA transcript analysis. This analysis, when paired with available sequence and expression data in other taxa, allowed us to assess the relationship between *otd* orthologs and paralogs among arthropods. Our analysis suggests an evolutionary scenario accounting for the gene duplication and expression characteristics associated with *otd* and is informative with regard to reconstructing the ancestral state of the most recent common ancestor between crustaceans and insects, a marine arthropod stem species existing during the Cambrian ~550 mya.

## Materials and methods

### Amphipod culture

*P. hawaiensis* is maintained as reported in Browne et al. (2005) with the following modifications. The breeding colony is kept in interconnected shallow plastic trays at ~30°C at the Kewalo Marine Lab. Recirculating seawater is provided via magnetic drive pump (Iwaki). Fresh, filtered seawater is added at regular intervals. Animals are fed a liquefied mixture of algae, plankton, fatty acids (Selco), and vitamins (Kent Marine).

### Cloning and sequence analysis

Total RNA was isolated with TRIzol Reagent (Gibco BRL) from a pool of mixed-stage *P. hawaiensis* embryos. First strand cDNA was generated with the SuperScript Preamplification System (Gibco BRL). Initial degenerate PCR was completed using conserved nested primers to aligned *otd* proteins (forward primer 5′-GAR MGN CAN CAN TTY AC-3′, reverse outer primer 5′-NCK NCK RTT YTT RAA CCA-3′, and reverse inner primer 5′-CAR YTN GAY GTN YTN GA-3′).

For isolation of additional 5′ sequences, polyA RNA was isolated using the Micro Poly(A)Pure Small Scale mRNA Purification Kit (Ambion) from a pool of *P. hawaiensis* embryos. First strand cDNAs were generated with the SMART PCR cDNA Synthesis Kit (BD Biosciences). Initial degenerate PCR was completed using the forward primer 5′-CAG MGG MGG GAR MGI ACI ACI TTY AC -3′ and reverse primer 5′-GC CCK CCK RTT YTT RAA CCA IAC YTG-3′. Sequences 5′ of the homeodomain were obtained via 5′RACE (SMART RACE cDNA Amplification Kit, BD Biosciences) with nested sequence-specific primers obtained from the initial degenerate PCR. Sequences 3′ of the homeodomain were obtained by performing 3′RACE (Gibco BRL) with sequence-specific nested primers obtained from the initial degenerate PCR (primers available upon request).

### Phylogenetic analysis

Maximum parsimony (MP) and maximum likelihood (ML) analyses were executed with PAUP* 4.0Beta (Swofford 2003). Bayesian phylogenetic inference analysis was executed with MrBayes 3.1.1 (Ronquist and Huelsenbeck 2003). Relevant *orthodenticle* sequences were selected from those available via genbank; all available arthropod sequences were used. Searches were rooted with the *D. melanogaster paired* gene and resulting trees were visualized with TreeView X (Page 1996).

A mutational saturation plot (Philippe et al. 1994; Philippe and Forterre 1999) was generated using a MP analysis with ten random sequence addition MP heuristic searches. The best tree was used to generate a patristic distance matrix. In additional parsimony analyses, MP heuristic searches were subjected to 2,000 bootstrap replicates, each with a random addition MP heuristic search.

For ML analyses, 1,000 random sequence addition ML heuristic searches were run. Searches were then subjected to 2,000 bootstrap replicates, each with a random addition ML heuristic search. The molecular evolution model, HKY + G, was selected using Model-Test 3.7 (Posada and Crandall 1998) for ML analysis.

For Bayesian analyses, all parameters were unlinked except topology. Two different codon partition schemes were generated: (1) codon position partition 1, 2, 3 or (2) codon position partition 1+2, 3. Multiple searches were run from 2 million generations to 5 million generations and trees were sampled every 100 generations. Posterior estimates from runs were analyzed by eye with Tracer v1.3 (Rambaut and Drummond 2003, Tracer v1.3, http://www.evolve.zoo.ox.ac.uk) to determine the number of generations to burn in and to assess convergence of data sets. Consensus trees from independent runs within each partition scheme were compared to assess convergence and topology congruence of data sets.

### In situ hybridization

*P. hawaiensis* embryos were removed from the ventral brood pouch and allowed to develop to the desired stage at 26°C (Browne et al. 2005), then dissected with tungsten wire needles and fixed for 60 min in filtered seawater mixed 9:1 with 37% formaldehyde at room temperature. Fixed embryos were then washed with PT (1× PBS, pH 7.4, and 0.1% Triton X-100) and stored in 100% MeOH at −20°C. Whole-mount in situ hybridization using digoxigenin-labeled RNA was performed as reported in Finnerty et al. (2003), with the following changes. Before hybridization, fixed embryos were incubated for 60 min in a sodium dodecyl sulfate (SDS)-based detergent solution [1% SDS, 0.5% Tween-20, 50 mM Tris–HCl (pH 7.5), 1.0 mM EDTA (pH 8.0), and 150 mM NaCl] at 65°C. After hybridization, embryos were intensively washed with 2× SSC solution at 65°C to remove hybridization buffer and free probe. After the 65°C washes, embryos were then washed at room temperature in PBT (1× PBS containing 0.1% bovine serum albumin and 0.1% Triton X-100). Embryos were then incubated in sheep anti-digoxigenin-AP (1:3,000) overnight at 4°C and subsequently washed with PBT before performing BCIP/NBT color reaction.

After color reaction, embryos were washed with PT and incubated overnight at 4°C in 50% glycerol + 1 mg/ml Hoechst dye. Embryos were stored in 70% glycerol at −20°C. Images were generated from dissected and mounted embryos using an Axioskop2 (Zeiss) and Axio Imager Z1 (Zeiss) and photographed with either an Axiocam (Zeiss) or an ORCA ER (Hamamatsu) using MAC-compatible Openlab and Velocity (Improvision) imaging software.

### Antibody staining

Staged *P. hawaiensis* embryos (Browne et al. 2005) were fixed in the same way as for in situ but for a shorter period of time (15 min). Antibody detection was performed as reported in Patel (1994). All antibody incubations were done overnight at 4°C at the following concentrations: mouse anti-acetylated tubulin primary antibody (Hybridoma Bank, clone E7) 1:1 with PBT, alexa-594 conjugated goat anti-mouse secondary antibody (Molecular Probes) 1:250 with PBT, and HRP conjugated goat anti-mouse secondary antibody (Jackson ImmunoResearch). Embryos were then washed and incubated overnight at 4°C in 50% glycerol + 1 mg/ml Hoechst solution and stored in 70% glycerol at −20°C. Images were generated from dissected and mounted embryos as described above.

## Results

### Isolation and characterization of *P. hawaiensis otd* orthologs

Two *P. hawaiensis otd* sequences, *Photd1* and *Photd2*, were recovered from repeated degenerate PCR screens on independent pools of *Parhyale* cDNA. The isolated fragments (103 bp) were sufficient for unambiguous orthology identification. Nested non-overlapping primers were designed to the two initial degenerate PCR fragments and used to amplify the remaining mRNA transcript sequences both 5′ and 3′ of the initial homeodomain fragments. All of the *Photd1* sequences are identical with minor variations in the length of the recovered transcripts. The longest recovered complete *Photd1* cDNA was 2 kb. In contrast, the organization of the *Photd2* paralog is significantly more complex. The *Photd2* gene has multiple splice variants both 5′ and 3′ of an invariant homeodomain core sequence (supplemental data). RACE yielded five unique open reading frame sequences upstream of the homeodomain and three unique open reading frame sequences downstream of the homeodomain. Therefore, the *Photd2* gene appears to be capable of generating potentially 15 unique transcripts; however, the true number of *Photd2* transcripts produced from these recovered alternative splice variants is unknown.

### Phylogenetic analysis

The *orthodenticle* homeodomains show extreme amino acid conservation whereas regions both upstream and downstream of the homeodomain are too variable among orthologs to reliably align in distantly related taxa. In addition, due to *Photd2* possessing multiple splice variants both upstream and downstream of the homeodomain, we restricted our phylogenetic analyses of *orthodenticle* genes in disparate taxa to the homeodomain region. Our survey of metazoan taxa revealed different complements of *otd* paralogs in different groups (Fig. 1). In cases in which two paralogs exist, the paralogs appear more similar to each other than to orthologs in other species. We propose two readily distinguishable modes of gene expansion are likely to have occurred to generate the range of *otd* paralogs across metazoans. In the lineage leading to the vertebrates, whole-genome duplication has generated additional paralogous genes (Germot et al. 2001) which, in most cases, are found on different chromosomes. In contrast, in the cnidarian (Finnerty et al., submitted for publication) and arthropod lineages for which genomic information is available, tandem gene duplication appears to have generated paralogous genes in close proximity to one another. However, it is unclear whether duplications occurred independently in each taxon or at particular nodes in the metazoan tree. Specifically in the arthropod lineage leading to the Tetraconata (crustaceans + insects) (Dohle 2001), it is likely that a single duplication of an ancestral *otd* gene occurred to give rise to two paralogs in descendant taxa (Fig. 1). Within any given lineage arising after this duplication, *otd* paralogs are likely to have been subjected to the effects of two homogenizing phenomena: (1) increased probability of gene conversion (mechanism by which mismatched heteroduplexed DNA basepairs between gene copies are repaired) due to the proximity of the paralogs and (2) the effects of mutational saturation. Both of these phenomena would contribute to *otd* paralogs becoming more similar to one another than to *otd* orthologs in related taxa. As a test for mutation saturation among *otd* paralogs, we generated a mutational saturation plot (supplemental data). The presence of a mutational plateau indicates that saturation for mutational changes has occurred, therefore indicating a loss of phylogenetic signal necessary to accurately resolve branch lengths (Philippe et al. 1994; Philippe and Forterre 1999). Thus, node relationships could not be recovered to generate a comprehensive, reliable, gene tree indicating orthology among *otd* paralogs using traditional phylogenetic tools (data not shown). In summary, the orthology of the two *Parhyale* genes to *orthodenticle* is clear; however, whether multiple independent gene duplications have occurred in different lineages or whether an early metazoan ancestor already had separate *otd* paralogs cannot be ruled out based on available evidence (Fig. 1).

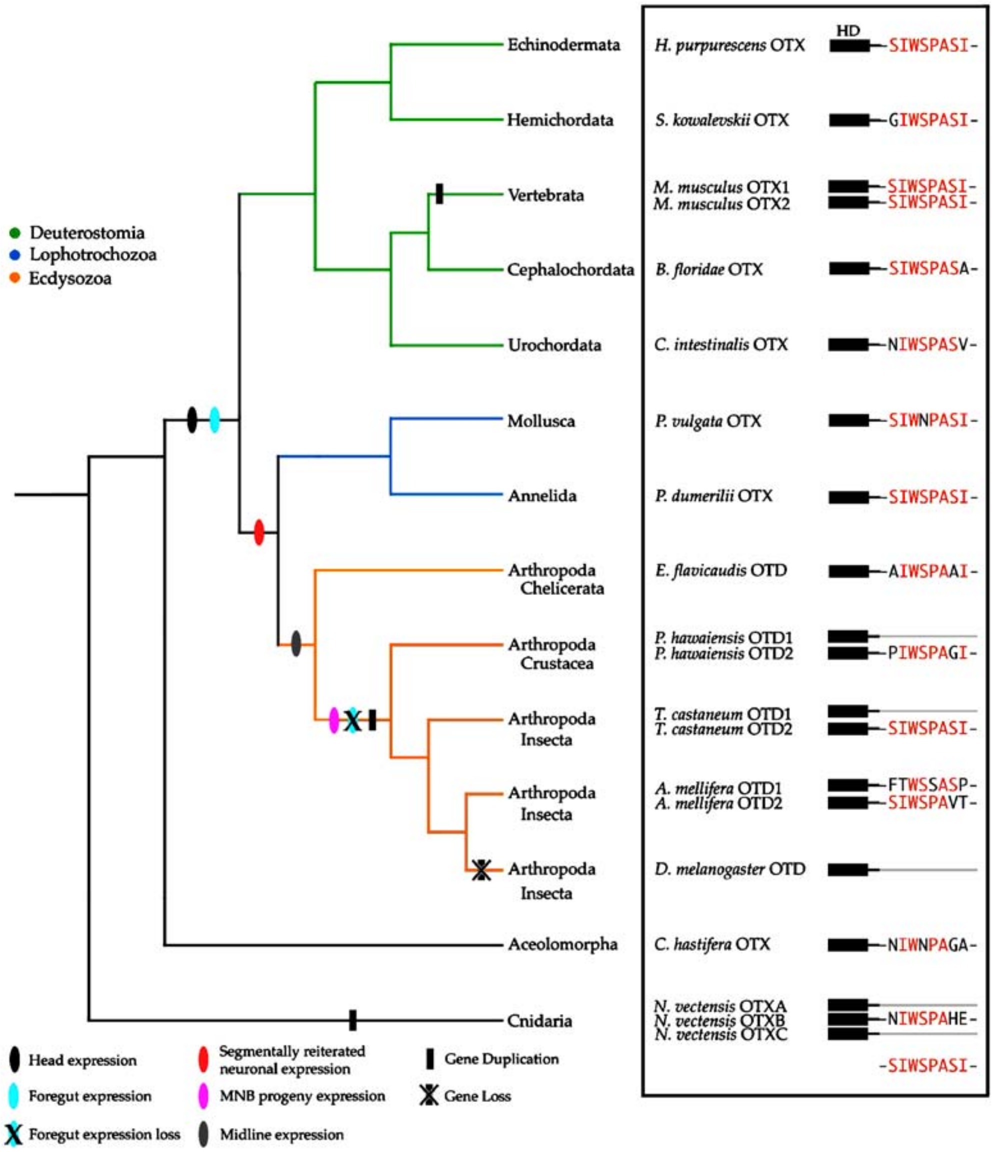


**Fig. 1** A model proposing evolutionary duplication and loss of *orthodenticle* paralogs in metazoans. Character legend at the *left bottom* of figure. *Green* associated with deuterostome lineages, *blue* associated with lophotrochozoan lineages, *orange* associated with ecdysozoan lineages. Cladogram (modified from Adoutte et al. 2000) on the *left-hand side* indicates evolutionary relationship between species grouped in *box* on the *right-hand* side. No implication of polarity is associated with the order of multiple characters occurring on the same branch. For each taxon, orthologs of OTD are included with a schematic of the C-terminal end of each protein that includes the sequence of the short WSP amino acid motif; absence of this motif is indicated with a *grey line*. Beneath the WSP sequences is the amino acid consensus for the motif. For all taxa, retention of consensus amino acids is indicated with *red letters*. The WSP motif is detected in at least one *orthodenticle* paralog of all examined taxa with the notable exception of *D. melanogaster*. In this model, an ancestral *orthodenticle* gene was present in the lineage leading to cnidarians + bilaterians and has experienced at least three independent duplication events via two dissimilar duplication processes (*black bars*). In the lineage leading to the vertebrates, *orthodenticle* gene duplication occurred via entire genome duplication. In the cnidarian lineage leading to *Nematostella vectensis*, *orthodenticle* expansion occurred via local gene duplication generating three closely linked paralogs (*NvOtx* sequences, K. Pang, personal communication). Available data in the arthropod lineage indicates an *orthodenticle* expansion also occurred via local gene duplication in the lineage leading to the crustaceans + insects (Tetraconata). In taxa for which data is available, at least one *otd* paralog retains an identifiable WSP motif. In the insects, *T. castaneum* and *A. mellifera*, the duplicate paralogs are known to be in close proximity to one another. The dipteran *D. melanogaster* has lost the *otd* paralog containing the WSP motif (*crossed black bar*). The expression of *orthodenticle* orthologs in the head occurred before the lineage including the Eubilateria (*black oval*). This expression character is currently placed after divergence of the Aceolmorpha lineage (*ChOtx* sequences, A. Hejnol, personal communication). Expression of *orthodenticle* in the foregut also appears to be an ancient feature, also occurring before the divergence of the deuterostome and protostome lineages (*blue oval*). Expression of *orthodenticle* appears in segmentally reiterated neurons in the lineage leading to Lophotrochozoa + Ecdysozoa (*red oval*). Available data in the arthropod lineage indicates expression of *orthodenticle* in the ventral midline arose before the divergence of the chelicerate + Tetraconata lineage (*grey oval*). Foregut expression is lost in the lineage leading to the Tetraconata (*crossed blue oval*). Additional data indicate that *orthodenticle* expression in median neuroblast progeny (*MNB*) predates the divergence Tetraconata lineage (*purple oval*). Full species names are *Apis mellifera*, *Branchiostoma floridae*, *Ciona intestinalis*, *Convolutriloba hastifera*, *Drosophila melanogaster*, *Euscorpius flavicaudis*, *Holopneustes purpurescens*, *Mus musculus*, *Nematostella vectensis*, *Parhyale hawaiensis*, *Patella vulgata*, *Platyneris dumerilii*, *Saccoglossus kowalevskii*, and *Tribolium castaneum*

## Expression of the *P. hawaiensis Photd1* paralog

The *Photd1* paralog is first detectable in descendants from each of the three ectoderm clones El, Er, and Ep (Gerberding et al. 2002; Browne et al. 2005) at gastrulation in the early stage 8 embryo (Fig. 2a). There is no maternal *Photd1* transcript contribution to oocytes nor is there detectable *Photd1* zygotic transcription before gastrulation. As the cells that form the germ disc aggregate in the late stage 8 embryo, *Photd1* expression spreads, increases in intensity, and is detected anteriorly in an arc as the head lobes begin to condense laterally. The head lobes are composed of cells from the El (ectoderm left) and Er (ectoderm right) ectoderm clones (Gerberding et al. 2002). *Photd1* expression also remains present in the future ventral midline cells of the Ep (ectoderm posterior) clone as they converge into a single, tightly associated cluster, at the posterior end of the germ disc (Fig. 2b). In the following stage, stage 9, the anterior lateral-most *Photd1*-expressing cells are beginning to organize into the characteristic bilaterally symmetric head lobe arcs. A persistent arc of less intense staining connects the head lobes across the midline. This staining is associated with cells at the anterior-most region of the developing germ disc. Posteriorly, the medial cluster of *Photd1*-expressing cells have organized as a half circle with a sharply demarcated anterior boundary (Fig. 2c). The formation of this boundary occurs as the posterior ectoderm of the germband begins to organize into a characteristic gridded array (Browne et al. 2005). As parasegment precursor rows (PPRs) begin to

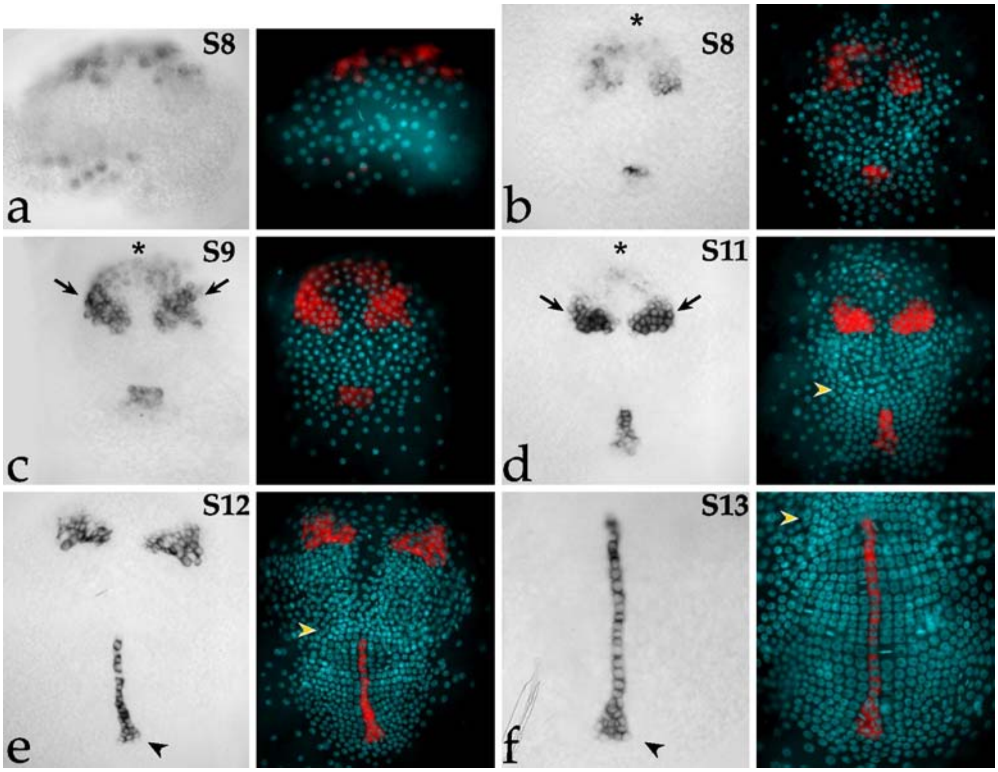


**Fig. 2** Expression of *P. hawaiensis otd1*, gastrulation through germband formation. All embryos are mounted ventral side up with the anterior end oriented towards the top. For each panel, a brightfield image is on the *left* (*Photd1* label is *black*), and a Hoechst label of the same embryo is on the *right* (nuclei are *blue*, *Photd1* is false-colored *red*). In panels **d**–**f**, the *yellow arrowhead* marks the position of the Mn segment. **a** Early stage 8 embryo (*S8*), *Photd1*, is present in descendants of El, Er, and Ep blastomeres at gastrulation. **b** Late Stage 8 embryo (*S8*); anteriorly and medially, the field of El and Er cells expressing *Photd1* has expanded, forming an arch (*asterisk*). Posteriorly, the few Ep cells expressing *Photd1* have converged medially into a tightly associated cluster. **c** Stage 9 embryo (*S9*); anteriorly, the lateral-most *Photd1* expressing cells are beginning to organize into the characteristic bilaterally symmetric head lobe arcs (*black arrows*). Posteriorly, the medial cluster of *Photd1*-expressing cells organizes with a sharply demarcated anterior boundary. Formation of this boundary occurs as the posterior ectoderm of the germband begins to organize into a stereotypic gridded array. **d** Stage 11 embryo (*S11*); *Photd1*-expressing cells have condensed anteriorly into two bilaterally symmetric groups largely coincident with the head lobe arcs (*black arrows*); the more anterior medial expression is fading (*asterisk*). Posteriorly, *Photd1*-expressing cells begin to form the ventral midline by converging into a single column of cells medially. The anterior boundary of the *Photd1*-expressing midline cells is at the level of PPR1 which is immediately posterior of the Mn segment (*yellow arrowhead*) and corresponds with the first maxillary segment, Mx1. **e** Stage 12 embryo (*S12*); anterior medial expression of *Photd1* is undetectable. Posteriorly, as the germband extends, expression in the single cell ventral midline column is maintained via both cell division within forming parasegments as well as convergence and intercalation of additional cells at the posterior extreme of the midline as new PPR rows organize in the region of the wedge-shaped cluster of *Photd1*-expressing cells (*black arrowhead*). **f** Stage 13 embryo (*S13*), view of the posterior germband. The posterior-most wedge of *Photd1* cells are indicated with an *arrowhead*. The ventral midline cells of the single cell column are all derived from the Ep blastomere and express *Photd1* and bisect and maintain the boundary between the posterior El and Er clones in the developing germband

organize in the developing germband in the stage 11 embryo, anterior lateral *Photd1*-expressing cells have condensed into two bilaterally symmetric groups coincident with the developing head lobes; the weaker anterior medial arc of expression has begun to diminish (Fig. 2d). Posteriorly, *Photd1*-expressing cells begin to form the ventral midline by converging medially and extending along the developing ventral midline as a single column of cells (Fig. 2d). The anterior-most *Photd1*-expressing midline cell is positioned at PPR1, this corresponds to the anterior limit of the Ep clone midline and is coincident with the position of the future first maxillary segment, Mx1. The first ectodermal row to form in the germband is PPR0 which corresponds to the future Mn segment (yellow arrowhead, Fig. 2d). As germband extension continues in the stage 12 embryo, the anterior medial arc of *Photd1* expression is no longer detectable. Posteriorly, *Photd1* expression within the extending single cell ventral midline column is retained via both maintenance of expression in PPR progeny and by convergence and intercalation of cells from more posterior PPRs as they organize at the posterior-most extreme of the ventral midline column (black arrowhead, Fig. 2e,f). Throughout germband development, *Photd1* expression in the ventral midline is confined to cells derived from the Ep blastomere. The cells of the Ep midline lineage are distinct from all other ectodermal cells and are observed to both bisect and maintain a precise physical boundary between the posterior El and Er clones in the developing germband from PPR1-posterior (Fig. 2f; Gerberding et al. 2002; Browne et al. 2005). El and Er blastomere progeny intermix considerably across the midline in regions anterior of the Ep midline lineage (Gerberding et al. 2002).

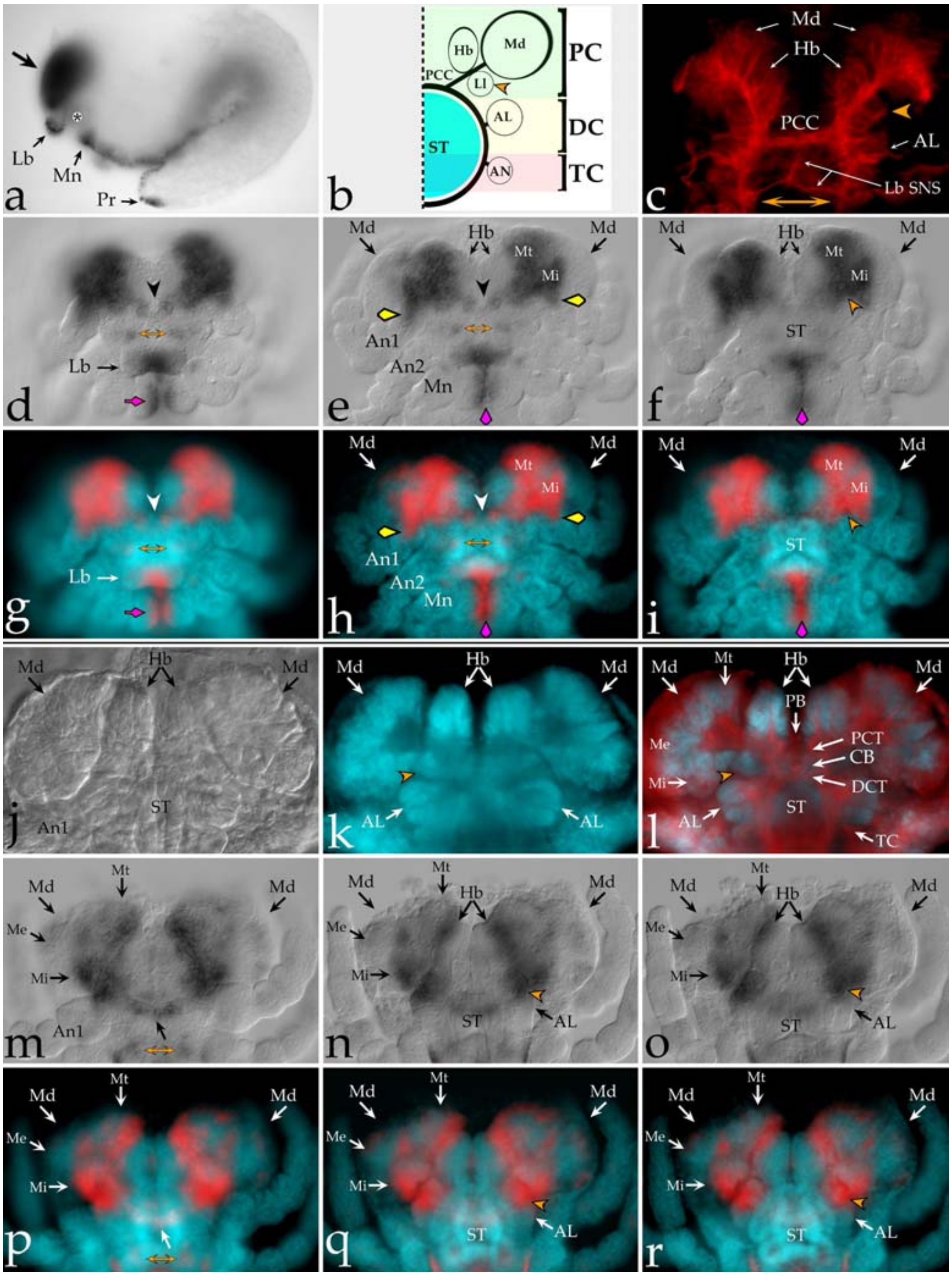
By stage 21, segmentation is complete and the stomodeum is visible medially in the anterior region of the An2 segment (Fig. 3f,i). Expression of *Photd1* appears strongly in the anterior head ectoderm and developing brain, medially in the labrum, and *Photd1* expression in the ventral midline has expanded into ectoderm of the mandibular (Mn) segment anterior of Ep clone descendant cells (Fig. 3a). The region of medial ectoderm between the labrum and the Mn segment has invaginated to form the stomodeum and the lining of the foregut (Fig. 3a; Browne et al. 2005). In the developing anterior head, *Photd1* expression is detected across the midline in the region associated with the formation of the preoral protocerebral commisure (PCC) and the future protocerebral bridge (black arrowhead Fig. 3d,e; white arrowhead Fig. 3g,i).

The *Parhyale* supraesophageal ganglia are composed of the protocerebral neuromere (PC), the deutocerebral neuromere (DC), and the tritocerebral neuromere (TC) (Fig. 3b). These three neuromeres together constitute the brain of *Parhyale*. Expression of *Photd1* is clearly regionalized within the developing PC (Fig. 3d–f). In the medulla complex of the PC, *Photd1* expression is detected throughout the medulla terminalis (Mt), in the medulla interna (Mi), and strongly in a few large cell bodies forming the anlagen of the lateral lobes at the proximal base of the Medulla complex (Fig. 3e,f,h,i). The anterior domain of *Photd1* extends into the proximal base of the An1 antennal appendages that innervate the DC (yellow arrowhead, Fig. 3e,h). *Photd1* is detected neither in the developing anterior medial hemiellipsoid bodies (Hb) nor in the medulla externa (Me). More posteriorly, a de novo domain of *Photd1* is detected in a small bilateral cluster of four to six cells at the base of the labrum and flanking the stomodeum that appear to be positionally associated with receiving/sending axons to the labral nerve of the stomatogastric nervous system (SNS) via the TC neuromere (Fig. 3d,e,g,h). At the posterior extreme of the ventral midline, *Photd1* is also expressed strongly in the developing proctodeum and hindgut ectoderm (Fig. 3a).

By stage 23, the embryonic CNS is well developed, the VNC possesses a full complement of segmentally reiterated anterior and posterior commissures, and many structural landmarks are clearly identifiable in the developing brain and VNC (Figs. 3j–l and 4a–c; Browne et al. 2005). In the developing medulla complex of the PC, *Photd1* expression has retracted (compare Fig. 3e,n). *Photd1* is diminished in the Mt and only remains prominent in cells along the proximal medial boundary adjacent to the Hb (Fig. 3m–r). In contrast, expression of *Photd1* remains strong in the Mi and in the Ll of the PC as these clusters of cells continue to expand (compare Fig. 3f,wk–l,n–r). *Photd1* expression is absent in the Hb and anterior DC (Fig. 3m–r). A narrow band of weak ectodermal expression is detected associated with the anterior border of the labrum (Fig. 3m,p). *Photd1* expression remains detectable in the small bilateral cluster of Lb SNS cells flanking the stomodeum (Fig. 3m,p). These clusters most likely innervate the TC.

## Expression of the *P. hawaiensis Photd1* paralog in the VNC

Expression of *Photd1* at stage 23 in the VNC includes a defined group of segmentally reiterated cells (Fig. 4d,e). These *Photd1*-expressing cells are associated with the anterior commissure (ACom) and the region between the ACom and posterior commissure (PCom) (Fig. 4c–e). The presence of several *Photd1*-expressing cells arrayed along the anterior ACom appears to be a unique attribute not yet reported in any other arthropod (Fig. 4d). The midline glial cells that ensheath the ACom (Gerberding et al. 2001) appear to express *Photd1* (Fig. 4c–e). In *Drosophila*, the positionally analogous glia would be the MGM and MGA glia (Klämbt et al. 1991). The MGM and MGA glia in *Drosophila* do not express *otd*. However, the array of *Photd1*-positive VNC cells shares some features with that of *Drosophila otd* VNC expression; for example, the central cluster of *Photd1* cells between the ACom and the PCom are median neuroblast (MNB) progeny (Fig. 4d, e; Browne et al. 2005). *Photd1* expression in both the MNB progeny and the ACom glia are predictable based on PPR midline lineage studies in a related amphipod (Gerberding et al. 1999, 2001). The expression of *Photd1* in cells associated with the anterior border of the ACom most



likely represents de novo expression flanking the ventral midline lineage. It is also noteworthy that not all ventral midline cell lineages continue to express *Photd1*; for example, expression is absent in the midline glia that ensheath the PCom (Gerberding et al. 2001). Thus, *Photd1* expression is not maintained in all progeny generated by the *Parhyale* Ep blastomere ventral midline lineage.

In summary, the expression of the *Photd1* gene is zygotic, begins during gastrulation, and is associated with the initial organization of the anterior head and ventral midline (Fig. 2). Later in development, *Photd1* becomes restricted anteriorly to defined regions of the PC, anterior DC, and a SNS cell cluster that correlates positionally with innervation of the labral nerves via the TC at the lateral junction of the labrum and stomodeum (Fig. 3). In the developing VNC, *Photd1* is associated with a subset of ventral midline progeny giving rise to glia ensheathing the ACom, the MNB progeny neurons, and de novo expression in a group of cells arrayed along the anterior border of the ACom (Fig. 4d).

## Expression of the *P. hawaiensis otd2* paralog

The temporal and spatial expression dynamics of the second *otd* paralog, *Photd2*, are very different from that of *Photd1*, and many aspects have no direct parallel in other arthropod taxa. Three in situ probes corresponding to

◀ **Fig. 3** *P. hawaiensis otd1* brain expression. All panels (except panel **a** in which embryo anterior is left) are oriented anterior towards the top. **a**–**i** Stage 21 (*S21*) embryos; **j**–**r** stage 23 (*S23*) embryos. **d**–**f** and **m**–**o** Nomarski images though progressively more dorsal focal planes of a S21 embryo and a S23 embryo, respectively, with *Photd1* label in black. **g**–**i** and **p**–**r** Matching focal plane overlays for Hoechst DNA counterstain (nuclei are *blue*) and *Photd1* (false-colored *red*). **a** Segmentation is complete by S21; anteriorly, *Photd1* appears strongly in the anterior head ectoderm (*black arrow*) and medially in the labrum (*Lb*). Ventral midline expression has expanded anteriorly to include the mandibular segment (*Mn*) (expansion into the mandibular segment is first detected at stage 18, data not shown). The region of medial ectoderm between the labrum and the Mn segment invaginates to form the opening of the stomodeum and lumen of the foregut (*asterisk*). Continuity of ventral midline expression is maintained to the posterior extreme of the embryo where *Photd1* is expressed strongly in the ectoderm of the developing proctodeum (*Pr*) and hindgut. **b** A schematic of the *Parhyale* supraesophageal ganglia showing the neuromere position of the major neuropil substructures in a S21 embryo; the midline is indicated by a *vertical dashed line* and the stomodeum (*ST*) is indicated in *blue* and surrounded by a circum-oral neuropil ring of axon tracts (*in black*). The protocerebral neuromere (*PC*) is indicated in *green* and includes the preoral protocerebral commissure (*PCC*) medially, followed by the hemiellipsoid body (*Hb*) and lateral lobe (*Ll*, *orange arrow*) and by the medulla complex (*Md*) laterally. The deutocerebral neuromere (*DC*) is indicated in *yellow* and includes the antennal lobe (*AL*) neuropil. The tritocerebral neuromere (*TC*) is indicated in *red* and includes the antenna 2 neuropil (*AN*). **c** S21 embryo stained for acetylated tubulin to detect axon tract morphology in the developing brain. Medially, the PCC is detected. Lateral of the PCC and projecting anteriorly along the neuroaxis, axon tracts are detected innervating the developing bilateral Hb. Posterior of the Hb are axon tracts innervating the developing lateral lobes of the PC (*orange arrowhead*, also in **b**, **f**, **i**). Just posterior and lateral of the PCC, the AL of the anterior DC can be seen projecting axons into the circumoesophageal ring. The two major axon tracts associated with the labral stomatogastric nervous system (*Lb SNS*) are visible as a prominent medial decending tract, the nervus connectivus, from the PCC and a more posterior lateral tract, the labral nerve, innervating the TC. For panels **d**–**i**, in the developing anterior head, *Photd1* expression (*black staining*) crosses the midline in the region associated with the formation of the future protocerebral bridge (*black arrowhead* in **d**, **e** and *white arrowhead* in **g**, **h**). Bilaterally near the base of the labrum, at the level of the An2 segment and the tritocerebrum (*TC*), *Photd1* is expressed in a field of four to six cells (*orange double-arrow* in **d**, **g** and **e**, **h**) associated with the development of components of the labral stomatogastric nervous system (*Lb SNS*). Ventral midline expression of *Photd1* in the ectoderm is now present in the Mn segment (*purple arrowhead* in **e**, **h** and **f**, **i**). *Photd1* is also present in the distal medial-most region of the developing Lb (**d**, **g**) and distal tips of the developing mandibles (*Mn*) (*purple arrow* in **d**, **g**). *Photd1* expression in the Md of the PC is detected throughout the proximal most region, the medulla terminalis (*Mt*) (**e**, **h** and **f**, **i**), and also in the medulla interna (*Mi*) (**e**, **h** and **f**, **i**). *Photd1* expression is particularly strong in cell bodies of the forming lateral lobes of the PC (*orange arrowheads* in **f**, **i**). Lateral expression of the anterior *Photd1* expression domain extends posteriorly to the proximal base of the antennae (*An1*) that innervate the anterior DC (*yellow arrowheads* in **e**, **h**). **j**–**l** Nomarski (**j**), Hoechst DNA counterstain (**k**, nuclei are *blue*), and matching focal planes for Hoechst and acetylated tubulin (**l**, nuclei in *blue*, acetylated tubulin in *red*) to show morphology and position of axon tracts in the developing S23 *Parhyale* brain. Focal plane bisects the neuroaxis of the supraesophageal ganglia. The anterior-most supraesophageal neuromere, the protocerebrum (*PC*) is composed of a number of distinct substructures. At the anterior midline is the protocerebral bridge (*PB*); this is followed posteriorly by a thick bundle of axons forming the protocerebral commissure tract (*PCT*). Immediately posterior of the PCT is the central body (*CB*) (**l**). Proceeding laterally, the first anterior bilateral neuropils are the hemiellipsiod bodies (*Hb*) (**j**–**l**). Posterior of the PCT are the lateral lobe neuropils (*orange arrowhead*) (**k**, **l**). The lateral-most structure of the PC is the medulla complex (*Md*) (**j**–**l**). The Md complex can be further subdivided into the proximal-most medulla terminalis (*Mt*), lateral medulla interna (*Mi*), and distal-most medulla externa (*Me*) neuropils (**l**). The second supraesophageal neuromere, the deutocerebrum (*DC*), is positioned anteriorly and laterally of the circumesophageal ring surrounding the stomodeum (*ST*) and includes a thick bundle of axons forming the deutocerebral commissure tract (*DCT*) (**l**). Lateral and posterior of the DCT are the antennal lobes (*AL*) that receive nerves from the first antennae (*An1*) (**j**–**l**). The third supraesophageal neuromere, the tritocerebrum (*TC*), is positioned along the lateral posterior portion of the circumesophogeal ring and receives nerves from the second antennae (*An2*) (**l**). **m**–**r** *Photd1* expression in the Md of the PC has retracted to the proximal-most region of the Mt in cells lying adjacent to the lateral border of the Hb and *Photd1* expression remains pervasive throughout the developing Mi. Expression also remains pervasive in developing lateral lobes of the PC (*orange arrowhead*, **n**, **q** and **o**, **r**). A transverse ectodermal band of *Photd1* is also detected at the anterior base of the labrum (*black arrow* **m**, *white arrow* **p**). *Photd1* expression near the base of the labrum, at the level of the TC, associated with the development of the stomatogastric nervous system (*SNS*) is retained (*orange double-arrow*; **m**, **p**)

unique *Photd2* splice variants were made, one to a 5′ splice variant and two to unique 3′ splice variants. No variation in expression profile was detected between probes to different splice variants. Expression is shown for the 3′ splice variant possessing the WSP motif (Fig. 5). *Photd2* first appears at stage 12 during germband formation in two bilaterally symmetric anterior domains coincident with *Photd1* (compare Fig. 2e with Fig. 5a,b). The initiation of *Photd2* expression, after the anterior ectoderm has been patterned into distinct head lobes, implies that *Photd2*, in contrast to *Photd1*, does not play a major role in patterning the anterior head ectoderm. No *Photd2* expression is detected in the formation ventral midline cell lineage of the Ep clone during germband development (Fig. 5a,b). Later in embryonic development (beginning at stage 22), anterior expression of *Photd2* is highly restricted in the anterior head and only appears in the developing PC in a small number of cells associated with the base of the Md complex (Fig. 5c,d). *Photd2* also begins to appear in a restricted set of ventral midline cells (Fig. 5c–g). At the level of the Mn segment, *Photd2* expression is largely coincident with *Photd1* midline expression (Fig. 5c,d). However, in the Mx1 and all posterior segments, ventral midline expression of *Photd2* is mutually exclusive of *Photd1* expression (compare Fig. 4 with Fig. 5e–g). *Photd2* cells in the ventral midline appear to have two distinct fates (Fig. 5e,f). At stage 22, the two ventral-most *Photd2*-expressing cells occur in the middle of developing segments and appear to be ectodermal (Fig. 5e), whereas more dorsally, three *Photd2*-expressing cells at the midline junctions of each segment may be glia (Fig. 5f; Gerberding et al. 2001).

In summary, *Photd2* is expressed zygotically and first appears after the formation of the head lobes in an anterior domain in germband embryos (Fig. 5a,b). Late in embryogenesis, the anterior *Photd2* domain has retracted to a few cells at the base of the medulla (Fig. 5c,d). No expression

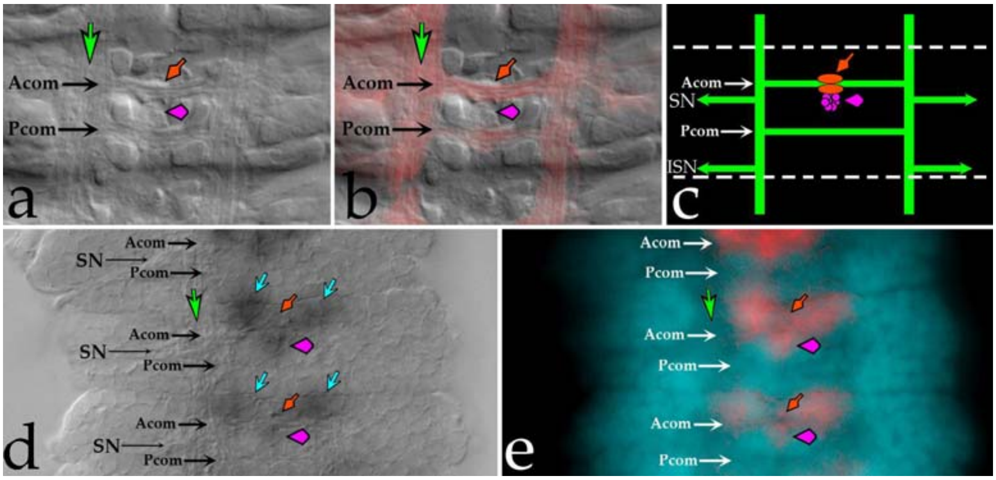


**Fig. 4** Ventral nerve cord expression of *P. hawaiensis otd1*. All panels are dorsal views bisecting a stage 23 embryonic VNC and are oriented anterior towards the top. **a**–**c** Nomarski (**a**), matching planes for nomarski and acetylated tubulin (**b**, acetylated tubulin in *red*), and a VNC schematic (**c**). **d** and **e** Nomarski with *Photd1* in *black* (**d**) and matching planes for Hoechst DNA counterstain and *Photd1* (**e**, nuclei are *blue*; *Photd1* is false-colored *red*). The position of the longitudinal axon tracts is indicated with a *green arrow*. The position of the anterior commissure (*ACom*) and of the posterior commissure (*PCom*) are indicated (**a**–**e**). The position of the segmental nerves (*SN*) is indicated in **c** and **d**. The position of the intersegmental nerves (*ISN*) is indicated in **c**. A group of cells located immediately anterior of the ACom that express *Photd1* can be identified (*blue arrows*, **d**). *Photd1* is detected in glia ensheathing the ACom (*red arrow*) and in the cluster of median neuroblast (*MNB*) progeny between the anterior and posterior commissures (*purple arrowhead*)

of *Photd2* is detected in the ventral midline until late in embryogenesis. Ventral midline expression of *Photd2* is mutually exclusive of its paralog *Photd1*. The fates of these *Photd2*-expressing midline cells are likely to be both ectodermal in the middle of each segment and glial in more dorsal cells at segment boundaries (Fig. 5e–g).

## Discussion

The problems imposed on animals by both their environment and respective genetic history have resulted in a diversity of differences in nervous system organization. Considerable variation is observed in the development and overall organization of the brain and nervous systems among insects alone (e.g., Condron and Zinn 1994; Reichert and Boyan 1997; Urbach and Technau 2003b). In insects and other arthropods, understanding the evolution, development, and segmental homologies of the head and brain has posed long-standing problems in the fields of morphology, paleontology, evolution, and molecular genetics, resulting in what has been described as ‘the endless debate’ (e.g., Rempel 1975; Scholtz and Edgecombe 2005; Browne et al. 2005; Maxmen et al. 2005). In this study, we provide new data from a crustacean that helps clarify the organization and development of the arthropod head and brain.

### Expression of paralogous *Parhyale otd* genes in the anterior head

Available evidence suggests that a single *orthodenticle* gene was likely expressed in the head and foregut of the eubilaterian ancestor (Fig. 1; Bruce and Shankland 1998; Harada et al. 2000; Arendt et al. 2001; Lowe et al. 2003). In the lineage leading to the Lophotrochozoa and Ecdysozoa, *orthodenticle* expression was acquired in segmentally reiterated neurons along the anterior–posterior axis of the body (Fig. 1; Bruce and Shankland 1998). At some point before the divergence of the Tetraconata, the *orthodenticle* gene experienced a number of changes. The number of *orthodenticle* genes expanded via a tandem duplication event (Fig. 1; Li et al. 1996), expression was gained in MNB progeny neurons, and expression was lost in the foregut (Fig. 1). We expect that more comprehensive analysis of *orthodenticle* ortholog expression in taxa diverging before the Tetraconata, including the lophotrochozoans, may reveal that some of these general expression attributes may have first appeared at deeper evolutionary nodes.

Gene duplication is often accompanied by divergence in both gene expression and function. The expression of the *Parhyale otd* paralogs are clearly divergent from *otd* orthologs in both *Tribolium* and *Drosophila*. In contrast to the earliest expression in *Tribolium*, the earliest expression in *Parhyale* (and *Drosophila*) is zygotic and has no maternal component (Fig. 2; Finkelstein and Perrimon 1990). Neither paralog in *Tribolium* is expressed in the DC or TC (Li et al. 1996). The *Parhyale Photd1* paralog is clearly expressed in the anterior DC (as is Drosophila *otd*) as well as in a restricted region of the TC (Fig. 3). This is a feature, thus far, unique to *Parhyale*. The restriction of the *de novo* TC domain to cell bodies associated with innervation of the labral SNS, coupled with the presence of a prominent nervus connectivus linking the labral SNS to the PCC (Fig. 3c), makes it tempting to speculate that ancestral labral innervation was restricted to the PC in combination with an anterior stomodeum (Rempel 1975; Eriksson and Budd 2000; Browne et al. 2005; Maxmen et al. 2005). This is important

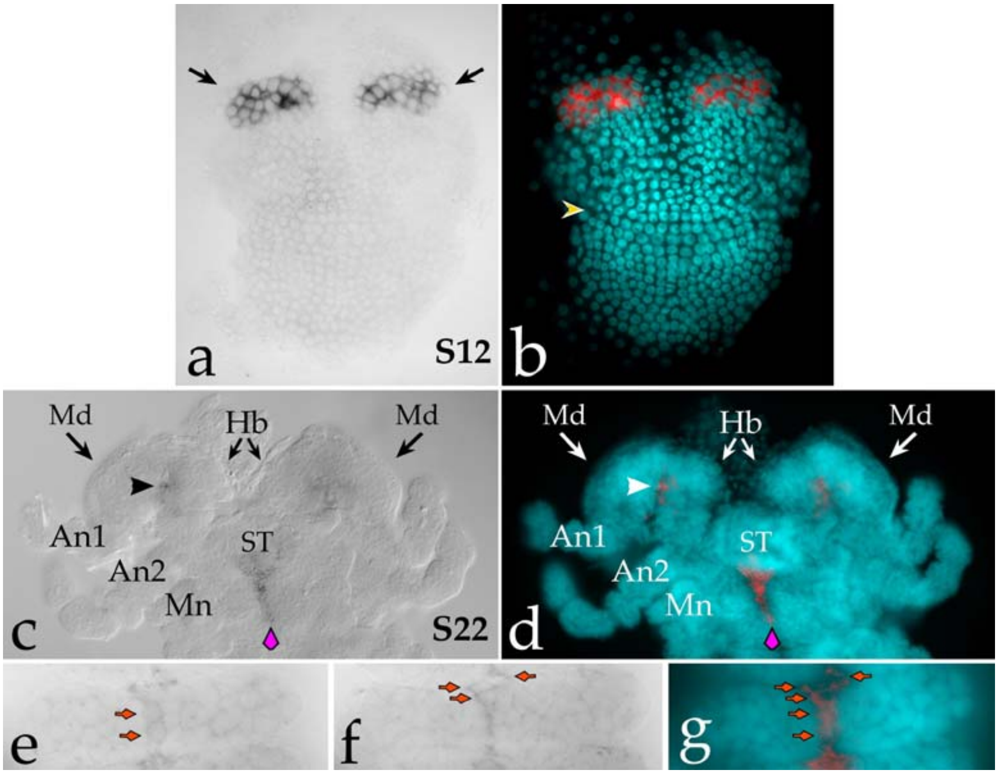


**Fig. 5** Expression of *P. hawaiensis otd2*. All embryos are mounted with the anterior end oriented towards the top. **a**, **b** A stage 12 embryo is mounted ventral side up with a nomarski image on the left and *Photd2* signal in *black*. The image on the *right* is overlay of matching focal planes for Hoechst DNA counterstain with *Photd2* (nuclei are *blue*; *Photd2* is false-colored *red*). *Photd2* is first detected at this stage anteriorly in the condensation of cells in the head lobes (*black arrows*) and is coincident with the anterior *Photd1* domain at this stage of development (compare with Fig. 2e). The *yellow arrowhead* marks the position of the Mn segment. *Photd2* is not expressed in the ventral midline lineage until much later in stage 22 embryos. **c**–**g** *Photd2* expression in a stage 22 embryo. **c**, **d** Embryo mounted dorsal side up, with nomarski image on the *left* and *Photd2* signal in *black*. The image on the *right* is overlay of matching focal planes for Hoechst DNA counterstain with *Photd2* on the right (nuclei are *blue*; *Photd2* is false-colored *red*). In the developing anterior head, *Photd2* expression is restricted to a small number of cells at the base of the Md. Ventral midline expression of *Photd2* can now be detected in the midline ectoderm of the Mn segment (*purple arrowhead*). **e**–**g** A stage 22 embryonic thoracic segment mounted ventral side up. **e** Ventral nomarski focal plane with *Photd2* signal in *black*. **f** Dorsal nomarski focal plane with *Photd2* signal in *black*. **g** Collapsed and merged ventral and dorsal focal planes for Hoechst and *Photd2* (nuclei are *blue*; *Photd2* is false-colored *red*). *Photd2* is expressed in VNC midline cells exclusive of *Photd1* (compare with Fig. 4). The two midline cells in **e** appear to be ectodermal. The three slightly more dorsal cells in **f** are at the segmental border and may be glial cells

because the positional origin of the labrum, as the most prominent anterior ectodermal structure in most arthropods, has played a central role in attempts to understand underlying segmental patterning of the arthropod head. Expression analysis of *otd* orthologs in onychophorans would prove particularly useful in helping to resolve questions regarding ancestral features of head patterning and brain development in the Ecdysozoa and may settle the debate regarding the positional origin and innervation patterns associated with the appendicular labrum.

A universal mechanism for patterning both the invertebrate and vertebrate brain has been proposed to account for their tripartite brain organization (Hirth et al. 2003). Based on expression patterns in *Drosophila*, this model proposes that the interface between the posterior boundary of *orthodenticle* expression and anterior expression domain of the gene *unpg* maintains the boundary between the DC and TC via mutual repression. In *Tribolium*, neither *otd* paralog is expressed in the DC during brain development. In *Parhyale*, *Photd1* is expressed only transiently in the anterior DC (Fig. 3e,h). In light of this variation in *otd* expression, it would be informative to validate this mechanism by observing the relationship between *Photd1* and *Parhyale unpg* orthologs in the maintenance of gene expression boundaries correlating with the DC–TC neuromere interface.

The expression of *Photd2* in the head is radically divergent both from its paralog *Photd1* as well as from *otd* orthologs in divergent arthropod taxa. The initial anterior domain of *Photd2* is coincident with *Photd1* but does not appear until after the head lobes are organized (Fig. 5). As development progresses, the anterior *Photd2* expression domain decays and, late in development, is only detectable in a few cells of the PC (Fig. 5). *Photd2* is consistently expressed at a much lower level than *Photd1*, a phenomenon also observed in *Tribolium*.

## Expression of paralogous *Parhyale otd* genes in the ventral midline

Current data suggest *orthodenticle* expression was acquired in the ventral midline in arthropods, (Figs. 1, 2, 4, and 5; Finkelstein et al. 1990; Li et al. 1996; Telford and Thomas 1998). In *Parhyale*, as in *Drosophila* and *Tribolium*, expression of *otd* in the midline cells is apparent at an early stage of development (Fig. 2). The ontological origin of *Tribolium* midline cells are undescribed; however, midline cell lineages are known in *Drosophila* (Klämbt et al. 1991; Bossing and Technau 1994; Schmid et al. 1999), the grasshopper *Schistocerca* (Condron and Zinn 1994), as well as in a related amphipod crustacean *Orchestia* (Gerberding and Scholtz 1999, 2001). Our data suggests some similarities with *otd* expression in insect midline lineages as well as a number of significant differences.

The ectodermal ventral midline cell lineage in the closely related amphipod, *Orchestia*, generates a specific set of glial and neuronal progeny. For each segment, two glial cells are produced that ensheath the anterior commissure, two glial cells are produced to ensheath the posterior commissure, two to three glia are produced that are positioned at the midline of the intersegmental border, and ~10 median neuroblast progeny are produced that cluster between the ACom and PCom (Gerberding and Scholtz 2001). The midline of *Parhyale* develops in an analogous fashion to that of *Orchestia*. In *Drosophila*, the midline develops from a unique arrangement of mesectodermal cells. The midline lineage produces both the MNB progeny neural cluster and the glia that ensheath the anterior commissure (Klämbt et al. 1991). In insects, the PCom is pioneered ahead of the ACom (Klämbt et al. 1991; Myers and Bastiani 1993), whereas in *Parhyale* the ACom is pioneered ahead of the PCom. *Photd1*-positive midline cells appear in both the MNB progeny neural cluster and in the glia that ensheath the ACom, expression features shared with both *Drosophila* and grasshopper (Fig. 4c–e). The position of the *Parhyale* MNB neural cluster is displaced anteriorly relative to *Drosophila* and grasshopper, but this is likely a pleiotropic effect due largely to the observed heterochrony in commissure formation (reversed order of commissure formation).

In addition to the conserved features, *Photd1* is also uniquely expressed in a group of cells arrayed along the anterior border of the ACom (Fig. 4d). The presence of *Photd1* in these cells represents de novo expression flanking the ventral midline lineage, as well as expression in a group of cells not yet observed in other arthropods. In addition, the late embryonic expression of *Photd2* in a group of midline cells mutually exclusive of *Photd1* has no parallel in other taxa. At least three of these reiterated cells that are found dorsally at the midline junction of segments may represent glial descendants of the ventral midline lineage (Fig. 5e–g; Gerberding and Scholtz 2001).

## Implications regarding the evolution and duplication of the *orthodenticle* gene

Understanding the history of *orthodenticle* gene duplication during the course of metazoan evolution presents an interesting dilemma. Expansion of *otd* genes within the vertebrates via genome amplification is straightforward (Germot et al. 2001). However, in all other instances of duplication, *otd* paralogs are in close proximity, suggesting tandem duplication, and the paralogs appear more similar to one another than to *otd* orthologs in other species. This suggests two possible scenarios: (1) *otd* paralogs in each case represent independent duplications or (2) *otd* duplications are very ancient, occurred at deeper nodes, and paralogs in each lineage appear to be different from their orthologs. The history of *otd* duplication could be sufficiently complex that we will not be able to determine the relationships between paralogs from different lineages.

We suggest that two well-known genetic mechanisms, gene conversion and mutational saturation, can be used as support for scenario 2 based on currently available genomic information in the taxa analyzed. Gene conversion is a process by which paralogous genes acquire identical nucleotide sequences via mismatch repair of heteroduplexed DNA (Radding 1982). The likelihood of gene conversion increases with the proximity of the paralogs. Gene conversion would become a problem when attempting to reconstruct the orthology relationships between paralogs due to regions between paralogs becoming more similar to each other than to their orthologs in other lineages and, thus, loosing useful phylogenetic signal. The second mechanism, mutational saturation, also results in misleading phylogenetic signal among paralogous sequences (Philippe and Forterre 1999). A manifestation of mutational saturation, for example, is the process of codon usage bias governing the pool of available tRNAs, again acting to make paralogs more similar to one another than to their respective orthologs in other lineages, in this instance via purifying selection (Nei and Kumar 2000).

Among the taxa we sampled, the beetle *T. castaneum* genome contains two *otd* paralogs within ~60 kb of one another (BeetleBase, http://www.bioinformatics.ksu.edu/beetlebase). The hymenopteran *Apis mellifera* genome contains two *otd* paralogs within ~40 kb of one another (BeeBase, http://racerx00.tamu.edu/bee_resources.html). Preliminary information from the crustacean *Daphnia pulex* genome has revealed the presence of two *otd* paralogs within ~24 kb of one another (F. Poulin and N. Patel, personal communication). Sampling among the chelicerates has only revealed a single *otd* ortholog. Thus, the current evidence suggests a single tandem duplication of an ancestral *otd* gene in the arthropod lineage leading to the Tetraconata (Fig. 1). Within the Tetraconata, two things appear to have happened to *otd* paralogs: (1) loss of one *otd* duplicate in the lineage leading to the dipterans as represented by *Drosophila* (Fig. 1) and (2) physical proximity of the gene duplicates may have led to sequence homogenization within a given lineage via the two distinct genetic mechanisms outlined above, gene conversion, and/or mutational saturation. Large-scale genomic resources are not yet available for *Parhyale*; however, we predict that the two *P. hawaiensis orthodenticle* paralogs are physically linked.

The test for mutational saturation suggests that base pair saturation has played a role in *orthodenticle* gene divergence in different taxa, making recovery of paralog orthology among the tandem duplicates present in Tetraconata (crustaceans + insects) particularly problematic with currently available phylogenetic tools. Direct evidence of gene conversion could possibly be demonstrated by investigating *orthodenticle* paralog sequences at the level of populations, revealing instances of allelic variation resulting from heteroduplex DNA mismatch repair. It is tantalizing to hypothesize that the unique retention of the WSP domain in both *A. mellifera* paralogs may represent evidence for this mechanism downstream of the homeodomain within this lineage (Fig. 1).

In summary, the data presented here represent additional evidence for a single ancestral *orthodenticle* like gene in the common ancestor of the lineage leading to the cnidarians + bilaterians (Fig. 1; Li et al. 1996). In most cases, we also recover an eight-amino-acid diagnostic sequence, the WSP motif, downstream of the homeodomain in at least one *otd* paralog (Fig. 1). It is highly probable that this motif existed in the single ancestral *orthodenticle* gene before the divergence of the cnidarians and the bilaterians. This ancestral *orthodenticle* gene was independently duplicated in lineages leading to the vertebrates (genome duplication), the cnidarians (three linked copies), and the Tetraconata (tandem duplication). One paralog was subsequently lost in the lineage leading to the dipterans (Fig. 1).

## Prospects

Our analysis of *otd* paralog gene expression in *Parhyale* has indicated some of the general boundaries/structures associated with the developing supraesophageal neuromeres that comprise the brain. We are hopeful that, when used in combination with other markers, they will contribute to understanding the distribution and number of neural stem cells that initially specify the *Parhyale* brain (e.g., Urbach and Technau 2003a,b). This type of neuroblast gene expression data will prove to be important in crustaceans as, in sharp contrast to insects, crustacean neuroblasts do not delaminate from the ectoderm and, thus, are often not clearly distinguishable from surrounding non-neural ectodermal cells. In addition, crustacean neuroblasts may switch fates as they divide, generating both neural-fated and non-neural fated ectoderm progeny. Thus, the best means for discriminating between ectodermal cells, neuroblasts, neurons, and glia in crustaceans will be with panels of molecular markers used in concert with neuronal morphology.

Comprehensive characterization of gene expression and, subsequently, genetic function (e.g. Pavlopoulos and Averof 2005) in the *Parhyale* brain and nervous system will begin to provide us with a framework and new set of tools for exploring diverse aspects related to developmental patterning at three levels: the basic underpinnings of neurogenesis in metazoans, aspects of developmental mechanisms inherent to neurogenesis in arthropods, and specific details related to the unique development of the nervous system of diverse crustaceans.

**Acknowledgements** This work has been supported by the NSF (WEB, DBI-0310269), the NIH (WEB, NCRR-P20RR16467), the Boehringer Ingelheim Foundation (EAW), the Deutsche Forschungsgemeinschaft (EAW, DFG Wi 1797/2-2), the European Community's Marie Curie Research Training Network ZOONET under contract MRTN-CT-2004-005624 (EAW). EAW also acknowledges support from the EMBO Young Investigator Programme. We thank Kevin Pang and Andreas Hejnol for *Otx* sequences, Frank Poulin and Nipam H. Patel for *D. pulexotd* information, and Casey Dunn for computation advice. We also thank the following for providing critical comments that have significantly improved this communication: Elaine Seaver, Amy Maxmen, Andreas Hejnol, Casey Dunn, Nipam H. Patel, and two anonymous reviewers.

## References

Acampora D, Avantaggiato V, Tuorto F, Barone P, Reichert H, Finkelstein R, Simeone A (1998) Murine *Otx1* and *Drosophila otd* genes share conserved genetic functions required in invertebrate and vertebrate brain development. Development 125:1691–1702

Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, de Rosa R (2000) The new animal phylogeny: reliability and implications. Proc Natl Acad Sci USA 97:4453–4456

Arendt D, Nubler-Jung K (1996) Common ground plans in early brain development in mice and flies. Bioessays 18:255–259

Arendt D, Technau U, Wittbrodt J (2001) Evolution of the bilaterian larval foregut. Nature 409:81–85

Bossing T, Technau GM (1994) The fate of the CNS midline progenitors in *Drosophila* as revealed by a new method for single cell labelling. Development 120:1895–1906

Browne WE, Price AL, Gerberding M, Patel NH (2005) Stages of embryonic development in the amphipod crustacean, *Parhyale hawaiensis*. Genesis 42:124–149

Bruce AEE, Shankland M (1998) Expression of the head gene *Lox22-Otx* in the leech *Helobdella* and the origin of the bilaterian body plan. Dev Biol 201:101–112

Cohen S, Jürgens G (1990) Mediation of *Drosophila* head development by gap-like segmentation genes. Nature 346:482–485

Cohen S, Jürgens G (1991) *Drosophila* headlines. Trends Genet 7:267–272

Condron BG, Zinn K (1994) The Grasshopper median neuroblast is a multipotent progenitor cell that generates glia and neurons in distinct temporal phases. J Neurosci 14:5766–5777

Cook CE, Yue Q, Akam M (2005) Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. Proc R Soc Lond B 272:1295–1304

Dalton D, Chadwick R, McGinnis W (1989) Expression and embryonic function of *empty spiracles*: a *Drosophila* homeo box gene with two patterning functions on the anterior–posterior axis of the embryo. Genes Dev 3:1940–1956

Dohle W (2001) Are the insects terrestrial crustaceans? A discussion of some new facts and arguments and the proposal of the proper name 'Tetraconata' for the monophyletic unit Crustacea plus Hexapoda. Ann Soc Entomol Fr 37:85–103

Duman-Scheel M, Patel NH (1999) Analysis of molecular marker expression reveals neuronal homology in distantly related arthropods. Development 126:2327–2334

Eriksson BJ, Budd GE (2000) Onychophoran cephalic nerves and their bearing on our understanding of head segmentation and stem-group evolution of Arthropoda. Arthropod Struct Dev 29:197–209

Finkelstein R, Perrimon N (1990) The *orthodenticle* gene is regulated by *bicoid* and *torso* and specifies *Drosophila* head development. Nature 346:485–488

Finkelstein R, Smouse D, Capaci TM, Spradling AC, Perrimon N (1990) The *orthodenticle* gene encodes a novel homeo domain protein involved in the development of the *Drosophila* nervous system and ocellar visual structures. Genes Dev 4:1516–1527

Finnerty JR, Paulson D, Burton P, Pang K, Martindale MQ (2003) Early evolution of a homeobox gene: the parahox gene *Gsx* in the Cnidaria and the Bilateria. Evolut Develop 5:331–345

Friedrich M, Tautz D (1995) Ribosomal DNA phlogeny of the major extant arthropod classes and the evolution of myriapods. Nature 376:165–167

Gallitano-Mendel A, Finkelstein R (1998) Ectopic *orthodenticle* expression alters segment polarity gene expression but not head segment identity in the *Drosophila* embryo. Dev Biol 199: 125–137

Gerberding M (1997) Germ band formation and early neurogenesis of *Leptodora kindti* (Cladocera): first evidence for neuroblasts in the entomostracan crustaceans. Invertebr Reprod Dev 32: 63–73

Gerberding M, Scholtz G (1999) Cell lineage of the midline cells in the amphipod crustacean *Orchestia cavimana* (Crustacea, Malacostraca) during formation and separation of the germ band. Dev Genes Evol 209:91–102

Gerberding M, Scholtz G (2001) Neurons and glia in the midline of the higher crustacean *Orchestia cavimana* are generated via an invariant cell lineage that comprises a median neuroblast and glial progenitors. Dev Biol 235:397–409

Gerberding M, Browne WE, Patel NH (2002) Cell lineage analysis of the amphipod crustacean *Parhyale hawaiensis* reveals an early restriction of cell fates. Development 129:5789–5801

Germot A, Lecointre G, Plouhinec J-L, Le Mentec C, Girardot F, Mazan S (2001) Structural evolution of *Otx* Genes in Craniates. Mol Biol Evol 18:1668–1678

Giribet G, Edgecombe GD, Wheeler WC (2001) Arthropod phylogeny based on eight molecular loci and morphology. Nature 413:157–161

Giribet G, Richter S, Edgecombe GD, Wheeler WC (2005) The position of crustaceans within Arthropoda—evidence from nine molecular loci and morphology. In: Koenemann S, Jenner RA (eds) Crustacea and arthropod relationships. CRC, Boca Raton, pp 307–352

Grossniklaus U, Cadigan KM, Gehring WJ (1994) Three maternal coordinate systems cooperate in the patterning of the *Drosophila* head. Development 120:3155–3171

Hanstrom B (1928) Vergleichende Anatomie Des Nervensystems Der Wirbellosen Tiere Unter Berucksichtigung Seiner Funktion. Springer, Berlin Heidelberg New York

Harada Y, Okai N, Taguchi S, Tagawa K, Humphreys T, Satoh N (2000) Developmental expression of the hemichordate *otx* ortholog. Mech Dev 91:337–339

Harzsch S (2003) Ontogeny of the ventral nerve cord in malacostracan crustaceans: a common plan for neuronal development in Crustacea, Hexapoda and other Arthropoda? Arthropod Struct Dev 32:17–37

Hirth F, Therianos S, Loop T, Gehring WJ, Reichert H, Furukubo-Tokunaga K (1995) Developmental defects in brain segmentation caused by mutations of the homeobox genes *orthodenticle* and *empty spiracles* in *Drosophila*. Neuron 15: 769–778

Hirth F, Kammermeier L, Frei E, Walldorf U, Noll M, Reichert H (2003) An urbilaterian origin of the tripartite brain: developmental genetic insights from *Drosophila*. Development 130:2365–2373

Hwang UW, Friedrich M, Tautz D, Park CJ, Kim W (2001) Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature 413:154–157

Klämbt C, Jacobs JR, Goodman CS (1991) The midline of the *Drosophila* central nervous system: a model for the genetic analysis of cell fate, cell migration, and growth cone guidance. Cell 64:801–815

Li Y, Brown SJ, Hausdorf B, Tautz D, Denell RE, Finkelstein R (1996) Two *orthodenticle*-related genes in the short-germ beetle *Tribolium castaneum*. Dev Genes Evol 206:35–45

Lichtneckert R, Reichert H (2005) Insights into the urbilaterian brain: conserved genetic patterning mechanisms in insect and vertebrate brain development. Heredity 94:465–477

Lowe CJ, Wu M, Salic A, Evans L, Lander E, Stange-Thomann N, Gruber CE, Gerhart J, Kirschner M (2003) Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. Cell 113:853–865

Maxmen A, Browne WE, Martindale MQ, Giribet G (2005) Neuroanatomy of sea spiders implies an appendicular origin of the protocerebral segment. Nature 437:1144–1148

Myers PZ, Bastiani MJ (1993) Cell–cell interactions during the migration of an identified commissural growth cone in the embryonic grasshopper. J Neurosci 13:115–126

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford, Oxford

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Patel N (1994) Imaging neuronal subsets and other cell types in whole mount *Drosophila* embryos and larvae using antibody probes. Methods Cell Biol 44:445–487

Pavlopoulos A, Averof M (2005) Establishing genetic transformation for comparative developmental studies in the crustacean *Parhyale hawaiensis*. Proc Natl Acad Sci USA 102:7888–7893

Philippe H, Sorhannus U, Baroin A, Perasso R, Gasse F, Adoutte A (1994) Comparison of molecular and paleontological data in diatoms suggest a major gap in the fossil record. J Evol Biol 7:247–265

Philippe H, Forterre P (1999) The rooting of the universal tree of life is not reliable. J Mol Evol 49:509–523

Posada D, Crandall K (1998) ModelTest: testing the model of DNA substitution. Bioinformatics 14:817–818

Radding (1982) Homologous pairing and strand exchange in genetic recombination. Annu Rev Genet 16:405–437

Regier JC, Shultz JW, Kambic RE (2005) Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc R Soc Lond B 272:395–401

Reichert H, Boyan G (1997) Building a brain: developmental insights in insects. Trends Neurosci 20:258–264

Rempel JG (1975) The evolution of the insect head: the endless dispute. Quaest Entomol 11:7–25

Richter S (2002) The Tetraconata concept: hexapod–crustacean relationships and the phylogeny of Crustacea. Org Divers Evol 2:217–237

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574

Royet J, Finkelstein R (1995) Pattern formation in *Drosophila* head development: the role of the *orthodenticle* homeobox gene. Development 121:3561–3572

Sandeman D, Sandeman R, Derby C, Schmidt M (1992) Morphology of the brain of crayfish, crabs, and spiny lobsters: a common nomenclature for homologous structures. Biol Bull 183:304–326

Schmid A, Chiba A, Doe CQ (1999) Clonal analysis of *Drosophila* embryonic neuroblasts: neural cell types, axon projections and muscle targets. Development 126:4653–4689

Schmidt-Ott U, Gonzalez-Gaitan M, Jäckle H, Technau GM (1994) Number, identity, and sequence of the *Drosophila* head segments as revealed by neural elements and their deletion patterns in mutants. Proc Natl Acad Sci USA 91:8363–8367

Scholtz G, Edgecombe GD (2005) Heads, Hox, and the phylogenetic position of trilobites. In: Koenemann S, Jenner RA (eds) Crustacea and arthropod relationships. CRC, Boca Raton, pp 139–165

Schröder R (2003) The genes *orthodenticle* and *hunchback* substitute for bicoid in the beetle *Tribolium*. Nature 422:621–625

Sharman AC, Brand M (1998) Evolution and homology of the nervous system: cross-phylum rescues of *otd/Otx* genes. Trends Genet 14:211–214

Strausfeld NJ (1998) Crustacean–insect relationships: the use of brain characters to derive phylogeny amongst segmented invertebrates. Brain Behav Evol 52:186–206

Strausfeld NJ, Hildebrand JG (1999) Olfactory systems: common design, uncommon origins? Curr Opin Neurobiol 9:634–639

Swofford D (2003) PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4: 4.0Beta edn. Sinauer, Sunderland, Massachusetts

Telford MJ, Thomas RH (1998) Expression of homeobox genes shows chelicerate arthorpods retain their deutocerebral segment. Proc Natl Acad Sci USA 95:10671–10675

Urbach R, Technau GM (2003a) Molecular markers for identified neuroblasts in the developing brain of *Drosophila*. Development 130:3621–3637

Urbach R, Technau GM (2003b) Early steps in building the insect brain: neuroblast formation and segmental patterning in the developing brain of different insect species. Arthropod Struct Dev 32:103–123

Walldorf U, Gehring WJ (1992) *Empty spiracles*, a gap gene containing a homeobox involved in *Drosophila* head development. EMBO J 11:2247–2259

Whitington PM (1996) Evolution of neural development in the arthropods. Sem Cell Dev Biol 7:605–614

Whitington PM, Leach D, Sanderman R (1993) Evolutionary change in neural development within the arthropods: axonogenesis in the embryos of two crustaceans. Development 118:449–461

Wieschaus E, Perrimon N, Finkelstein R (1992) *Orthodenticle* activity is required for the development of medial structures in the larval and adult epidermis of *Drosophila*. Development 115:801–811

Wimmer EA, Jäckle H, Pfeifle C, Cohen SM (1993) A *Drosophila* homologue of human *Sp1* is a head-specific segmentation gene. Nature 366:690–694

Wimmer EA, Cohen SM, Jäckle H, Desplan C (1997) *buttonhead* does not contribute to a combinatorial code proposed for *Drosophila* head development. Development 124:1509–1517

Younossi-Hartenstein A, Green P, Liaw G-J, Rudolph K, Lengyel J, Hartenstein V (1997) Control of early neurogenesis of the *Drosophila* brain by the head gap genes *tll*, *otd*, *ems*, and *btd*. Dev Biol 182:270–283

[PhD_A1_published_data]

**“Browne_et_al,2006_supplement.doc”**

# Supplement to *Parhyale hawaiensis otd2*

## Supplementary Figure A

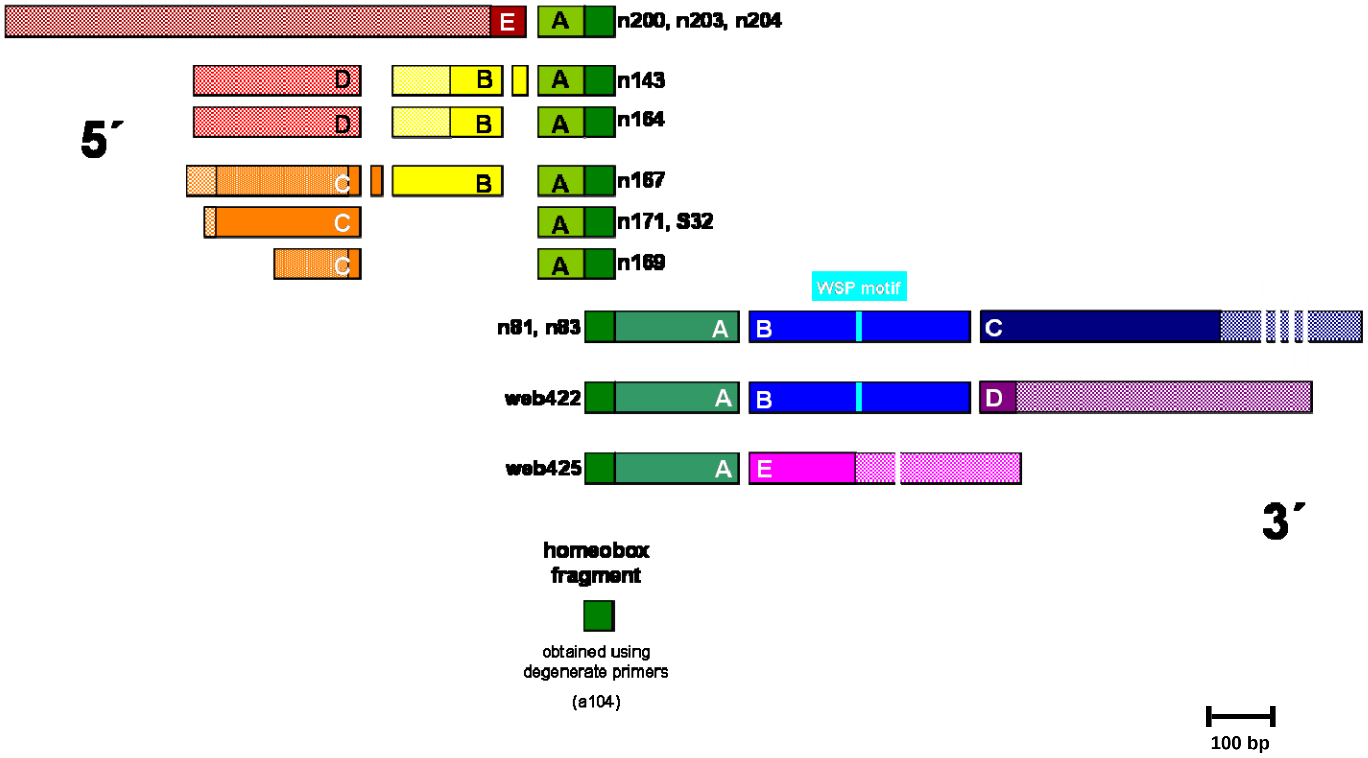


**Figure A. Schematic overview over 5' and 3' RACE clones of *Parhyale hawaiensis otd*2 and their putative splicing structure.** A schematic alignment of the nucleic acid sequences of 5' RACE clones (top of panel), 3' RACE clones (middle of panel) and the short *otd*2 homeobox fragment obtained by PCR using degenerate primers ("fishing" PCR, bottom of panel). Individual putative splicing units are characterized by respective colors and letters: 5'E splicing unit in dark red, 5'D in bright red, 5'C orange, 5'B yellow, 5'A light green; homeobox core fragment green, 3'A splicing unit pale green, 3'B blue, 3'C indigo, 3'D purple and 3'E pink. Each different clone (or group of clones) is comprised of a different combination of respective splicing units.

Open reading frames within the obtained RACE clones were determined by the 5'-most ATG (5' RACE clones) or first 3' stop codon (3' RACE clones, respectively). Coding sequences are in solid colours, non-coding sequences are in hatched colours. A WSP-motif which is present in translated 3'B splicing units is indicated in cyan. For more detailed information about individual RACE clones, see supplementary figure C.

All putative splicing units are displayed in their correct relative scale. Each white perpendicular stroke in clones n81/n83 3'C and clone web425 3'E corresponds to 500 bp of noncoding sequence.

5'C of n169 appears to be truncated as compared to 5'C of n171/S32. In n167, the lack of one adenosine at the 5' end of the sequence changes the open reading frame and thus brings the putative start codon out of frame as compared to n171/S32.

# Supplementary Figure B

**Figure B. Protein alignments of translated 5' and 3' *otd*2 RACE clones.** Translated protein sequences that correspond to putative splice units are marked in their respective colours (see also Figure A). The first alignment shows translated 5'RACE sequences (incl. homeodomain of clone a104); the second alignment shows translated 3' RACE sequences (incl. homeodomain of clone a104).

Alignment of translated 5' RACE sequences

```
n203
n143
n164
n167
n171                                    MPWESHLSFSLKLQSRKKHTHH
n169
a104


n203
n143
n164
n167
n171    HLSASLLVLTHIVRRLKCPKQVQSYSSLRTCSDLPHLSSSRGRNPEACTA
n169
a104


n203
n143                                                   MNG
n164                                                   MNG
n167       MWASAQGGGGGGGVAFPGFGGSCGAGGSAAAAAGVYLKTPAPYTMNG
n171    DPVMWASAQ
n169       MWASAQ
a104

n203                   MQGTHLLTTRSTQARALTIAATPGRKQRRERTTYT
n143    LGLGVGMGVDSLHPSMGYPPPTDYLYSEYKYSYMAATPGRKQRRERTTYT
n164    LGLGVGMGVDSLHPSMGYPPPTDYLY        TATPGRKQRRERTTYT
n167    LGLGVGMGVDSLHPSMGYPPPTDYLY        TATPGRKQRRERTTYT
n171                                      AATPGRKQRRERTTYT
n169                                      AATPGRKQRRERTTYT
a104


n203    RAQLDILETL
n143    RAQLDILETL
n164    RAQLDILETL
n167    RAQLDILETL
n171    RAQLDILETL
n169    RAQLDILETL
a104    RAQLDILETLFGKTRYPDIFMREEVAIKINLPESRI
```

# Alignment of translated 3' RACE sequences

```
a104      RAQLDILETLFGKTRYPDIFMREEVAIKINLPESRI
n81                              EVAIKINLPESRIQVWFKNRRAKCRQQ
web422                              IKINLPESRIQVWFKNRRAKCRQQ
web425                              IKINLPESRIQVWFKNRRAKCRQQ

a104
n81       QKQQAANGEKAPRTKKVNKSPPPSTNNTITNSSLIGSGAGNNAATTNHHH
web422    QKQQAANGEKAPRTKKVNKSPPPSTNNTITNSSLIGSGAGNNAATTNHHH
web425    QKQQAANGEKAPRTKKVNKSPPPSTNNTITNSSLIGSGAGNNAATITTTT

a104
n81       HHHTTPNNNNNNNSNSSNSSTTLTSPLTTSHGARDASPPSPPPSVSSGSS
web422    HHHTTPNNNNNNNSNSSNSSTTLTSPLTTSHGARDASPPSPPPSVSSGSS
web425    IITITIKVNRKRANITIKRKTESTVQTRKTRVRSSRWRRFTTAPKKIERP

a104
n81       LSLPPSTPYNPIWSPAGIPHQSSTLPAMGDLMGGGGLDRGGYGSACYQGY
web422    LSLPPSTPYNPIWSPAGIPHQSSTLPAMGDLMGGGGLDRGGYGSACYQGY
web425    AIWA

a104
n81       GTPSYYSNMDYLAPMTHSQMSSMTSLNPMTSAHHQHLGAHHHSAAAGLSP
web422    GTPSYYSNMDYLAPMTHSQVNTFAHNIDLSILIMLTCICK
web425

a104
n81       SPLGSSLGSSPLSSSHHLVSSLSSTSLNSGSHAALTHPPLTHSSSLNQDS
web422
web425

a104
n81       PRSPSGAPGSLTPLPQDCLDGYPEDKSAIIAPNSAAAASAAAAWKNYQGF
web422
web425

a104
n81       QSL
web422
web425
```

# Supplementary Figure C

**Figure C. Complete List of DNA-sequences of aforementioned 5' and 3' RACE clones of *Parhyale hawaiensis otd*2 along with their translated protein sequence.** Respective putative splice units are coloured as in Figure A.

```
  1     ggtgtgttattgtgatggtgagagcagcacagtttgtcgactgcaagtgaacacaag   57

 58  tgaatgaggtagaaggtgtgaaagtgaatatgaacacaagtgaacaatgttgtgattata  117

118  aaatggactgatggtagtgaatgtggatacaagtgaaataggtagggttgttcatgcaat  177

178  gcgactggaagtggtgaatgtggacacaactgcgctaggtagccgcctcgaacgtggctg  237

238  gctaccgataatgatagtacatgaaagtggggacgataatcagggtgcgaacttggtgga  297

298  agtgacaaaaaacgaaaattgcggacatagccagatttgttgcttaagaaactgaaagag  357

358  agaagaacaggtgaagaaaacagatatgggtttattgggaacaatattgaagtggtgcaa  417

418  gtgcgctgacactggtagttgttggaaattgtgaagatactcgtaaaaagaagtgcggaa  477

478  gattatttaacagcgactgaaatttccagtgtgaaaatacatataaatatgagcagtgt  537

538  agggcagtgacttttttaacattaggtctttgattgtactgtgcagataggtatgcggttg  597

598  agcttcctgttaggcgcgagaaaactgaatcatttcggctagaacatgtacaagttaaca  657

658  acataagaaattcggccggtgtttcagaaatatagtttggttgtttacaaagctgaactt  717

718  gagggcccactcctgcaaagaagattctgctcaagcctcgaaaataatctttgtttttc  777

                                                    M  Q  G  T  H
778  taagcaacaaacagacctctaacagaaatttaactgctaatagcgATGCAAGGAACGCAT  837

      L  L  T  T  R  S  T  Q  A  R  A  L  T  I  A  A  T  P  G  R
838  CTATTAACAACAAGAAGCACGCAGGCAAGAGCACTGACCATTGCGGCCACTCCAGGCCGG  897

      K  Q  R  R  E  R  T  T  Y  T  R  A  Q  L  D  I  L  E  T  L
898  AAGCAGCGGCGGGAGCGCACGACGTACACCAGAGCGCAACTGGACATCCTGGAGACCCTG  957
```

DNA-sequence of **Ph_*otd*2n203** (*Parhyale hawaiensis otd*2 clone n203, exemplary for clones n200, n203 and n204), obtained from *otd*2 5' RACE. Ph_*otd*2n203 consists of consensus 5' splicing units 5'C and 5'A. Start of open reading frame is marked as ATG.

```
  1                                    ggagcggggcccttctagtcgagtgt   26

 27  ggtggtggtgatgatggttagtggtggcagtattgttaccagcgccgctgcctatgagcg   86

 87  acgagttcgtaattgtgttgtaaggggttgatggaggcagagacagggatgaccccgagc  146

147  taacggacggagggggtgacggtggggacgcgtctctagctccatgggacgtcgtcaaag  206

207  gcgacgtcagggttgttgagctgttactgctgttactgttattgttgttattattgttgg  266

267  gggtggtgtgatggtggggtcgcattccccgggttcgggggcagttgcggggctgggggg  326

                                                          M  N  G
327  tcagcagcagcagccgctggggtctatctcaaaacccccgccccctacaccATGAACGGA  386

      L  G  L  G  V  G  M  G  V  D  S  L  H  P  S  M  G  Y  P  P
387  CTTGGACTCGGTGTCGGCATGGGGGTAGACAGCCTGCACCCAAGCATGGGCTATCCTCCC  446

      P  T  D  Y  L  Y  S  E  Y  K  Y  S  Y  M  A  A  T  P  G  R
447  CCTACCGATTATTTATACAGTGAGTATAAATACTCGTACATGGCGGCCACCCCAGGCCGG  506

      K  Q  R  R  E  R  T  T  Y  T  R  A  Q  L  D  I  L  E  T  L
507  AAGCAGCGGCGGGAGCGCACTACGTACACCAGAGCGCAACTGGACATCCTGGAGACCCTG  566
```

DNA-sequence of **Ph_*otd*2n143** (*Parhyale hawaiensis otd*2 clone n143), obtained from *otd*2 5' RACE.
Ph_*otd*2n143 consists of consensus 5' splicing units 5'D, 5'B and 5'A. Start of open reading frame is marked as ATG.

```
  1                                                             gg   2

  3  agcggggcccttctagtcgagtgtggtggtggtgatgatggttagtggtggcagtattgt   62

 63  taccagcgccgctgcctatgagcgacgagttcgtaattgtgttgtaaggggttgatggag  122

123  gcagagacagggatgaccccgagctaacggacggagggggtgacggtggggacgcgtctc  182

183  tagctccatgggacgtcgtcaaaggcgacgtcagggttgttgagctgttactgctgttac  242

243  tgttattgttgttattattgttgggggtggtgtgatggtggggtcgcattccccgggttc  302

303  gggggcagttgcggggctggggggtcagcagcagcagccgctggggtctatctcaaaacc  362

                     M  N  G  L  G  L  G  V  G  M  G  V  D  S  L
363  cccgccccctacaccATGAACGGACTTGGACTCGGTGTCGGCATGGGGGTAGACAGCCTG  422

      H  P  S  M  G  Y  P  P  P  T  D  Y  L  Y  T  A  T  P  G  R
423  CACCCAAGCATGGGCTATCCTCCCCCTACCGATTATTTATACACGGCCACTCCAGGCCGG  482

      K  Q  R  R  E  R  T  T  Y  T  R  A  Q  L  D  I  L  E  T  L
483  AAGCAGCGGCGGGAGCGCACTACGTACACCAGAGCGCAACTGGACATCCTGGAGACCCTG  542
```

DNA-sequence of **Ph_*otd*2n164** (*Parhyale hawaiensis otd*2 clone n164), obtained from *otd*2 5' RACE. Ph_*otd*2n164 consists of consensus 5' splicing units 5'D, 5'B and 5'A. Start of open reading frame is marked as ATG.

```
  1                                  gattgaaaactttgccgtggtgtctctctt   30

 31  tcaattccagcaagttttaatgccgtgggagtcacatctaagcttctccctaaagttgca   90

 91  atcaagaaaaaacacacacaccaccaccttagtgcaagcttgcttgtgttaactcatata  150

151  gtgcgtcgcttgaagtgcccaaagcaggtacaaagttattcatctctacgtacgtgctct  210

211  gatctgcctcatctatccagcagccgcggccgaaacccagaggcctgcacagcggaccca  270

         M  W  A  S  A  Q  G  G  G  G  G  G  G  V  A  F  P  G  F
271  gtaATGTGGGCCTCTGCTCAGGGTGGCGGTGGGGGAGGCGGGGTCGCATTCCCCGGGTTC  330

      G  G  S  C  G  A  G  G  S  A  A  A  A  A  G  V  Y  L  K  T
331  GGGGGCAGTTGCGGGGCTGGGGGGTCAGCAGCAGCAGCCGCTGGGGTCTATCTCAAAACC  390

      P  A  P  Y  T  M  N  G  L  G  L  G  V  G  M  G  V  D  S  L
391  CCCGCCCCCTACACCATGAACGGACTTGGACTCGGTGTCGGCATGGGGGTAGACAGCCTG  450

      H  P  S  M  G  Y  P  P  P  T  D  Y  L  Y  T  A  T  P  G  R
451  CACCCAAGCATGGGCTATCCTCCCCCTACCGATTATTTATACACGGCCACTCCAGGCCGG  510

      K  Q  R  R  E  R  T  T  Y  T  R  A  Q  L  D  I  L  E  T  L
511  AAGCAGCGGCGGGAGCGCACTACGTACACCAGAGCGCAACTGGACATCCTGGAGACCCTG  570
```

DNA-sequence of **Ph_*otd*2n167** (*Parhyale hawaiensis otd*2 clone n167), obtained from *otd*2 5' RACE. Ph_*otd*2n167 consists of consensus 5' splicing units 5'C, 5'B and 5'A. Start of open reading frame is marked as ATG.

```
                                             M  P  W  E  S  H  L
  1                       gaattccagcaagttttaATGCCGTGGGAGTCACATCTA    39

      S  F  S  L  K  L  Q  S  R  K  K  H  T  H  H  H  L  S  A  S
 40  AGCTTCTCCCTAAAGTTGCAATCAAGAAAAAAACACACACACCACCACCTTAGTGCAAGC    99

      L  L  V  L  T  H  I  V  R  R  L  K  C  P  K  Q  V  Q  S  Y
100  TTGCTTGTGTTAACTCATATAGTGCGTCGCTTGAAGTGTCCAAAGCAGGTACAAAGTTAT  159

      S  S  L  R  T  C  S  D  L  P  H  L  S  S  S  R  G  R  N  P
160  TCATCTCTACGTACGTGCTCTGATCTGCCTCATCTATCCAGCAGCCGCGGCCGAAACCCA  219

      E  A  C  T  A  D  P  V  M  W  A  S  A  Q  A  A  T  P  G  R
220  GAGGCCTGCACAGCGGACCCAGTAATGTGGGCCTCTGCTCAGGCGGCCACTCCAGGCCGG  279

      K  Q  R  R  E  R  T  T  Y  T  R  A  Q  L  D  I  L  E  T  L
280  AAGCAGCGGCGGGAGCGCACTACGTACACCAGAGCGCAACTGGACATCCTGGAGACCCTG  339
```

DNA-sequence of **Ph_*otd*2n171** (*Parhyale hawaiensis otd*2 clone n171, exemplarily for clones n171 and S32), obtained from *otd*2 5' RACE. Ph_*otd*2n171 consists of consensus 5' splicing units 5'C and 5'A. Start of open reading frame is marked as ATG.

```
  1                     gagtgcgtcgcttgaagtgtccaaagcaggtacaaagttat   41

 42  tcatctctacgtacgtgctctgatctgcctcatctatccagcagccgcggccgaaaccca  101

                              M  W  A  S  A  Q  A  A  T  P  G  R
102  gaggcctgcacagcggacccagtaATGTGGGCCTCTGCTCAGGCGGCCACTCCAGGCCGG  161

      K  Q  R  R  E  R  T  T  Y  T  R  A  Q  L  D  I  L  E  T  L
162  AAGCAGCGGCGGGAGCGCACTACGTACACCAGAGCGCAACTGGACATCCTGGAGACCCTG  221
```

DNA-sequence of **Ph_*otd*2n169** (*Parhyale hawaiensis otd*2 clone n169), obtained from *otd*2 5' RACE. Ph_*otd*2n169 consists of consensus 5' splicing units 5'C and 5'A. Start of open reading frame is marked as ATG.

```
                                   R  A  Q  L  D  I  L  E  T  L
 1                               CAGAGCGCAACTGGACATCCTGGAGACCCTG   31

     F  G  K  T  R  Y  P  D  I  F  M  R  E  E  V  A  I  K  I  N
32  TTCGGCAAGACGCGGTATCCCGACATCTTTATGAGAGAGGAAGTTGCGATCAAGATCAAT   91

     L  P  E  S  R  I
92  CTCCCTGAGAGTAGAATA                                            109
```

DNA-sequence of **Ph_*otd*2a104** (*Parhyale hawaiensis otd*2 clone a104, exemplary for clones a104 and a105), obtained from *otd*2 "fishing" PCR using degenerate primers. Ph_*otd*2a104 consists of a part of the *otd*2 homeobox. Note that the *otd*2 homeobox sequence obtained from "fishing" PCR overlaps with 5'A and 3'A sequences.

```
         E  V  A  I  K  I  N  L  P  E  S  R  I  Q  V  W  F  K  N  R
    1  GGAAGTTGCGATCAAGATCAATCTCCCTGAGAGTAGAATACAGGTTTGGTTTAAGAATCG    60

         R  A  K  C  R  Q  Q  Q  K  Q  Q  A  A  N  G  E  K  A  P  R
   61  TAGAGCAAAGTGCCGGCAGCAACAGAAGCAGCAGGCCGCCAACGGCGAGAAGGCTCCACG   120

         T  K  K  V  N  K  S  P  P  P  S  T  N  N  T  I  T  N  S  S
  121  CACCAAGAAAGTGAACAAATCTCCTCCTCCATCCACCAACAACACAATTACGAACTCGTC   180

         L  I  G  S  G  A  G  N  N  A  A  T  T  N  H  H  H  H  H  H
  181  GCTCATAGGCAGCGGCGCTGGTAACAATGCTGCCACCACTAACCACCATCACCACCATCA   240

         T  T  P  N  N  N  N  N  N  N  S  N  S  S  N  S  S  T  T  L
  241  CACCACCCCCAACAATAATAACAACAACAACAGTAACAGCAGTAACAGCTCAACAACTCT   300

         T  S  P  L  T  T  S  H  G  A  R  D  A  S  P  P  S  P  P  P
  301  GACGTCGCCTCTGACGACGTCCCATGGAGCTAGAGACGCGTCCCCACCGTCACCCCCTCC   360

         S  V  S  S  G  S  S  L  S  L  P  P  S  T  P  Y  N  P  I  W
  361  GTCCGTTAGCTCGGGGTCATCCCTGTCTCTGCCTCCATCAACCCCTTACAACCCCATTTG   420

         S  P  A  G  I  P  H  Q  S  S  T  L  P  A  M  G  D  L  M  G
  421  GTCTCCAGCTGGGATACCCCACCAGTCCTCCACGTTACCCGCAATGGGGGACTTGATGGG   480

         G  G  G  L  D  R  G  G  Y  G  S  A  C  Y  Q  G  Y  G  T  P
  481  TGGTGGTGGGTTGGACAGAGGAGGGTACGGGAGCGCTTGCTACCAGGGCTACGGTACTCC   540

         S  Y  Y  S  N  M  D  Y  L  A  P  M  T  H  S  Q  M  S  S  M
  541  ATCCTATTACTCGAACATGGACTACCTCGCACCTATGACCCATTCTCAGATGAGTTCAAT   600

         T  S  L  N  P  M  T  S  A  H  H  Q  H  L  G  A  H  H  H  S
  601  GACCTCATTGAACCCTATGACCTCGGCTCACCACCAGCATCTTGGTGCTCATCACCACTC   660

         A  A  A  G  L  S  P  S  P  L  G  S  S  L  G  S  S  P  L  S
  661  AGCGGCTGCGGGTCTGAGTCCCTCACCACTCGGTTCATCACTGGGCTCCTCCCCTCTGAG   720

         S  S  H  H  L  V  S  S  L  S  S  T  S  L  N  S  G  S  H  A
  721  CTCATCACATCACCTCGTCTCATCACTAAGCTCAACATCCCTCAACAGCGGTAGTCACGC   780

         A  L  T  H  P  P  L  T  H  S  S  S  L  N  Q  D  S  P  R  S
  781  TGCCCTCACTCACCCTCCCCTCACCCATTCTTCGTCTTTGAATCAAGACAGCCCTCGTAG   840

         P  S  G  A  P  G  S  L  T  P  L  P  Q  D  C  L  D  G  Y  P
  841  TCCGTCTGGAGCCCCAGGGTCGCTTACCCCTCTGCCCCAGGACTGCCTGGACGGTTATCC   900

         E  D  K  S  A  I  I  A  P  N  S  A  A  A  A  S  A  A  A  A
  901  AGAAGATAAGTCAGCCATTATTGCACCGAACTCCGCGGCCGCAGCATCAGCGGCAGCGGC   960

         W  K  N  Y  Q  G  F  Q  S  L Stop
  961  GTGGAAGAACTACCAAGGATTTCAGAGTTTGTGAagtaacgaaggtcacgtaatagtgac  1020

 1021  gaattctttagccatgaaaacaatgtatagtttgtaattattttatgatgtttcatgaac  1080

 1081  gcccgcttactcttgaagagtttctgtaaaactattgcgatttttttccagtaatattagc 1140

 1141  ttatgactgttcaaatattatcgaacagtaattaatgtggctgatattcaatacgaataa  1200

 1201  tttttgtagatgaacttgggaagcaacacagtgtaacctacacgatcacggaggacagtt  1260

 1261  ttccatatggtaggctaaatttggaaatctaacgcggatctgggatacgtcacgtggaac  1320

 1321  acgagggaaattcattagccttagccatatataacgtccaggggggtctaatagggaacgt  1380

 1381  ttcatattaacaatgctcttactccgggaaacggaagtttcaattggctactacatagtt  1440
```

```
1441  aaaacaaatccaacttgaaagtaaatctcatgagtgcctccgtgctacttcaatgtgaag  1500

1501  ttagcacttttggtattgaaaaataactgttggatgccactacacagtgcttagagcgcc  1560

1561  gacactcttaattactttggtgttggtgtaactgtacatgcaataacaataagtacatca  1620

1621  gaaatacgctaatgggagggaaaacttgtcactgattctactacagccctcaccacagta  1680

1681  tgtgcttagcaatgttattatttttgtttttacatacaattccaatagttccctgttgtt  1740

1741  atagagttggcctggaacatacttttgctgcaagaaattcttcattttcaaaatagcgtg  1800

1801  aacaagattacttataccacgatgaaagcttagcacttgaaaatacatcgttccacatca  1860

1861  aggatctattggcgagacccttctttggtcacaacgtgcgctgacaaaacataacgtgaa  1920

1921  gatgagatgagagacagaatcgtgagtggcttggttgaccgttaagagttataagagtcc  1980

1981  ccatgaaaacagcagcagccttgcgtgatggactcaggccctagtgctgctaggtgcgat  2040

2041  acgaaactaggcatgcaagctatttgcgacactaggagcactaaatccgcggagataggt  2100

2101  atgcactgctgccctggaggcacttaccttcttcagaggtggggagagtaggggtgctgt  2160

2161  tcgttgaaccttccgtcgatgacgctatagttcttgatgacttatgaccgactgcagccg  2220

2221  tacttgctcctctggggcaagttggtttatgtctgcagcacgaaccagtgcctgtggatt  2280

2281  tcccagttaggggtggggtcaagttgactttacctttgcccgcagtggggcatgggagag  2340

2341  gccatgcataggctatagggttaatttagtcaaattcgctcatactatgttatggggagc  2400

2401  agcgccttcaggggaactgtggaggttaccttaactatcagttagtctttcgggacgtct  2460

2461  gtccgagcccactgggcttcttgtctatttcatacttctacctgatgttactttcgtctg  2520

2521  tttttccttgccgatatgctctctctgagcgttgatcctgtgtgggaacacggtcgtgca  2580

2581  tggatcgattccacgtttatcttagcttacaatatgtcccttccttggccgctcttagaa  2640

2641  cagctttgtaccgatttagggcggcttaaactatcttatatacctatacctagttaagca  2700

2701  agttcccagccttgttaccagatggaaagtgaaagctactctgctacctaactcttcagc  2760

2761  ttactttccagtctccgtcacggtccaggaggctagtgccctaccaggaataaaaggtgc  2820

2821  gtaacgttgcgtacgtccagtagatgtggtggcgggttgcgtacgtccagtaggtgtggc  2880

2881  ggcaggttctgaccatgcacttcaccccaccctgtgtatcagttaggctgtactactcat  2940

2941  tcggatgtgggaggaggttgcgagaaaattgttgtggaaatactggagacactaaagtga  3000

3001  tgtattaacggcagaaatgtcaaacgctgaactctttagatactgacggtgtcttctgct  3060

3061  aaaatcagtttctttttcttcatcagcagcgactggtttccttctttgatgtttctgat  3120

3121  ggcttctctaatggccaacccgaaaccgttgacaaattttgctatttattgcagctggta  3180

3181  tcggcacaatacgagttattgtaataaaatagatgcattggctctgtaataccaaaaaaa  3240

3241  aaaaaaaaaaaaaaa                                               3255
```

DNA-sequence of **Ph_*otd*2n81** (*Parhyale hawaiensis otd*2 clone n81, exemplary for clones n81 and n83), obtained from *otd*2 3' RACE. Ph_*otd*2n81 consists of consensus 3' splicing units 3'A, 3'B and 3'C. Stop of open reading frame is marked as TGA.

```
                       I  K  I  N  L  P  E  S  R  I  Q  V  W  F  K
   1                  ATCAAGATCAATCTCCCTGAGAGTAGAATACAGGTTTGGTTTAA    60

        N  R  R  A  K  C  R  Q  Q  Q  K  Q  Q  A  A  N  G  E  K  A
  61  GAATCGTAGAGCAAAGTGCCGGCAGCAACAGAAGCAGCAGGCCGCCAACGGCGAGAAGGC   120

        P  R  T  K  K  V  N  K  S  P  P  P  S  T  N  N  T  I  T  N
 121  TCCACGCACCAAGAAAGTGAACAAATCTCCTCCTCCATCCACCAACAACACAATTACAAA   180

        S  S  L  I  G  S  G  A  G  N  N  A  A  T  T  N  H  H  H  H
 181  CTCGTCGCTCATAGGCAGCGGCGCTGGTAACAATGCTGCCACCACTAACCATCATCACCA   240

        H  H  T  T  P  N  N  N  N  N  N  N  S  N  S  S  N  S  S  T
 241  CCATCACACCACCCCCAACAATAATAACAACAATAACAGTAACAGCAGTAACAGCTCAAC   300

        T  L  T  S  P  L  T  T  S  H  G  A  R  D  A  S  P  P  S  P
 301  AACCCTGACGTCGCCTCTGACGACGTCCCATGGAGCTAGAGACGCGTCCCCACCGTCACC   360

        P  P  S  V  S  S  G  S  S  L  S  L  P  P  S  T  P  Y  N  P
 361  CCCTCCGTCCGTTAGCTCGGGGTCATCCCTGTCTCTGCCTCCATCAACCCCTTACAACCC   420

        I  W  S  P  A  G  I  P  H  Q  S  S  T  L  P  A  M  G  D  L
 421  CATTTGGTCTCCAGCGGGGATACCCCACCAGTCCTCCACGTTACCTGCTATGGGGGACCT   480

        M  G  G  G  G  L  D  R  G  G  Y  G  S  A  C  Y  Q  G  Y  G
 481  GATGGGTGGTGGTGGGTTGGACAGAGGAGGGTACGGGAGCGCTTGCTACCAGGGCTACGG   540

        T  P  S  Y  Y  S  N  M  D  Y  L  A  P  M  T  H  S  Q  V  N
 541  TACTCCATCCTATTACTCGAACATGGACTACCTCGCACCTATGACCCATTCTCAGGTAAA   600

        T  F  A  H  N  I  D  L  S  I  L  I  M  L  T  C  I  C  K Stop
 601  TACCTTTGCTCATAATATAGATCTTTCCATATTGATAATGCTAACGTGTATTTGCAAATA   660

 661  Ataacaatttaagtttcaatattttagcacacgcacgttcataaccctcagtgcctttct   720

 721  taaaaatgtgatcagttgtaatgcgttgcaatatgaaaaagataggtaataaattgtccc   780

 781  tattttgtttctgatatctttcacgtatcagtataaacttaaaatgctccctctgctgcg   840

 841  aaatcgttgttccactccaaaaatttcattctgatttagcagtttcatctcagccaaatt   900

 901  attttactgtttacattaacaggttagtgcgtatcattacgtacttcagccgtggccatt   960

 961  actaacaaacctgttacgtacacagcaaataatttggtttatatcgaactgctataatac  1020

1021  agaatgaaattttggctgtagaacaactattttggaacagaggggtataaaggaatgg  1080

1081  cacgtggacctcattgcattacatttattgcaaagttgactatggagataaactctgaaa  1140

1141  cgttttcgtctaagaaaaaaaaaaaaaaaaaa                              1172
```

DNA-sequence of *Parhyale hawaiensis otd*2 clone **web422**, obtained from *otd*2 3' RACE. **web422** consists of consensus 3' splicing units 3'A, 3'B and 3'D. Stop of open reading frame is marked as TAA.

```
                I  K  I  N  L  P  E  S  R  I  Q  V  W  F  K  N  R
   1            ATCAAGATCAATCTCCCTGAGAGTAGAATACAGGTTTGGTTTAAGAATCG     60

        R  A  K  C  R  Q  Q  Q  K  Q  Q  A  A  N  G  E  K  A  P  R
  61  TAGAGCAAAGTGCCGGCAGCAACAGAAGCAGCAGGCCGCCAACGGCGAGAAGGCTCCACG    120

        T  K  K  V  N  K  S  P  P  P  S  T  N  N  T  I  T  N  S  S
 121  CACCAAGAAAGTGAACAAATCTCCTCCTCCATCCACCAACAACACAATTACGAACTCGTC    180

        L  I  G  S  G  A  G  N  N  A  A  T  I  T  T  T  T  I  I  T
 181  GCTCATAGGCAGCGGCGCTGGTAACAATGCTGCCACCATCACCACCACCACCATCATCAC    240

        I  T  I  K  V  N  R  K  R  A  N  I  T  I  K  R  K  T  E  S
 241  CATCACCATAAAAGTAAATCGAAAGAGGGCAAACATCACCATAAAAAGAAAGACAGAAAG    300

        T  V  Q  T  R  K  T  R  V  R  S  S  R  W  R  R  F  T  T  A
 301  CACAGTTCAGACAAGGAAAACGCGCGTAAGGTCCTCAAGGTGGAGGAGGTTTACGACTGC    360

        P  K  K  I  E  R  P  A  I  W  A Stop
 361  TCCGAAGAAGATCGAACGTCCCGCGATATGGGCGTGAgcccgacggaagcgcccttcgat    420

 421  gaggaagccagcgacggcgccgccaaagaggttcctgtgtgcccgggcgaggacgagtat    480

 481  gacctccccgcggtgggtcgggacgagaagagtgaccgcgtcagcggcgacatggcgtcg    540

 541  tcagagagcgaagccatgagtcccaccaagaaaaagaagagcaaacataaatcaaagcat    600

 601  cacagtaagcaccacggaaaacataaggagaaagaggggaaagtaaaagacaaagaaggt    660

 661  aaaagtaaagcaaagatgggattgctggtggaaggcaaagcaggcgaagagaagagcaag    720

 721  gttaataaagtcaagctaggcaaggccaaagagaacaaaaagacaagcaaaagcaacaaa    780

 781  aagaaattatctgtgaagaacgttgctgctattgatgccctgtctgtcgccacagagcag    840

 841  acattgaaggacatcacgtttatgctggacgctcctggcgtggcgcggctaagtcctgac    900

 901  gaggagctgcgggaggtgatgcccgatgatctctaccagccgcccgtggtggacgacggc    960

 961  cacagcgcggtgctgcagctgacgccagatgccctttgatagagcaggagtaccggcag   1020

1021  aggctcaagatggatatggtggagcagatgcgccccaccgactgtgctggtgtcgaagtt   1080

1081  actgaagactgcgctgccaaaaatcagcataagaagagaaaacgaaagcgggacgacgtg   1140

1141  gagaccggcgggtgcttcagtgaaggagaagctaagaaaaaaaaaaaaaaaa           1192
```

DNA-sequence of *Parhyale hawaiensis otd*2 clone **web425**, obtained from *otd*2 3' RACE. **web425** consists of consensus 3' splicing units 3'A and 3'E. Stop of open reading frame is marked as TGA.

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.1_Ph_otd1]

### [PhD_A2.1.1_Ph_otd1_sequences]

### **"Ph_otd1_general_read_me.txt"**

```
General aspects:

(i)       All sequences are in FASTA format.

(ii)      [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)     Comments are preceded by an apostrophe (')

(iv)      All sequences without olignonucleotide sequence are indicated (e.g. "w/o 1013, 1016"). In these cases, only novel
Parhyale sequence is presented.


Parhyale hawaiensis otd1 terminology:


(i)       In general, sequence denotation begins with: ">Ph_otd1_"

(ii)      The above (i) is followed by: "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                         "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                         "ORF"               for coherent ORF sequences
                                         "cDNA"              for complete cDNA sequences

                                         "ref"               if reference sequence is presented
                                         "HD"                if homeobox/homeodomain fragment is presented

(iii)     The above (ii) is followed by clone id.



Parhyale hawaiensis otd1 synopsis:

cDNA:  2036 bp
5' UTR: 255 bp (1-255)
ORF:   1191 bp (256-1446), 396 aa
3'UTR:  590 bp (1447-2036)
HD:     180 bp (517-696), 60 aa


>Ph_otd1_HD [bp] 'homeobox
CAACGCAGGGAGCGCACCACCTTCACTAGAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGT
AGAGTGCAGGTTTGGTTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAAC

>Ph_otd1_ref [bp] 'source: Ph_otd1_cDNA25, extended 5' by 30 bp of Ph_otd1_5R01 and 57 bp of Ph_otd1_3R12; nuc exch
(1439G>A,1502G>A,1519 G>A, 1885T>-, 1949 A>G)
AGTTGTGCGAACGTAAGCTTGTCGAATACGTTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCA
GTAATCTCCCTTTGCGCCTGTGCCAGTGAACGTACCTATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGAGACATCTTTGATCATCCACCACCGATG
CACATGAACGGAGATCCCATGGCCTACCATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACC
CATCATCCCCACTCCATGGTGGACCCTACGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCACAACCAGGCCGCAAG
CAACGCAGGGAGCGCACCACCTTCACTAGAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGT
AGAGTGCAGGTTTGGTTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACC
GCTACCAGCAACACAGCCAGCACTGTCGGCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTAGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACC
GCCACCAACACTACATCTTCTGCTGTGCCATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCA
ACCCCTACCCAACCCTATGCTGGGTTCCAGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCG
AACTGTATGCAGACTACGGGACACTACACGCCGATGCCTCCTCATGCCAAGACCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTC
AACACGAGCCACATGACCAACTTTTCTCAGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACC
GCAGCCGAAAGCTTTCACAATCTTTGAACGGCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGAGTTAGCGGATTGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGA
GTTTCTGTTGTGGAGACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCCAAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTA
AGGCGAGGATAGAGGAGCATACCTGCGGAAAGATGAGTACTTGTGTGAACTGCTGTGGCAATTCTGATAACAATCGCTGAACTGGTAATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCA
TGTTCGTTTTTTTGTTATCTACCGGAATGCATGGATGAATTTTTGCATAAAACACTTTTCATTTTTTTTTTTTTTTTGAGAAGGTAAGATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTT
GTTTCGTATAAGTGGACACGAAATTATACTTTCTTTTTTTTTTCCACTGTTACGATATGTTTGACCTTTTCTACTTGCGAGAAAAAAAAAAAAAAAAAAAA

>Ph_otd1_HD [aa] 'homeodomain
QRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKNRRAKCRQQN

>Ph_otd1_ref [aa]
MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPE
SRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTT
PTPTQPYAGFQGAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGG
TAAESFHNL&
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.1_Ph_otd1]

### [PhD_A2.1.1_Ph_otd1_sequences]

**“Ph_otd1_cDNA_sequences.txt”**

```
>Ph_otd1_ref [bp] 'source: Ph_otd1_cDNA25, extended 5' by 30 bp of Ph_otd1_5R01 and 57 bp of Ph_otd1_3R12; nuc exch
(1439G>A,1502G>A,1519 G>A, 1885T>-, 1949 A>G)
AGTTGTGCGAACGTAAGCTTGTCGAATACGTTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCA
GTAATCTCCCTTTGCGCCTGTGCCAGTGAACGTACCTATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGAGACATCTTTGATCATCCACCACCGATG
CACATGAACGGAGATCCCATGGCCTACCATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACC
CATCATCCCCACTCCATGGTGGACCCTACGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCACAACCAGGCCGCAAG
CAACGCAGGGAGCGCACCACCTTCACTAGAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGT
AGAGTGCAGGTTTGGTTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACC
GCTACCAGCAACACAGCCAGCACTGTCGGCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTAGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACC
GCCACCAACACTACATCTTCTGCTGTGCCATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCA
ACCCCTACCCAACCCTATGCTGGGTTCCAGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCG
AACTGTATGCAGACTACGGGACACTACACGCCGATGCCTCCTCATGCCAAGACCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTC
AACACGAGCCACATGACCAACTTTTCTCAGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACC
GCAGCCGAAAGCTTTCACAATCTTTGAACGGCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGAGTTAGCGGATTGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGA
GTTTCTGTTGTGGAGACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCCAAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTA
AGGCGAGGATAGAGGAGCATACCTGCGGAAAGATGAGTACTTGTGTGAACTGCTGTGGCAATTCTGATAACAATCGCTGAACTGGTAATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCA
TGTTCGTTTTTTTGTTATCTACCGGAATGCATGGATGAATTTTTGCATAAAACACTTTTCATTTTTTTTTTTTTTTTGAGAAGGTAAGATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTT
GTTTCGTATAAGTGGACACGAAATTATACTTTCTTTTTTTTTTCCACTGTTACGATATGTTTGACCTTTTCTACTTGCGAGAAAAAAAAAAAAAAAAAAAA

>Ph_otd1_cDNA25 [bp] 'w/o 1013, 1016
TTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCAGTAATCTCCCTTTGCGCCTGTGCCAGTGAA
CGTACCTATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGAGACATCTTTGATCATCCACCACCGATGCACATGAACGGAGATCCCATGGCCTACCAT
GGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACCCATCATCCCCACTCCATGGTGGACCCTACG
CTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCACAACCAGGCCGCAAGCAACGCAGGGAGCGCACCACCTTCACTAGA
GCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCCGA
GCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCGGC
AGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTAGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACTACATCTTCTGCTGTGCCA
TCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCCAG
GGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACACG
CCGATGCCTCCTCATGCCAAGACCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTCAG
ATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCGAAAGCTTTCACAGTCTTTGAACG
GCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGGGTTAGCGGATTGCAATGGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTGTTGTGGAGACGCTCCGTTGTTTG
TTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCCAAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTAAGGCGAGGATAGAGGAGCATACCTGCGGAA
AGATGAGTACTTGTGTGAACTGCTGTGGCAATTCTGATAACAATCGCTGAACTGGTAATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCATGTTCGTTTTTTTGTTATCTACCGGAATGC
ATGGATGAATTTTTGCATAAAACACTTTTCATTTTTTTTTTTTTTTTTGAGAAGGTAAGATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTTGTTTCGTATAAGTAGACACGAAATTATAC
TTTCTTTTTTTTTTC

>Ph_otd1_cDNA26 [bp] 'w/o 1013, 1016
TTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCAGTAATCTCCCTTTGCGCCTGTGCCAGTGAA
CGTACCTATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGTGACATCTTTGATCATCCACCACCGATGCACATGAACGGAGATCCCATGGCCTACCAT
GGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCAAACGCCACAGCTGGGTTGCCCATGGGCATGGCAACCCGTCATCCCCACTCCATGGTGGACCCTACG
CTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCGCAACCAGGTCGCAAGCAACGCAGGGAGCGCACCACCTTCACTAGA
GCTCAGCTCGACGTACTAGAAACCCTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAGATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCCGA
GCGAAGTGCCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCGGC
AGTGTGGCTTCTGGCACCAGCACGACGGACAGTACTCGGGTCGCCGCTCCTAGGGCCAAGAAATCCACAGCCAAACCCAGCACCCCACCTGTAGCTACCGCCACTAACACCACATCTTCTGCTGTGCCA
TCCACGGCCAAATCTGAACCCAGGACCCAATCCCCCGCGGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCCAG
GGAGCTGCCCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGATCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACACG
CCGATGCCTCCTCATGCCAAGACCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGCAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTCAG
ATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACGGCAGCCCAGCAGTACAGAGTGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCGAAAGCTTTCACAATCTTTGAACG
GCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGAGTTAGCAGATCGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTGTTGTGGAGACGCTCCGTTGTTTG
TTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCCAAGTTCAGCAGCAGTAATGCGCTGTAAAACAATAAAAGGTTAAGGCGAGGATAGAGGAGCATACCTGCGGAA
AGATGAGTACTTGTGTGAACTGCTGTGGTAATTCTGATAACAATCGCTGAACTCGAAATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCATGTTCGTTTTTTGTTATCTACCGCAATGCA
TGGATGTATTTTTGCATAAAACACTTTTCATTTTTTTTTTGGAGAAGGTAAGATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTTGTTTCGTGTAAGTGGACACGAAATTATACTTTCTTT
TTTTTC

>Ph_otd1_cDNA27 [bp] 'w/o 1013, 1016
TTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCAGTAATCTCCCTTTGCGCCTGTGCCAGTGAA
CGTACCTATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGAGACATCCTTGATCATCCACCACCGATGCACATGAACGGAGATCCCATGGCCTACCAT
GGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACCCATCATCCCCACTCCATGGTGGACCCTACG
CTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAAGTGAGTGCACAGCAGTGCCACCACAACCAGGCCGCAAGCAACGCAGGGAGCGC
ACCACCTTCACTAGAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGG
TTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACA
GCCAGCACTGTCGGCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTAGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACTACA
TCTTCTGCTGTGCCATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCC
TATGCTGGGTTCCAGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACT
ACGGGACACTACACGCCGATGCCTCCTCATGCCAAGACCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATG
ACCAACTTTTCTCAGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCGAAAGCTTT
CACAATCTTTGAACGGCATTCTTATAGATACGCCCTTCAAATATTCTGTTTATTACTTGCTGTTTAGAGTTAGCAGATCGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTGTTGTGGAG
ACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCCAAGTTCAGCAGCAGTAATGCGCTGTAAAACAATAAAAGGTTAAGGCGAGGATAGAGG
AGCATACCTGCGGAAAGATGAGTACTTGTGTGAACTGCTGTGGTAATTCTGATAACAATCGCTGAACTCGAAATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCATGTTCGTTTTTTGTT
ATCTACCGCAATGCATGGATGTATTTTTGCATAAAACACTTTTCATTTTTTTTTGGAGAAGGTAAGATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTTGTTTCGTGTAAGTGGACACGAA
ATTATACTTTCTTTTTTTTC

>Ph_otd1_cDNA29 [bp] 'w/o 1013, 1016, id to 32
```

```
TTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCAGTAATCTCCCTTTGCGCCTGTGCCAGTGAA
CGTACCTATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGTGACATCTTTGATCATCCACCACCGATGCACATGAACGGAGATCCCATGGCCTACCAT
GGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCAAACGCCACAGCTGGGTTGCCCATGGGCATGGCAACCCATCATCCCCACTCCATGGTGGACCCTACG
CTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAACTCTTAACGACAATTACTACGAAGAAAACTCACAGACGCTTACCGAAAATTCA
GAAGCCTATGCTAATCATACCCAACAAAATTACCCGCAGTACCGGTACTTTCCATACCCTCCCCAAATAACTCAACAGCTGCCACATAACCAGCAGTGGAATCAGTTTACTGGCGGTTCTTCAAAATCA
GGCTATGAACCTGTTTCATACAATGAATTTAAAATGCCACCGCAACCAGGTCGCAAGCAACGCAGGGAGCGCACCACCTTCACTAGAGCTCAGCTCGACGTACTAGAAACCCTGTTCGGCAAGACGCGG
TATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAGATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGC
AACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCGGCAGTGTGGCTTCTGGCACCAGCACGACGGACAGTACTCGGGCC
GCCGCTCCTGGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACCACATCTTCTGCTGTGCCATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCGGCA
TACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCCAGGGAGCTGCCCATACCCTTTCGACGAACGTGGTCGGCCCCAAC
GCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACACGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAG
AGTTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTCAGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCA
GCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCGAAAGCTTTCACAATCTTTGAACGGCATTCTTATAGATACGCCCTTCAAGTATTTTGTTTATTACT
TGCTGTTTAGCGTTAGCGGATCGCAATAGGGATTGGATTGCTTGGTGTGTTGATGGTTTAGAGTTTCTGTTGTGGAGACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCG
ACATGACTTTGTGCTGCTCCAAGTTCTGCTGCAGCAATGCGCTGTAAAACAATAAAAGGTTAAAGCGATGATAGAGGAGCATACTTGTGGAAATATGAGTACTTGTGTGAACTGCTGTGGTAATTTGGA
AATTTGGATAACAGTCGCTGAGCTGGTAATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCATGTTCGTTTTTTGTTATCTACCGCAATGCATGGATGTATTTCTGCATAAAACACTTTTC
ATTTTTTTTTTGGAGAAGGTAAGATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTTGTTTCGTGTAAGTGGACACAAAATCATACTTTCTTTTTTTTC

>Ph_otd1_cDNA34 [bp] 'w/o 1013, 1016
TTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCAGTAATCTCCCTTTGCGCCTGTGCCAGTGAA
CGTACCTATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGAGACATCCTTGATCATCCACCACCGATGCACATGAACGGAGATCCCATGGCCTACCAT
GGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACCCATCATCCCCACTCCATGGTGGACCCTACG
CTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAACTCATAACCCAAATTTCTACGAAGAAAACTCACAGACGCCTACGGGAAATTCA
GAAGGCTATGCTAATCATACCCAACAAAATTACCCGAAGTACCAGTACTTTCCATACCCTCCCCAAGTAACTCAACAGCTGCCACATAACCAGCAGTGGAATCAGTTTACTGGCGGTTCTTCAAAATTA
GGTTACGAACCTATAACTTACAATGAATTTAAAATGCCACCACAACCAGGCCGCAAGCAACGCAGGGAGCGCACCACCTTCACTAGAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGG
TATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGC
AACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCGGCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCC
GCCGCTCCTGGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACTACATCTTCTGCTGTGCCATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCA
TACAAACCCCCAACCACCTCTCCTGCGTCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCCAGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCACC
GCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACTGCGGGACACTACACGCCGATGCCTCCTCATACCAAGACCCCCCCTTCATGGAACCAG
AGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTCAGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCA
GCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCGAAAGCTTTCACAATCTTTGAACGGCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACT
TGCTGTTTAGAGTTAGCGGATTGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTGTTGTGGAGACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATG
ACTTTGTGCTGCTCCAAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTAAGGCGAGGATAGAGGAGCATACCTGCGGAAAGATGAGTACTTGTGTGAACTGCTGTGGCAATTCTGATAACA
ATCGCTGAACTGGTAATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCATGTTCGTTTTTTTGTTATCTACCGGAATGCATGGATGAATTTTTGCATAAAACACTTTTCATTTTTTTTTTT
TTTTTGAGAAGGTAAGATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTTGTTTCGTATAAGTGGACACGAAATTATACTTTCTTTTTTTTTTC

>Ph_otd1_ORF17 [bp] 'w/o 1014, 1097
ATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCAAACGCCACAGCTGGGTTGCCCATGGGCATGGCAACCCATCATCCCCACTCCATGGTGGACCCTA
CGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCGCAACCAGGTCGCAAGCAACGCAGGGAGCGCACCACCTTCACTA
GAGCTCAGCTCGACGTACTAGAAACCCTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAGATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCC
GAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCG
GCAGTGTGGCTTCTGGCACCAGCACGACGGACAGTACTCGGGCCGCCGCTCCTGGGGCCAAGAAATCCACAGCCAAACCCAGCACCCCACCTGTAGCTACCGCCACTAACACCACATCTTCTGCTGTGC
CATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCGGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCC
AGGGAGCTGCCCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGATCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACA
CGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGCAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAGCTTTTCTC
AGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACGGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTAC

>Ph_otd1_ORF21 [bp] 'w/o 1014, 1097
ATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACCCATCATCCCCACTCCATGGTGGACCCTA
CGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCACAACCAGGCCGCAAGCAACGCAGGGAGCGCACCACCTTCACTA
GAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCC
GAGCGAAGTGTCGCCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCG
GCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTGGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACTACATCTTCTGCTGTGC
CATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCC
AGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTCCCAGCCTATGGCAATCGACAGCTTTACGGTCGGGCCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACA
CGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTC
AGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTAC

>Ph_otd1_ORF201 [bp] 'w/o 1014, 1097
ATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCAAACGCCACAGCTGGGTTGCCCATGGGCATGGCAACTCATCATCCCCACTCCATGGTGGACCCCA
CGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCGCAACCAGGTCGCAAGCAACGCAGGGAGCGCACCACCTTCACTG
GAGCTCAGCTCGACGTACTAGAAACCCTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAGATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCC
GAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCG
GCAGTGTGGCTTCTGGCACCAGCACGACGGACAGTACTCGGGCCGCCGCTCCTAGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACCACATCTTCTGCTGTGC
CATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCGGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCC
AGGGAGCTGCCCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACA
CACCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGTTACACCCCCTCATACTACGGTAACATTGAGTACCTTCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTC
AGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTAC

>Ph_otd1_ORFs01 [bp] 'w/o 1014, 1097
ATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACCCGTCATCCCCACTCCATGGTGGACCCTA
CGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCACAACCAGGCCGCAAGCAACGCAGGGAGCGCACCACCTTCACTA
GAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCC
GAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGCAGCAGCGACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCG
GCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTGGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACTACATCTTCTGCTGTGC
CATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCC
AGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACA
CGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTC
AGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTAC

>Ph_otd1_ORFs02 [bp] 'w/o 1014, 1097
ATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACCCATCATCCCCACTCCATGGTGGACCCTA
CGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCACAACCAGGCCGCAAGCAACGCAGGGAGCGCACCACCTTCACTA
GAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGGACCGCC
GAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAACAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCG
GCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTAGGGCCAAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACTACATCTTCTGCTGTGC
CATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCC
AGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACA
```

```
CGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTC
AGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTAC

>Ph_otd1_ORFs06 [bp] 'w/o 1014, 1097
ATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCAAACGCCACAGCTGGGTTGCCCATGGGCATGGCAACCCATCATCCCCACTCCATGGTGGACCCTA
CGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCGCAACCAGGTCGCAAGCAACGCAGGGAGCGCACCACCTTCACTA
GAGCTCAGCTCGACGTACTAGAAACCCTGTTCGGCAAGACGCGGTATCCCGACATCTTCATAAGAGAGGAAGTTGCGATCAAGATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCC
GAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAACAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCG
GCAGTGTGGCTTCTGGCACCAGCACGACGGACAGTACTCGGGCCGCCGCTCCTAGGGCCAAGAAATCCACAGCCAAACCCAGCACCCCACCTGTAGCTACCGCCACTAACACCACATCTTCTGCCGTGC
TATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCGGCATACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCC
AGGGAGCTGCCCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGATCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACA
CGCCGATGCCTCCTCATGCCATGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGCAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTC
AGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACGGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTAC

>Ph_otd1_5R01 [bp] 'w/o 1005
AGTTGTGCGAACGTAAGCTTGTCGAATACGTTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCA
GTAATCTCCCTTTGCGCCTGTGCCAGTGAACGTACCTATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGATCTTCAGCGCCTCTTTCGAGACATCCTTGATCATCCACCACCGATG
CACATGAACGGAGATCCCATGGCCTACCATGGTTCATTTGTGACTCAGTCACAAACTCTGGATAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCTATGGGCATGGCAACC
CATCATCCCCACTCCATGGTGGACCCTACGCTGCACATGAGCTACCATTCCACAGCAGAGTCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCGCAACCAGGTCGCAAG
CAACGCAGGGAGCGCACCACCTTCACTAGAGCTCAGCTCGACGTGCTGGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCAT

>Ph_otd1_5R03 [bp] 'w/o 1005
AGTTGTGCGAACGTAAGCTTGTCGAATACGTTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCA
GTAATCTCCCTTTGCGCCTGTGCCAGTGAACGTACCTATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGTGACATCTTTGATCATCCACCACCGATG
CACATGAACAGAGATCCCATGGCCTACCATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCAAACGCCACAGCTGGGTTGCCCATGGGCATGGCAACC
CATCATCCCCACTCCATGGTGGACCCTACGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCGCAACCAGGTCGCAAG
CAACGCAGGGAGCGCACCACCTTCACTAGAGCTCAGCTCGACGTACTAGAAACCCTGTTCGGCAAGACGCGGTATCCCGACATCTTCAT

>Ph_otd1_5R04 [bp] 'w/o 1005
AGTTGTGCGAACGTAAGCTTGTCGAATACGTTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGTGATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCA
GTAATCTCCCTTTGCGCCTGTGCCAGTGAACGTACCTATTCCAGAAGCTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGAGACATCCTTGATCATCCACCACCGATG
CACATGAACGGAGATCCCATGGCCTACCATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCCATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACC
CATCATCCCCACTCCATGGTGGACCCTACGCTGCACATGAGCTACCATTCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCACAACCAGGCCGCAAG
CAACGCAGGGAGCGCACCACCTTCACTAGAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCAT

>Ph_otd1_vsdP_QW12 [bp] 'w/o fw_QRR, re_QVW
TAGAGCTCAGCTCGACGTACCAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTG

>Ph_otd1_vsdP_Phad04 [bp] 'w/o fw_QRR, re_QVW
TAGAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCAGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTG

>Ph_otd1_vsdP_Phad05 [bp] 'w/o fw_QRR, re_QVW
TAGAGCTCAGCTCGACGTACTAGAAACCCTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAAGTTGCGATCAAGATCAACCTACCAGAGAGTAGAGTG

>Ph_otd1_3R09 [bp] 'w/o 1000
ATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGC
AGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCGGCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTAGGGCC
AAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACTACATCTTCTGCTGTGCCATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACC
ACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCCAGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGC
CTATGGCAATCGACAGCTTTACGGGCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACACGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCA
TACTACGGTAACATTGAGTACCTTCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTCAGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGA
GCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCGAAAGCTTTCACAATCTTTGAACGGCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGAGTTA
GCGGATTGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTGTTGTGGAGACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCC
AAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTAAGGCGAGGATAGAGGAGCATACCTGCGGAAAGATGAGTACTTGTGTGAACTGCTGTGGCAATTCTGATAACAATCGCTGAACTGGTA
ATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCATGTTCGTTTTTTTGTTATCTACCGGAATGCATGGATGAATTTTTGCATAAAACACTTTTCATTTTTTTTTTTTTTTTTGAGAAGGTAA
GATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTTGTTTCGTATAAGTGGACACGAAATTATACTTTCTTTTTTTTTTTCCACTGTTACGATATGTTTGACCTTTTCTACTTGCGAGAATAA
AAAA

>Ph_otd1_3R11 [bp] 'w/o 1000
ATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGC
AGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCGGCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTAGGGCC
AAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACTACATCTTCTGCTGTGCCATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACC
ACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCCAGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGC
CTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACACGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCA
TACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTCAGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGA
GCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCGAAAGCTTTCACAATCTTTGAATGGCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGAGTTA
GCGGATTGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTGTTGTGGAGACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCC
AAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTAAGGCGAGGATAGAGGAGCATACCTGCGGAAAGATGAGTACTTGTGTGAACTGCTGTGGCAATTCTGATAACAATCGCTGAACTGGTA
ATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCATGTTCGTTTTTTTGTTATCTACCGGAATGCATGGATGAATTTTTGCATAAAACACTTTTCATTTTTTTTTTTTTTTTTGAGAAGGTAA
GATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTTGTTTCGTATAAGTGGACACGAAATTATACTTTCTTTTTTTTTTTCCACTGTTACGATATGTTTGACCTTTTCTACTTGCAAAAAAAAA
AAAAAAAAAAAAAAAAAAAA

>Ph_otd1_3R12 [bp] 'w/o 1000
ATGAGAGAGGAAGTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAACACCAACAACAATAACAACAGCAACAGTAACAACAGC
AGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGCACTGTCGGCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTAGGGCC
AAGAAATCCACAGCCAAACCAAGCACCCCACCTGTAGCTACCGCCACCAACACTACATCTTCTGCTGTGCCATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCATACAAACCCCCAACC
ACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCCAGGGAGCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGC
CTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACACGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCA
TACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTCAGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGA
GCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCGAAAGCTTTCACAATCTTTGAACGGCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGAGTTA
GCGGATTGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTGTTGTGGAGACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCC
AAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTAAGGCGAGGATAGAGGAGCATGCCTGCGGAAAGATGAGTACTTGTGTGAACTGCTGTGGCAATTCTGATAACAATCGCTGAACTGGTA
ATGACCTTTGATTAAATGATGTCCATATTGAGAAGTTTTCCATGTTCGTTTTTTTGTTATCTACCGGAATGCATGGATGAATTTTTGCATAAAACACTTTTCATTTTTTTTTTTTTTTTGAGAAGGTAAG
ATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTTGTTTCGTATAAGTGGACACGAAATTATACTTTCTTTTTTTTTTTCCACTGTTACGATATGTTTGACCTTTTCTACTTGCGAGAAAAAAA
AAAAAAAAAAAAA


>Ph_otd1_ref [aa]
MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPE
SRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTT
```

```
PTPTQPYAGFQGAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGG
TAAESFHNL&

>Ph_otd1_cDNA25 [aa]
MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPE
SRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTT
PTPTQPYAGFQGAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGG
TAAESFHSL&

>Ph_otd1_cDNA26 [aa]
MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATRHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPE
SRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRVAAPRAKKSTAKPSTPPVATATNTTSSAVPSTAKSEPRTQSPAAYKPPTTSPASAHTPPNTT
PTPTQPYAGFQGAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRVAPTKQECMDYGG
TAAESFHNL&

>Ph_otd1_cDNA27 [aa]
MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRSECTAVPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIK
INLPESRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHT
PPNTTPTPTQPYAGFQGAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQEC
MDYGGTAAESFHNL&

>Ph_otd1_cDNA29 [aa]' id to 32
MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRTLNDNYYEENSQTLTENSEAYANHTQQNYPQYRYFPYPPQITQQLPHNQQ
WNQFTGGSSKSGYEPVSYNEFKMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVA
SGTSTTDSTRAAAPGAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTNVVGPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMP
PHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHNL&

>Ph_otd1_cDNA34 [aa]
MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRTHNPNFYEENSQTPTGNSEGYANHTQQNYPKYQYFPYPPQVTQQLPHNQQ
WNQFTGGSSKLGYEPITYNEFKMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVA
SGTSTTDSTRAAAPGAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTNVVDPTAYSSSLWQSTALRSDLGTPNCMQTAGHYTPMP
PHTKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHNL&

>Ph_otd1_ORF17 [bp] 'w/o 1014, 1097
GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKNRR
AKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPGAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQ
GAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTSFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGG

>Ph_otd1_ORF21 [bp] 'w/o 1014, 1097
GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKNRR
AKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPGAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQ
GAAHTLSTNVVDPNAYSPSLWQSTALRSGLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGG

>Ph_otd1_ORF201 [bp] 'w/o 1014, 1097
GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTGAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKNRR
AKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQ
GAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGG

>Ph_otd1_ORFs01 [bp] 'w/o 1014, 1097
GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATRHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKNRR
AKCRQQNTNNNNNSNSNNSSSDNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPGAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQ
GAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGG

>Ph_otd1_ORFs02 [bp] 'w/o 1014, 1097
GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKDRR
AKCRQQNTNNNNNNNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQ
GAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGG

>Ph_otd1_ORFs06 [bp] 'w/o 1014, 1097
GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFIREEVAIKINLPESRVQVWFKNRR
AKCRQQNTNNNNNSNSNNNSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVLSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQ
GAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAMTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYRAAPTKQECMDYGG

>Ph_otd1_5R01 [aa] 'w/o 1005
MHMNGDPMAYHGSFVTQSQTLDNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIF

>Ph_otd1_5R03 [aa] 'w/o 1005
MHMNRDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIF

>Ph_otd1_5R04 [aa] 'w/o 1005
MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLLGMSSFPSLRMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIF

>Ph_otd1_vsdP_QW12 [aa] 'w/o fw_QRR, re_QVW
RAQLDVPETLFGKTRYPDIFMREEVAIKINLPESRV

>Ph_otd1_vsdP_Phad04 [aa] 'w/o fw_QRR, re_QVW
RAQLDVLETLFGKTQYPDIFMREEVAIKINLPESRV

>Ph_otd1_vsdP_Phad05 [aa] 'w/o fw_QRR, re_QVW
RAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRV

>Ph_otd1_3R09 [aa] 'w/o 1000
MREEVAIKINLPESRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPT
TSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTNVVDPNAYSSSLWQSTALRADLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYR
AAPTKQECMDYGGTAAESFHNL&

>Ph_otd1_3R11 [aa] 'w/o 1000
MREEVAIKINLPESRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPT
TSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYR
AAPTKQECMDYGGTAAESFHNL&
```

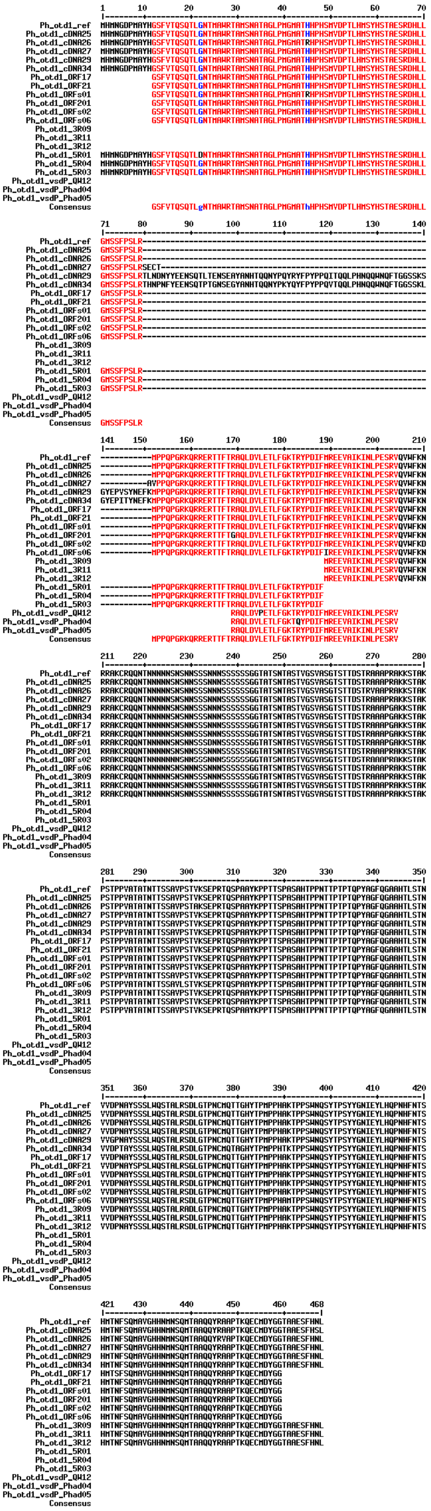
```
>Ph_otd1_3R12 [aa] 'w/o 1000
MREEVAIKINLPESRVQVWFKNRRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAKPSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPT
TSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTNVVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTSHMTNFSQMAVGHHNMNSQMTAAQQYR
AAPTKQECMDYGGTAAESFHNL&
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.1_Ph_otd1]

[PhD_A2.1.2_Ph_otd1_alignments]

## “alignment_Ph_otd1_all_aa_sequences.gif”



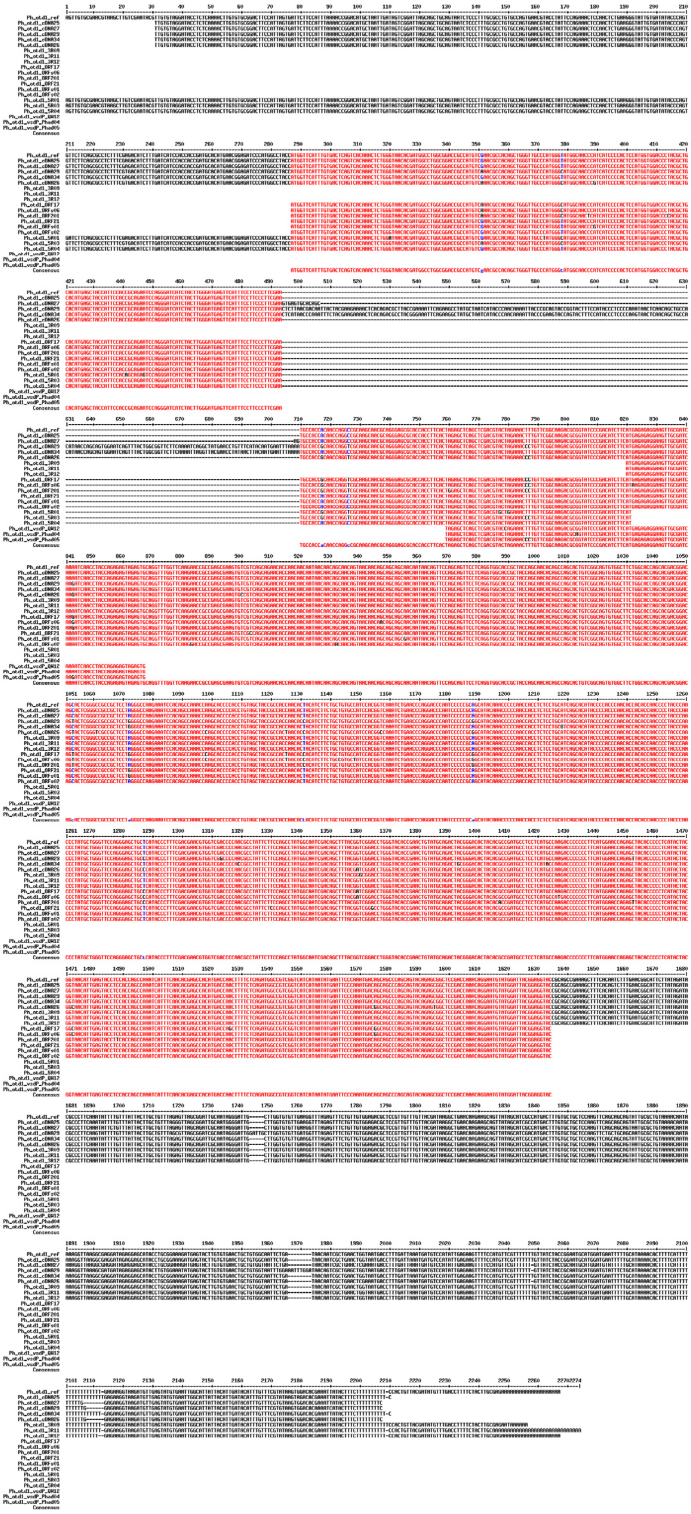
```
                     1       10        20        30        40        50        60        70
                     |--------+---------+---------+---------+---------+---------+---------|
        Ph_otd1_ref  MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
     Ph_otd1_cDNA25  MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
     Ph_otd1_cDNA26  MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATRHPHSMVDPTLHMSYHSTAESRDHLL
     Ph_otd1_cDNA27  MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
     Ph_otd1_cDNA29  MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
     Ph_otd1_cDNA34  MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
      Ph_otd1_ORF17             GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
      Ph_otd1_ORF21             GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
     Ph_otd1_ORFs01             GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATRHPHSMVDPTLHMSYHSTAESRDHLL
     Ph_otd1_ORF201             GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
     Ph_otd1_ORFs02             GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
     Ph_otd1_ORFs06             GSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
       Ph_otd1_3R09
       Ph_otd1_3R11
       Ph_otd1_3R12
       Ph_otd1_5R01  MHMNGDPMAYHGSFVTQSQTLDNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
       Ph_otd1_5R04  MHMNGDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
       Ph_otd1_5R03  MHMNRDPMAYHGSFVTQSQTLGNTMAWRTAMSNATAGLPMGMATHHPHSMVDPTLHMSYHSTAESRDHLL
  Ph_otd1_vsdP_QW12
Ph_otd1_vsdP_Phad04
Ph_otd1_vsdP_Phad05
          Consensus             GSFVTQSQTLgNTMAWRTAMSNATAGLPMGMAThHPHSMVDPTLHMSYHSTAESRDHLL

                     71      80        90        100       110       120       130       140
                     |--------+---------+---------+---------+---------+---------+---------|
        Ph_otd1_ref  GMSSFPSLR-------------------------------------------------------------
     Ph_otd1_cDNA25  GMSSFPSLR-------------------------------------------------------------
     Ph_otd1_cDNA26  GMSSFPSLR-------------------------------------------------------------
     Ph_otd1_cDNA27  GMSSFPSLRSECT---------------------------------------------------------
     Ph_otd1_cDNA29  GMSSFPSLRTLNDNYYEENSQTLTENSEAYANHTQQNYPQYRYFPYPPQITQQLPHNQQWNQFTGGSSKS
     Ph_otd1_cDNA34  GMSSFPSLRTHNPNFYEENSQTPTGNSEGYANHTQQNYPKYQYFPYPPQVTQQLPHNQQWNQFTGGSSKL
      Ph_otd1_ORF17  GMSSFPSLR-------------------------------------------------------------
      Ph_otd1_ORF21  GMSSFPSLR-------------------------------------------------------------
     Ph_otd1_ORFs01  GMSSFPSLR-------------------------------------------------------------
     Ph_otd1_ORF201  GMSSFPSLR-------------------------------------------------------------
     Ph_otd1_ORFs02  GMSSFPSLR-------------------------------------------------------------
     Ph_otd1_ORFs06  GMSSFPSLR-------------------------------------------------------------
       Ph_otd1_3R09
       Ph_otd1_3R11
       Ph_otd1_3R12
       Ph_otd1_5R01  GMSSFPSLR-------------------------------------------------------------
       Ph_otd1_5R04  GMSSFPSLR-------------------------------------------------------------
       Ph_otd1_5R03  GMSSFPSLR-------------------------------------------------------------
  Ph_otd1_vsdP_QW12
Ph_otd1_vsdP_Phad04
Ph_otd1_vsdP_Phad05
          Consensus  GMSSFPSLR

                     141     150       160       170       180       190       200       210
                     |--------+---------+---------+---------+---------+---------+---------|
        Ph_otd1_ref  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
     Ph_otd1_cDNA25  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
     Ph_otd1_cDNA26  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
     Ph_otd1_cDNA27  ----------AVPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
     Ph_otd1_cDNA29  GYEPVSYNEFKMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
     Ph_otd1_cDNA34  GYEPITYNEFKMPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
      Ph_otd1_ORF17  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
      Ph_otd1_ORF21  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
     Ph_otd1_ORFs01  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
     Ph_otd1_ORF201  -----------MPPQPGRKQRRERTTFTGAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKN
     Ph_otd1_ORFs02  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRVQVWFKD
     Ph_otd1_ORFs06  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFIREEVAIKINLPESRVQVWFKN
       Ph_otd1_3R09                                                  MREEVAIKINLPESRVQVWFKN
       Ph_otd1_3R11                                                  MREEVAIKINLPESRVQVWFKN
       Ph_otd1_3R12                                                  MREEVAIKINLPESRVQVWFKN
       Ph_otd1_5R01  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIF
       Ph_otd1_5R04  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIF
       Ph_otd1_5R03  -----------MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIF
  Ph_otd1_vsdP_QW12                              RAQLDVPETLFGKTRYPDIFMREEVAIKINLPESRV
Ph_otd1_vsdP_Phad04                              RAQLDVLETLFGKTQYPDIFMREEVAIKINLPESRV
Ph_otd1_vsdP_Phad05                              RAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRV
          Consensus             MPPQPGRKQRRERTTFTRAQLDVLETLFGKTRYPDIFMREEVAIKINLPESRV

                     211     220       230       240       250       260       270       280
                     |--------+---------+---------+---------+---------+---------+---------|
        Ph_otd1_ref  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAK
     Ph_otd1_cDNA25  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAK
     Ph_otd1_cDNA26  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRVAAPRAKKSTAK
     Ph_otd1_cDNA27  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAK
     Ph_otd1_cDNA29  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPGAKKSTAK
     Ph_otd1_cDNA34  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPGAKKSTAK
      Ph_otd1_ORF17  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPGAKKSTAK
      Ph_otd1_ORF21  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPGAKKSTAK
     Ph_otd1_ORFs01  RRAKCRQQNTNNNNNSNSNNSSSDNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPGAKKSTAK
     Ph_otd1_ORF201  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAK
     Ph_otd1_ORFs02  RRAKCRQQNTNNNNNNNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAK
     Ph_otd1_ORFs06  RRAKCRQQNTNNNNNSNSNNNSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAK
       Ph_otd1_3R09  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAK
       Ph_otd1_3R11  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAK
       Ph_otd1_3R12  RRAKCRQQNTNNNNNSNSNNSSSNNNSSSSSSGGTATSNTASTVGSVASGTSTTDSTRAAAPRAKKSTAK
       Ph_otd1_5R01
       Ph_otd1_5R04
       Ph_otd1_5R03
  Ph_otd1_vsdP_QW12
Ph_otd1_vsdP_Phad04
Ph_otd1_vsdP_Phad05
          Consensus

                     281     290       300       310       320       330       340       350
                     |--------+---------+---------+---------+---------+---------+---------|
        Ph_otd1_ref  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
     Ph_otd1_cDNA25  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
     Ph_otd1_cDNA26  PSTPPVATATNTTSSAVPSTAKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
     Ph_otd1_cDNA27  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
     Ph_otd1_cDNA29  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
     Ph_otd1_cDNA34  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
      Ph_otd1_ORF17  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
      Ph_otd1_ORF21  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
     Ph_otd1_ORFs01  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
     Ph_otd1_ORF201  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
     Ph_otd1_ORFs02  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
     Ph_otd1_ORFs06  PSTPPVATATNTTSSAVLSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
       Ph_otd1_3R09  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
       Ph_otd1_3R11  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
       Ph_otd1_3R12  PSTPPVATATNTTSSAVPSTVKSEPRTQSPAAYKPPTTSPASAHTPPNTTPTPTQPYAGFQGAAHTLSTN
       Ph_otd1_5R01
       Ph_otd1_5R04
       Ph_otd1_5R03
  Ph_otd1_vsdP_QW12
Ph_otd1_vsdP_Phad04
Ph_otd1_vsdP_Phad05
          Consensus

                     351     360       370       380       390       400       410       420
                     |--------+---------+---------+---------+---------+---------+---------|
        Ph_otd1_ref  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
     Ph_otd1_cDNA25  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
     Ph_otd1_cDNA26  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
     Ph_otd1_cDNA27  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
     Ph_otd1_cDNA29  VVGPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
     Ph_otd1_cDNA34  VVDPTAYSSSLWQSTALRSDLGTPNCMQTAGHYTPMPPHTKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
      Ph_otd1_ORF17  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
      Ph_otd1_ORF21  VVDPNAYSPSLWQSTALRSGLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
     Ph_otd1_ORFs01  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
     Ph_otd1_ORF201  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
     Ph_otd1_ORFs02  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
     Ph_otd1_ORFs06  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAMTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
       Ph_otd1_3R09  VVDPNAYSSSLWQSTALRADLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
       Ph_otd1_3R11  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
       Ph_otd1_3R12  VVDPNAYSSSLWQSTALRSDLGTPNCMQTTGHYTPMPPHAKTPPSWNQSYTPSYYGNIEYLHQPNHFNTS
       Ph_otd1_5R01
       Ph_otd1_5R04
       Ph_otd1_5R03
  Ph_otd1_vsdP_QW12
Ph_otd1_vsdP_Phad04
Ph_otd1_vsdP_Phad05
          Consensus

                     421    430       440       450       460     468
                     |--------+---------+---------+---------+-------|
        Ph_otd1_ref  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHNL
     Ph_otd1_cDNA25  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHSL
     Ph_otd1_cDNA26  HMTNFSQMAVGHHNNMNSQMTAAQQYRVAPTKQECMDYGGTAAESFHNL
     Ph_otd1_cDNA27  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHNL
     Ph_otd1_cDNA29  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHNL
     Ph_otd1_cDNA34  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHNL
      Ph_otd1_ORF17  HMTSFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGG
      Ph_otd1_ORF21  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGG
     Ph_otd1_ORFs01  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGG
     Ph_otd1_ORF201  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGG
     Ph_otd1_ORFs02  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGG
     Ph_otd1_ORFs06  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGG
       Ph_otd1_3R09  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHNL
       Ph_otd1_3R11  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHNL
       Ph_otd1_3R12  HMTNFSQMAVGHHNNMNSQMTAAQQYRAAPTKQECMDYGGTAAESFHNL
       Ph_otd1_5R01
       Ph_otd1_5R04
       Ph_otd1_5R03
  Ph_otd1_vsdP_QW12
Ph_otd1_vsdP_Phad04
Ph_otd1_vsdP_Phad05
          Consensus
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.1_Ph_otd1]

### [PhD_A2.1.2_Ph_otd1_alignments]

**“alignment_Ph_otd1_all_bp_sequences.gif”**

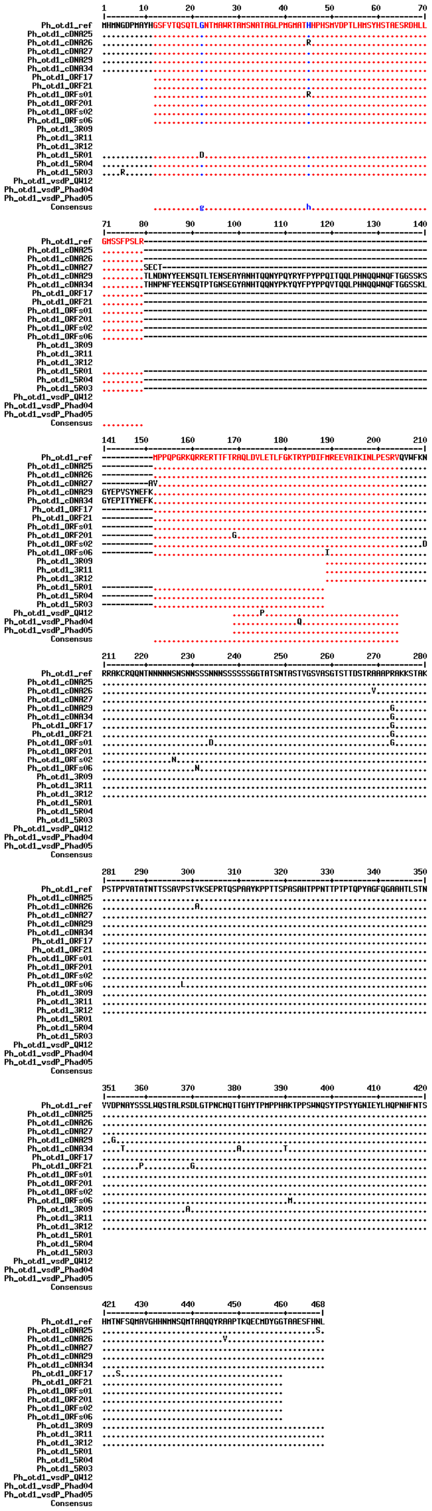


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.1_Ph_otd1]

[PhD_A2.1.2_Ph_otd1_alignments]

## “difference_alignment_Ph_otd1_all_aa_sequences.gif”

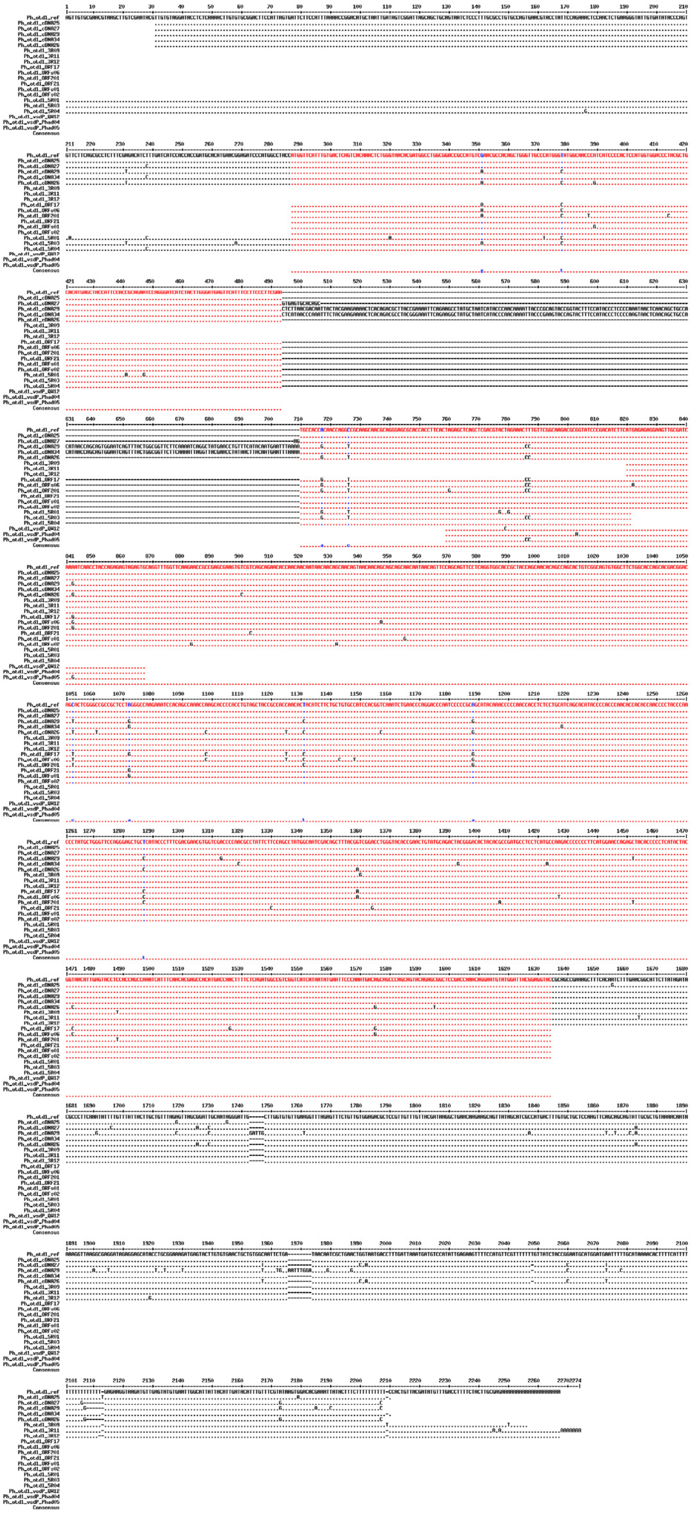


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.1_Ph_otd1]

[PhD_A2.1.2_Ph_otd1_alignments]

## “difference_alignment_Ph_otd1_all_bp_sequences.gif”



[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.2_Ph_otd2]

## “Ph_otd2_read_me.txt”

```
For a detailed list of Parhyale hawaiensis otd2 (Ph otd2) sequences and the corresponding alignments, see 3.2.1.4 as well as
Browne et al., 2006 (supplementary data).
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.3_Ph_hbn]

[PhD_A2.3.1_Ph_hbn_sequences[

## “Ph_hbn_general_read_me.txt”

```
General aspects:

(i)       All sequences are in FASTA format.

(ii)      [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)     Comments are preceded by an apostrophe (')

(iv)      All sequences without olignonucleotide sequence are indicated (e.g. "w/o 1512, 1516"). In these cases, only novel
Parhyale sequence is presented.


Parhyale hawaiensis hbn terminology:


(i)       In general, sequence denotation begins with: ">Ph_hbn_"

(ii)      The above (i) is followed by: "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                        "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                        "ORF"               for coherent ORF sequences
                                        "cDNA"              for complete cDNA sequences

                                        "ref"               if reference sequence is presented
                                        "HD"                if homeobox/homeodomain fragment is presented
                                        "GEH"               if GEH/octapeptide (encoding) fragment is presented

(iii)     The above (ii) is followed by clone id.


Parhyale hawaiensis hbn synopsis:

cDNA:               2651 bp
5'UTR:               171 bp (1-171)
ORF:                1482 bp (172-1653), 493 aa
3'UTR:               998 bp (1654-2651)
HD:                  180 bp (694-873),   60 aa
octapeptide/GEH   24 bp (517-540),    8 aa


>Ph_hbn_HD [bp] 'homeobox
GTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAGTATCCAGACGTCTTCACCCGAGAGGAGTTGGCACTCAACCTTGACCTATCCGAGGCA
CGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAG

>Ph_hbn_GEH [bp] 'encoding octapeptide/GEH domain
TACACCATAGAAAATATACTCGCC

>Ph_hbn_ref [bp] 'source: Ph_hbn_cDNA01, extended 5' 60bp Ph_hbn_5Rn25; 3' 32bp Ph_hbn_3RHH357; exact poly(A) not supported
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAGTCGAGGAGATACTTACCAAAGTGTAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGA
AGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAA
ACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACA
ATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATT
TACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCC
TCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAGTATCCAGAC
GTCTTCACCCGAGAGGAGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCG
TACATTGCTGGCGCTCACAACGATGCAATCATATCCAGTCTGCCAGATATTGTCGGTATACCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCT
TCCACTGCAGATTTCTTTAAGCTGCAGGCGTTCCACGCTCTGCAACTCAATACCTTCCTGCCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTT
CTTCCATCTTCCCTCCCCATAATCCCCAACCATCCACCTATCTTTCCAGCTCTCCCAATATCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACG
GCACCTCAATTTCCCAGCTGCAGTACTTCCTCTGATCTTCCGAACGTTTCCGCACACAATGTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGACCACC
ACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGGCTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCC
GGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAGATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATA
ACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCATGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACA
CGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGTGCTTCAGCGTCCAGTCGATCGTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATT
TACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAAGCCAGTTATTTTGGGCAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGT
TGCAATGTCTGCCAGTTTACGATTCAGTCAGCGAGCTATGTTACGGTGCTCGTGCTCATCCGGCAACAAATAAGTTTGACCATCTACTAGTTTGTCTCTCAATGTAAAACCTTTGTGACGATGTCTAGC
AGCTTGTTTCTCATCCAGCTAAATTATAACTCATCCAACGACCAGTGTTACATAAATCGCATGCGTCGGGTTTTCCTTTATCGTGCCAGCTATGAAACGCTGTCTGCCGCTCATCCAGTCAGATAAATT
GAACCAGTTGACAGACTTGCACCCATCTAGCCCGCCATTTTGAAGGCTCTAATACGAAGCAGATGCTACCTCTTCTCACAATGAAAGGGTTCGTTTTTTAGCGCTTAGCAGCAACCGCTTGTGATCTAA
CGAGCGTTGGAGCGTTGCCACCCATGTTGTACCGAGGGATGGTGTGATAGCACGGTACCAATCAATGGTATATGGTACTAATATAGCGCATAACACCACATAAATTATTAGAGTACACTGTTTTGCACT
AACTTCGTCTTGCTTATCGGCAAAGATTTGTTGTAAGAGATCGTGATCAAGGGATAAACCCCATGACATAC


>Ph_hbn_GEH [aa] 'octapeptide/GEH domain
YTIENILA

>Ph_hbn_HD [aa] 'homeodomain
VRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKRE
```

```
>Ph_hbn_ref [aa]
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSL
PDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLP
NVSAHNVAVINPLTSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.3_Ph_hbn]

[PhD_A2.3.1_Ph_hbn_sequences]

## “Ph_hbn_cDNA_sequences.txt”

```
>Ph_hbn_ref [bp] 'source: Ph_hbn_cDNA01, extended 5' 60bp Ph_hbn_5Rn25; 3' 32bp Ph_hbn_3RHH357; exact poly(A) not supported
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAGTCGAGGAGATACTTACCAAAGTGTAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGA
AGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAA
ACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACA
ATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATT
TACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCC
TCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAGTATCCAGAC
GTCTTCACCCGAGAGGAGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCG
TACATTGCTGGCGCTCACAACGATGCAATCATATCCAGTCTGCCAGATATTGTCGGTATACCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCT
TCCACTGCAGATTTCTTTAAGCTGCAGGCGTTCCACGCTCTGCAACTCAATACCTTCCTGCCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTT
CTTCCATCTTCCCTCCCCATAATCCCCAACCATCCACCTATCTTTCCAGCTCTCCCAATATCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACG
GCACCTCAATTTCCCAGCTGCAGTACTTCCTCTGATCTTCCGAACGTTTCCGCACACAATGTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGACCACC
ACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGGCTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCC
GGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAGATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATA
ACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCATGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACA
CGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGTGCTTCAGCGTCCAGTCGATCGTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATT
TACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAAGCCAGTTATTTTGGGCAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGT
TGCAATGTCTGCCAGTTTACGATTCAGTCAGCGAGCTATGTTACGGTGCTCGTGCTCATCCGGCAACAAATAAGTTTGACCATCTACTAGTTTGTCTCTCAATGTAAAACCTTTGTGACGATGTCTAGC
AGCTTGTTTCTCATCCAGCTAAATTATAACTCATCCAACGACCAGTGTTACATAAATCGCATGCGTCGGGTTTTCCTTTATCGTGCCAGCTATGAAACGCTGTCTGCCGCTCATCCAGTCAGATAAATT
GAACCAGTTGACAGACTTGCACCCATCTAGCCCGCCATTTTGAAGGCTCTAATACGAAGCAGATGCTACCTCTTCTCACAATGAAAGGGTTCGTTTTTTAGCGCTTAGCAGCAACCGCTTGTGATCTAA
CGAGCGTTGGAGCGTTGCCACCCATGTTGTACCGAGGGATGGTGTGATAGCACGGTACCAATCAATGGTATATGGTACTAATATAGCGCATAACACCACATAAATTATTAGAGTACACTGTTTTGCACT
AACTTCGTCTTGCTTATCGGCAAAGATTTGTTGTAAGAGATCGTGATCAAGGGATAAACCCCATGACATAC

>Ph_hbn_cDNA01 [bp] 'w/o 1512, 1516
TAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGAAGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACC
ATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAAACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTT
TCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACAATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCC
ATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATTTACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGAC
AGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCCTCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCC
CGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAGTATCCAGACGTCTTCACCCGAGAGGAGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTGCTGGTG
TGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCGTACATTGCTGGCGCTCACAACGATGCAATCATATCCAGTCTGCCAGATATTGTCGGTATA
CCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCTTTAAGCTGCAGGCGTTCCACGCTCTGCAACTCAATACCTTCCTG
CCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTTCTTCCATCTTCCCTCCCCATAATCCCCAACCATCCACCTATCTTTCCAGCTCTCCCAATA
TCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTTCCCAGCTGCAGTACTTCCTCTGATCTTCCGAACGTTTCCGCACACAAT
GTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGACCACCACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTG
GATCTGCTGCGGCTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCT
GGGACTCTAAAGATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTT
TCCTTTCGTTCATGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGTGCTTCAGC
GTCCAGTCGATCGTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAA
GCCAGTTATTTTGGGCAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGTTGCAATGTCTGCCAGTTTACGATTCAGTCAGCGAGCTATGTTACGGTGCTCGTGCTCATC
CGGCAACAAATAAGTTTGACCATCTACTAGTTTGTCTCTCAATGTAAAACCTTTGTGACGATGTCTAGCAGCTTGTTTCTCATCCAGCTAAATTATAACTCATCCAACGACCAGTGTTACATAAATCGC
ATGCGTCGGGTTTTCCTTTATCGTGCCAGCTATGAAACGCTGTCTGCCGCTCATCCAGTCAGATAAATTGAACCAGTTGACAGACTTGCACCCATCTAGCCCGCCATTTTGAAGGCTCTAATACGAAGC
AGATGCTACCTCTTCTCACAATGAAAGGGTTCGTTTTTTAGCGCTTAGCAGCAACCGCTTGTGATCTAACGAGCGTTGGAGCGTTGCCACCCATGTTGTACCGAGGGATGGTGTGATAGCACGGTACCA
ATCAATGGTATATGGTACTAATATAGCGCATAACACCACATAAATTATTAGAGTACACTGTTTTGCACTAACTTCGTCTTGCTTATCGGCAAAGATTTGTTGTAAGAG


>Ph_hbn_cDNA05 [bp] 'w/o 1512, 1516
TAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGAAGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACC
ATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAAACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTT
TCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACAATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCC
ATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATTTACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGAC
AGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCCTCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCC
CGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAGTATCCAGACGTCTTCACCCGAGAGGAGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTGCAGGTG
TGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCGTACATTGCTGGCGCTCACAACGATGCAATCATATCCAGTCTGCCAGATATTGTCGGTATA
CCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCTTTAAGCTGCAGGCGTTCCACGCTCTGCAACTCAATACCTTCCTG
CCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTTCTTCCATCTTCCCTCCCCATAATCCCCAACCATCCACCTATCTTTCCAGCTCTCCCAATA
TCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTTCCCAGCTGCAGTACTTCCTCTGATCTTCCGAACGTTTCCGCACACAAT
GTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGACCACCACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTG
GATCTGCTGCGGCTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCT
GGGACTCTAAAGATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTT
TCCTTTCGTTCATGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGTGCTTCAGC
GTCCAGTCGATCGTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAA
GCCAGTTATTTTGGGCAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGTTGCAATGTCTGCCAGTTTACGATTCAGTCAGCGAGCTATGTTACGGTGCTCGTGCTCATC
CGGCAACAAATAAGTTTGACCATCTACTAGTTTGTCTCTCAATGTAAAACCTTTGTGACGATGTCTAGCAGCTTGTTTCTCATCCAGCTAAATTATAACTCATCCAACGACCAGTGTTACATAAATCGC
ATGCGTCGGGTTTTCCTTTATCGTGCCAGCTATGAAACGCTGTCTGCCGCTCATCCAGTCAGATAAATTGAACCAGTTGACAGACTTGCACCCATCTAGCCCGCCATTTTGAAGGCTCTAATACGAAGC
AGATGCTACCTCTTCTCACAATGAAAGGGTTCGTTTTTTAGCGCTTAGCAGCAACCGCTTGTGATCTAACGAGCGTTGGAGCGTTGCCACCCATGTTGTACCGAGGGATGGTGTGATAGCACGGTACCA
ATCAATGGTATATGGTACTAATATAGCGCATAACACCACATAAATTATTAGAGTACACTGTTTTGCACTAACTTCGTCTTGCTTATCGGCAAAGATTTGTTGTAAGAG


>Ph_hbn_cDNA11 [bp] 'w/o 1512, 1516
TAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGAAGCTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACC
ATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAAACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTT
TCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACAATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCC
ATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATTTACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGAC
AGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCCTCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCC
```

```
CGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAGTATCCAGACGTCTTCACCCGAGAGGAGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTGCAGGTG
TGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGGGAAGGCGATGGGCAGGGAGTCGCCCTCGCCGTACATTGCTGGCGCTCACAACGATGCAATCATATCCAGTCTGCCAGATATTGTCGGTATA
CCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCTTTAAGCTGCAGGCGTTCCACGCTCTGCAACTCAATACCTTCCTG
CCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTTCTTCCATCTTCCCTCCCCATAATCCCCAACCATCCACCTATCTTTCCAGCTCTCCCAATA
TCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTTCCCAGCTGCAGTACTTCCTCTGATCTTCCGAACGTTTCCGCACACAAT
GTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGACCACCACCTCGGTCGCGTCTACCACCCCTCTACTATGCACCGCCGACTCACGAGGGCAATCCGTG
GATCTGCTGCGGCTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCT
GGACTCTAAAGATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTT
CCTTTCGTTCATGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGTGCTTCAGCG
TCCAGTCGATCGTTTGCAGAGTCTGTTCGCATTCTTGCACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAAG
CCAGTTATTTTGGGCAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGTTGCAATGTCTGCCAGTTTACGATTCAGTCAGCGAGCTATGTTACGGTGCTCGTGCTCATCC
GGCAACAAATAAGTTTGACCATCTACTAGTTTGTCTCTCAATGTAAAACCTTTGTGACGATGTCTAGCAGCTTGTTTCTCATCCAGCTAAATTATAACTCATCCAACGACCAGTGTTACATAAATCGCA
TGCGTCGGGTTTTCCTTTATCGTGCCAGCTATGAAACGCTGTCTGCCGCTCATCCAGTCAGATAAATTGAACCAGTTGACAGACTTGCACCCATCTAGCCCGCCATTTTGAAGGCTCTAATACGAAGCA
GATGCTACCTCTTCTCACAATGAAAGGGTTCGTTTTTTAGCGCTTAGCAGCAACCGCTTGTGATCTAACGAGCGTTGGAGCGTTGCCACCCATGTTGTACCGAGGGATGGTGTGATAGCACGGTACCAA
TCAATGGTATATGGTACTAATATAGCGCATAACACCACATAAATTATTAGAGTACACTGTTTTGCACTAACTTCGTCTTGCTTATCGGCAAAGATTTGTTGTAAGAG

>Ph_hbn_ORF01 [bp] 'w/o 1514, 1515
AAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAAACCTGTAGCAAAATTATGCGAAGCCCTTCGCAC
CAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACAATGGCAGGTGATACAGTTAATGTACCAGCCTCA
GAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATTTACACCATAGAAAATATACTCGCCCTCAAGACC
AGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCCTCCGACCACGCCGTGGAGGGAACGGACTTGGAG
AAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAATATCCTGACGTCTTCACTCGAGAGGAGTTGGCACTCAACCTT
GACCTATCCGAGGCACGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCTTACATTGCTGGCGCTCACAACGATGCAATCATA
TCCAGTCTGCCAGATATTGTCGGTATACCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCTTTAAGCTGCAGGCGTTC
CACGCTCTGCAACTCAATACCTTCCTGCCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTTCTTCCATCTTCCCTCCCCATAATCCCCAACCAT
CCACCTATCTTTCCAGCTCTCCCAATATCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTTCCCAGCTGCAGTACTTCCTCT
GATCTTCCGAACGTTTCCGCACACAATGTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGTCCACCACCTCGGTCGCGTCTACCACCCCTCTACTATAC
ACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGGCTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGG
ACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAGATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTC

>Ph_hbn_ORF02 [bp] 'w/o 1514, 1515
AAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAAACCTGTAGCAAAATTATGCGAAGCCCTTCGCAC
CAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACAATGGCAGGTGATACAGTTAATGTACCAGCCTCA
GAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATTTACACCATAGAAAATATACTCGCCCTCAAGACC
AGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCCGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCCTCCGACCACGCCGTGGAGGGAACGGACTTGGAG
AAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAATGTCCTGACGTCTTCACTCGAGAGGAGTTGGCACTCAACCTT
GACCTATCCGAGGCACGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCGTACATTGCTGGCGCTCACAACGATGCAATCATA
TCCAGTCTGCCAGATATTGTCGGTATACCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCTTTAAGCTGCAGGCGTTC
CACGCTCTGCAACTCAATACCTTCCTGCCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTTCTTCCATCTTCCCTCCCCATAATCCCCAACCAT
CCACCTATCTTTCCAGCTCTCCCAATATCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTTCCCAGCTGCAGTACTTCCTCT
GATCTTCCGAACGTTTCCGCACACAATGTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGTCCACCACCTCGGTCGCGTCTACCACCCCTCTACTATAC
ACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGGCTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGG
ACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAGATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTC

>Ph_hbn_ORF05 [bp] 'w/o 1514, 1515
AAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAAACCTGTAGCAAAATTATGCGAAGCCCTTCGCAC
CAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACAATGGCAGGTGATACAGTTAATGTACCAGCCTCA
GAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATTTACACCATAGAAAATATACTCGCCCTCAAGACC
AGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCCTCCGACCACGCCGTGGAGGGAACGGACTTGGAG
AAACCCCGCAGGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAATATCCTGACGTCTTCACTCGAGAGGAGTTGGCACTCAACCTT
GACCTATCCGAGGCACGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCGTACATTGCTGGCGCTCACAACGATGCAATCATA
TCCAGTCTGCCAGATATTGTCGGTATACCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCTTTAAGCTGCAGGCGTTC
CACGCTCTGCAACTCAATACCTTCCTGCCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTTCTTCCATCTTCCCTCCCCATAATCCCCAACCAT
CCACCTATCTTTCCAGCTCTCCCAATATCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTTCCCAGCTGCAGTACTTCCTCT
GATCTTCCGAACGTTTCCGCACACAATGTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGTCCACCACCTCGGTCGCGTCTACCACCCCTCTACTATAC
ACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGGCTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGG
ACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAGATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTC

>Ph_hbn_5RHH51 [bp] 'w/o 1505
GGAATTTTCACACTTCCCTTTACCGAGTTTTACTGAGTCGAGGAGATACTTACCAAAGTGTAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGA
AGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAA
AGCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACA
ATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATT
TACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCC
TCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAGTATCCAGAC
GTCTTCACCCGAGAGGAGTTGG

>Ph_hbn_5RHH53 [bp] 'w/o 1505
GGAATTTTCACACTTCCCTTTACCGAGTTTTACTGAGTCGAGGAGATACTTACCAAAGTGTAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGA
AGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAA
ACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACA
ATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCAATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATT
TACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCC
TCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAGTATCCAGAC
GTCTTCACCCGAGAGGAGTTGG

>Ph_hbn_5RHH54 [bp] 'w/o 1505
GGATTTTACCAATCACCCCTTTACCAAGCGTTTACTGAGTCGAGGAGATACTTACCAAAGTGTAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCT
GAAGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATC
AAACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTA
CAATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAA
TTTACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTT
CCTCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAATATCCTG
ACGTCTTCACTCGAGAGGAGTTGG

>Ph_hbn_5Rn25 [bp] 'w/o 1507
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAGTCGAGGAGATACTTACCAAAGTGTAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGA
AGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAA
ACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACA
ATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATT
```

```
TACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCC
TCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACT

>Ph_hbn_5Rn26 [bp] 'w/o 1507
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAGTCGAGGAGATACTTACCAAAGTGTAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGA
AGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAA
ACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACA
ATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATT
TACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCC
TCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACT

>Ph_hbn_5Rn28 [bp] 'w/o 1507
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAGTCGAGGAGATACTTACCAAAGTGTAAAAATGAGGAAAAATTGAACTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGA
AGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACCATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAA
ACCTGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAATACAGTCACGGTAGACATGTCTCCCAGTAACTGTACA
ATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATTTCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATT
TACACCATAGAAAATATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCGCCCAAGTGTCCCCCGTCTCCTTCC
TCCGACCACGCCATGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCCCGTACAACGTTCACAACGTTCCAACT

>Ph_hbn_vsdPms078 [bp] 'w/o fw_QRR, re_QVW
AACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAATATCCTGACGTCTTCACTCGAGAGGAGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTG

>Ph_hbn_vsdPms093 [bp] 'w/o fw_QRR, re_QVW
AACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAATATCCTGACGTCTTCACTCGAGAGGAGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTG

>Ph_hbn_vsdPms109 [bp] 'w/o fw_QRR, re_QVW
AACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAATATCCTGACGTCTTCACTCGAGAGGAGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTG

>Ph_hbn_3RHH313 [bp] 'premature poly(A), w/o 1501
AGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCGTACATTGCTGGCGCTC
ACAACGATGCAATCATATCCAGTCTGCCAGATATTGTCGGTATACCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCT
TTAAGCTGCAGGCGTTCCACGCTCTGCAACTCAATACCTTCCTGCCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTTCTTCCATCTTCCCTCC
CCATAATCCCCAACCATCCACCTATCTTTCCAGCTCTCCCAATATCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTTCCCA
GCTGCAGTACTTCCTCTGATCTTCCGAACGTTTCCGCACACAATGTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGTCCCA

>Ph_hbn_3RHH314 [bp] 'premature poly(A), w/o 1501
AGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCGTACATTGCTGGCGCTC
ACAACGATGCAATCATATCCAGTCTGCCAGATATTGTCGGTATACCAATAACATCGTCTGTTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCT
TTAAGCTGCAGGCGTTCCACGCTCTGCAACTCAATACCTTCCTGCCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTTCTTCCATCTTCCCTCC
CCATAACCCCCAACCATCCACCTATCTTTCCAGCTCTCCCAATATCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTTCCCA
GCTGCAGTACTTCCTCTGATCTTCCGAACGTTTCCGCACACAATGTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGTCCCA

>Ph_hbn_3RHH315 [bp] 'premature poly(A), w/o 1501
AGTTGGCACTCAACCTTGACCTATCCGAGGCACGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAGTCGCCCTCGCCGTACATTGCTGGCGCTC
ACAACGATGCAATCATATCCAGTCTGCCAGATATTGTCGGTATACCAATAACATCGTCTGCTCCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCT
TTAAGCTGCAGGCGTTCCACGCTCTGCAACTCAATACCTTCCTGCCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTTCTTCCATCTTCCCTCC
CCATAATCCCCAACCATCCACCTATCTTTCCAGCTCTCCCAATATCCCCAATGGGGTCTCCAGATGTGGTCTCCACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTCCCCA
GCTGCAGTACTTCCTCTGATCTTCCGAACGTTTCCGCACACAATGTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGTCCCA

>Ph_hbn_3RHH336 [bp] 'extended RACE; still premature, poly(A) - novel site, w/o 1504
AACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGTCCACCACCTCGGTCGCGTCTACCACCCCTTTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATTTGCTGCGG
CTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGTAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAG
ATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCA
TGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGCGCTTCAGCGTCCAGTCGATC
GTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAAAAA

>Ph_hbn_3RHH356 [bp] 'incomplete seq, only fragment, w/o 1504
AACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGACCACCACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGG
CTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAG
ATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCA
TGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGTGCTTCAGCGTCCAGTCGATC
GTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAAGCCAGTTATTTT
GGGCAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGTTGCAATGTCTGCCAGTTTA

>Ph_hbn_3RHH357 [bp] 'w/o 1504
AACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGACCACCACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGG
CTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAG
ATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCA
TGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGTGCTTCAGCGTCCAGTCGATC
GTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAAGCCAGTTATTTT
GGGCAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGTTGCAATGTCTGCCAGTTTACGATTCAGTCAGCGAGCTATGTTACGGTGCTCGTGCTCATCCGGCAACAAATA
AGTTTGACCATCTACTAGTTTGTCTCTCAATGTAAAACCTTTGTGACGATGTCTAGCAGCTTGTTTCTCATCCAGCTAAATTATAACTCATCCAACGACCAGTGTTACATAAATCGCATGCGTCGGGTT
TTCCTTTATCGTGCCAGCTATGAAACGCTGTCTGCCGCTCATCCAGTCAGATAAATTGAACCAGTTGACAGACTTGCACCCATCTAGCCCGCCATTTTGAAGGCTCTAATACGAAGCAGATGCTACCTC
TTCTCACAATGAAAGGGTTCGTTTTTTAGCGCTTAGCAGCAACCGCTTGTGATCTAACGAGCGTTGGAGCGTTGCCACCCATGTTGTACCGAGGGATGGTGTGATAGCACGGTACCAATCAATGGTATA
TGGTACTAATATAGCGCATAACACCACATAAATTATTAGAGTACACTGTTTTGCACTAACTTCGTCTTGCTTATCGGCAAAGATTTGTTGTAAGAGATCGTGATCAAGGGATAAACCCCATGACATACA
AAAAAAAAAAAAAAAAAAAA

>Ph_hbn_3RHH358 [bp] 'one direction usable seq only, w/o 1504
AACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCGTGGATCTGACCACCACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGG
CTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAG
ATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCA
TGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGTGCTTCAGCGTCCAGTCGATC
GTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAAGCCAGTTATTTT
GGGCAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGTTGCAATGT

>Ph_hbn_3RHH361 [bp] 'seq gap 35 bp 'N' added, w/o 1504
AACCCATTAACATCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGTCCACCACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGG
CTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAG
ATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCA
TGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGCGCTTCGGCGTCCAGTCGATC
```

```
GTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAAGCCAGTTATTTT
GGACAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGTTGCAATG
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
ACGGTGCTCGTGCTCATCCGGCAACAAATAAGTTTGACCATCTACTAGTTTGTCTCTCAATGTAAAACCTTTGTGACGATGTCTAGCAGCTTGTTTCTCATCCAGCTAAATTATAACTCATCCAACGAC
CAGTGTTACATAAATCGCATGCGTCGGGTTTTCCTTTATCGTGCCAGCTATGAAACGCTGTCTGCCGCTCATCCAGTCAGATAAATTGAACCAGTTGACAGACTTGCACCCATCTAGCCCGCCATTTTG
AAGGCTCTAATACGAAGCAGATGCTACCTCTTCTCACAATGAAAGGGTTCGTTTTTTAGCGCTTAGCAGCAACCGCTTGTGATCTAACGAGCGTTGGAGCGTTGCCACCCATGTTGTACCGAGGGATGG
TGTGATAGCACGGTACCAATCAATGGTATATGGTACTAATATAGCGCATAACACCACATAAATTATTAGAGTACACTGTTTTGCACTAACTTCGTCTTGCTTATCGGCAAAGATTTGTTGTAAGAGATC
GTGATCAAGGGATAAACCCCATGACATACTAAAAGAAGAGATACAGTAATGATAGGGCTGCAGGCTGGAGAATAACATGCCTTTAGGCAGAGTGCCCAAAAGCGACCTACGTCACTTACATAATGTTGC
ATTGCGTAAATTCTTCGGTTATATATCAGCCCGTATGAAACTAGGAATATCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_hbn_3RHH363 [bp] 'one direction usable seq only, w/o 1504
AACCCATTAACTTCCTTAGCCCACATGACTGCTGCACCTTCCATGGATCTGTCCACCACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATCAGCTGCGG
CTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAG
ATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCA
TGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGCGCTTCAGCGTCCAGTCGATC
GTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAAGCCAGTTATTTT
GGACAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGTTGCAATG

>Ph_hbn_3RHH366 [bp] 'one direction usable seq only, w/o 1504
AACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGTCCACCACCTCGGTCGCGTCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGG
CTCAAGGCGCGCCAGCACCAGGTCCTGCTGGAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGGACTCTAAAG
ATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAGGCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCA
TGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTCACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGCGCTTCAGCGTCCAGTCGATC
GTTTGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACATTTCCTTTCCGCCAGAAGTACACAAGCCAGTTATTTT
GGACAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAATGGTTGCAATGTCTG


>Ph_hbn_ref [aa]
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSL
PDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLP
NVSAHNVAVINPLTSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_cDNA01 [aa] 'w/o 1512, 1516
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVLVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSL
PDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLP
NVSAHNVAVINPLTSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_cDNA05 [aa] 'w/o 1512, 1516
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSL
PDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLP
NVSAHNVAVINPLTSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_cDNA11 [aa] 'w/o 1512, 1516
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKRGKAMGRESPSPYIAGAHNDAIISSL
PDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLP
NVSAHNVAVINPLTSLAHMTAAPSMDLTTTSVASTTPLLCTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGL#

>Ph_hbn_ORF01 [aa] 'w/o 1514, 1515
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSL
PDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLP
NVSAHNVAVINPLTSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_ORF02 [aa] 'w/o 1514, 1515
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGRPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQCPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSL
PDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLP
NVSAHNVAVINPLTSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_ORF05 [aa] 'w/o 1514, 1515
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRRQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSL
PDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLP
NVSAHNVAVINPLTSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_5RHH51 [aa] 'w/o 1505
MSVNSTIASVISTPPGSKSTTADHLARDQSCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREEL

>Ph_hbn_5RHH54 [aa] 'w/o 1505
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREEL

>Ph_hbn_5RHH53 [aa] 'w/o 1505
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREEL

>Ph_hbn_5Rn25 [aa] 'w/o 1507
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQ

>Ph_hbn_5Rn26 [aa] 'w/o 1507
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAVEGTDLEKPRKQVRRSRTTFTTFQ
```

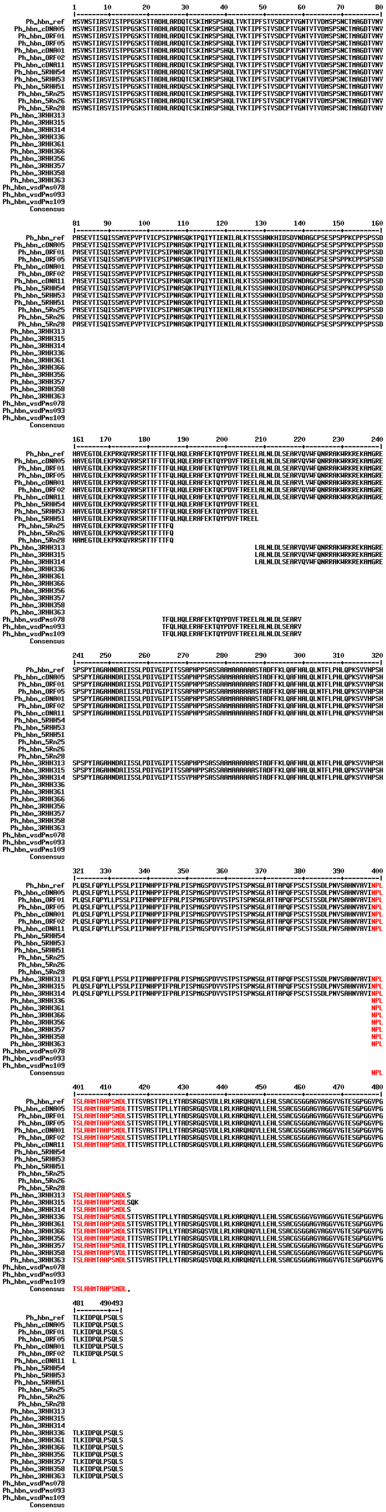
```
>Ph_hbn_5Rn28 [aa] 'w/o 1507
MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNVPASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSS
HNKHIDSDVNDAGCPSESPSPPKCPPSPSSDHAMEGTDLEKPRKQVRRSRTTFTTFQ

>Ph_hbn_vsdPms078 [aa] 'w/o fw_QRR, re_QVW
TFQLHQLERAFEKTQYPDVFTREELALNLDLSEARV

>Ph_hbn_vsdPms093 [aa] 'w/o fw_QRR, re_QVW
TFQLHQLERAFEKTQYPDVFTREELALNLDLSEARV

>Ph_hbn_vsdPms109 [aa] 'w/o fw_QRR, re_QVW
TFQLHQLERAFEKTQYPDVFTREELALNLDLSEARV

>Ph_hbn_3RHH313 [aa] 'premature poly(A), w/o 1501
LALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLP
IIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPLTSLAHMTAAPSMDLS

>Ph_hbn_3RHH314 [aa] 'premature poly(A), w/o 1501
LALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSLPDIVGIPITSSVPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLP
ITPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPLTSLAHMTAAPSMDLS

>Ph_hbn_3RHH315 [aa] 'premature poly(A), w/o 1501
LALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSHPLQSLFQPYLLPSSLP
IIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPLTSLAHMTAAPSMDLSQK

>Ph_hbn_3RHH336 [aa] 'extended RACE; still premature, poly(A) - novel site, w/o 1504
NPLTSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGVGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_3RHH356 [aa] 'incomplete seq, only fragment, w/o 1504
NPLTSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_3RHH357 [aa] 'w/o 1504
NPLTSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_3RHH358 [aa] 'one direction usable seq only, w/o 1504
NPLTSLAHMTAAPSVDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_3RHH361 [aa] 'seq gap 35 bp 'N' added, w/o 1504
NPLTSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_3RHH363 [aa] 'one direction usable seq only, w/o 1504
NPLTSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDQLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#

>Ph_hbn_3RHH366 [aa] 'one driection usable seq only, w/o 1504
NPLTSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPGTLKIDPQLPSQLS#
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.3_Ph_hbn]

### [PhD_A2.3.2_Ph_hbn_alignments]

**“alignment_Ph_hbn_all_aa_sequences.gif”**



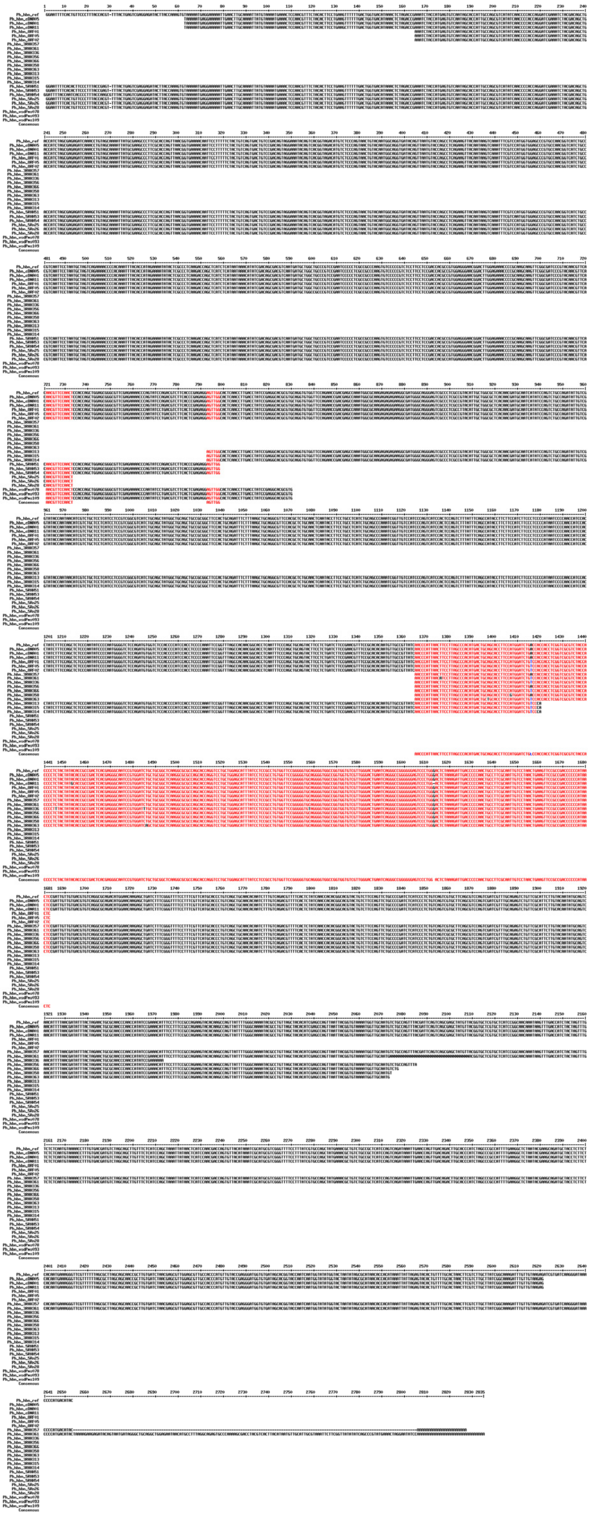
```
                 1       10        20        30        40        50        60        70        80
                 |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
   Ph_hbn_cDNA05 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
    Ph_hbn_ORF01 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
    Ph_hbn_ORF05 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
   Ph_hbn_cDNA01 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
    Ph_hbn_ORF02 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
   Ph_hbn_cDNA11 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
   Ph_hbn_5RHH54 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
   Ph_hbn_5RHH53 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
   Ph_hbn_5RHH51 MSVNSTIASVISTPPGSKSTTADHLARDQSCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
    Ph_hbn_5Rn25 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
    Ph_hbn_5Rn26 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
    Ph_hbn_5Rn28 MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
  Ph_hbn_3RHH313
  Ph_hbn_3RHH315
  Ph_hbn_3RHH314
  Ph_hbn_3RHH336
  Ph_hbn_3RHH361
  Ph_hbn_3RHH366
  Ph_hbn_3RHH356
  Ph_hbn_3RHH357
  Ph_hbn_3RHH358
  Ph_hbn_3RHH363
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus

                 81      90       100       110       120       130       140       150       160
                 |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
   Ph_hbn_cDNA05 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
    Ph_hbn_ORF01 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
    Ph_hbn_ORF05 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
   Ph_hbn_cDNA01 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
    Ph_hbn_ORF02 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGRPSESPSPPKCPPSPSSD
   Ph_hbn_cDNA11 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
   Ph_hbn_5RHH54 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
   Ph_hbn_5RHH53 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
   Ph_hbn_5RHH51 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
    Ph_hbn_5Rn25 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
    Ph_hbn_5Rn26 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
    Ph_hbn_5Rn28 PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
  Ph_hbn_3RHH313
  Ph_hbn_3RHH315
  Ph_hbn_3RHH314
  Ph_hbn_3RHH336
  Ph_hbn_3RHH361
  Ph_hbn_3RHH366
  Ph_hbn_3RHH356
  Ph_hbn_3RHH357
  Ph_hbn_3RHH358
  Ph_hbn_3RHH363
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus

                 161    170       180       190       200       210       220       230       240
                 |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRE
   Ph_hbn_cDNA05 HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRE
    Ph_hbn_ORF01 HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRE
    Ph_hbn_ORF05 HAVEGTDLEKPRRQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRE
   Ph_hbn_cDNA01 HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVLVWFQNRRAKWRKREKAMGRE
    Ph_hbn_ORF02 HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQCPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRE
   Ph_hbn_cDNA11 HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKRGKAMGRE
   Ph_hbn_5RHH54 HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREEL
   Ph_hbn_5RHH53 HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREEL
   Ph_hbn_5RHH51 HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREEL
    Ph_hbn_5Rn25 HAVEGTDLEKPRKQVRRSRTTFTTFQ
    Ph_hbn_5Rn26 HAVEGTDLEKPRKQVRRSRTTFTTFQ
    Ph_hbn_5Rn28 HAMEGTDLEKPRKQVRRSRTTFTTFQ
  Ph_hbn_3RHH313                                                LALNLDLSEARVQVWFQNRRAKWRKREKAMGRE
  Ph_hbn_3RHH315                                                LALNLDLSEARVQVWFQNRRAKWRKREKAMGRE
  Ph_hbn_3RHH314                                                LALNLDLSEARVQVWFQNRRAKWRKREKAMGRE
  Ph_hbn_3RHH336
  Ph_hbn_3RHH361
  Ph_hbn_3RHH366
  Ph_hbn_3RHH356
  Ph_hbn_3RHH357
  Ph_hbn_3RHH358
  Ph_hbn_3RHH363
Ph_hbn_vsdPms078                        TFQLHQLERAFEKTQYPDVFTREELALNLDLSEARV
Ph_hbn_vsdPms093                        TFQLHQLERAFEKTQYPDVFTREELALNLDLSEARV
Ph_hbn_vsdPms109                        TFQLHQLERAFEKTQYPDVFTREELALNLDLSEARV
       Consensus

                 241    250       260       270       280       290       300       310       320
                 |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
   Ph_hbn_cDNA05 SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
    Ph_hbn_ORF01 SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
    Ph_hbn_ORF05 SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
   Ph_hbn_cDNA01 SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
    Ph_hbn_ORF02 SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
   Ph_hbn_cDNA11 SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
   Ph_hbn_5RHH54
   Ph_hbn_5RHH53
   Ph_hbn_5RHH51
    Ph_hbn_5Rn25
    Ph_hbn_5Rn26
    Ph_hbn_5Rn28
  Ph_hbn_3RHH313 SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
  Ph_hbn_3RHH315 SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
  Ph_hbn_3RHH314 SPSPYIAGAHNDAIISSLPDIVGIPITSSVPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
  Ph_hbn_3RHH336
  Ph_hbn_3RHH361
  Ph_hbn_3RHH366
  Ph_hbn_3RHH356
  Ph_hbn_3RHH357
  Ph_hbn_3RHH358
  Ph_hbn_3RHH363
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus

                 321    330       340       350       360       370       380       390       400
                 |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
   Ph_hbn_cDNA05 PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
    Ph_hbn_ORF01 PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
    Ph_hbn_ORF05 PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
   Ph_hbn_cDNA01 PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
    Ph_hbn_ORF02 PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
   Ph_hbn_cDNA11 PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
   Ph_hbn_5RHH54
   Ph_hbn_5RHH53
   Ph_hbn_5RHH51
    Ph_hbn_5Rn25
    Ph_hbn_5Rn26
    Ph_hbn_5Rn28
  Ph_hbn_3RHH313 PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
  Ph_hbn_3RHH315 PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
  Ph_hbn_3RHH314 PLQSLFQPYLLPSSLPITPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
  Ph_hbn_3RHH336                                                                              NPL
  Ph_hbn_3RHH361                                                                              NPL
  Ph_hbn_3RHH366                                                                              NPL
  Ph_hbn_3RHH356                                                                              NPL
  Ph_hbn_3RHH357                                                                              NPL
  Ph_hbn_3RHH358                                                                              NPL
  Ph_hbn_3RHH363                                                                              NPL
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus                                                                              NPL

                 401    410       420       430       440       450       460       470       480
                 |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref TSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
   Ph_hbn_cDNA05 TSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
    Ph_hbn_ORF01 TSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
    Ph_hbn_ORF05 TSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
   Ph_hbn_cDNA01 TSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
    Ph_hbn_ORF02 TSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
   Ph_hbn_cDNA11 TSLAHMTAAPSMDLTTTSVASTTPLLCTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
   Ph_hbn_5RHH54
   Ph_hbn_5RHH53
   Ph_hbn_5RHH51
    Ph_hbn_5Rn25
    Ph_hbn_5Rn26
    Ph_hbn_5Rn28
  Ph_hbn_3RHH313 TSLAHMTAAPSMDLS
  Ph_hbn_3RHH315 TSLAHMTAAPSMDLSQK
  Ph_hbn_3RHH314 TSLAHMTAAPSMDLS
  Ph_hbn_3RHH336 TSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGVGVAGGVVGTESGPGGVPG
  Ph_hbn_3RHH361 TSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
  Ph_hbn_3RHH366 TSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
  Ph_hbn_3RHH356 TSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
  Ph_hbn_3RHH357 TSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
  Ph_hbn_3RHH358 TSLAHMTAAPSVDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
  Ph_hbn_3RHH363 TSLAHMTAAPSMDLSTTSVASTTPLLYTADSRGQSVDQLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus TSLAHMTAAPSMDL.

                 481    490493
                 |--------+--|
      Ph_hbn_ref TLKIDPQLPSQLS
   Ph_hbn_cDNA05 TLKIDPQLPSQLS
    Ph_hbn_ORF01 TLKIDPQLPSQLS
    Ph_hbn_ORF05 TLKIDPQLPSQLS
   Ph_hbn_cDNA01 TLKIDPQLPSQLS
    Ph_hbn_ORF02 TLKIDPQLPSQLS
   Ph_hbn_cDNA11 L
   Ph_hbn_5RHH54
   Ph_hbn_5RHH53
   Ph_hbn_5RHH51
    Ph_hbn_5Rn25
    Ph_hbn_5Rn26
    Ph_hbn_5Rn28
  Ph_hbn_3RHH313
  Ph_hbn_3RHH315
  Ph_hbn_3RHH314
  Ph_hbn_3RHH336 TLKIDPQLPSQLS
  Ph_hbn_3RHH361 TLKIDPQLPSQLS
  Ph_hbn_3RHH366 TLKIDPQLPSQLS
  Ph_hbn_3RHH356 TLKIDPQLPSQLS
  Ph_hbn_3RHH357 TLKIDPQLPSQLS
  Ph_hbn_3RHH358 TLKIDPQLPSQLS
  Ph_hbn_3RHH363 TLKIDPQLPSQLS
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.3_Ph_hbn]

[PhD_A2.3.2_Ph_hbn_alignments]

**“alignment_Ph_hbn_all_bp_sequences.gif”**

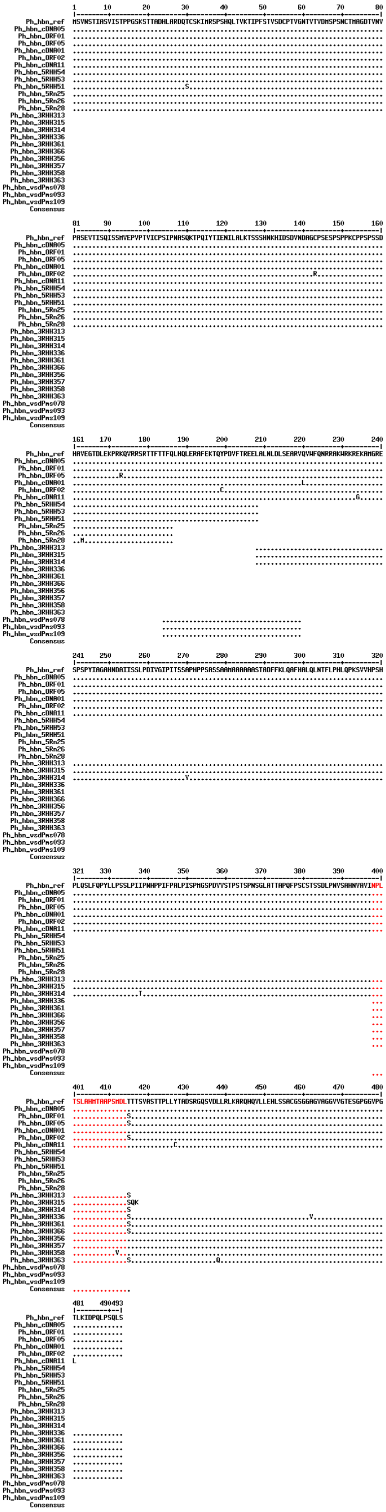


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.3_Ph_hbn]

[PhD_A2.3.2_Ph_hbn_alignments]

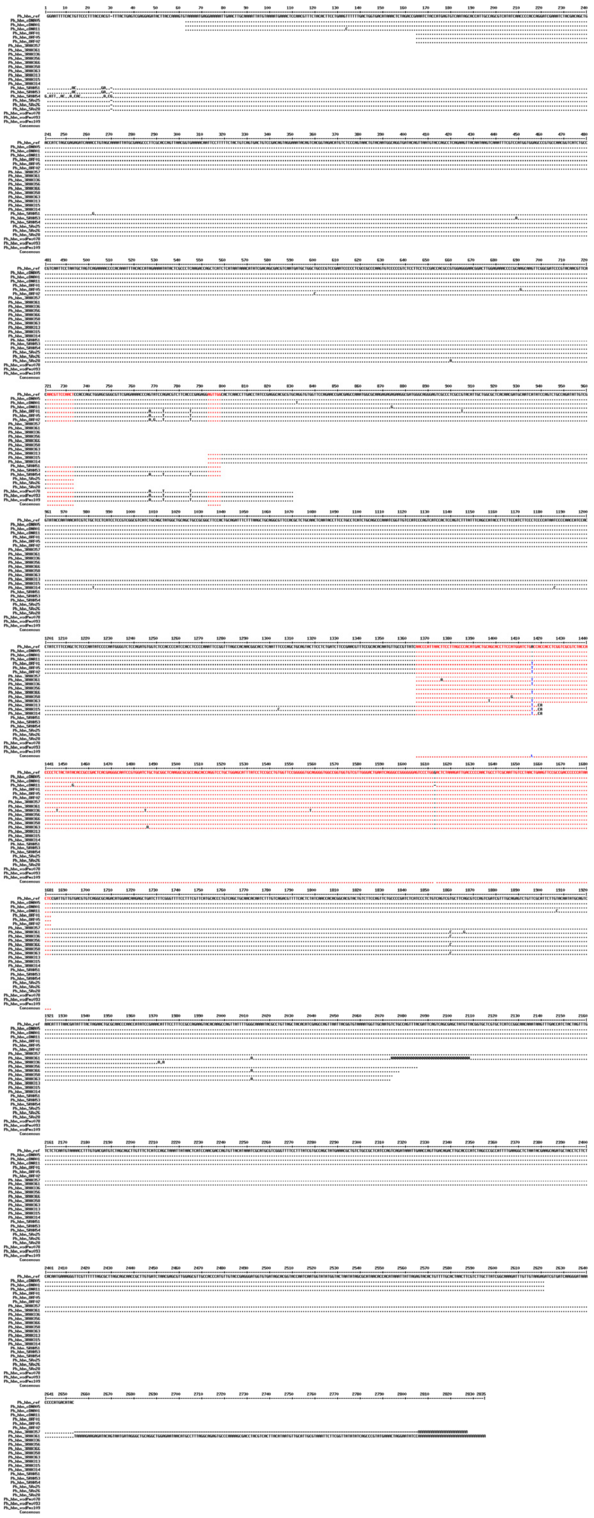
## “difference_alignment_Ph_hbn_all_aa_sequences.gif”



```
                  1       10        20        30        40        50        60        70        80
                  |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref  MSVNSTIASVISTPPGSKSTTADHLARDQTCSKIMRSPSHQLTVKTIPFSTVSDCPTVGNTVTVDMSPSNCTMAGDTVNV
   Ph_hbn_cDNA05  ................................................................................
    Ph_hbn_ORF01  ................................................................................
    Ph_hbn_ORF05  ................................................................................
   Ph_hbn_cDNA01  ................................................................................
    Ph_hbn_ORF02  ................................................................................
   Ph_hbn_cDNA11  ................................................................................
   Ph_hbn_5RHH54  ................................................................................
   Ph_hbn_5RHH53  ................................................................................
   Ph_hbn_5RHH51  .............................S..................................................
    Ph_hbn_5Rn25  ................................................................................
    Ph_hbn_5Rn26  ................................................................................
    Ph_hbn_5Rn28  ................................................................................
  Ph_hbn_3RHH313
  Ph_hbn_3RHH315
  Ph_hbn_3RHH314
  Ph_hbn_3RHH336
  Ph_hbn_3RHH361
  Ph_hbn_3RHH366
  Ph_hbn_3RHH356
  Ph_hbn_3RHH357
  Ph_hbn_3RHH358
  Ph_hbn_3RHH363
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus

                  81      90       100       110       120       130       140       150       160
                  |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref  PASEVTISQISSMVEPVPTVICPSIPNASQKTPQIYTIENILALKTSSSHNKHIDSDVNDAGCPSESPSPPKCPPSPSSD
   Ph_hbn_cDNA05  ................................................................................
    Ph_hbn_ORF01  ................................................................................
    Ph_hbn_ORF05  ................................................................................
   Ph_hbn_cDNA01  ................................................................................
    Ph_hbn_ORF02  ..............................................................R.................
   Ph_hbn_cDNA11  ................................................................................
   Ph_hbn_5RHH54  ................................................................................
   Ph_hbn_5RHH53  ................................................................................
   Ph_hbn_5RHH51  ................................................................................
    Ph_hbn_5Rn25  ................................................................................
    Ph_hbn_5Rn26  ................................................................................
    Ph_hbn_5Rn28  ................................................................................
  Ph_hbn_3RHH313
  Ph_hbn_3RHH315
  Ph_hbn_3RHH314
  Ph_hbn_3RHH336
  Ph_hbn_3RHH361
  Ph_hbn_3RHH366
  Ph_hbn_3RHH356
  Ph_hbn_3RHH357
  Ph_hbn_3RHH358
  Ph_hbn_3RHH363
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus

                  161    170       180       190       200       210       220       230       240
                  |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref  HAVEGTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRE
   Ph_hbn_cDNA05  ................................................................................
    Ph_hbn_ORF01  ................................................................................
    Ph_hbn_ORF05  ............R...................................................................
   Ph_hbn_cDNA01  ...........................................................L....................
    Ph_hbn_ORF02  ......................................C.........................................
   Ph_hbn_cDNA11  .........................................................................G......
   Ph_hbn_5RHH54  ................................................
   Ph_hbn_5RHH53  ................................................
   Ph_hbn_5RHH51  ................................................
    Ph_hbn_5Rn25  ..........................
    Ph_hbn_5Rn26  ..........................
    Ph_hbn_5Rn28  ..M.......................
  Ph_hbn_3RHH313                                                 .................................
  Ph_hbn_3RHH315                                                 .................................
  Ph_hbn_3RHH314                                                 .................................
  Ph_hbn_3RHH336
  Ph_hbn_3RHH361
  Ph_hbn_3RHH366
  Ph_hbn_3RHH356
  Ph_hbn_3RHH357
  Ph_hbn_3RHH358
  Ph_hbn_3RHH363
Ph_hbn_vsdPms078                         ....................................
Ph_hbn_vsdPms093                         ....................................
Ph_hbn_vsdPms109                         ....................................
       Consensus

                  241    250       260       270       280       290       300       310       320
                  |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref  SPSPYIAGAHNDAIISSLPDIVGIPITSSAPHPPSASSAAMAAAAAASTADFFKLQAFHALQLNTFLPHLQPKSVVHPSH
   Ph_hbn_cDNA05  ................................................................................
    Ph_hbn_ORF01  ................................................................................
    Ph_hbn_ORF05  ................................................................................
   Ph_hbn_cDNA01  ................................................................................
    Ph_hbn_ORF02  ................................................................................
   Ph_hbn_cDNA11  ................................................................................
   Ph_hbn_5RHH54
   Ph_hbn_5RHH53
   Ph_hbn_5RHH51
    Ph_hbn_5Rn25
    Ph_hbn_5Rn26
    Ph_hbn_5Rn28
  Ph_hbn_3RHH313  ................................................................................
  Ph_hbn_3RHH315  ................................................................................
  Ph_hbn_3RHH314  .............................V..................................................
  Ph_hbn_3RHH336
  Ph_hbn_3RHH361
  Ph_hbn_3RHH366
  Ph_hbn_3RHH356
  Ph_hbn_3RHH357
  Ph_hbn_3RHH358
  Ph_hbn_3RHH363
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus

                  321    330       340       350       360       370       380       390       400
                  |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref  PLQSLFQPYLLPSSLPIIPNHPPIFPALPISPMGSPDVVSTPSTSPNSGLATTAPQFPSCSTSSDLPNVSAHNVAVINPL
   Ph_hbn_cDNA05  ................................................................................
    Ph_hbn_ORF01  ................................................................................
    Ph_hbn_ORF05  ................................................................................
   Ph_hbn_cDNA01  ................................................................................
    Ph_hbn_ORF02  ................................................................................
   Ph_hbn_cDNA11  ................................................................................
   Ph_hbn_5RHH54
   Ph_hbn_5RHH53
   Ph_hbn_5RHH51
    Ph_hbn_5Rn25
    Ph_hbn_5Rn26
    Ph_hbn_5Rn28
  Ph_hbn_3RHH313  ................................................................................
  Ph_hbn_3RHH315  ................................................................................
  Ph_hbn_3RHH314  .................T..............................................................
  Ph_hbn_3RHH336                                                                               ...
  Ph_hbn_3RHH361                                                                               ...
  Ph_hbn_3RHH366                                                                               ...
  Ph_hbn_3RHH356                                                                               ...
  Ph_hbn_3RHH357                                                                               ...
  Ph_hbn_3RHH358                                                                               ...
  Ph_hbn_3RHH363                                                                               ...
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus                                                                               ...

                  401    410       420       430       440       450       460       470       480
                  |--------+---------+---------+---------+---------+---------+---------+---------|
      Ph_hbn_ref  TSLAHMTAAPSMDLTTTSVASTTPLLYTADSRGQSVDLLRLKARQHQVLLEHLSSACGSGGAGVAGGVVGTESGPGGVPG
   Ph_hbn_cDNA05  ................................................................................
    Ph_hbn_ORF01  ..............S.................................................................
    Ph_hbn_ORF05  ..............S.................................................................
   Ph_hbn_cDNA01  ................................................................................
    Ph_hbn_ORF02  ..............S.................................................................
   Ph_hbn_cDNA11  ..........................C.....................................................
   Ph_hbn_5RHH54
   Ph_hbn_5RHH53
   Ph_hbn_5RHH51
    Ph_hbn_5Rn25
    Ph_hbn_5Rn26
    Ph_hbn_5Rn28
  Ph_hbn_3RHH313  ..............S
  Ph_hbn_3RHH315  ..............SQK
  Ph_hbn_3RHH314  ..............S
  Ph_hbn_3RHH336  ..............S..............................................V..................
  Ph_hbn_3RHH361  ..............S.................................................................
  Ph_hbn_3RHH366  ..............S.................................................................
  Ph_hbn_3RHH356  ................................................................................
  Ph_hbn_3RHH357  ................................................................................
  Ph_hbn_3RHH358  ...........V....................................................................
  Ph_hbn_3RHH363  ..............S......................Q..........................................
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus  ...............

                  481    490493
                  |--------+--|
      Ph_hbn_ref  TLKIDPQLPSQLS
   Ph_hbn_cDNA05  .............
    Ph_hbn_ORF01  .............
    Ph_hbn_ORF05  .............
   Ph_hbn_cDNA01  .............
    Ph_hbn_ORF02  .............
   Ph_hbn_cDNA11  L
   Ph_hbn_5RHH54
   Ph_hbn_5RHH53
   Ph_hbn_5RHH51
    Ph_hbn_5Rn25
    Ph_hbn_5Rn26
    Ph_hbn_5Rn28
  Ph_hbn_3RHH313
  Ph_hbn_3RHH315
  Ph_hbn_3RHH314
  Ph_hbn_3RHH336  .............
  Ph_hbn_3RHH361  .............
  Ph_hbn_3RHH366  .............
  Ph_hbn_3RHH356  .............
  Ph_hbn_3RHH357  .............
  Ph_hbn_3RHH358  .............
  Ph_hbn_3RHH363  .............
Ph_hbn_vsdPms078
Ph_hbn_vsdPms093
Ph_hbn_vsdPms109
       Consensus
```

[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.3_Ph_hbn]
[PhD_A2.3.2_Ph_hbn_alignments]

**“difference_alignment_Ph_hbn_all_bp_sequences.gif”**



# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.4_Ph_al1]

### [PhD_A2.4.1_Ph_al1_sequences]

### “Ph_al1_general_read_me.txt”

```
General aspects:

(i)       All sequences are in FASTA format.

(ii)      [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)     Comments are preceded by an apostrophe (')

(iv)      All sequences without olignonucleotide sequence are indicated (e.g. "w/o 1213, 1216"). In these cases, only novel
Parhyale sequence is presented.


Parhyale hawaiensis al1 terminology:


(i)       In general, sequence denotation begins with: ">Ph_al1_"

(ii)      The above (i) is followed by: "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                        "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                        "ORF"               for coherent ORF sequences
                                        "cDNA"              for complete cDNA sequences

                                        "ref"               if reference sequence is presented
                                        "HD"                if homeobox/homeodomain fragment is presented
                                        "OAR"               if OAR (encoding) fragment is presented

(iii)     The above (ii) is followed by clone id.


Parhyale hawaiensis al1 synopsis:

cDNA:              1849 bp
5'UTR:     169 bp (1-169)
ORF:               1533 bp (170-1702), 510 aa
3'UTR:     147 bp (1703-1849)
HD:        180 bp (791-970),   60 aa
OAR                   63 bp (1529-1591), 21 aa


>Ph_al1_HD [bp]
CAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCC
CGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAG

>Ph_al1_OAR [bp]
CAAGAACGACGGGCTTCTTCTATAGCGGAGCTGCGACAAAGAGCCAGAGAGCATGAGAAAAGA

>Ph_al1_ref [bp] 'Ph_al1_cDNA01, extended 5' by 35 bp of Ph_al1_5Rn2, 3' by 62 bp of Ph_al1_3Rn8, 1 nuc exchange (1048 T>C)
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAA
ATTATAAGCAGTCAAAATTGAATCATTCAAGTGAAGAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAG
TGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAGCAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTC
TCCAAGTTTCAAACCTGCAAACGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACCAAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCC
ACCTACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAAGCATAAGAATAAAAAAGAATGAAATTGATTACGACAGTAATGAGGAATCGCAAACGAGATCCCTTGCGTCAAAATCCAC
TGAATTTTCATCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCAC
CCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAAT
AGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGTCGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCA
TCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTACGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCCTACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGGGT
GCCTCGACTACCTGGCTACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCTAGGCGGGTTGGGCGCTGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGA
TTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACTTCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCT
TCGTCCCCCGTCAATCCCCGAACTCAACGACTATAGCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGCATCGGATGTTGGACCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTC
TATAGCGGAGCTGCGACAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGATTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCCCGCGATGTTC
ACCCAAAAGAGAGATCAAAACATGAATAAAATGGTAACAATTGTCAGAGAACAACATAAAACATCGATGAAAAATATTCTCGAAAAGGAGTCATATATGTAAATAGTAATGCTGGTTAGACTAAACATC
AAAGAAAACACTTCCCCAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_al1_HD [aa]
QRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQE

>Ph_al1_OAR [aa]
QERRASSIAELRQRAREHEKR

>Ph_al1_ref [aa]
MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGKRSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQ
EDSSIRIKKNEIDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQN
RRAKWRKQEKVGPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPL
RGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRSPRCSPKREIKT&
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.4_Ph_al1]

[PhD_A2.4.1_Ph_al1_sequences]

## “Ph_al1_cDNA_sequences.txt“

```
>Ph_al1_ref [bp] 'Ph_al1_cDNA01, extended 5' by 35 bp of Ph_al1_5Rn2, 3' by 62 bp of Ph_al1_3Rn8, 1 nuc exchange (1048 T>C)
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAA
ATTATAAGCAGTCAAAATTGAATCATTCAAGTGAAGAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAG
TGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAGCAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTC
TCCAAGTTTCAAACCTGCAAACGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACCAAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCC
ACCTACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAAGCATAAGAATAAAAAAGAATGAAATTGATTACGACAGTAATGAGGAATCGCAAACGAGATCCCTTGCGTCAAAATCCAC
TGAATTTTCATCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCAC
CCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAAT
AGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGTCGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCA
TCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTACGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCCTACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGGGT
GCCTCGACTACCTGGCTACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCTAGGCGGGTTGGGCGCTGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGA
TTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACTTCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCT
TCGTCCCCCGTCAATCCCCGAACTCAACGACTATAGCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGCATCGGATGTTGGACCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTC
TATAGCGGAGCTGCGACAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGATTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCCCGCGATGTTC
ACCCAAAAGAGAGATCAAAACATGAATAAAATGGTAACAATTGTCAGAGAACAACATAAAACATCGATGAAAAATATTCTCGAAAAGGAGTCATATATGTAAATAGTAATGCTGGTTAGACTAAACATC
AAAGAAAACACTTCCCCAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_al1_cDNA01 [bp] 'w/o 1213, 1216
AATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAAATTATAAGCAGTCAAAATTGAATCATTCAAGTGAA
GAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAGTGATCCAGCGACGGGTCCTCGACTCCACCCTTCAA
GCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAGCAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCTCCAAGTTTCAAACCTGCAAACGAACAAGAAACCG
CCACCCAAACCGAGGTGCAAACTTCAATACACAAACCAAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCTACCCCGCACATCAATGTAGTCGATGATGACG
ATCAAGAAGACTCAAGCATAAGAATAAAAAAGAATGAAATTGATTACGACAGTAATGAGGAATCGCAAACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCATCTCAAAGCAGTTCAAAGGAATTGC
AGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCACCCCTAACAAGAAAAAACAACGTCGATATAGAACAA
CGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCC
AAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGTCGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCATCCTCCTCCACTTCTTTCCTCTTTTATTGGCACTA
CGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCCTACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGGGTGCCTCGACTACCTGGCTACCTCACCGGTCTTGTGC
CGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCTAGGCGGGTTGGGCGCTGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCC
CACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACTTCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATCCCCGAACTCAACGACTATA
GCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGCATCGGATGTTGGACCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAGCTGCGACAAAGAGCCAGAGAGCATG
AGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGATTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCCCGCGATGTTCACCCAAAAGAGAGATCAAAACATGAATAAAATGGT
AACAATTGTCAGAGAACAACATAAAACATCGATGAAAAATATTCTCGAAAAGGAGTCATATATGTAAATAGTAAT

>Ph_al1_cDNA04 [bp] 'w/o 1213, 1216
AATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTGCTGCTGTATAAAAAAATTATAAGCAGTCAAAATTGAATCATTCAAGTGAA
GAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAGTGATCCAGCGACGGGTCCTCGACTCCACCCTTCAA
GCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAGCAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCTCCAAGTTTCAAACCTGCAAACGAACAAGAAACCG
CCACCCAAACCGAGGTGCAAACTTCAATACACAAACCAAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCTACCCCGCACATCAATGTAGTCGATGATGACG
ATCAAGAAGACTCAAGCATAAGAATAAAAAAGAATGAAATTGATTACGACAGTAATGAGGAATCGCAAACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCATCTCAAAGCAGTTCAAAGGAATTGC
AGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCACCCCTAACAAGAAAAAACAACGTCGATATAGAACAA
CGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCC
AAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGTCGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCATCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTA
CGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCCTACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGGGTGCCTCGACTACCTGGCTACCTCACCGGTCTTGTGC
CGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCTAGGCGGGTTGGGCGCTGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCC
CACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACTTCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATCCCCGAACTCAACGACTACA
GCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGCATCGGGTGTTGGACCCAATAATTCTGACAGCAAAGAACAAGATCGACGGGCTTCTTCTATAGCGGAGCTGCGACAAAGAGCCAGAGAGCATG
AGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGATTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCCCGCGATGTTCACCCAAAAGAGAGATCAAAACATGAATAAAATGGT
AACAATTGTCAGAGAACAACATAAAACATCGATGAAAAATATTCTCGAAAAGGAGTCATATATGTAAATAGTAAT

>Ph_al1_cDNA05 [bp] 'w/o 1213, 1216
AATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAAATTATAAGCAGTCAAAATTGAATCATTCAAGTGAA
GAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAGTGATCCAGCGACGGGTCCTCGACTCCACCCTTCAA
GCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAGCAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCTCCAAGTTTCAAACCTGCAAACGAACAAGAAACCG
CCACCCAAACCGAGGTGCAAACTTCAATACACAAACCAAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCTACCCCGCACATCAATGTAGTCGATGATGACG
ATCAAGAAGACTCAAGCATAAGAATAAAAAAGAATGAAATTGATTACGACAGTAATGAGGAATCGCAAACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCATCTCAAAGCAGTTCAAAGGAATTGC
AGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCACCCCTAACAAGAAAAAACAACGTCGATATAGAACAG
CGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCC
AAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGTCGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCATCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTA
CGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCCTACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGGGTGCCTCGACTACCTGGCTACCTCACCGGTCTTGTGT
CGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCTAGGCGGGTTGGGCGCTGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCC
CACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACTTCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATCCCCGAACTCAACGACTATA
GCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGCATCGGATGTTGGACCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCCATAGCGGAGCTGCGACAAAGAGCCAGAGAGCATG
AGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGATTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCCCGCGATGTTCACCCAAAAGAGAGATCAAAACATGAATAAAATGGT
AACAATTGTCAGAGAACAACATAAAACATCGATGAAAAATATTCTCGAAAAGGAGTCATATATGTAAATAGTAAT

>Ph_al1_ORF13 [bp] 'w/o 1214, 1215
CGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACTTCCCCCATCACCAGTGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAA
CAATCACAACAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCCCCAAGTTTCAAACCCGCAAGCGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACC
AAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCAACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAGGTGTAAGAATAAATAAGAATGA
AATCGATTACGACAGTAATGAGGATTCACAAACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCGTCTCAAAGCAGTTCAAAGGAATTGCAAAAAGACGCAGAAGATTGTACAGAAGATGACGCCGA
AATTGATATCAATGGAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGCCACCCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGC
GTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGT
CGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCTCTTCATCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTACATACACTTCACCATTCGCTCCTCCACCTCATCCTCC
TACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTTCCGATACTTCCCGGGGTGCCCCGACTACCTGGCTACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCCTGGGAGGTCTCTCCAGTGT
AGGCGGGTTGGGCGCCGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACT
```

```
TCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATCCCCGAACTCAACGACTATAGTTCCATCTTGGCGAGCTCGGCCGCTCTTTCGACGGC
ATCGGATGTTGGGCCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAACTGAGACAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGA
TTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACTCGGAGATCCC

>Ph_al1_ORF14 [bp] 'w/o 1214, 1215
CGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACTTCCCCCATCACCAGTGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAA
CAATCACAACAACGACAATTTAAATAACAACTTAACCGGTGACGGTAAAAGATCTTCCCCAAGTTTCAAACCCGCAAGCGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACC
AAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCAACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAGGTGTAAGAATAAATAAGAATGA
AATCGATTACGACAGTAATGAGGATTCACAAACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCGTCTCAAAGCAGTTCAAAGGAATTGCAAAAAGACGCAGAAGATTGTACAGAAGATGACGCCGA
AATTGATATCAATGGAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGCCACCCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGC
GTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGT
CGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCTCTTCATCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTACATACACTTCACCATTCGCTCCTCCACCTCATCCTCC
TACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTTCCGATACTTCCCGGGGTGCCCCGACTACCTGGCTACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCCTGGGAGGTCTCTCCAGTGT
AGGCGGGTTGGGCGCCGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACT
TCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATCCCCGAACTCAACGACTATAGTTCCATCTTGGCGAGCTCGGTCGCTCTTTCGACGGC
ATCGGATGTTGGGCCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAACTGAGACAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGA
TTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACTCGGAGATCCC

>Ph_al1_ORF15 [bp] 'w/o 1214, 1215
CGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACTTCCCCCATCACCAGTGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAA
CAATCACAACAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCCCCAAGTTTCAAACCCGCAAGCGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACC
AAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCAACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAGGTGTAAGAATAAATAAGAATGA
AATCGATTACGACAGTAATGAGGATTCACAAACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCGTCTCAAAGCAGTTCAAAGGAATTGCAAAAAGACGCAGAAGATTGTACAGAAGATGACGCCGA
AATTGATATCAATGGAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGCCACCCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGC
GTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGT
CGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCTCTTCATCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTACATACACTTCACCATTCGCTCCTCCACCTCATCCTCC
TACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTTCCGATACTTCCCGGGGTGCCCCGACTACCTGGCTACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCCTGGGAGGTCTCTCCAGTGT
AGGCGGGTTGGGCGCCGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACT
TCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATCCCCGAACTCAACGACTATAGTTCCATCTTGGCGAGCTCGGCCGCTCTTTCGACGGC
ATCGGATGTTGGGCCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAACTGAGACAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGA
TTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACTCGGAGATCCC

>Ph_al1_ORF16 [bp] 'w/o 1214, 1215
CGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAGTGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAA
CAATCACAGCAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCTCCAAGTTTCAAACCTGCAAACGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACC
AAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCTACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGGCTCAAGCATAAGAATAAAAAAGAATGA
AATTGATTACGACAGTAATGAGGAATCGCAAACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCATCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGA
AATTGATATCAACGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCACCCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGC
GTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGT
CGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCATCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTACGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCC
TACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGGGTGCCTCGACTACCTGGCTACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCT
AGGCGGGTTGGGCGCTGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACT
TCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTATTCCTGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATCCCCGAACTCAACGACTATAGCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGC
ATCGGATGTTGGACCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAGCTGCGACAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGA
TTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCC

>Ph_al1_ORF17 [bp] 'w/o 1214, 1215
CGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAGTGATTCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAA
CAATCACAGCAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCTCCAAGTTTCAAACCTGCAAACGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACC
AAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCTACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAAGCATAAGAATAAAAAAGAATGA
AATTGATTACGACAGTAATGAGGAATCGCAAACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCATCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGA
AATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCACCCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGC
GTTCTCCAGAACACATTACCCAGATGTCTTCACGAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGT
CGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCATCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTACGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCC
TACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGGGTGCCTCGACTACCTGGCTACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCT
AGGCGGGTTGGGCGCTGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAGCGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACT
TCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATCCCCGAACTCAACGACTATAGCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGC
ATCGGATGTTGGACCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAGCTGCGACAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGA
TTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCC

>Ph_al1_ORF18 [bp] 'w/o 1214, 1215
CGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACTTCCCCCATCACCAGTGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAA
CAATCACAACAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCCCCAAGTTTCAAACCCGCAAGCGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACC
AAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCAACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAGGTGTAAGAATAAATAAGAATGA
AATCGATTACGACAGTAATGAGGATTCACAAACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCGTCTCAAAGCAGTTCAAAGGAATTGCAAAAAGACGCAGAAGATTGTACAGAAGATGGCGCCGA
AATTGATATCAATGGAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGCCACCCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGC
GTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGT
CGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCTCTTCATCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTACATACACTTCACCATTCGCTCCTCCACCTCATCCTCC
TACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTTCCGATACTTCCCGGGGTGCCCCGACTACCTGGCTACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCCTGGGAGGTCTCTCCAGTGT
AGGCGGGTTGGGCGCCGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACT
TCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATCCCCGAACTCAACGACTATAGTTCCATCTTGGCGAGCTCGGCCGCTCTTTCGACGGC
ATCGGATGTTGGGCCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAACTGAGACAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGA
TTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACTCGGAGATCCC

>Ph_al1_5Rn2 [bp] 'w/o 1208
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAA
ATTATAAGCAGTCAAAATTGAATCATTCAAGTGAAGAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACTTCCCCCATCACCAG
TGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCTCAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAACAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTC
CCCAAGTTTCAAACCCGCAAGCGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACCTAATAACAAACGAAGTTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCC
ACCAACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAGGCATAAGAATAAAAAAAGAATGAAATTGATTACGACAGTAATGAGGATTCACAAACGAGATCCCTTGCGTCAAAATCCA
CTGAATTTTCGTCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGAAGATGAAGCCA
CCCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAA

>Ph_al1_5Rn3 [bp] 'w/o 1208
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAA
TTATAAGCAGTCAAAATTGAATCATTCAAGTGAAGAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACTTCCCCCATCACCAGT
GATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAACAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCC
CCAAGTTTCAAACCCGCAAGCGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACCTAATAACAAACGAAGTTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCA
CCAACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAGGCATAAGAATAAAAAAGAATGAAATTGATTACGACAGTAATGAGGATTCACAAACGAGATCCCTTGCGTCAAAATCCACT
```

```
GAATTTTCGTCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGAAGATGAAGCCACC
CCTAACAAGAAAAGACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAA

>Ph_al1_5Rn4 [bp] 'w/o 1208
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAA
ATTATAAGCAGTCAAAATTGAATCATTCAAGTGAAGAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACTTCCCCCATCACCAG
TGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAACAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTC
CCCAAGTTTCAAACCCGCAAGCGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACCTAATAACAAACGAAGTTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCC
ACCAACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAGGCATAAGAATAAAAAAAGAATGAAATTGATTACGACAGTAATGAGGATTCACAAACGAGATCCCTTGCGTCAAAATCCA
CTGAATCTTCGTCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGAAGATGAAGCCA
CCCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAA

>Ph_al1_5Rn118 [bp] 'w/o 1208
GAATTTTCACTGTTCCCTTTACCACGTTTTACTGAATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAAA
TTATAAGCAGTCAAAATTGAATCATTCAAGTGAAGAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAGT
GATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAGCAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCT
CCAAGTTTCAAACCTGCAAACGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACCAAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCA
CCTACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAAGCATAAGAATAAAAAAGAATGAAATTGATTACGACAGTAATGAGGAATCGCAAACGAGATCCCTTGCGTCAAAATCCACT
GAATTTTCATCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCACC
CCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAA

>Ph_al1_5Rn119 [bp] 'w/o 1208, id to Ph_al1_5Rn149
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTTTAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAA
ATTATAAGCAGTCAAAATTGAATCATTCAAGTGAAGAAATATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAG
TGATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAGCAACGACAATTTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTC
TCCAAGTTTCAAACCTGCAAACGAACAAGAAACCGCCACCCAAACCGAGGTGCAAACTTCAATACACAAACCAAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCC
ACCTACCCCGCACATCAATGTAGTCGATGATGACGATCAAGAAGACTCAAGCATAAGAATAAAAAAGAATGAAATTGATTACGACAGTAATGAGGAATCGCAAACGAGATCCCTTGCGTCAAAATCCAC
TGAATTTTCATCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGATGACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCAC
CCCTAACAAGAAAAAACAACGTCGATATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAA

>Ph_al1_vsdP_Phad16 [bp] 'w/o fw_QRR, re_QVW
GTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATC

>Ph_al1_vsdP_Pha092 [bp] 'w/o fw_QRR, re_QVW
GTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATC

>Ph_al1_vsdP_Pha097 [bp] 'w/o fw_QRR, re_QVW
GTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATC

>Ph_al1_vsdP_Pha102 [bp] 'w/o fw_QRR, re_QVW
GTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAACTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATC

>Ph_al1_3Rn5 [bp] 'w/o 1203
CGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGTCGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCATCCTCCTCCACTTCCT
TCCTCTTTTATTGGCACTACGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCCTACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGGGTGCCTCGACTACCTGGC
TACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCTAGGCGGGTTGGGCGCTGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTC
AGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACTTCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATC
CCCGAACTCAACGACTATAGCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGCATCGGATGTTGGACCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAGCTGCGA
CAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAGGAACAAAGATTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCCCGCGATGTTCACCCAAAAGAGAGATC
AAAACATGAATAAAATGGTAACAATTGTCAGAGAACAACATAAAACATCGATGAAAAATATTCTCGAAAAGGAGTCATATATGTAAATAGTAATGCTGGTTAGACTAAACATCAAAGAAAACACTTCCC
CAAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_al1_3Rn7 [bp] 'w/o 1203
CGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGTCGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCGTCTTCGGTCCCTCTTCATCCTCCTCCACTTCCT
TCCTCTTTTATTGGCACTACGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCCTACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTTCCGATACTTCCCGGGGTGCCTCGACTACCTGGC
TACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCTTGGCGGGTTGGGCGCTGGTGCACTGAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTT
AGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACTTCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCGGGTCTTCGTCCTCCGTCAATT
CCCGAACTCAACGACTATAGTTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGCATCGGATGTTGGGCCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAGCTGCGA
CAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGATTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCCCGCGATGTTCACCCAAAAGAGAGATC
AAAACATAAATAAAATGGTAACAATTTGCAGAGAGCAACATAAAACATCGATGAATAATATTCTCGAAAAGGAGTCATAATGTAAATAGTAATGTTGCTTAGACTAAACATCAAAGAAAACACTTCCCC
AAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_al1_3Rn8 [bp] 'w/o 1203
CGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGTCGGACCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCATCCTCCTCCACTTCCT
TCCTCTTTTATTGGCACTACGTACACTTCACCGTTCGCTCCTCCACCTCATCCTCCTACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGGGTGCCTCGACTACCTGGC
TATCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCTAGGCGGGTTGGGCGCTGGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTC
AGAAACGATCTGGGGAGCCCACTGAGAGGAGATTTAGCTCCACGACTTCCCCTACTTCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCGTCAATC
CCCGAACTCAACGACTATAGCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGCATCGGATGTTGGACCCAATAATTCTGACAGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAGCTGCGA
CAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGGAAGAACAAAGATTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCCCGCGATGTTCACCCAAAAGAGAGATC
AAAACATGAATAAAATGGTAACAATTGTCAGAGAACAACATAAAACATCGATGAAAAATATTCTCGAAAAGGAGTCATATATGTAAATAGTAATGCTGGTTAGACTAAACATCAAAGAAAACACTTCCC
CAAAAAAAAAAAAAAAAAAAAAAAAAA


>Ph_al1_ref [aa]
MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGKRSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQ
EDSSIRIKKNEIDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQN
RRAKWRKQEKVGPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPL
RGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRSPRCSPKREIKT&

>Ph_al1_cDNA01 [aa] 'w/o 1213, 1216
MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGKRSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQ
EDSSIRIKKNEIDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQN
RRAKWRKQEKVGPSHHLAGPFPSSSSVPLHPPPLLSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPL
RGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRSPRCSPKREIKT&

>Ph_al1_cDNA04 [aa] 'w/o 1213, 1216
MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGKRSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQ
EDSSIRIKKNEIDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQN
RRAKWRKQEKVGPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPL
```

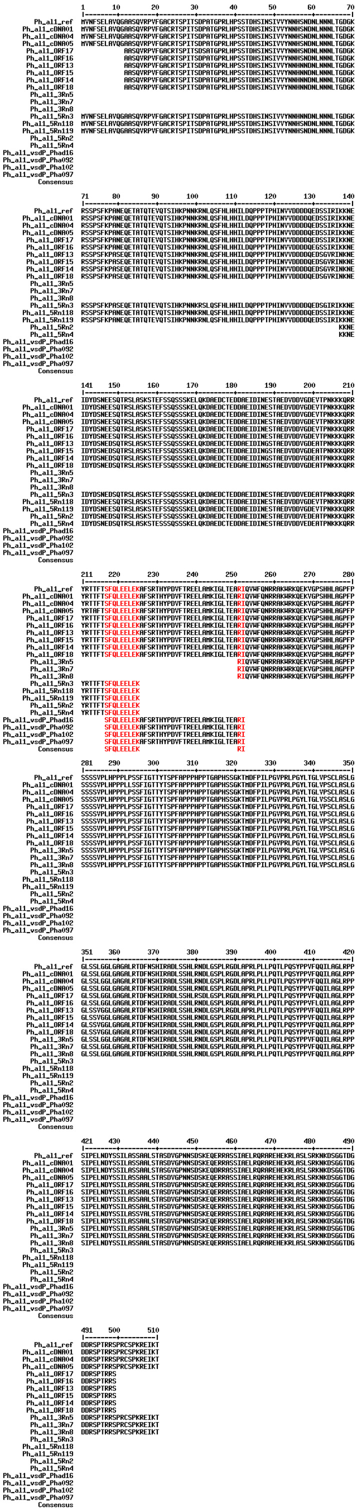
```
RGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASGVGPNNSDSKEQDRRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRSPRCSPKREIKT&

>Ph_al1_cDNA05 [aa] 'w/o 1213, 1216
MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGKRSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQ
EDSSIRIKKNEIDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRRYRTAFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQN
RRAKWRKQEKVGPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVSSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPL
RGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRSPRCSPKREIKT&

>Ph_al1_ORF13 [aa] 'w/o 1214, 1215
AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGKRSSPSFKPASEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSGVRINKNE
IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINGSTAEDVDDVGDEATPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKV
GPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSVGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLL
PQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRS

>Ph_al1_ORF14 [aa] 'w/o 1214, 1215
AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGKRSSPSFKPASEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSGVRINKNE
IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINGSTAEDVDDVGDEATPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKV
GPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSVGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLL
PQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSVALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRS

>Ph_al1_ORF15 [aa] 'w/o 1214, 1215
AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGKRSSPSFKPASEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSGVRINKNE
IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINGSTAEDVDDVGDEATPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKV
GPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSVGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLL
PQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRS

>Ph_al1_ORF16 [aa] 'w/o 1214, 1215
AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGKRSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEGSSIRIKKNE
IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKV
GPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLL
PQTLPQSYPPVFLQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRS

>Ph_al1_ORF17 [aa] 'w/o 1214, 1215
AASQVRPVFGACRTSPITSDSATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGKRSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSSIRIKKNE
IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKV
GPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHLRSDLGSPLRGDLAPRLPLL
PQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRS

>Ph_al1_ORF18 [aa] 'w/o 1214, 1215
AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGKRSSPSFKPASEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSGVRINKNE
IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDGAEIDINGSTAEDVDDVGDEATPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKV
GPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSVGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLL
PQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRS

>Ph_al1_5Rn2 [aa] 'w/o 1208
KKNEIDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVEDEATPNKKKQRRYRTTFTSFQLEELEK

>Ph_al1_5Rn3 [aa] 'w/o 1208
MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGKRSSPSFKPASEQETATQTEVQTSIHKPNNKRSLQSFHLHHILDQPPPTPHINVVDDDDQ
EDSGIRIKKNEIDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVEDEATPNKKRQRRYRTTFTSFQLEELEK

>Ph_al1_5Rn4 [aa] 'w/o 1208
KKNEIDYDSNEDSQTRSLASKSTESSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVEDEATPNKKKQRRYRTTFTSFQLEELEK

>Ph_al1_5Rn118 [aa] 'w/o 1208
MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGKRSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQ
EDSSIRIKKNEIDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRRYRTTFTSFQLEELEK

>Ph_al1_5Rn119 [aa] 'w/o 1208, id to Ph_al1_5Rn149
MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGKRSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQ
EDSSIRIKKNEIDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRRYRTTFTSFQLEELEK

>Ph_al1_vsdP_Phad16 [aa] 'w/o fw_QRR, re_QVW
SFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARI

>Ph_al1_vsdP_Pha092 [aa] 'w/o fw_QRR, re_QVW
SFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARI

>Ph_al1_vsdP_Pha097 [aa] 'w/o fw_QRR, re_QVW
SFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARI

>Ph_al1_vsdP_Pha102 [aa] 'w/o fw_QRR, re_QVW
SFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARI

>Ph_al1_3Rn5 [aa] 'w/o 1203
RIQVWFQNRRAKWRKQEKVGPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHL
RNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRRNKDSGGTDGDDRSPTRRSPRCSPKREI
KT

>Ph_al1_3Rn7 [aa] 'w/o 1203
RIQVWFQNRRAKWRKQEKVGPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHL
RNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRSPRCSPKREI
KT

>Ph_al1_3Rn8 [aa] 'w/o 1203
RIQVWFQNRRAKWRKQEKVGPSHHLAGPFPSSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLGGLSSLGGLGAGALRTDFNSHIRADLSSHL
RNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPPSIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDGDDRSPTRRSPRCSPKREI
KT
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.4_Ph_al1]

### [PhD_A2.4.2_Ph_al1_alignments]

#### **“alignment_Ph_al1_all_aa_sequences.gif”**



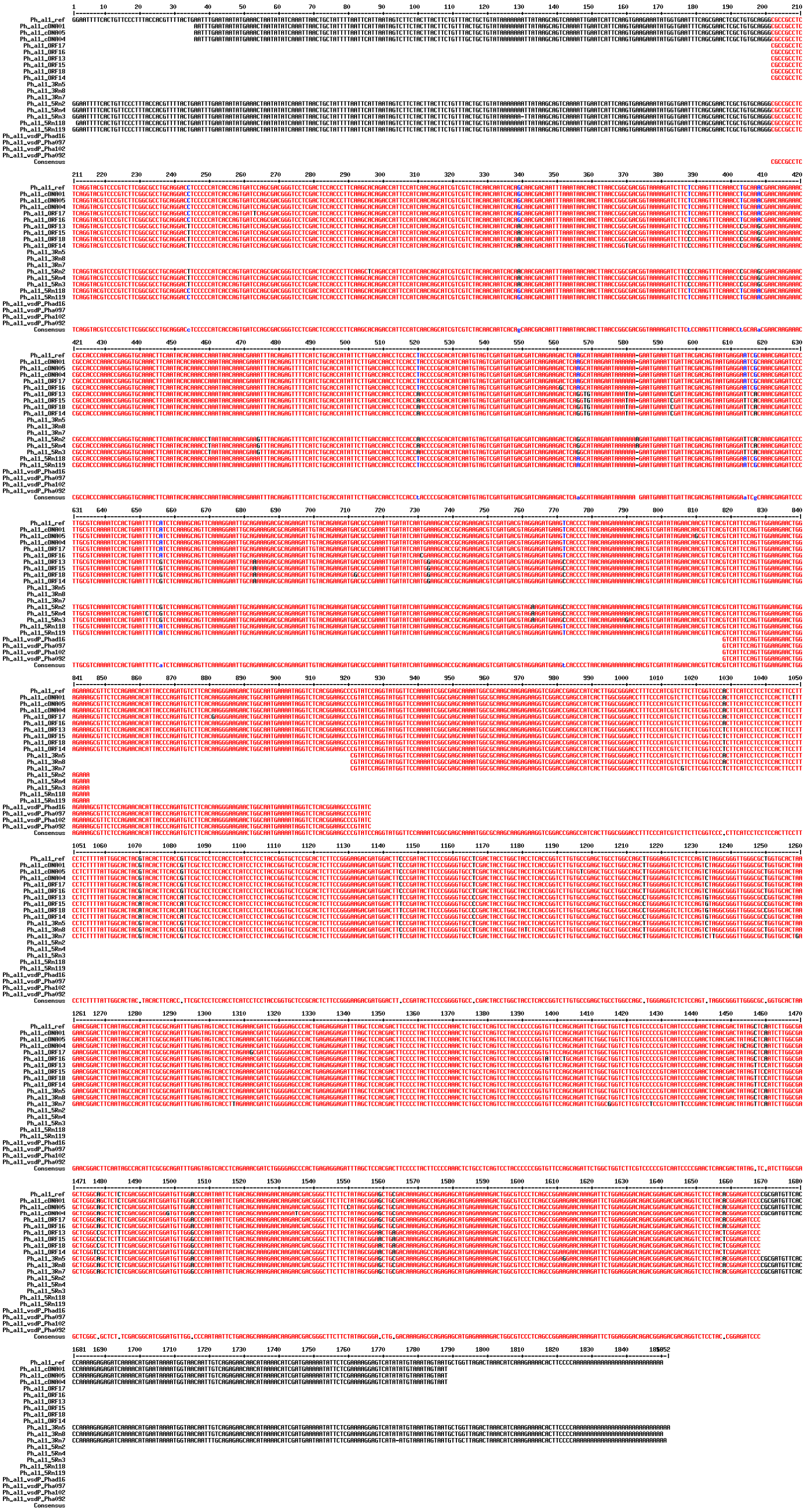
```
                   1       10        20        30        40        50        60        70
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGK
     Ph_al1_cDNA01 MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGK
     Ph_al1_cDNA04 MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGK
     Ph_al1_cDNA05 MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGK
      Ph_al1_ORF17            AASQVRPVFGACRTSPITSDSATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGK
      Ph_al1_ORF16            AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGK
      Ph_al1_ORF13            AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGK
      Ph_al1_ORF15            AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGK
      Ph_al1_ORF14            AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGK
      Ph_al1_ORF18            AASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGK
       Ph_al1_3Rn5
       Ph_al1_3Rn7
       Ph_al1_3Rn8
       Ph_al1_5Rn3 MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHNNDNLNNNLTGDGK
     Ph_al1_5Rn118 MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGK
     Ph_al1_5Rn119 MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGK
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha092
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
         Consensus

                   71      80        90       100       110       120       130       140
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref RSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSSIRIKKNE
     Ph_al1_cDNA01 RSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSSIRIKKNE
     Ph_al1_cDNA04 RSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSSIRIKKNE
     Ph_al1_cDNA05 RSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSSIRIKKNE
      Ph_al1_ORF17 RSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSSIRIKKNE
      Ph_al1_ORF16 RSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEGSSIRIKKNE
      Ph_al1_ORF13 RSSPSFKPASEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSGVRINKNE
      Ph_al1_ORF15 RSSPSFKPASEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSGVRINKNE
      Ph_al1_ORF14 RSSPSFKPASEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSGVRINKNE
      Ph_al1_ORF18 RSSPSFKPASEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSGVRINKNE
       Ph_al1_3Rn5
       Ph_al1_3Rn7
       Ph_al1_3Rn8
       Ph_al1_5Rn3 RSSPSFKPASEQETATQTEVQTSIHKPNNKRSLQSFHLHHILDQPPPTPHINVVDDDDQEDSGIRIKKNE
     Ph_al1_5Rn118 RSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSSIRIKKNE
     Ph_al1_5Rn119 RSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSSIRIKKNE
       Ph_al1_5Rn2                                                                   KKNE
       Ph_al1_5Rn4                                                                   KKNE
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha092
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
         Consensus

                   141     150       160       170       180       190       200       210
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRR
     Ph_al1_cDNA01 IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRR
     Ph_al1_cDNA04 IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRR
     Ph_al1_cDNA05 IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRR
      Ph_al1_ORF17 IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRR
      Ph_al1_ORF16 IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRR
      Ph_al1_ORF13 IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINGSTAEDVDDVGDEATPNKKKQRR
      Ph_al1_ORF15 IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINGSTAEDVDDVGDEATPNKKKQRR
      Ph_al1_ORF14 IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINGSTAEDVDDVGDEATPNKKKQRR
      Ph_al1_ORF18 IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDGAEIDINGSTAEDVDDVGDEATPNKKKQRR
       Ph_al1_3Rn5
       Ph_al1_3Rn7
       Ph_al1_3Rn8
       Ph_al1_5Rn3 IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVEDEATPNKKRQRR
     Ph_al1_5Rn118 IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRR
     Ph_al1_5Rn119 IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRR
       Ph_al1_5Rn2 IDYDSNEDSQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVEDEATPNKKKQRR
       Ph_al1_5Rn4 IDYDSNEDSQTRSLASKSTESSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVEDEATPNKKKQRR
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha092
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
         Consensus

                   211     220       230       240       250       260       270       280
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
     Ph_al1_cDNA01 YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
     Ph_al1_cDNA04 YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
     Ph_al1_cDNA05 YRTAFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
      Ph_al1_ORF17 YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
      Ph_al1_ORF16 YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
      Ph_al1_ORF13 YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
      Ph_al1_ORF15 YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
      Ph_al1_ORF14 YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
      Ph_al1_ORF18 YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
       Ph_al1_3Rn5                                         RIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
       Ph_al1_3Rn7                                         RIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
       Ph_al1_3Rn8                                         RIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
       Ph_al1_5Rn3 YRTTFTSFQLEELEK
     Ph_al1_5Rn118 YRTTFTSFQLEELEK
     Ph_al1_5Rn119 YRTTFTSFQLEELEK
       Ph_al1_5Rn2 YRTTFTSFQLEELEK
       Ph_al1_5Rn4 YRTTFTSFQLEELEK
Ph_al1_vsdP_Phad16       SFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARI
Ph_al1_vsdP_Pha092       SFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARI
Ph_al1_vsdP_Pha102       SFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARI
Ph_al1_vsdP_Pha097       SFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARI
         Consensus       SFQLEELEK                         RI

                   281     290       300       310       320       330       340       350
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
     Ph_al1_cDNA01 SSSSVPLHPPPLLSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
     Ph_al1_cDNA04 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
     Ph_al1_cDNA05 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVSSCLASLG
      Ph_al1_ORF17 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
      Ph_al1_ORF16 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
      Ph_al1_ORF13 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
      Ph_al1_ORF15 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
      Ph_al1_ORF14 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
      Ph_al1_ORF18 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
       Ph_al1_3Rn5 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
       Ph_al1_3Rn7 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
       Ph_al1_3Rn8 SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
       Ph_al1_5Rn3
     Ph_al1_5Rn118
     Ph_al1_5Rn119
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha092
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
         Consensus

                   351     360       370       380       390       400       410       420
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref GLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
     Ph_al1_cDNA01 GLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
     Ph_al1_cDNA04 GLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
     Ph_al1_cDNA05 GLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
      Ph_al1_ORF17 GLSSLGGLGAGALRTDFNSHIRADLSSHLRSDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
      Ph_al1_ORF16 GLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFLQILAGLRPP
      Ph_al1_ORF13 GLSSVGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
      Ph_al1_ORF15 GLSSVGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
      Ph_al1_ORF14 GLSSVGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
      Ph_al1_ORF18 GLSSVGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
       Ph_al1_3Rn5 GLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
       Ph_al1_3Rn7 GLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
       Ph_al1_3Rn8 GLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
       Ph_al1_5Rn3
     Ph_al1_5Rn118
     Ph_al1_5Rn119
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha092
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
         Consensus

                   421     430       440       450       460       470       480       490
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
     Ph_al1_cDNA01 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
     Ph_al1_cDNA04 SIPELNDYSSILASSAALSTASGVGPNNSDSKEQDRRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
     Ph_al1_cDNA05 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
      Ph_al1_ORF17 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
      Ph_al1_ORF16 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
      Ph_al1_ORF13 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
      Ph_al1_ORF15 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
      Ph_al1_ORF14 SIPELNDYSSILASSVALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
      Ph_al1_ORF18 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
       Ph_al1_3Rn5 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRRNKDSGGTDG
       Ph_al1_3Rn7 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
       Ph_al1_3Rn8 SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
       Ph_al1_5Rn3
     Ph_al1_5Rn118
     Ph_al1_5Rn119
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha092
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
         Consensus

                   491     500       510
                   |--------+---------|
        Ph_al1_ref DDRSPTRRSPRCSPKREIKT
     Ph_al1_cDNA01 DDRSPTRRSPRCSPKREIKT
     Ph_al1_cDNA04 DDRSPTRRSPRCSPKREIKT
     Ph_al1_cDNA05 DDRSPTRRSPRCSPKREIKT
      Ph_al1_ORF17 DDRSPTRRS
      Ph_al1_ORF16 DDRSPTRRS
      Ph_al1_ORF13 DDRSPTRRS
      Ph_al1_ORF15 DDRSPTRRS
      Ph_al1_ORF14 DDRSPTRRS
      Ph_al1_ORF18 DDRSPTRRS
       Ph_al1_3Rn5 DDRSPTRRSPRCSPKREIKT
       Ph_al1_3Rn7 DDRSPTRRSPRCSPKREIKT
       Ph_al1_3Rn8 DDRSPTRRSPRCSPKREIKT
       Ph_al1_5Rn3
     Ph_al1_5Rn118
     Ph_al1_5Rn119
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha092
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
         Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.4_Ph_al1]

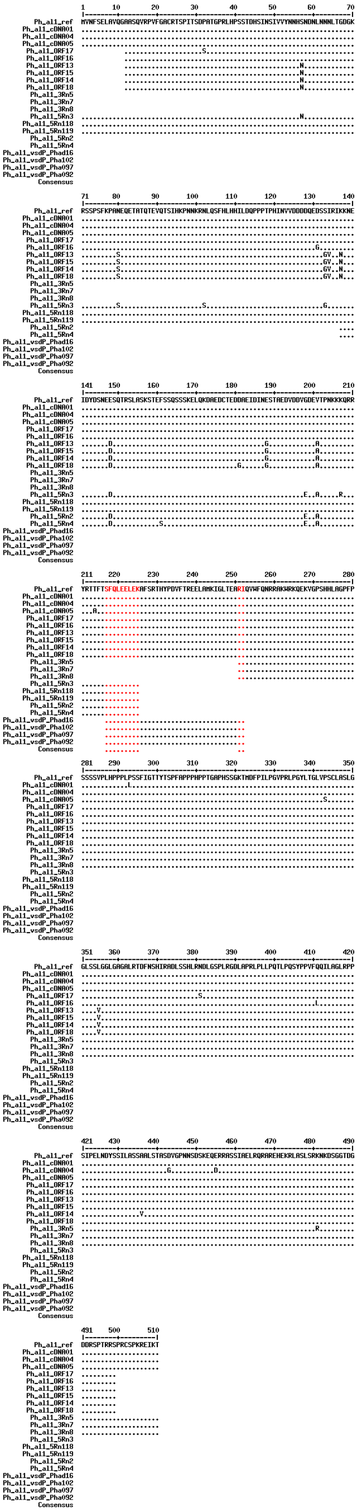
[PhD_A2.4.2_Ph_al1_alignments]

## “alignment_Ph_al1_all_bp_sequences.gif”



[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.4_Ph_al1]
[PhD_A2.4.2_Ph_al1_alignments]

## **“difference_alignment_Ph_al1_all_aa_sequences.gif”**



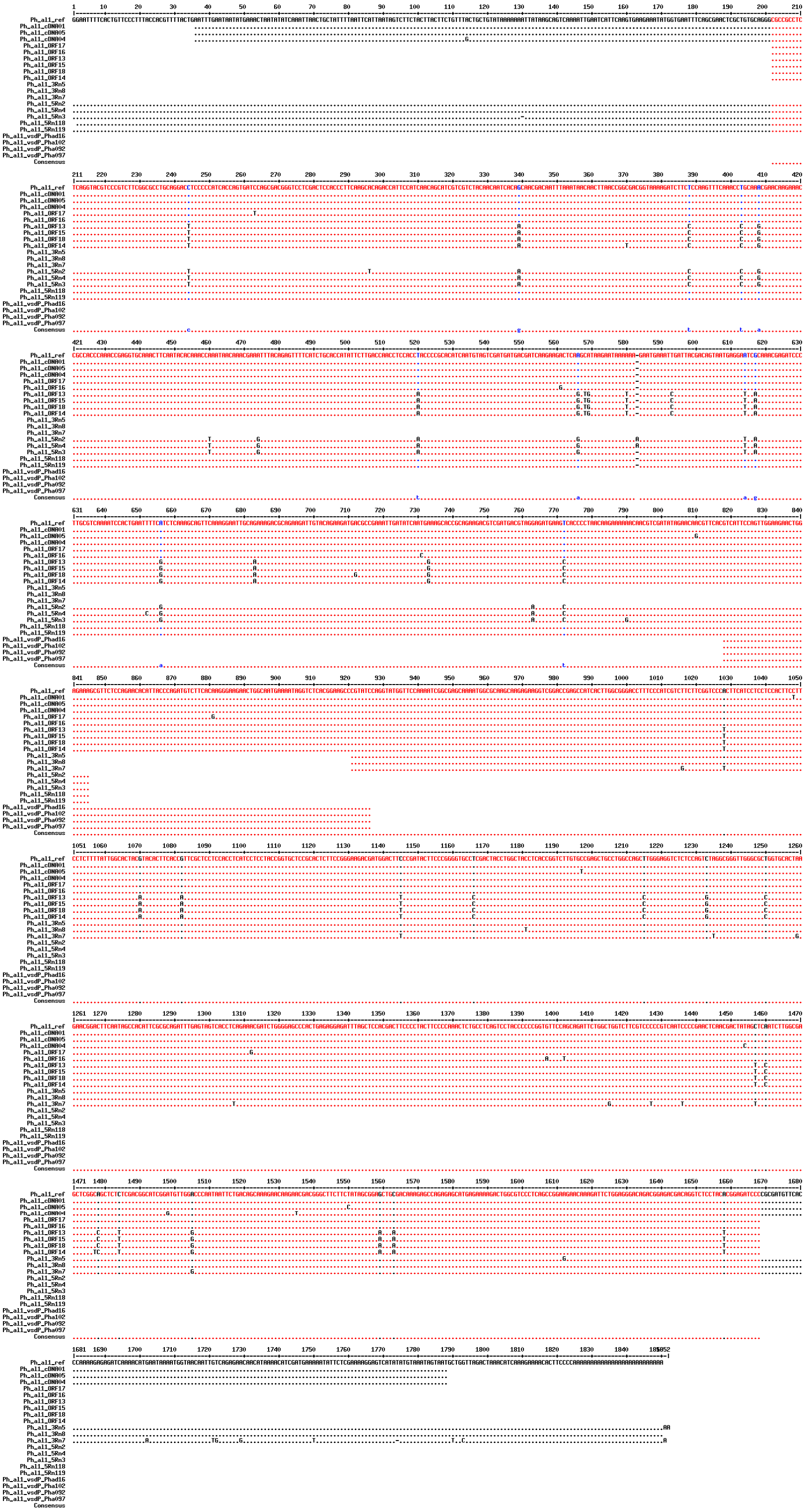
```
                   1       10        20        30        40        50        60        70
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref MVNFSELAVQGAASQVRPVFGACRTSPITSDPATGPRLHPSSTDHSINSIVVYNNHSNDNLNNNLTGDGK
     Ph_al1_cDNA01 ......................................................................
     Ph_al1_cDNA04 ......................................................................
     Ph_al1_cDNA05 ......................................................................
      Ph_al1_ORF17            ....................S......................................
      Ph_al1_ORF16            ...........................................................
      Ph_al1_ORF13            .............................................N.............
      Ph_al1_ORF15            .............................................N.............
      Ph_al1_ORF14            .............................................N.............
      Ph_al1_ORF18            .............................................N.............
       Ph_al1_3Rn5
       Ph_al1_3Rn7
       Ph_al1_3Rn8
       Ph_al1_5Rn3 ........................................................N.............
     Ph_al1_5Rn118 ......................................................................
     Ph_al1_5Rn119 ......................................................................
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
Ph_al1_vsdP_Pha092
         Consensus

                   71      80        90        100       110       120       130      140
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref RSSPSFKPANEQETATQTEVQTSIHKPNNKRNLQSFHLHHILDQPPPTPHINVVDDDDQEDSSIRIKKNE
     Ph_al1_cDNA01 ......................................................................
     Ph_al1_cDNA04 ......................................................................
     Ph_al1_cDNA05 ......................................................................
      Ph_al1_ORF17 ......................................................................
      Ph_al1_ORF16 ............................................................G.........
      Ph_al1_ORF13 .........S....................................................GV..N...
      Ph_al1_ORF15 .........S....................................................GV..N...
      Ph_al1_ORF14 .........S....................................................GV..N...
      Ph_al1_ORF18 .........S....................................................GV..N...
       Ph_al1_3Rn5
       Ph_al1_3Rn7
       Ph_al1_3Rn8
       Ph_al1_5Rn3 .........S.....................S..............................G.......
     Ph_al1_5Rn118 ......................................................................
     Ph_al1_5Rn119 ......................................................................
       Ph_al1_5Rn2                                                                   ....
       Ph_al1_5Rn4                                                                   ....
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
Ph_al1_vsdP_Pha092
         Consensus

                   141     150       160       170       180       190       200      210
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref IDYDSNEESQTRSLASKSTEFSSQSSSKELQKDAEDCTEDDAEIDINESTAEDVDDVGDEVTPNKKKQRR
     Ph_al1_cDNA01 ......................................................................
     Ph_al1_cDNA04 ......................................................................
     Ph_al1_cDNA05 ......................................................................
      Ph_al1_ORF17 ......................................................................
      Ph_al1_ORF16 ......................................................................
      Ph_al1_ORF13 .......D.......................................G............A.........
      Ph_al1_ORF15 .......D.......................................G............A.........
      Ph_al1_ORF14 .......D.......................................G............A.........
      Ph_al1_ORF18 .......D................................G......G............A.........
       Ph_al1_3Rn5
       Ph_al1_3Rn7
       Ph_al1_3Rn8
       Ph_al1_5Rn3 .......D.................................................E..A.....R...
     Ph_al1_5Rn118 ......................................................................
     Ph_al1_5Rn119 ......................................................................
       Ph_al1_5Rn2 .......D.................................................E..A.........
       Ph_al1_5Rn4 .......D............S....................................E..A.........
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
Ph_al1_vsdP_Pha092
         Consensus

                   211     220       230       240       250       260       270      280
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref YRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLAGPFP
     Ph_al1_cDNA01 ......................................................................
     Ph_al1_cDNA04 ......................................................................
     Ph_al1_cDNA05 ...A..................................................................
      Ph_al1_ORF17 ......................................................................
      Ph_al1_ORF16 ......................................................................
      Ph_al1_ORF13 ......................................................................
      Ph_al1_ORF15 ......................................................................
      Ph_al1_ORF14 ......................................................................
      Ph_al1_ORF18 ......................................................................
       Ph_al1_3Rn5                                         ..............................
       Ph_al1_3Rn7                                         ..............................
       Ph_al1_3Rn8                                         ..............................
       Ph_al1_5Rn3 ...............
     Ph_al1_5Rn118 ...............
     Ph_al1_5Rn119 ...............
       Ph_al1_5Rn2 ...............
       Ph_al1_5Rn4 ...............
Ph_al1_vsdP_Phad16       ....................................
Ph_al1_vsdP_Pha102       ....................................
Ph_al1_vsdP_Pha097       ....................................
Ph_al1_vsdP_Pha092       ....................................
         Consensus       .........                         ..

                   281     290       300       310       320       330       340      350
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref SSSSVPLHPPPLPSSFIGTTYTSPFAPPPHPPTGAPHSSGKTMDFPILPGVPRLPGYLTGLVPSCLASLG
     Ph_al1_cDNA01 ............L.........................................................
     Ph_al1_cDNA04 ......................................................................
     Ph_al1_cDNA05 ..............................................................S.......
      Ph_al1_ORF17 ......................................................................
      Ph_al1_ORF16 ......................................................................
      Ph_al1_ORF13 ......................................................................
      Ph_al1_ORF15 ......................................................................
      Ph_al1_ORF14 ......................................................................
      Ph_al1_ORF18 ......................................................................
       Ph_al1_3Rn5 ......................................................................
       Ph_al1_3Rn7 ......................................................................
       Ph_al1_3Rn8 ......................................................................
       Ph_al1_5Rn3
     Ph_al1_5Rn118
     Ph_al1_5Rn119
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
Ph_al1_vsdP_Pha092
         Consensus

                   351     360       370       380       390       400       410      420
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref GLSSLGGLGAGALRTDFNSHIRADLSSHLRNDLGSPLRGDLAPRLPLLPQTLPQSYPPVFQQILAGLRPP
     Ph_al1_cDNA01 ......................................................................
     Ph_al1_cDNA04 ......................................................................
     Ph_al1_cDNA05 ......................................................................
      Ph_al1_ORF17 ..............................S.......................................
      Ph_al1_ORF16 ............................................................L.........
      Ph_al1_ORF13 ....V.................................................................
      Ph_al1_ORF15 ....V.................................................................
      Ph_al1_ORF14 ....V.................................................................
      Ph_al1_ORF18 ....V.................................................................
       Ph_al1_3Rn5 ......................................................................
       Ph_al1_3Rn7 ......................................................................
       Ph_al1_3Rn8 ......................................................................
       Ph_al1_5Rn3
     Ph_al1_5Rn118
     Ph_al1_5Rn119
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
Ph_al1_vsdP_Pha092
         Consensus

                   421     430       440       450       460       470       480      490
                   |--------+---------+---------+---------+---------+---------+---------|
        Ph_al1_ref SIPELNDYSSILASSAALSTASDVGPNNSDSKEQERRASSIAELRQRAREHEKRLASLSRKNKDSGGTDG
     Ph_al1_cDNA01 ......................................................................
     Ph_al1_cDNA04 ......................G...........D...................................
     Ph_al1_cDNA05 ......................................................................
      Ph_al1_ORF17 ......................................................................
      Ph_al1_ORF16 ......................................................................
      Ph_al1_ORF13 ......................................................................
      Ph_al1_ORF15 ......................................................................
      Ph_al1_ORF14 ...............V......................................................
      Ph_al1_ORF18 ......................................................................
       Ph_al1_3Rn5 ............................................................R.........
       Ph_al1_3Rn7 ......................................................................
       Ph_al1_3Rn8 ......................................................................
       Ph_al1_5Rn3
     Ph_al1_5Rn118
     Ph_al1_5Rn119
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
Ph_al1_vsdP_Pha092
         Consensus

                   491     500      510
                   |--------+---------|
        Ph_al1_ref DDRSPTRRSPRCSPKREIKT
     Ph_al1_cDNA01 ....................
     Ph_al1_cDNA04 ....................
     Ph_al1_cDNA05 ....................
      Ph_al1_ORF17 .........
      Ph_al1_ORF16 .........
      Ph_al1_ORF13 .........
      Ph_al1_ORF15 .........
      Ph_al1_ORF14 .........
      Ph_al1_ORF18 .........
       Ph_al1_3Rn5 ....................
       Ph_al1_3Rn7 ....................
       Ph_al1_3Rn8 ....................
       Ph_al1_5Rn3
     Ph_al1_5Rn118
     Ph_al1_5Rn119
       Ph_al1_5Rn2
       Ph_al1_5Rn4
Ph_al1_vsdP_Phad16
Ph_al1_vsdP_Pha102
Ph_al1_vsdP_Pha097
Ph_al1_vsdP_Pha092
         Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.4_Ph_al1]

[PhD_A2.4.2_Ph_al1_alignments]

## “difference_alignment_Ph_al1_all_bp_sequences.gif”



# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.5_Ph_al2]

### [PhD_A2.5.1_Ph_al2_sequences]

# "Ph_al2_general_read_me.txt"

```
General aspects:

(i)       All sequences are in FASTA format.

(ii)      [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)     Comments are preceded by an apostrophe (')

(iv)      All sequences without olignonucleotide sequence are indicated (e.g. "w/o 1314, 1315"). In these cases, only novel
Parhyale sequence is presented.


Parhyale hawaiensis al2 terminology:


(i)       In general, sequence denotation begins with: ">Ph_al2_"

(ii)      The above (i) is followed by: "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                        "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                        "ORF"               for coherent ORF sequences
                                        "cDNA"              for complete cDNA sequences

                                        "ref"               if reference sequence is presented
                                        "HD"                if homeobox/homeodomain fragment is presented
                                        "OAR"               if OAR (encoding) fragment is presented

(iii)     The above (ii) is followed by clone id.


Parhyale hawaiensis al2 synopsis:

cDNA:              2594 bp
5'UTR:     566 bp (1-566)
ORF:               1761 bp (567-2327), 586 aa
3'UTR:     567 bp (2328-2594)
HD:        180 bp (1464-1643), 60 aa
OAR                  63 bp (2226-2288), 21 aa


>Ph_al2_HD [bp]
CAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAAGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACCAGGGAGGAGTTGGCAATGAAGATAGGCCTGACGGAGGCA
CGCATACAGGTCTGGTTCCAGAACCGACGTGCAAAGTGGCGCAAGCAAGAA

>Ph_al2_OAR [bp]
GCAGAGCGGCGACAGGCGTCTCTCCACGACCTCAGACTTAAGGCACGAAATCAGGAAGTGCGA

>Ph_al2_ref 'Ph_al2_ORF05, extended 5' by 593bp of Ph_al2_5Rn11 and 3' by 294 bp of Ph_arx3_3Rfwx05; exch nuc (1103 T>C, 1193
A>G, 1208 T>C, 1232 T>C, 1344 A>T, 1478 G>A, 1723 C>T, 1862 C>G, 1892 A>G, 1898 T>C, 2045 C>T, 2087 T>G, 2243 T>G, 2264 C>T, 2267
A>G, 2270 G>A)
GAGTTTACCACCTGGCTTCGTTCTGGAAGCACTGCGCGTATTGCTCCATTGAACTGCAAGGCAATTATCACCTGGGCAGTGTGTCTTGTTTCGCAGTTATTTGGTACCTTACGGGGTTGAATTAAAAAC
AAACTTTTAATATAGGTCTTACCCTCAGGAACAGAACTGAAATATTAATTGCTGTGCTCATACTGAATTTTTTTCTAAAAGTGTATGTGTTGAAGCTTGATCTAAAATTTCAATGTGACGAGACTGGTA
TTTTTCGAAAGTTAGCTACTTGAGATAAACGATTTTCTTCTTTTGTGGTGTCAGTGTCACACTGTACATGGGATCTTAAATTATTCTTCGGAACTGCGCGAGAAAAGTCCTCTGAATTTATCAACAGTC
TGGAGACCCTCATCCGAAACCCAGATGACTACAAGATTAACAGCCTCAGCTGCACCTACACAGGACTACTGAATTCTTCACTGGAGAGGTTTGGCATACAGTGACTTCACTCAAACTTCTATTTCTGAA
AGTAAAATTCTCTTCTTACCTGCAATAGATTGAGCAAGTTACTTAAAACAATGGGAATTCTGGAAGAGTGCAGCGTCTCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGG
CCTCACCAACCATGCTCGCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTGTAGACCCTGAGATGACGAGCACTACAGTGTCTGACCTCACCTCGGGTGTCATTTCCCACAACA
TTCACAGCATTGCAAACCTCAGCTCGCAACATTTCGGCCAACAGGGTGGTCGAACATCTCCCAGAGGTAGAAGCTCACCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACGAGCG
AAACAATTTCCAGAGCACAGTCCATCCTCAGCATTCACAACCTGACAAAGCAGCTGTCTCCCCCGCTAAACAAAAAACTCCATGACACCCTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCATCTA
CCACGGCTTCGACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCTCTCACCTCCGTTCAGTCCAGCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCTATGC
TTCATCAAGGGAGCGGCGGACCCGCGCACAGGAATACGAAGAAAAACAGGTCGGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACCGAATCATCCT
CAACTGTTTCAAAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCACATCTGGAAGGGTGTGATGGAGGGGAAGAAGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGACTGTGGAG
AGGGTGGGAGCGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAAGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGT
TTACCAGGGAGGAGTTGGCAATGAAGATAGGCCTGACGGAGGCACGCATACAGGTCTGGTTCCAGAACCGACGTGCAAAGTGGCGCAAGCAAGAAAAAGTTGGACCCTCAAGTCACCCGTACGGAGGAG
CCTTCCCACCGACGTCCCCCGTTCCTCTACATCCTCCACCTCTTCTCCCATCATTTGGAAGCACCGGCGGTGCTTACCCTCACCCGTTCCCACCCCACCCCTCCATACGGGGAGGCAAGCCCCTGGAAA
CCCCTCGTCTGTCCCCGTACCTCCCCTCACCGCCGCCATCTTTGCTACACGCATCGTATCTCGCAGGAGGTCTGCCGGGGCTCGCGACCCCCACCGGCATGGCTTCTCTGCTGCTCAGGGAGCTTCCCT
CAGCTTATCCTCGACCAGCGCTACCTGCTCTCTTACCCTCACAAACGCCATCTTTTACAACCGTGTTAGCGCAACTGTCACACCGATCCAAAATAGACCCATCAGCTTATTATCTACCACCCTACTTGG
CCCATCTCACTGCTCTACCCCCGTCAACTCTCCCCTCTTCTTTGTCCCCTGCCCTCACCCCACCTTCACCCTCGTCCACCTTTCCCTCTCCTATCCCTTCGACATCTTCTCCACCGATAAATGGCCTGG
GACTTGCGACTCCCCCAGGCCTCAGTGTAGAGGCAGAGCGGCGACAGGCGTCTCTCCACGACCTCAGACTTAAGGCACGAAATCAGGAAGTGCGACTTGAGATGCTACGAAAAGCTGGCGAAATTACAG
CGTGAGGTAGAAGTAGTTTACTGGGTTCTTCAACGTAACTTTAGGGAACAACCATAAAGAATCAATACCATAGTTGCACCATAATATCAGTTATAAATTTTTCTGTAATGGTATATGCTATACTTTGTT
TCCAACTCCAGCAATATCATACGGAAAAATGATTGCCAATAGCTCGTCTACGATATTTAAGCCTTTAATATTCGTTTTTAATGATTCATTCCGCTCTCTGCTCAAATCATATCATTAGAAAAAAAAAAA
AAAAAAAAAAAAAA


>Ph_al2_HD [aa]
QRRYRTTFTSYQLEELEKVFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQE
```

```
>Ph_al2_OAR [aa]
AERRQASLHDLRLKARNQEVR

>Ph_al2_ref [aa]
MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLTSGVISHNIHSIANLSSQHFGQQGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTK
QLSPPLNKKLHDTLVQMVTCAQSMPSTTASTSTSSPHQNLHSRRALSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMEEDPSVVADGS
SSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLP
SFGSTGGAYPHPFPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVLAQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSP
ALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQASLHDLRLKARNQEVRLEMLRKAGEITA&
```

## [PhD_A2_Parhyale_goi_sequences]

### [PhD_A2.5_Ph_al2]

#### [PhD_A2.5.1_Ph_al2_sequences]

**“Ph_al2_cDNA_sequences.txt”**

```
>Ph_al2_ref 'Ph_al2_ORF05, extended 5' by 593bp of Ph_al2_5Rn11 and 3' by 294 bp of Ph_arx3_3Rfwx05; exch nuc (1103 T>C, 1193
A>G, 1208 T>C, 1232 T>C, 1344 A>T, 1478 G>A, 1723 C>T, 1862 C>G, 1892 A>G, 1898 T>C, 2045 C>T, 2087 T>G, 2243 T>G, 2264 C>T, 2267
A>G, 2270 G>A)
GAGTTTACCACCTGGCTTCGTTCTGGAAGCACTGCGCGTATTGCTCCATTGAACTGCAAGGCAATTATCACCTGGGCAGTGTGTCTTGTTTCGCAGTTATTTGGTACCTTACGGGGTTGAATTAAAAAC
AAACTTTTAATATAGGTCTTACCCTCAGGAACAGAACTGAAATATTAATTGCTGTGCTCATACTGAATTTTTTTCTAAAAGTGTATGTGTTGAAGCTTGATCTAAAATTTCAATGTGACGAGACTGGTA
TTTTTCGAAAGTTAGCTACTTGAGATAAACGATTTTCTTCTTTTGTGGTGTCAGTGTCACACTGTACATGGGATCTTAAATTATTCTTCGGAACTGCGCGAGAAAAGTCCTCTGAATTTATCAACAGTC
TGGAGACCCTCATCCGAAACCCAGATGACTACAAGATTAACAGCCTCAGCTGCACCTACACAGGACTACTGAATTCTTCACTGGAGAGGTTTGGCATACAGTGACTTCACTCAAACTTCTATTTCTGAA
AGTAAAATTCTCTTCTTACCTGCAATAGATTGAGCAAGTTACTTAAAACAATGGGAATTCTGGAAGAGTGCAGCGTCTCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGG
CCTCACCAACCATGCTCGCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTGTAGACCCTGAGATGACGAGCACTACAGTGTCTGACCTCACCTCGGGTGTCATTTCCCACAACA
TTCACAGCATTGCAAACCTCAGCTCGCAACATTTCGGCCAACAGGGTGGTCGAACATCTCCCAGAGGTAGAAGCTCACCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACGAGCG
AAACAATTTCCAGAGCACAGTCCATCCTCAGCATTCACAACCTGACAAAGCAGCTGTCTCCCCCGCTAAACAAAAAACTCCATGACACCCTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCATCTA
CCACGGCTTCGACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCTCTCACCTCCGTTCAGTCCAGCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCTATGC
TTCATCAAGGGAGCGGCGGACCCGCGCACAGGAATACGAAGAAAAACAGGTCGGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACCGAATCATCCT
CAACTGTTTCAAAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCACATCTGGAAGGGTGTGATGGAGGGGAAGAAGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGACTGTGGAG
AGGGTGGGAGCGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAAGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGT
TTACCAGGGAGGAGTTGGCAATGAAGATAGGCCTGACGGAGGCACGCATACAGGTCTGGTTCCAGAACCGACGTGCAAAGTGGCGCAAGCAAGAAAAAGTTGGACCCTCAAGTCACCCGTACGGAGGAG
CCTTCCCACCGACGTCCCCCGTTCCTCTACATCCTCCACCTCTTCTCCCATCATTTGGAAGCACCGGCGGTGCTTACCCTCACCCGTTCCCACCCCACCCCTCCATACGGGGAGGCAAGCCCCTGGAAA
CCCCTCGTCTGTCCCCGTACCTCCCCTCACCGCCGCCATCTTTGCTACACGCATCGTATCTCGCAGGAGGTCTGCCGGGGCTCGCGACCCCCACCGGCATGGCTTCTCTGCTGCTCAGGGAGCTTCCCT
CAGCTTATCCTCGACCAGCGCTACCTGCTCTCTTACCCTCACAAACGCCATCTTTTACAACCGTGTTAGCGCAACTGTCACACCGATCCAAAATAGACCCATCAGCTTATTATCTACCACCCTACTTGG
CCCATCTCACTGCTCTACCCCCGTCAACTCTCCCCTCTTCTTTGTCCCCTGCCCTCACCCCACCTTCACCCTCGTCCACCTTTCCCTCTCCTATCCCTTCGACATCTTCTCCACCGATAAATGGCCTGG
GACTTGCGACTCCCCCAGGCCTCAGTGTAGAGGCAGAGCGGCGACAGGCGTCTCTCCACGACCTCAGACTTAAGGCACGAAATCAGGAAGTGCGACTTGAGATGCTACGAAAAGCTGGCGAAATTACAG
CGTGAGGTAGAAGTAGTTTACTGGGTTCTTCAACGTAACTTTAGGGAACAACCATAAAGAATCAATACCATAGTTGCACCATAATATCAGTTATAAATTTTTCTGTAATGGTATATGCTATACTTTGTT
TCCAACTCCAGCAATATCATACGGAAAAATGATTGCCAATAGCTCGTCTACGATATTTAAGCCTTTAATATTCGTTTTTAATGATTCATTCCGCTCTCTGCTCAAATCATATCATTAGAAAAAAAAAAA
AAAAAAAAAAAAAA

>Ph_al2_ORF03 [bp] 'w/o 1314, 1315
TCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGGCCTCACCAACCATGCTCGCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTGTAGACCCTGAG
ATGACGAGCACTACAGTGTCTGACCTCCCCTCGGGTGTCATTTCCCACAACATTCACAGCATTGCAAACCTCAGCTCGCAACATTTCGGCCAACAGGGTGGTCGAACATCTCCTAGAGGTAGAAGCTCA
CCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACGAGCGAAACAATTTCCAGAGCACAATCCATCCTCAGCATTCACAACTTGACAAAGCAGCTGTCTCCCCCGCTAAACAAAAAA
CTCCATGACACCTTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCATCTACCACGGCTTCGACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCTCTCACCTCCGTTCAGTCCA
GCAAATGCTGGTGACGAACCACAGCCCTTAGCTCACATGAAGCTGCCTATGCTTCATCAAGGGAGCGGCGGACCCGCACACGGGAATACGAAGAAAAACAGGTCGGCCCAGCCAGACCACGACCTGGAA
TTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACCGAATCATCCTCAACTGTTTCAAAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCACATCTGGAAGGGTGTGATGGA
GGGGAAGAAGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGACTGTGGAGAGGGTGGGAGCGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATACAGAACTACCTTCACTTCTTAC
CAACTCGAAGAGTTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACCAGGGAGGAGTTGGCAATGAAGATAGGCCTGACGGAGGCACGCATACAGGTCTGGTTCCAGAACCGACGTGCA
AAGTGGCGCAAGCAAGAAAAAGTTGGACCCTCAAGTCACCCGTACGGAGGAGCCTTCCCACCGACGTCCCCCGTTCCTCTACATCCTCCACCTCTTCTCCCATCATTTGGAAGCACCGGCGGTGCTTAC
CCTCACCCGTTCCCACCCCACCCCTCCATACGGGGAGGCAAGCCCCTGGAAACCCCTCGTCTGTCCCCGTACCTCCCCTCACCGCCGCCATCTTTGCTGCACGCATCGTATCTCGCAGGAGGTCTGCCG
GGGCTCGCGACCCCCACCGGCATGGCTTCTCTGCTGCTCAGGGAGCTTCCCTCAGCTTATCCTCGACCAGCGCTACCTGCTCTCTTACCCTCACAAACGCCATCTTTTACAACCGTGTTAGCGCAACTG
TCACACCGATCCAAAATAGACCCATCAGCTTATTATCTACCACCCTACTTGGCCCATCTCACTGCTCTACCCCCGTCAACTCTCCCCTCTTCTTTGTCCCCTGCCCTCACCCCACCTTCACCCTCGTCC
ACCTTTCCCTCTCCTATCCCTTCGACATCTTCTCCACCGATAAATGGCCTGGGACTTGCGACTCCCCCAGGCCTCAGTGTAGAGGCAGAGCGGCGGCAGGCGTCTCTCCACGACCTCAGACTTAAGGCA
CGAAATCAGGAAGTGCGACTTGAGATGCTA

>Ph_al2_ORF04 [bp] 'w/o 1314, 1315
TCCTCACAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGGCCTCACCAACCATGCTCGCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTGTAGACCCTGAG
ATGACGAGCACTACAGTGTCTGACCTCCCCTCGGGTGTCATTTCCCACAACATTCACAGCATTGCAAACCTCAGCTCGCAACATTTCGGCCAACAGGGTGGTCGAACATCTCCTAGAGGTAGAAGCTCA
CCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACGAGCGAAACAATTTCCAGAGCACAATCCATCCTCAGCATTCACAACTTGACAAAGCAGCTGTCTCCCCCGCTAAACAAAAAA
CTCCATGACACCTTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCATCTACCACGGCTTCGACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCTCTCACCTCCGTTCAGTCCA
GCAAATGCTGGTGACGAACCACAGCCCTTAGCTCACATGAAGCTGCCTACGCTTCATCAAGGGAGCGGCGGACCCGCACACAGGAATACGAAGAAAAACAGGTCGGCCCAGCCAGACCACGACCTGGAA
TTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACCGAATCATCCTCAACTGTTTCAAAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCACATCTGGAAGGGTGTGATGGA
GGGGAAGAAGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGACTGTGGAGAGGGTGGGAGCGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATACAGAACTACCTTCACTTCTTAC
CAACTCGAAGAGTTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACCAGGGAGGAGTTGGCAATGTAGATAGGCCTGACGGAGGCACGCATACAGGTCTGGTTCCAGAACCGACGTGCA
AAGTGGCGCAAGCAAGAAAAAGTTGGACCCTCAAGTCACCCGTACGGAGGAGCCTTCCCACCGACGTCCCCCGTTCCTCTACATCCTCCACCTCTTCTCCCATCATCTGGAAGCACCGGCGGTGCTTAC
CCTCACCCGCTCCCACCCCACCCCTCCATACGGGGAGGCAAGCCCCTGGAAACCCCTCGTCTGTCCCCGTACCTCCCCTCACCGCCGCCATCTTTGCTACACGCATCGTATCTCGCAGGAGGTCTGCCG
GGGCTCGCGACCCCCACCGGCATGGCTTCTCTGCTGCTCAGGGAGCTTCCCTCAGCTTATCCTCGACCAGCGCTACCTGCTCTCTTACCCTCACAAACGCCATCTTTTACAACCGTGTTAGCGCAACTG
TCACACCGATCCAAAATAGACCCATCAGCTTATTATCTACCACCCTACTTGGCCCATCTCACTGCTCTACCCCCGTCAACTCTCCCCTCTTCTTTGTCCCCTGCCCTCACCCCACCTTCACCCTCGTCC
ACCTTTCCCTCTCCTATCCCTTCGACATCTTCTCCACCGATAAATGGCCTGGGACTTGCGACTCCCCCAGGCCTCAGTGTAGAGGCAGAGCGGCGACAGGCGTCTCTCCACGACCTCAGACTTAAGGCA
CGAAATCAGGAAGTGCGACTTGAGATGCTA

>Ph_al2_ORF05 [bp] 'w/o 1314, 1315
TCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGGCCTCACCAACCATGCTCGCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTGTAGACCCTGAG
ATGACGAGCACTACAGTGTCTGACCTCACCTCGGGTGTCATTTCCCACAACATTCACAGCATTGCAAACCTCAGCTCGCAACATTTCGGCCAACAGGGTGGTCGAACATCTCCCAGAGGTAGAAGCTCA
CCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACGAGCGAAACAATTTCCAGAGCACAGTCCATCCTCAGCATTCACAACCTGACAAAGCAGCTGTCTCCCCCGCTAAACAAAAAA
CTCCATGACACCCTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCATCTACCACGGCTTCGACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCTCTCACCTCCGTTTAGTCCA
GCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCTATGCTTCATCAAGGGAGCGGCGGACCCGCGCACAGAAATACGAAGAAAAATAGGTCGGCCCAGCCAGACCACGATCTGGAA
TTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACCGAATCATCCTCAACTGTTTCAAAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTACACATCTGGAAGGGTGTGATGGA
GGGGAAGAAGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGACTGTGGAGAGGGTGGGAGCGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATATCGGACTACCTTCACTTCATAC
CAACTCGAAGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACCAGGGAGGAGTTGGCAATGAAGATAGGCCTGACGGAGGCACGCATACAGGTCTGGTTCCAGAACCGACGTGCA
AAGTGGCGCAAGCAAGAAAAAGTTGGACCCTCAAGTCACCCGTACGGAGGAGCCTTCCCACCGACGTCCCCCGTTCCTCTACATCCTCCACCTCTTCCCCCATCATTTGGAAGCACCGGCGGTGCTTAC
CCTCACCCGTTCCCACCCCACCCCTCCATACGGGGAGGCAAGCCCCTGGAAACCCCTCGTCTGTCCCCGTACCTCCCCTCACCGCCGCCATCTTTGCTACACGCATCCTATCTCGCAGGAGGTCTGCCG
GGGCTCGCAACCCCTACCGGCATGGCTTCTCTGCTGCTCAGGGAGCTTCCCTCAGCTTATCCTCGACCAGCGCTACCTGCTCTCTTACCCTCACAAACGCCATCTTTTACAACCGTGTTAGCGCAACTG
TCACACCGATCCAAAATAGACCCATCAGCTTACTATCTACCACCCTACTTGGCCCATCTCACTGCTCTACCCCCTTCAACTCTCCCCTCTTCTTTGTCCCCTGCCCTCACCCCACCTTCACCCTCGTCC
ACCTTTCCCTCTCCTATCCCTTCGACATCTTCTCCACCGATAAATGGCCTGGGACTTGCGACTCCCCCAGGCCTCAGTGTAGAGGCAGAGCGGCGACAGGCTTCTCTCCACGACCTCAGACTCAAAGCG
CGAAATCAGGAAGTGCGACTTGAGATGCTA

>Ph_al2_5Rg1 [bp] 'w/o 1305, id to Ph_al2_5Rg2-4
```

```
AAGGGAGCGGCGGACCCGCTCACAGGAATACGAAGAAAAACAGATCGGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACTGAATCATCCTCAACTG
TTTCAAAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCACATCTGGAAGGGTGTGATGGAGGGGAAGAAGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGATTGTGGAGAGGGTG
GGAGCGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAGGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACCA
GGGAGGAGTTGGCAAT

>Ph_al2_5Rp09 [bp] 'w/o 1305
GAGTTTACCACCTGGCTTCGTTCTGGAAGCACTGCGCGTATTGCTCCATTGAACTGCAAGGCAATTATCACCTGGGCAGTGTGTCTTGTTTCGCAGTTATTTGGTACCTTACGGGGTTGAATTAAAAAC
AAACTTTTAATATAGGTCTTACCCTCAGGAACAGAACTGAAATATTAATTGCTGTGCTCATACTGAATTTTTTTTCTAAAAGTGTATGTGTTGAAGCTTGATCTAAAATTTCAATGTGACAAGACTGGT
ATTTTTCGAAAGTTAGCTACTTGAGATAAACGATTTTCTTCTTTTGTGGTGTCAGTGTCACACTGTACATGGGATCTTAAATTATTCTTCGGAACTGCGCGAGAAAAGTCCTCTGAATTTATCAACAGT
CTGGAGACCCTCATCCGAAACCCAGATAACTACAAGATTAACAGCACAGGACTACTGAATTCTGCACTGGAGCGGTTTGGTACACACTGACTTCACTCAAACTTCTATTTCTGAGAGTAAAATTCTCTT
CTTACCTGCAATAGATTGAGCAAGTTACTTGAAACAATGGGAATTCTGGAAGAGTGCAGCGTCTCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGGCCTCACCAACCATG
CTCTCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTATAGACCCTGAGATGACGAGCACTACAGTGTCTGACCTCACCTCGGGTGTCATTTCCCACAACATTCACAGCATTGCA
AACCTCAGCTCACAACATTTCGGCCAACAGACAGGTGGTCGAACATCTCCCAGAGGTAGAAGCTCACCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACCAGTGAAACAATTTCC
AGAGCACAGTCCATCCTCAGCATACACAACCTGACAAAGCAGCTGTCTCCCCCGCTAAACAAAAAACTCCATGACACCCTGGTGCAAATGGTTACTTGTGCCTAGTCGATGCCATCTACCACGGCTTCG
ACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCCCTCACCTCCGTTCAGTCCAGCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCTATGCTTCACCAAGGG
AGCGGCGGACCCGCTCACAGGAATACGAAGAAAAACAGATCGGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACTGAATCATCCTCAACTGTTTCA
AAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCACATCTGGAAGGGTGTGATGGAGGGGAAGAAGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGATTGTGGAGAGGGTGGGAGC
GGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAGGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACTAGGGAG
GAGTTGGCAAT

>Ph_al2_5Rp11 [bp] 'w/o 1305
GAGTTTACCACCTGGCTTCGTTCTGGAAGCACTGCGCGTATTGCTCCATTGAACTGCAAGGCAATTATCACCTGGGCAGTGTGTCTTGTTTCGCAGTTATTTGGTACCTTACGGGGTTGAATTAAAAAC
AAACTTTTAATATAGGTCTTACCCTCAGGAACAGAACTGAAATATTAATTGCTGTGCTCATACTGAATTTTTTTTCTAAAAGTGTATGTGTTGAAGCTTGATCTAAAATTTCAATGTGACAAGACTGGTA
TTTTTCGAAAGTTAGCTACTTGAGATAAACGATTTTCTTCTTTTGTGGTGTCAGTGTCACACTGTACATGGGATCTTAAATTATTCTTCGGAACTGCGCGAGAAAAGTCCTCTGAATTTATCAACAGTC
TGGAGACCCTCATCCGAAACCCAGATAACTACAAGATTAACAGCACAGGACTACTGAATTCTGCACTGGAGCGGTTTGGTACACACTGACTTCACTCAAACTTCTATTTCTGAGAGTAAAATTCTCTTC
TTACCTGCAATAGATTGAGCAAGTTACTTGAAACAATGGGAATTCTGGGAAGATTGCAGCGTCTCCTCTCAGTCGCGGACAAACCCCCCTGGCTCTGGGACATGTGCATCAAGAGGGCCTCACCAACCA
TGCTCTCTCATAGCGATGCCGTAGCCAGCAGTTCAGTTCCAGCCTGATGCCTATAGACCCTGAGATGACGAGCACTACAGTGTCTGACCTCACCTCGGGTGTCATTTCCCACAACATTCACAGCATTGC
AAACCTCAGCTCACAACATTTCGGCCAACAGACAGGTGGTCGAACATCTCCCAGAGGTAGAAGCTCACCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACCAGTGAAACAATTTC
CAGAGCACAGTCCATCCTCAGCATACACAACCTGACAAAGCAGCTGTCTCCCCCGCTAAACAAAAAACTCCATGACACCCTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCATCTACCACGGCTTC
GACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCCCTCACCTCCGTTCAGTCCAGCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCTATGCTTCACCAAGG
GAGCGGCGGACCCGCTCACAGGAATACGAAGAAAAACAGATCGGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACTGAATCATCCTCAACTGTTTC
AAAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCACATCTGGAAGGGTGTGATGGAGGGGAAGAAGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGATTGTGGAGAGGGTGGGAG
CGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAGGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACTAGGGA
GGAGTTGGCAAT

>Ph_al2_5Rp13 [bp] 'w/o 1305
GAGTTTACCACCTGGCTTCGTTCTGGAAGCACTGCGCGTATTGCTCCATTGAACTGCAAGGCAATTATCACCTGGGCAGTGTGTCTTGTTTCGCAGTTATTTGGTACCTTACGGGGTTGAATTAAAAAC
AAACTTTTAATATAGGTCTTACCCTCAGGAACAGAACTGAAATATTAATTGCTGTGCTCATACTGAATTTTTTTTCTAAAAGTGTATGTGTTGAAGCTTGATCTAAAATTTCAATGTGACAAGACTGGTA
TTTTTCGAAAGTTAGCTACTTGAGATAAACGATTTTCTTCTTTTGTGGTGTCAGTGTCACACTGTACATGGGATCTTAAATTATTCTTCGGAACTGCGCGAGAAAAGTCCTCTGAATTTATCAACAGTC
TGGAGACCCTCATCCGAAACCCAGATAACTACAAGATTAACAGCACAGGACTACTGAATTCTGCACTGGAGCGGTTTGGTACACACTGACTTCACTCAAACTTCTATTTCTGAGAGTAAAATTCTCTTC
TTACCTGCAATAGATTGAGCAAGTTACTTGAAACAATGGGAATTCTGGAAGAGTGCAGCGTCTCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGGACATGTGCATCAAGAGGGCCTCACCAACCATG
CTCTCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTATAGACCCTGAGATGACGAGCACTACAGTGTCTGACCTCACCTCGGGTGTCATTTCCCACAACATTCACAGCATTGCA
AACCTCAGCTCACAACATTTCGGCCAACAGACAGGTGGTCGAACATCTCCCAGAGGTAGAAGCTCACCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACCAGTGAAACAATTTCC
AGAGCACAGTCCATCCTCAGCATACACAACCTGACAAAGCAGCTGTCTCCCCCGCTAAACAAAAAACTCCATGACACCCTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCATCTACCACGGCTTCG
ACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCCCTCACCTCCGTTCAGTCCAGCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCTATGCTTCACCAAGGG
AGCGGCGGACCCGCTCACAGGAATACGAAGAAAAACAGATCGGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACTGAATCATCCTCAACTGTTTCA
AAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCACATCTGGAAGGGTGTGATGGAGGGGAAGAAGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGATTGTGGAGAGGGTGGGAGC
GGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAGGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACTAGGGAG
GAGTTGGCAAT

>Ph_al2_5Rn09 [bp] 'w/o 1306
GAGTTTACCACCTGGCTTCGTTCTGGAAGCACTGCGCGTATTGCTCCATTGAACTGCAAGGCAATTATCACCTGGGCAGTGTGTCTTGTTTCGCAGTTATTTGGTACCTTACGGGGTTGAATTAAAAAC
AAACTTTTAATATAGGTCTTACCCTCAGGAACAGAACTGAAATATTAATTGCTGTGCTCATACTGAATTTTTTTTCTAAAAGTGTATGTGTTGAGGCTTGATCTAAAATTTCAATGTGACGAGACTGGTA
TTTTTCGAAAGTTAGCTACTTGAGATAAACGATTTTCTTCTTTTGTGGTGTCAGTGTCACACTGTACATGGGATCTTAAATTATTCTTCGGAACTGCGCGAGAAAAGTCCTCTGAATTTATCAACAGTC
TGGAGACCCTCATCCGAAACCCAGATGACTACAAGATTAACAGCCTCAGCTGCACCTACACAGGACTACTGAATTCTTCACTGGAGAGGTTTGGCATACAGTGACTTCACTCAAACTTCTATTTCTGAA
AGTAAAATTCTCTTCTTACCTGCAATAGATTGAGCAAGTTACTTAAAACAATGGGAATTCTGGAAGAGTGCAGCGTCTCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGG
CCTCACCAACCATGCTCGCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTATAGACCCTGAGATGACGAGCACTACAGTGTCTGACCTCACCTCGGGTGTTATTTCCCACAACA
TTCACAGCATTGCAAATCTCAGCTCGCAACATTTCGGCCAACAGACAGGTGGTCGAACATCTCCCAGAGGTAGAAGCTCACCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACGA
GCGAAACAATTTCCAGAGCACAGTCCATCCTCAGCATTCACAACCTGACAAAGCAGCTATCTCCCCCGCTAAACAAAAAACTCCATGACACCTTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCAT
CTACCACGGCTTCGACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCCCTCACCTCCGTTCAGTCCAGCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCAA
TGCTTCACCAAGGGAGCGGCGGACCCGCGCACAGGAATACGAAGAAAAACAGGTCCGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACCGAATCAT
CCTCAACTGTTTCAAAAATGGATGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCGCATCTGGAAGGGTGTGATGGAGGAGAAGAGGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGACTGTG
GAGAGGGCGGGAGCGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAAGAATTAGAGAAA

>Ph_al2_5Rn11 [bp] 'w/o 1306
GAGTTTACCACCTGGCTTCGTTCTGGAAGCACTGCGCGTATTGCTCCATTGAACTGCAAGGCAATTATCACCTGGGCAGTGTGTCTTGTTTCGCAGTTATTTGGTACCTTACGGGGTTGAATTAAAAAC
AAACTTTTAATATAGGTCTTACCCTCAGGAACAGAACTGAAATATTAATTGCTGTGCTCATACTGAATTTTTTTTCTAAAAGTGTATGTGTTGAAGCTTGATCTAAAATTTCAATGTGACGAGACTGGTA
TTTTTCGAAAGTTAGCTACTTGAGATAAACGATTTTCTTCTTTTGTGGTGTCAGTGTCACACTGTACATGGGATCTTAAATTATTCTTCGGAACTGCGCGAGAAAAGTCCTCTGAATTTATCAACAGTC
TGGAGACCCTCATCCGAAACCCAGATGACTACAAGATTAACAGCCTCAGCTGCACCTACACAGGACTACTGAATTCTTCACTGGAGAGGTTTGGCATACAGTGACTTCACTCAAACTTCTATTTCTGAA
AGTAAAATTCTCTTCTTACCTGCAATAGATTGAGCAAGTTACTTAAAACAATGGGAATTCTGGAAGAGTGCAGCGTCTCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGG
CCTCACCAACCATGCTCGCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTATAGACCCTGAGATGACGAGCACTACAGTGTCTGACCTCACCTCGGGTGTTATTTCCCACAACA
TTCACAGCATTGCAAATCTCAGCTCGCAACATTTCGGCCAACAGACAGGTGGTCGAACATCTCCCAGAGGTAGAAGCTCACCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACGA
GCGAAACAATTTCCAGAGCACAGTCCATCCTCAGCATTCACAACCTGACAAAGCAGCTATCTCCCCCGCTAAACAAAAAACTCCATGACACCTTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCAT
CTACCACGGCTTCGACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCCCTCACCTCCGTTCAGTCCAGCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCAA
TGCTTCACCAAGGGAGCGGCGGACCCGCGCACAGGAATACGAAGAAAAACAGGTCCGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACCGAATCAT
CCTCAACTGTTTCAAAAATGGATGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCGCATCTGGAAGGGTGTGATGGAGGAGAAGAGGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGACTGTG
GAGAGGGCGGGAGCGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAAGAATTAGAGAAA

>Ph_al2_5Rn12 [bp] 'w/o 1306
AGTTTACCACCTGGCTTCGTTCTGGAAGCACTGCGCGTATTGCTCCATTGAACTGCAAGGCAATTATCACCTGGGCAGTGTGTCTTGTTTCGCAGTTATTTGGTACCTTACGGGGTTGAATTAAAAACA
AACTTTTAATATAGGTCTTACCCTCAGGAACAGAACTGAAATATTAATTGCTGTGCTCATACTGAATTTTTTTTCTAAAAGTGTATGTGTTGAAGCTTGATCTAAAATTTCAATGTGACGAGACTGGTAT
TTTTCGAAAGTTAGCTACTTGAGATAAACGATTTTCTTCTTTTGTGGTGTCAGTGTCACACTGTACATGGGATCTTAAATTATTCTTCGGAACTGCGCGAGAAAAGTCCTCTGAATTTATCAACAGTCT
GGAGACCCTCATCCGAAACCCAGATGACTACAAGATTAACAGCCTCAGCTGCACCTACACAGGACTACTGAATTCTTCACTGGAGAGGTTTGGCATACAGTGACTTCACTCAAACTTCTATTTCTGAAA
GTAAAATTCTCTTCTTACCTGCAATAGATTGAGCAAGTTACTTAAAACAATGGGAATTCTGGAAGAGTGCAGCGTCTCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGGC
CTCACCAACCATGCTCGCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTATAGACCCTGAGATGACGAGCACTACAGTGTCTGACCTCACCTCGGGTGTTATTTCCCACAACAT
TCACAGCATTGCAAATCTCAGCTCGCAACATTTCGGCCAACAGACAGGTGGTCGAACATCTCCCAGAGGTAGAAGCTCACCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACGAG
CGAAACAATCTCCAGAGCACAGTCCATCCTCAGCATTCACAACCTGACAAAGCAGCTATCTCCCCCGCTAAACAAAAAACTCCATGACACCTTGGTGCAAATGGTTACTTGTGCCCAGTCGATGCCATC
TACCACGGCTTCGACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCCCTCACCTCCGTTCAGTCCAGCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCAAT
GCTTCACCAAGGGAGCGGCGGACCCGCGCACAGGAATACGAAGAAAAACAGGTCCGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACCGAATCATC
```

```
CTCAACTGTTTCAAAAATGGATGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCGCATCTGGAAGGGTGTGATGGAGGAGAAGAGGTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGACTGTGG
AGAGGGCGGGAGCGGAGGCGACACAGCTGGTCCTACCAAAGGGAAGCAAAGGAGATATCGAACTACCTTCACTTCATACCAACTCGAAGAATTAGAGAAA

>Ph_al2_vsdP_Pha089 [bp] 'w/o fw_QRR, re_QVW; id to Ph_al2_vsdP_Pha095
TTCATACCAACTCGAAGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACCAGGGAGGAGTTGGCAATGAAGATAGGCCTGACGGAGGCACGCATA

>Ph_al2_vsdP_Pha111 [bp] 'w/o fw_QRR, re_QVW
TTCATACCAACTCGAGGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACTAGGGAGGAGTTGGCAATGAAGATAGGCCTGACGGAGGCACGCATA

>Ph_al2_3Rn227 [bp] 'w/o 1301
ATACAGGTCTGGTTCCAGAACCGACGTGCAAAGTGGCGCAAGCAAGAAAAAGTTGGACCCTCAAGTCACCCGTACGGAGGAGCCTTCCCACCGACGTCCCCCGTTCCTCTACATCCTCCACCTCTTCTC
CCATCATTTGGAAGCACCGGCGGTGCTTACCCTCACCCGTTCCCACCCCACCCCTCCATACGGGGAGGCAAGCCCCTGGAAACCCCTCGTCTGTCCCCGTACCTCCCCTCACCGCCGCCATCTTTGCTA
CACGCATCGTATCTCGCAGGAGGTCTGCCGGGGCTCGCGACCCCCACCGGCATGGCTTCTCTGCTGCTCAGGGAGCTTCCCTCAGCTTATCCTCGACCAGCGCTACCTGCTCTCTTACCCTCACAAACG
CCATCTTTTACAACCGTGTTAGCGCAACTGTCACACCGATCCAAAATAGACCCATCAGCTTATTATCTACCACCCTACTTGGCCCATCTCACTGCTCTACCCCCGTCAACTCTCCCCTCTTCTTTGTCC
CCTGCCCTCACCCCACCTTCACCCTCGTCCACCTTTCCCTCTCCTATCCCTTCGACATCTTCTCCACCGATAAA

>Ph_al2_3Rfwx01 [bp] 'w/o 1304
ATCAGCTTATTATCTACCACCCTACTTGGCCCATCTCACTGCTCCACCCCCGTCAACTCTCCCCTCTTCTTTGTCCCCTGCCCTCACCCCACCTTCACCCTCGTCCACCTTTCCCTCTCCTATCCCTTC
GACATCTTCTCCACCGATAAATGGCCTGGGACTTGCGACTCCCCCAGGCCTCAGTGTAGAGGCAGAGCGGCGACAGGCGTCTCTCCACGACCTCAGACTTAAGGCACGAAATCAGGAAGTGCGACTTGA
GATGCTACGAAAAGCTGGCGAAATTACAGCGTGAGGTAGAAGTAGTTTACTGGGTTCTTCAACGTAACTTTAGGGAACAACCATAAAGAATCAATACCATAGTTGCACCATAATATCAGTTATAAATTT
TTCTGTAATGGTATATGCTATACTTTGTTTCCAACTCCAGCAATATCATACGAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_al2_3Rfwx05 [bp] 'w/o 1304
ATCAGCTTATTATCTACCACCCTACTTGGCCCATCTCACTGCTCTACCCCCGTCAACTCTCCCCTCTTCTTTGTCCCCTGCCCTCACCCCACCTTCACCCTCGTCCACCTTTCCCTCTCCTATCCCTTC
GACATCTTCTCCACCGATAAATGGCCTGGGACTTGCGACTCCCCCAGGCCTCAGTGTAGAGGCAGAGCGGCGACAGGCATCTCTCCACGACCTCAGACTTAAGGCACGAAATCAGGAAGTGCGACTTGA
GATGCTACGAAAAGCTGGCGAAATTACAGCGTGAGGTAGAAGTAGTTTACTGGGTTCTTCAACGTAACTTTAGGGAACAACCATAAAGAATCAATACCATAGTTGCACCATAATATCAGTTATAAATTT
TTCTGTAATGGTATATGCTATACTTTGTTTCCAACTCCAGCAATATCATACGGAAAAATGATTGCCAATAGCTCGTCTACGATATTTAAGCCTTTAATATTCGTTTTTAATGATTCATTCCGCTCTCTG
CTCAAATCATATCATTAGAAAAAAAAAAAAAAAAAAAAAAAAAA


>Ph_al2_ref [aa]
MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLTSGVISHNIHSIANLSSQHFGQQGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTK
QLSPPLNKKLHDTLVQMVTCAQSMPSTTASTSTSSPHQNLHSRRALSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMEEDPSVVADGS
SSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLP
SFGSTGGAYPHPFPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVLAQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSP
ALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQASLHDLRLKARNQEVRLEMLRKAGEITA&

>Ph_al2_ORF03 [aa] 'w/o 1314, 1315
SSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLPSGVISHNIHSIANLSSQHFGQQGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKK
LHDTLVQMVTCAQSMPSTTASTSTSSPHQNLHSRRALSPPFSPANAGDEPQPLAHMKLPMLHQGSGGPAHGNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMEEDPSVVADGSSSHLEGCDG
GEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLPSFGSTGGAY
PHPFPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVLAQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSS
TFPSPIPSTSSPPINGLGLATPPGLSVEAERRQASLHDLRLKARNQEVRLEML

>Ph_al2_ORF04 [aa] 'w/o 1314, 1315
SSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLPSGVISHNIHSIANLSSQHFGQQGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKK
LHDTLVQMVTCAQSMPSTTASTSTSSPHQNLHSRRALSPPFSPANAGDEPQPLAHMKLPTLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMEEDPSVVADGSSSHLEGCDG
GEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVFSRTHYPDVFTREELAM@IGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLPSSGSTGGAY
PHPLPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVLAQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSS
TFPSPIPSTSSPPINGLGLATPPGLSVEAERRQASLHDLRLKARNQEVRLEML

>Ph_al2_ORF05 [aa] 'w/o 1314, 1315
SSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLTSGVISHNIHSIANLSSQHFGQQGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKK
LHDTLVQMVTCAQSMPSTTASTSTSSPHQNLHSRRALSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMEEDPSVVADGSSTHLEGCDG
GEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLPPSFGSTGGAY
PHPFPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVLAQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSS
TFPSPIPSTSSPPINGLGLATPPGLSVEAERRQASLHDLRLKARNQEVRLEML

>Ph_al2_5Rg1 [aa] 'w/o 1305, id to 2-4
GSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVFSRTHYPDVFTR
EELA

>Ph_al2_5Rp9 [aa] 'w/o 1305
MGILEECSVSSQSRTNPLALDMCIKRASPTMLSHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQHFGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLT
KQLSPPLNKKLHDTLVQMVTCA@SMPSTTASTSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMEEDPSVVADG
SSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVFSRTHYPDVFTREELA

>Ph_al2_5Rp11 [aa] 'w/o 1305
MPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQHFGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTASTSTSSPHQNLHSRRA
PSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRY
RTTFTSYQLEELEKVFSRTHYPDVFTREELA

>Ph_al2_5Rp13 [aa] 'w/o 1305
MCIKRASPTMLSHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQHFGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTC
AQSMPSTTASTSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGE
DGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVFSRTHYPDVFTREELA

>Ph_al2_5Rn9 [aa] 'w/o 1306
MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQHFGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLT
KQLSPPLNKKLHDTLVQMVTCAQSMPSTTASTSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMDEDPSVVADG
SSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEK

>Ph_al2_5Rn11 [aa] 'w/o 1306
MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQHFGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLT
KQLSPPLNKKLHDTLVQMVTCAQSMPSTTASTSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMDEDPSVVADG
SSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEK

>Ph_al2_5Rn12 [aa] 'w/o 1306
```

MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQHFGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLT
KQLSPPLNKKLHDTLVQMVTCAQSMPSTTASTSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETESSSTVSKMDEDPSVVADG
SSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKGKQRRYRTTFTSYQLEELEK

>Ph_al2_vsdP_Pha089 [aa] 'w/o fw_QRR, re_QVW; id to Ph_al2_vsdP_Pha095
SYQLEELEKVFSRTHYPDVFTREELAMKIGLTEARI

>Ph_al2_vsdP_Pha111 [aa] 'w/o fw_QRR, re_QVW
SYQLEELEKVFSRTHYPDVFTREELAMKIGLTEARI

>Ph_al2_3Rn227 [aa] 'w/o 1301
IQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLPSFGSTGGAYPHPFPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQT
PSFTTVLAQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPI

>Ph_al2_3Rfwx01 [aa] 'w/o 1304
SAYYLPPYLAHLTAPPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQASLHDLRLKARNQEVRLEMLRKAGEITA&

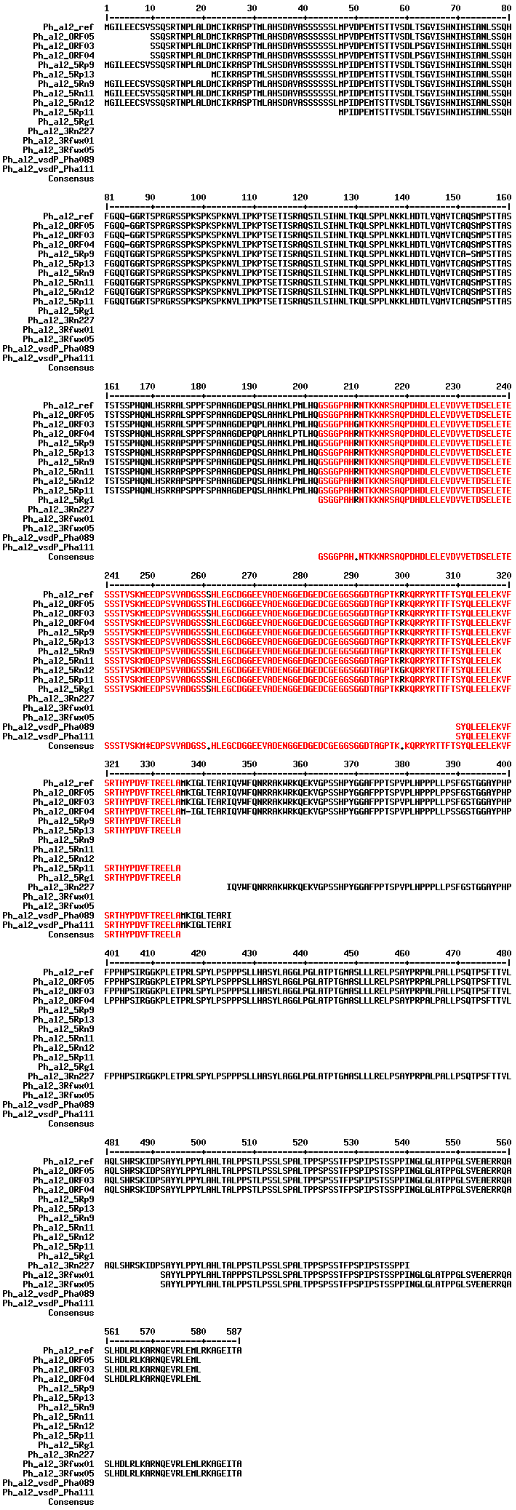
>Ph_al2_3Rfwx05 [aa] 'w/o 1304
SAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQASLHDLRLKARNQEVRLEMLRKAGEITA&

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.5_Ph_al2]

[PhD_A2.5.2_Ph_al2_alignments]

## “alignment_Ph_al2_all_aa_sequences.gif”



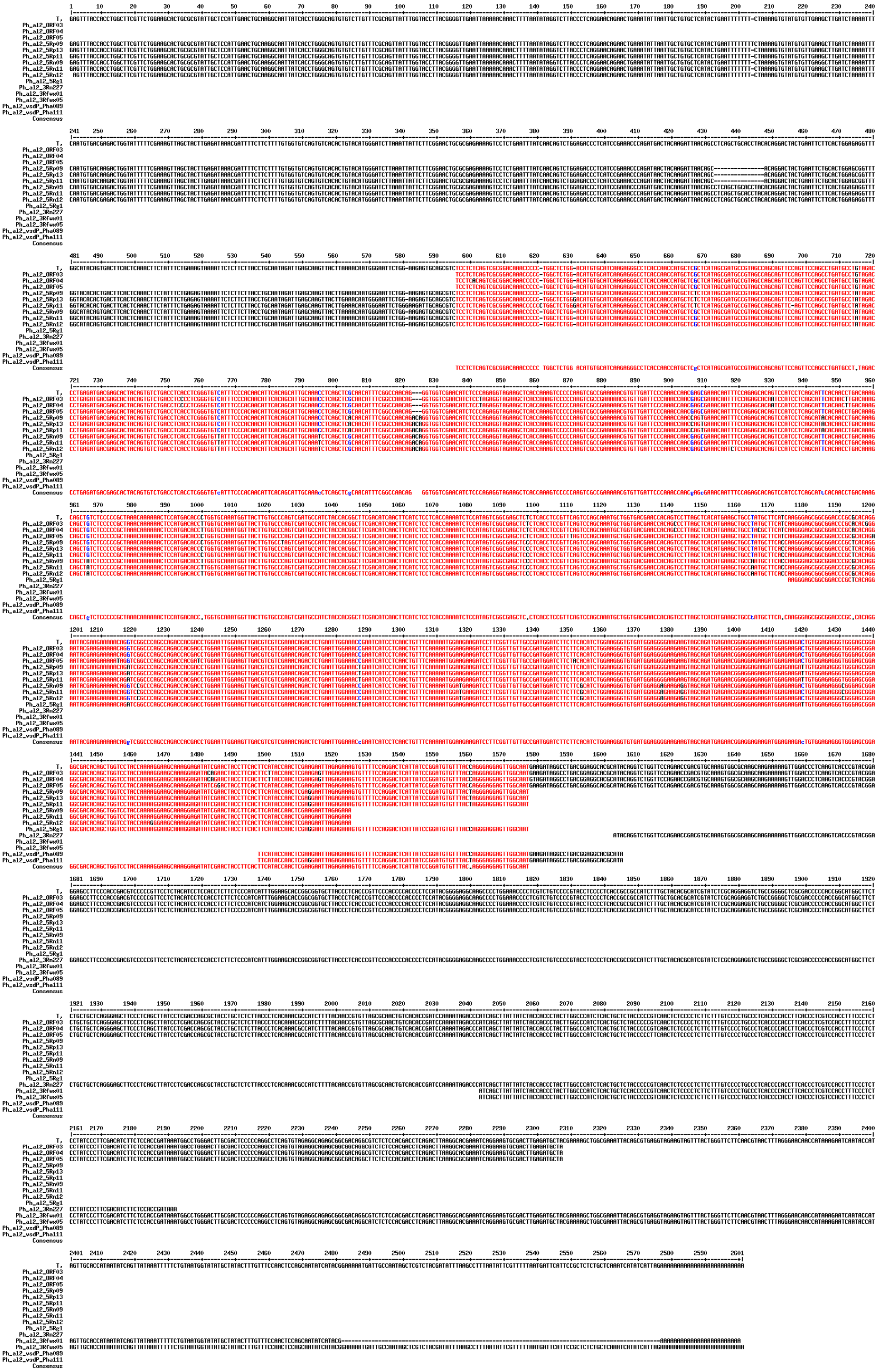
```
                   1       10        20        30        40        50        60        70        80
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLTSGVISHNIHSIANLSSQH
      Ph_al2_ORF05          SSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLTSGVISHNIHSIANLSSQH
      Ph_al2_ORF03          SSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLPSGVISHNIHSIANLSSQH
      Ph_al2_ORF04          SSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLPSGVISHNIHSIANLSSQH
       Ph_al2_5Rp9 MGILEECSVSSQSRTNPLALDMCIKRASPTMLSHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQH
      Ph_al2_5Rp13                      MCIKRASPTMLSHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQH
       Ph_al2_5Rn9 MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQH
      Ph_al2_5Rn11 MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQH
      Ph_al2_5Rn12 MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQH
      Ph_al2_5Rp11                                               MPIDPEMTSTTVSDLTSGVISHNIHSIANLSSQH
       Ph_al2_5Rg1
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus

                   81      90       100       110       120       130       140       150       160
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref FGQQ-GGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
      Ph_al2_ORF05 FGQQ-GGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
      Ph_al2_ORF03 FGQQ-GGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
      Ph_al2_ORF04 FGQQ-GGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
       Ph_al2_5Rp9 FGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCA-SMPSTTAS
      Ph_al2_5Rp13 FGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
       Ph_al2_5Rn9 FGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
      Ph_al2_5Rn11 FGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
      Ph_al2_5Rn12 FGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
      Ph_al2_5Rp11 FGQQTGGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
       Ph_al2_5Rg1
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus

                   161    170       180       190       200       210       220       230       240
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref TSTSSPHQNLHSRRALSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
      Ph_al2_ORF05 TSTSSPHQNLHSRRALSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
      Ph_al2_ORF03 TSTSSPHQNLHSRRALSPPFSPANAGDEPQPLAHMKLPMLHQGSGGPAHGNTKKNRSAQPDHDLELEVDVVETDSELETE
      Ph_al2_ORF04 TSTSSPHQNLHSRRALSPPFSPANAGDEPQPLAHMKLPTLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
       Ph_al2_5Rp9 TSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
      Ph_al2_5Rp13 TSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
       Ph_al2_5Rn9 TSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
      Ph_al2_5Rn11 TSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
      Ph_al2_5Rn12 TSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
      Ph_al2_5Rp11 TSTSSPHQNLHSRRAPSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
       Ph_al2_5Rg1                                           GSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus                                           GSGGPAH.NTKKNRSAQPDHDLELEVDVVETDSELETE

                   241    250       260       270       280       290       300       310       320
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref SSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVF
      Ph_al2_ORF05 SSSTVSKMEEDPSVVADGSSTHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVF
      Ph_al2_ORF03 SSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVF
      Ph_al2_ORF04 SSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVF
       Ph_al2_5Rp9 SSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVF
      Ph_al2_5Rp13 SSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVF
       Ph_al2_5Rn9 SSSTVSKMDEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEK
      Ph_al2_5Rn11 SSSTVSKMDEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEK
      Ph_al2_5Rn12 SSSTVSKMDEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKGKQRRYRTTFTSYQLEELEK
      Ph_al2_5Rp11 SSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVF
       Ph_al2_5Rg1 SSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVF
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089                                                                      SYQLEELEKVF
Ph_al2_vsdP_Pha111                                                                      SYQLEELEKVF
         Consensus SSSTVSKM#EDPSVVADGSS.HLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTK.KQRRYRTTFTSYQLEELEKVF

                   321    330       340       350       360       370       380       390       400
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref SRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLPSFGSTGGAYPHP
      Ph_al2_ORF05 SRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLPPSFGSTGGAYPHP
      Ph_al2_ORF03 SRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLPSFGSTGGAYPHP
      Ph_al2_ORF04 SRTHYPDVFTREELAM-IGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLPSSGSTGGAYPHP
       Ph_al2_5Rp9 SRTHYPDVFTREELA
      Ph_al2_5Rp13 SRTHYPDVFTREELA
       Ph_al2_5Rn9
      Ph_al2_5Rn11
      Ph_al2_5Rn12
      Ph_al2_5Rp11 SRTHYPDVFTREELA
       Ph_al2_5Rg1 SRTHYPDVFTREELA
     Ph_al2_3Rn227                         IQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLPSFGSTGGAYPHP
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089 SRTHYPDVFTREELAMKIGLTEARI
Ph_al2_vsdP_Pha111 SRTHYPDVFTREELAMKIGLTEARI
         Consensus SRTHYPDVFTREELA

                   401    410       420       430       440       450       460       470       480
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref FPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVL
      Ph_al2_ORF05 FPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVL
      Ph_al2_ORF03 FPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVL
      Ph_al2_ORF04 LPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVL
       Ph_al2_5Rp9
      Ph_al2_5Rp13
       Ph_al2_5Rn9
      Ph_al2_5Rn11
      Ph_al2_5Rn12
      Ph_al2_5Rp11
       Ph_al2_5Rg1
     Ph_al2_3Rn227 FPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVL
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus

                   481    490       500       510       520       530       540       550       560
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref AQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQA
      Ph_al2_ORF05 AQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQA
      Ph_al2_ORF03 AQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQA
      Ph_al2_ORF04 AQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQA
       Ph_al2_5Rp9
      Ph_al2_5Rp13
       Ph_al2_5Rn9
      Ph_al2_5Rn11
      Ph_al2_5Rn12
      Ph_al2_5Rp11
       Ph_al2_5Rg1
     Ph_al2_3Rn227 AQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPI
    Ph_al2_3Rfwx01            SAYYLPPYLAHLTAPPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQA
    Ph_al2_3Rfwx05            SAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQA
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus

                   561    570       580    587
                   |--------+---------+------|
        Ph_al2_ref SLHDLRLKARNQEVRLEMLRKAGEITA
      Ph_al2_ORF05 SLHDLRLKARNQEVRLEML
      Ph_al2_ORF03 SLHDLRLKARNQEVRLEML
      Ph_al2_ORF04 SLHDLRLKARNQEVRLEML
       Ph_al2_5Rp9
      Ph_al2_5Rp13
       Ph_al2_5Rn9
      Ph_al2_5Rn11
      Ph_al2_5Rn12
      Ph_al2_5Rp11
       Ph_al2_5Rg1
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01 SLHDLRLKARNQEVRLEMLRKAGEITA
    Ph_al2_3Rfwx05 SLHDLRLKARNQEVRLEMLRKAGEITA
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.5_Ph_al2]

[PhD_A2.5.2_Ph_al2_alignments]

## “alignment_Ph_al2_all_bp_sequences.gif”

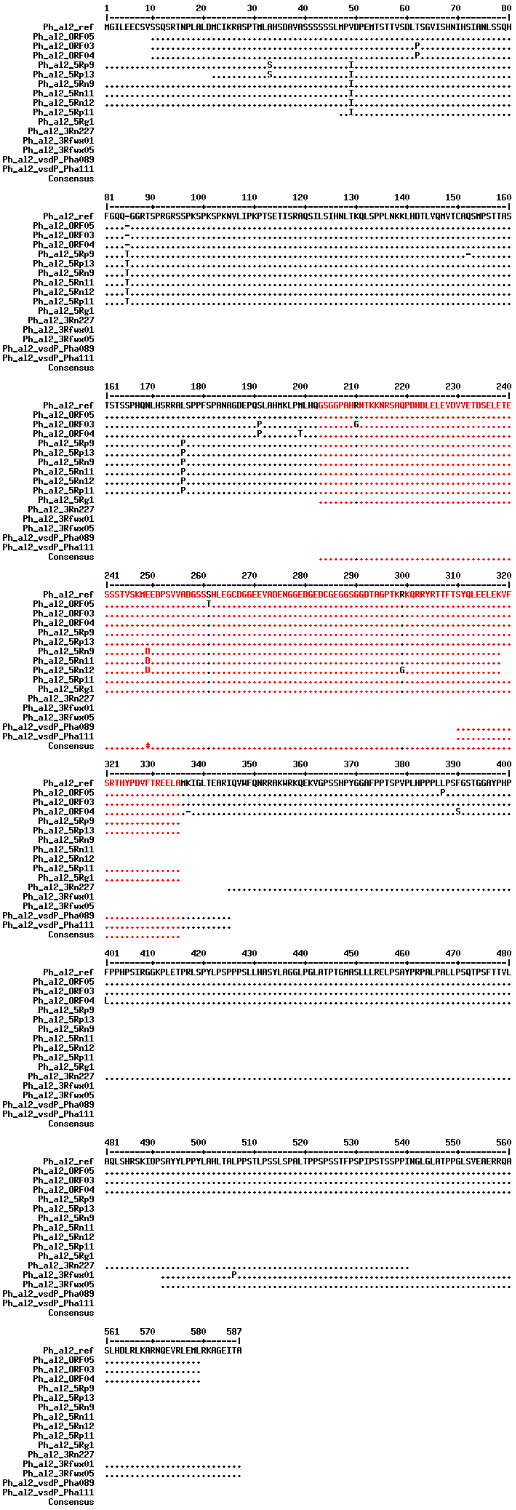


# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.5_Ph_al2]

### [PhD_A2.5.2_Ph_al2_alignments]

# “difference_alignment_Ph_al2_all_aa_sequences.gif”



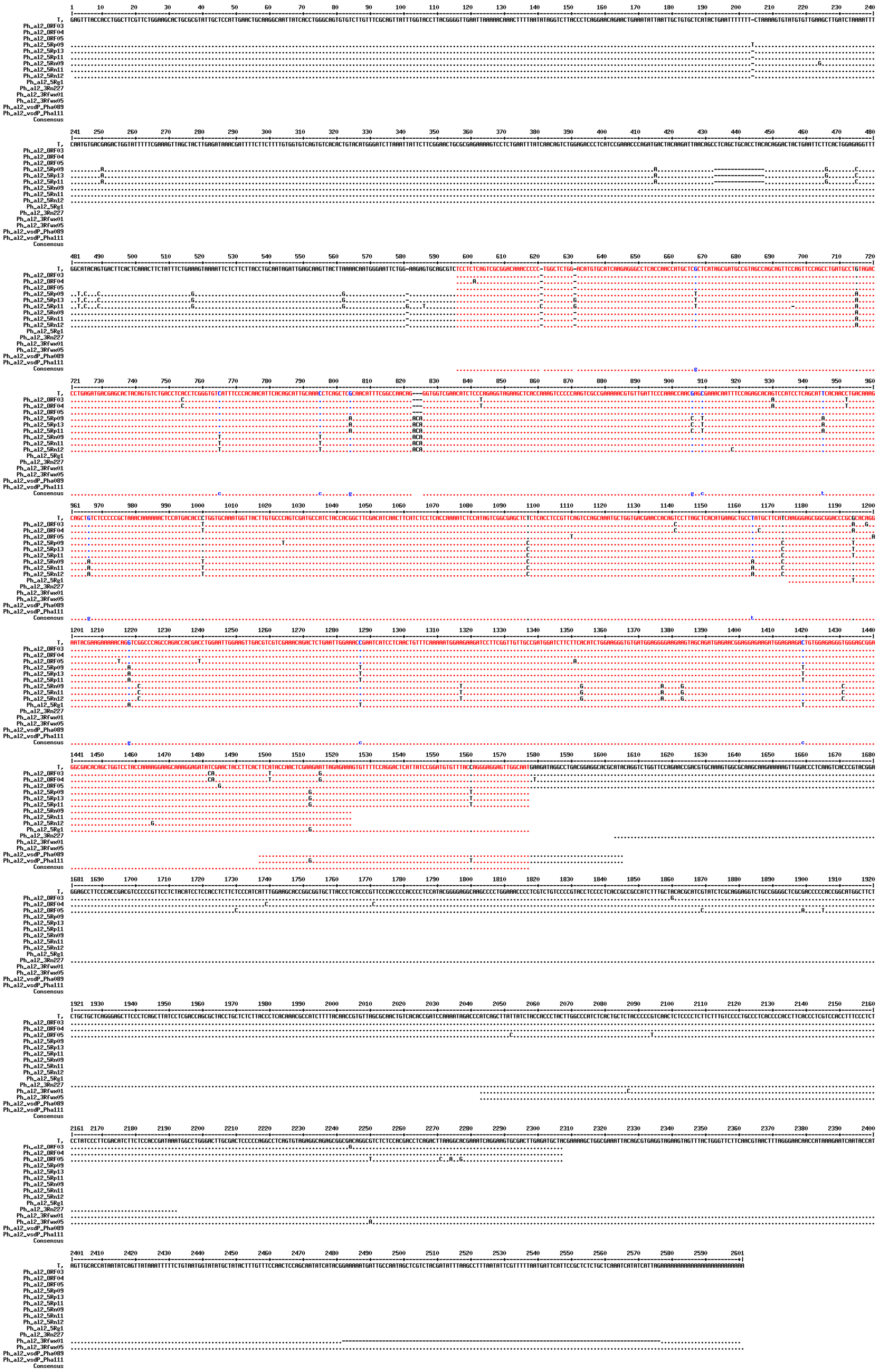
```
                   1       10        20        30        40        50        60        70        80
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref MGILEECSVSSQSRTNPLALDMCIKRASPTMLAHSDAVASSSSSSLMPVDPEMTSTTVSDLTSGVISHNIHSIANLSSQH
      Ph_al2_ORF05          .......................................................................
      Ph_al2_ORF03          ....................................................P..................
      Ph_al2_ORF04          ....................................................P..................
       Ph_al2_5Rp9 ................................S...............I...............................
      Ph_al2_5Rp13                      ...........S...............I...............................
       Ph_al2_5Rn9 ................................................I...............................
      Ph_al2_5Rn11 ................................................I...............................
      Ph_al2_5Rn12 ................................................I...............................
      Ph_al2_5Rp11                                               ..I...............................
       Ph_al2_5Rg1
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus

                   81      90        100       110       120       130       140       150       160
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref FGQQ-GGRTSPRGRSSPKSPKSPKNVLIPKPTSETISRAQSILSIHNLTKQLSPPLNKKLHDTLVQMVTCAQSMPSTTAS
      Ph_al2_ORF05 ....-...........................................................................
      Ph_al2_ORF03 ....-...........................................................................
      Ph_al2_ORF04 ....-...........................................................................
       Ph_al2_5Rp9 ....T..................................................................-........
      Ph_al2_5Rp13 ....T...........................................................................
       Ph_al2_5Rn9 ....T...........................................................................
      Ph_al2_5Rn11 ....T...........................................................................
      Ph_al2_5Rn12 ....T...........................................................................
      Ph_al2_5Rp11 ....T...........................................................................
       Ph_al2_5Rg1
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus

                   161     170       180       190       200       210       220       230       240
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref TSTSSPHQNLHSRRALSPPFSPANAGDEPQSLAHMKLPMLHQGSGGPAHRNTKKNRSAQPDHDLELEVDVVETDSELETE
      Ph_al2_ORF05 ................................................................................
      Ph_al2_ORF03 ..............................P..................G..............................
      Ph_al2_ORF04 ..............................P.......T.........................................
       Ph_al2_5Rp9 ...............P................................................................
      Ph_al2_5Rp13 ...............P................................................................
       Ph_al2_5Rn9 ...............P................................................................
      Ph_al2_5Rn11 ...............P................................................................
      Ph_al2_5Rn12 ...............P................................................................
      Ph_al2_5Rp11 ...............P................................................................
       Ph_al2_5Rg1                                           ......................................
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus                                           ......................................

                   241     250       260       270       280       290       300       310       320
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref SSSTVSKMEEDPSVVADGSSSHLEGCDGGEEVADENGGEDGEDCGEGGSGGDTAGPTKRKQRRYRTTFTSYQLEELEKVF
      Ph_al2_ORF05 ....................T...........................................................
      Ph_al2_ORF03 ................................................................................
      Ph_al2_ORF04 ................................................................................
       Ph_al2_5Rp9 ................................................................................
      Ph_al2_5Rp13 ................................................................................
       Ph_al2_5Rn9 ........D.....................................................................
      Ph_al2_5Rn11 ........D.....................................................................
      Ph_al2_5Rn12 ........D.................................................G...................
      Ph_al2_5Rp11 ................................................................................
       Ph_al2_5Rg1 ................................................................................
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089                                                                      ...........
Ph_al2_vsdP_Pha111                                                                      ...........
         Consensus ........#.......................................................................

                   321     330       340       350       360       370       380       390       400
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref SRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSSHPYGGAFPPTSPVPLHPPPLLPSFGSTGGAYPHP
      Ph_al2_ORF05 ..................................................................P.............
      Ph_al2_ORF03 ................................................................................
      Ph_al2_ORF04 ................-....................................................S..........
       Ph_al2_5Rp9 ...............
      Ph_al2_5Rp13 ...............
       Ph_al2_5Rn9
      Ph_al2_5Rn11
      Ph_al2_5Rn12
      Ph_al2_5Rp11 ...............
       Ph_al2_5Rg1 ...............
     Ph_al2_3Rn227                         ........................................................
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089 .........................
Ph_al2_vsdP_Pha111 .........................
         Consensus ...............

                   401     410       420       430       440       450       460       470       480
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref FPPHPSIRGGKPLETPRLSPYLPSPPPSLLHASYLAGGLPGLATPTGMASLLLRELPSAYPRPALPALLPSQTPSFTTVL
      Ph_al2_ORF05 ................................................................................
      Ph_al2_ORF03 ................................................................................
      Ph_al2_ORF04 L...............................................................................
       Ph_al2_5Rp9
      Ph_al2_5Rp13
       Ph_al2_5Rn9
      Ph_al2_5Rn11
      Ph_al2_5Rn12
      Ph_al2_5Rp11
       Ph_al2_5Rg1
     Ph_al2_3Rn227 ................................................................................
    Ph_al2_3Rfwx01
    Ph_al2_3Rfwx05
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus

                   481     490       500       510       520       530       540       550       560
                   |--------+---------+---------+---------+---------+---------+---------+---------|
        Ph_al2_ref AQLSHRSKIDPSAYYLPPYLAHLTALPPSTLPSSLSPALTPPSPSSTFPSPIPSTSSPPINGLGLATPPGLSVEAERRQA
      Ph_al2_ORF05 ................................................................................
      Ph_al2_ORF03 ................................................................................
      Ph_al2_ORF04 ................................................................................
       Ph_al2_5Rp9
      Ph_al2_5Rp13
       Ph_al2_5Rn9
      Ph_al2_5Rn11
      Ph_al2_5Rn12
      Ph_al2_5Rp11
       Ph_al2_5Rg1
     Ph_al2_3Rn227 ............................................................
    Ph_al2_3Rfwx01            ..............P......................................................
    Ph_al2_3Rfwx05            .....................................................................
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus

                   561     570       580    587
                   |--------+---------+------|
        Ph_al2_ref SLHDLRLKARNQEVRLEMLRKAGEITA
      Ph_al2_ORF05 ...................
      Ph_al2_ORF03 ...................
      Ph_al2_ORF04 ...................
       Ph_al2_5Rp9
      Ph_al2_5Rp13
       Ph_al2_5Rn9
      Ph_al2_5Rn11
      Ph_al2_5Rn12
      Ph_al2_5Rp11
       Ph_al2_5Rg1
     Ph_al2_3Rn227
    Ph_al2_3Rfwx01 ...........................
    Ph_al2_3Rfwx05 ...........................
Ph_al2_vsdP_Pha089
Ph_al2_vsdP_Pha111
         Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.5_Ph_al2]

[PhD_A2.5.2_Ph_al2_alignments]

## **“difference_alignment_Ph_al2_all_bp_sequences.gif”**



# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.6_Ph_pby1_pby2]

### [PhD_A2.6.1_Ph_pby1_pby2_sequences]

#### “Ph_pby1_general_read_me.txt”

```
General aspects:

(i)       All sequences are in FASTA format.

(ii)      [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)     Comments are preceded by an apostrophe (')

(iv)      All sequences without olignonucleotide sequence are indicated (e.g. "w/o 1614, 1615"). In these cases, only novel
Parhyale sequence is presented.


Parhyale hawaiensis pby1 terminology:


(i)       In general, sequence denotation begins with: ">Ph_pby1_"

(ii)      The above (i) is followed by: "A" or "B"          in case presented sequences represent isoform A or B
                                         "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                         "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                         "ORF"               for coherent ORF sequences
                                         "cDNA"              for complete cDNA sequences

                                         "ref"               if reference sequence is presented
                                         "HD"                if homeobox/homeodomain fragment is presented
                                         "PAX"               if paired/Pax domain (encoding fragment) is presented

(iii)     The above (ii) is followed by clone id.


Parhyale hawaiensis pby1 synopsis:

pby1 isoform A:

 cDNA: 2149 bp
5'UTR:   99 bp (1-99)
  ORF: 1653 bp (100-1752), 550 aa
3'UTR:  397 bp (1753-2149)
  PAX:  396 bp (250-645), 132 aa
   HD:  180 bp (727-906), 60 aa


pby1 isoform B:

 cDNA: 1791 bp
5'UTR:   99 bp (1-99)
  ORF: 1191 bp (100-1290), 396 aa
3'UTR:  501 bp (1291-1791)
  PAX:  396 bp (250-645), 132 aa
   HD:  180 bp (727-906), 60 aa


shared cDNA fraction of both isoforms:

 cDNA: 1174 bp
5'UTR:   99 bp (1-99)
  ORF: 1075 bp (100-1174), 358 aa
3'UTR: none
  PAX:  396 bp (250-645), 132 aa
   HD:  180 bp (727-906), 60 aa



>Ph_pby1_PAX [bp] 'shared by both isoforms
GCAAACAGCCGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTC
AAAGTGTCCCACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGT
ATCTGGTCATTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCC
TACAGGGAC

>Ph_pby1_HD [bp] 'shared by both isoforms
CAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCT
AGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCTAGC

>Ph_pby1A_ref [bp]
```

```
GGATTTTACAAAGAACCCCTTTACCAACTCTTAACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCAATGGAGTGCATGTATAGCTTTAGGCCTCAG
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAGGAACATCGCTGCTGTACCCACGGCAGCAGGTTGACGGCTTTGGCTTCAACTACAACACGCCGTACAACCCGGCAGTTCATCCTGGCTACCCATTTGGGGACCCTGGCTACCCAGGA
CAGATGTACTTGGGTCACGTGGCGGCCCAGCAGGCATTCCAGGCCCCGGGGAGGTCGGGTCAGAGTGAAGTCCAGACTCCCATGAGTGGTGATCGTCCCAGCGCTGGTCAGGAGCTAAACCAGCGCCAC
CGCTCCCGTGACAGCCTCCAGTTCCCTGGGCCCCAGAGCACTTACGGTTACCCTCAGCCCCTTAGCAACAACACTAGTGACGTGACAGCAACTAGGCTGTCGCCTCCCTGCACCAGTTACGCTGGAGCC
TCAGAGCAGCAGCAGCAGGGTCCTGTTACCACCTCAGCAGCCAACTTCTGGCAGCCCAGCAACCTAAAGCCCGAGCCTGGCACATCCTCACTTGGAAGCTCGAGCTTCGCCACTGCTGCCAATGGCGGA
TCTTTCTGTGAGAACCAAACGGGAACTTCTTCGAGCGCCGACCAGCAGCGACCGATGTCCGTGGGCCTCGCTTGAGGCCTTGCCCTCAGCTACAGCATGGTAGTCCTGTACTGTGTCATGCTACACCGT
ACTCTGCAGTACTGTAACACCAGTAATACTGCGCGTCGTAGCGCTGTAACTTACAGAACTGTGTTACATAGTACCATAGCCTGCACAAGTTGGACCCGTTGTTGTAGACCCCAGCTCATTGTACAGGAG
TGTACTGTAGTCTTGCTGAAGATTCTCAGAAGAATTTTTGCTTCCTCACGCAAGGTGAAGTTATTTAATTGGTGTTGTTGCTTCTAGCTGTGGACGGCGAAGATTGCTCTGCAGGGTGCTTCTATTTAT
GTTGTAAGCTTAGTTTATAAGTGAAATCCACTTACGATGGTTGGTCAAATGGGTCCAGGAAAAAGAAAAAAAAAAAAAAAAAAA

>Ph_pby1B_ref [bp]
GGATTTTACAAAGAACCCCTTTACCAACTCTTAACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCAATGGAGTGCATGTATAGCTTTAGGCCTCAG
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAGATGGTCCTGGATTCCTCTTTCCAATGCAAGGACAAGCTGCTGTGCTGAACATGTCTTTCTCGGATGGAGACGAACCTGAGAATTATAGCGGGAATCCCCCGGAAGCTTCTTATTAG
TTTTGTTTTACACGTCAACTTCCTCGCCACTTGGTGTTTCATATTGTAATGGCTCTAAATAAGGGAGTTGTTTTATGTTCTTCTAGCATTTCACTAAATTAAGCATGATTTGCTCCCCATAGAAGTAGC
AGTTTTCATATGATGTTTAGAGCGTGGGAACTGGGAACTACTGCCCCAGTAGTGCCAAGTCAGAAGTTGAATGGCTGAGCGCCACAATACGATTTAATAGATTCAGATGAACCCTTAGCGAACATCATT
GGTTGTATTTTATCGACTTCACTCAATCACTTTATAATATGATGCTACTTTTCTAAGCTGCCTAGTTTTAAACTTTTGACTACATCACTTCGTACTTACGCATTAACTTAACTTTGAGGACAGCCAGAG
TATCCTTCGTTTAAATTTTGAATATTATGGACTACTGTACAAGTGTGTTAATAGTACTGTAAATAAAGTAGTGGGAGCAATGACGAAAAAAAAAAAAAAAAAAAAAAAAAAAA


>Ph_pby1_shared_ref [bp] 'shared by both isoforms
GGATTTTACAAAGAACCCCTTTACCAACTCTTAACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCAATGGAGTGCATGTATAGCTTTAGGCCTCAG
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAG



>Ph_pby1_PAX [aa]
ANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRP
YRD

>Ph_pby1_HD [aa]
QRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKAS

>Ph_pby1A_ref [aa]
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRL
DERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWF
SNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPG
YPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQSEVQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQGPVTTSAANFWQPSNLKPEPGTSSLGS
SSFATAANGGSFCENQTGTSSSADQQRPMSVGLA&

>Ph_pby1B_ref [aa]
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRL
DERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWF
SNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENY
SGNPPEASY@

>Ph_pby1_shared_ref [aa]
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRL
DERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWF
SNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQ
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.6_Ph_pby1_pby2]

### [PhD_A2.6.1_Ph_pby1_pby2_sequences]

#### “Ph_pby1_cDNA_sequences.txt”

```
>Ph_pby1A_ref [bp] 'Ph_pby1_ORF16 w/o primers, extended 5' by Ph_pby1_5R0101 (129 bp); extended 3' by Ph_pby1_3R2219, 1597C>T
GGATTTTACAAAGAACCCCTTTACCAACTCTTAACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCAATGGAGTGCATGTATAGCTTTAGGCCTCAG
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAGGAACATCGCTGCTGTACCCACGGCAGCAGGTTGACGGCTTTGGCTTCAACTACAACACGCCGTACAACCCGGCAGTTCATCCTGGCTACCCATTTGGGGACCCTGGCTACCCAGGA
CAGATGTACTTGGGTCACGTGGCGGCCCAGCAGGCATTCCAGGCCCCGGGGAGGTCGGGTCAGAGTGAAGTCCAGACTCCCATGAGTGGTGATCGTCCCAGCGCTGGTCAGGAGCTAAACCAGCGCCAC
CGCTCCCGTGACAGCCTCCAGTTCCCTGGGCCCCAGAGCACTTACGGTTACCCTCAGCCCCTTAGCAACAACACTAGTGACGTGACAGCAACTAGGCTGTCGCCTCCCTGCACCAGTTACGCTGGAGCC
TCAGAGCAGCAGCAGCAGGGTCCTGTTACCACCTCAGCAGCCAACTTCTGGCAGCCCAGCAACCTAAAGCCCGAGCCTGGCACATCCTCACTTGGAAGCTCGAGCTTCGCCACTGCTGCCAATGGCGGA
TCTTTCTGTGAGAACCAAACGGGAACTTCTTCGAGCGCCGACCAGCAGCGACCGATGTCCGTGGGCCTCGCTTGAGGCCTTGCCCTCAGCTACAGCATGGTAGTCCTGTACTGTGTCATGCTACACCGT
ACTCTGCAGTACTGTAACACCAGTAATACTGCGCGTCGTAGCGCTGTAACTTACAGAACTGTGTTACATAGTACCATAGCCTGCACAAGTTGGACCCGTTGTTGTAGACCCCAGCTCATTGTACAGGAG
TGTACTGTAGTCTTGCTGAAGATTCTCAGAAGAATTTTTGCTTCCTCACGCAAGGTGAAGTTATTTAATTGGTGTTGTTGCTTCTAGCTGTGGACGGCGAAGATTGCTCTGCAGGGTGCTTCTATTTAT
GTTGTAAGCTTAGTTTATAAGTGAAATCCACTTACGATGGTTGGTCAAATGGGTCCAGGAAAAAGAAAAAAAAAAAAAAAAAAA

>Ph_pby1A_ORF09 [bp] 'w/o 1614, 1616
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCCCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCTTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCACGGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCCAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAGGAACATCGCTGCTGTACCCACGGCAGCAGGTTGACGGCTTTGGCTTCAACTACAACACGCCGTACAACCCGGCAGTTCATCCTGGCTACCCATTTGGGGACCCTGGCTACTCGGGA
CAGATGTACTTGGGTCACGTGGCGGCCCAGCAGGCATTCCAGGCCCCAGGGAGGTCGGGTCACAGTGAAGTCCAGACTCCCAGGAGTGGTGATCGTCCCAGCGTTGGTCAGGAGCTAAACCAGCGCAAC
CGCTCCCGTGACAGCCTCCAGTTCCCAGGGCCTCAGAGCACTTACGGTTACCCTCAGCCCCTTAGCAACAACACTAGTGACGTGACAGCTACTAGGCTGTCGCCTCCCTGCACCAGTTACGCTGGAGCC
TCAGAGCAGCAGCAGCAGGGTCCTGTTACCACCTCAGCAGCCAACTTCTGGCAGCCCAGCAACCTAAAGCCCGGGCCTGGCACATCCTCACTTGGAAGCTCGAGCTTCGCCACTGCTGCCAATGGCGGA
TCTTTCTGTGAGAACCAAACGGGAACTTCTTCGAGCCCCGACCAGCAGCGA

>Ph_pby1A_ORF13 [bp] 'w/o 1614, 1616
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCCCCCTTGCCATCCACCGCCAACCTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCATGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTCAACCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGCCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGGGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCCGAATTCACGGGAACGCCCGTTAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCTTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCACAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCCAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAACCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAGGAACATCGCTGCTGTACCCACGGCAGCAGGTTGACGGCTTTGGCTTCAACTACAACACGCCGTACAACCCGGCAGTTCATCCTGGCTACCCATTTGGGGACCCTGGCTACTCGGGA
CAGATGTACTTGGGTCACGTGGCGGCCCAGCAGGCATTCCAGGCCCCAGGGAGGTCGGGTCACAGTGAAGTCCAGACTCCCATGAGTGGTGATCGTCCCAGCGTTGGTCAGGAGCTAAACCAGCGCAAC
CGCTCCCGTGACAGCCTCCAGTTCCCAGGGCCTCAGAGCACTTACGGTTACCCTCAGCCCCTTAGCAACAACACTAGTGACGTGACAGCTACTAGGCTGTCCCCTCCCTGCACCAGTTACGCTGGAGCC
TCAGAGCAGCAGCAGCAGGGTCCTGTTACCACCTCAGCAGCCAACTTCTGGCAGCCCAGCAACCTAAAGCCCGAGCCTGGCACATCCTCACTTGGAAGCTCGAGCTTCGCCACTGCTGCCAATGGCGGA
TCTTTCTGTGAGAACCAAACGGGAACTTCTTCGAGCGCCGACCAGCAGCGA

>Ph_pby1A_ORF16 [bp] 'w/o 1614, 1616
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAGGAACATCGCTGCTGTACCCACGGCAGCAGGTTGACGGCTTTGGCTTCAACTACAACACGCCGTACAACCCGGCAGTTCATCCTGGCTACCCATTTGGGGACCCTGGCTACCCAGGA
CAGATGTACTTGGGTCACGTGGCGGCCCAGCAGGCATTCCAGGCCCCGGGGAGGTCGGGTCAGAGTGAAGTCCAGACTCCCATGAGTGGTGATCGTCCCAGCGCTGGTCAGGAGCTAAACCAGCGCCAC
CGCTCCCGTGACAGCCTCCAGTTCCCTGGGCCCCAGAGCACTTACGGTTACCCTCAGCCCCTTAGCAACAACACTAGTGACGTGACAGCAACTAGGCTGTCGCCTCCCTGCACCAGTTACGCTGGAGCC
TCAGAGCAGCAGCAGCAGGGTCCTGTTACCACCTCAGCAGCCAACTTCCGGCAGCCCAGCAACCTAAAGCCCGAGCCTGGCACATCCTCACTTGGAAGCTCGAGCTTCGCCACTGCTGCCAATGGCGGA
TCTTTCTGTGAGAACCAAACGGGAACTTCTTCGAGCGCCGACCAGCAGCGA

>Ph_pby1_5R0101 [bp] 'w/o 1607
GGATTTTACAAAGAACCCCTTTACCAACTCTTAACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCAATGGAGTGCATGTATAGCTTTAGGCCTCAG
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCCCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCATGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATCCTTAACCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTTGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCGGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTCACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAACT

>Ph_pby1_5n54 [bp] 'w/o 1607
GGATTTTACCAATCACCCCTTTACCAAGCGTTTACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCAATGGAGTGCATGTATAGCTTTAGGCCTCAG
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
```

```
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCT

>Ph_pby1_5n55 [bp] 'w/o 1607
GGATTTTACAAAGAACCCCTTTACCAACTCTTAACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCAATGGAGTGCATGTATAGCTTTAGGCCTCAG
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGTTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCT

>Ph_pby1_vsdP_PhQW11 [bp] 'w/o fw_QRR, re_QVW
TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTG

>Ph_pby1_vsdP_Phad02 [bp] 'w/o fw_QRR, re_QVW
TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTG

>Ph_pby1_vsdP_Phad06 [bp] 'w/o fw_QRR, re_QVW
TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGTGAGCACCTGGCAGAGAGGACTGGGCTCACTAAGGCTAGAGTG

>Ph_pby1_vsdP_Phad08 [bp] 'w/o fw_QRR, re_QVW
TGCAATTCAGCTGGAAATCCTGGAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTG

>Ph_pby1_vsdP_Phad14 [bp] 'w/o fw_QRR, re_QVW
TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGTGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGTTAGAGTA

>Ph_pby1A_3R2219 [bp] 'w/o 1602
GGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCTAGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTG
CCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGAATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGC
ATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCGCCTGCATACCAAGGAACATCGCTGCTGTACCCACGGCAGCAGGTTGACGGCTTTGGCTTCAACTACA
ACACGCCGTACAACCCGGCAGTTCATCCTGGCTACCCATTTGGGGACCCTGGCTACCCAGGACAGATGTACTTGGGTCACGTGGCGGCCCAGCAGGCATTCCAGGCCCCGGGGAGGTCGGGTCAGAATG
AAGTCCAGACTCCCATGAGTGGTGATCGTCCCAGCGCTGGTCAGGAGCTAAACCAGCGCCACCGCTCCCGTGACAGCCTCCAGTTCCCTGGGCCCCAGAGCACTTACGGTTACCCTCAGCCCCTTAGCA
ACAACACTTGTGACGTGACAGCAACTAGGCTGTCGCCTCCCTGCACCAGTTACGCTGGAGCCTCAGAGCAGCAGCAGCAGGGTCCTGTTACCACCTCAGCAGCCAACTTCTGGCAGCCCAGCAACCTAA
AGCCCGAGCCTGGCACATCCTCACTTGGAAGCTCGAGCTTCGCCACTGCTGCCAATGGCGGATCTTTCTGTGAGAACCAAACGGGAACTTCTTCGAGCGCCGACCAGCAGCGACCGATGTCCGTGGGCC
TCGCTTGAGGCCTTGCCCTCAGCTACAGCATGGTAGTCCTGTACTGTGTCATGCTACACCGTACTCTGCAGTACTGTAACACCAGTAATACTGCGCGTCGTAGCGCTGTAACTTACAGAACTGTGTTAC
ATAGTACCATAGCCTGCACAAGTTGGACCCGTTGTTGTAGACCCCAGCTCATTGTACAGGAGTGTACTGTAGTCTTGCTGAAGATTCTCAGAAGAATTTTTGCTTCCTCACGCAAGGTGAAGTTATTTA
ATTGGTGTTGTTGCTTCTAGCTGTGGACGGCGAAGATTGCTCTGCAGGGTGCTTCTATTTATGTTGTAAGCTTAGTTTATAAGTGAAATCCACTTACGATGGTTGGTCAAATGGGTCCAGGAAAAAGAA
AAAAAAAAAAAAAAAAA

>Ph_pby1A_3Rf27 [bp] 'w/o 1602
GGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCTAGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTG
CCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGAATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATACTGTTGTCCCCCCTGTGAAGATCGAGC
ATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCGCCTGCATACCAAGGAACATCGCTGCTGTACCCACGGCAGCAGGTTGACGGCTTTGGCTTCAACTACA
ACACGCCGTACAACCCGGCAGTTCATCCTGGCTACCCATTTGGGGACCCTGGCTACCCAGGACAGATGTACTTGGGTCACGTGGCGGCCCAGCAGGCATTCCAGGCCCCGGGGAGGTCGGGTCAGAGTG
AAGTCCAGACTCCCATGAGTGGTGATCGTCCCAGCGCTGGTCAGGAGCTAAACCAGCGCCACCGTTCCCGTGACAGCCTCCAGTTCCCTGGGCCCCAGAGCACTTACGGTTACCCTCAGCCCCTTAGCA
ACAACACTAGTGACGTGACAGCAACTAGGCTGTCGCCTCCCTGCACCAGTTACGCTGGAGCCTCAGAGCAGCAGCAGCAGGGTCCTGTTACCACCTCAGCAGCCAACTTCTGGCAGCCCAGCAACCTAA
AGCCCGAGCCTGGCACATCCTCACTTGGAAGCTCGAGCTTCGCCACTGCTGCCAATGGCGGATCTTTCTGTGAGAACCAAACGGGAACTTCTTCGAGCGCCGACCAGCAGCGACCGATGTCCGTGGGCC
TCGCTTGAGGCCTTGCCCTCAGCTACAGCATGGTAGTCCTGTACTGTGTCATGCTACACCGTACTCTGCAGTACTGTAACACCAGTAATACTGCGCGTCGTAGCGCTGTAACTTACAGAACTGTGTTAC
ATAGTACCATAGCCTGCACAAGTTGGACCCGTTGTTGTAGACCCCAGCTCATTGTACAGGAGTGTACTGTAGTCTTGCTGAAGATTCTCAGAAGAATTTTTGCTTCCTCACGCAAGGTGAAGTTATTTA
ATTGGTGGTGGTGCTTCTAGCTGTGGACGGCGAAGATTGCTCTGCAGGGTGCTTCTATTTATGTTGGAAGCTTAGTTTATAAGTGAAATCCACTTACGATGGTTGGTCAAATGGGTCAGGAAAAGAAAA
AAAAAAAAAAAAA

>Ph_pby1A_3Rf29 [bp] 'w/o 1602
GGCTCACTGAGGCTAGAGTGCAGGTTCGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCTAGCAACAGCCCGGAAGAGGCCGCACAGTGCCCTACCAACAACAGAAACGGCGGATCCCCAGTAGCTG
CCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGAATGTTCAACAGATACGGTCTGCAGGCTCCTGCCAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGC
ATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCGCCTGCATACCAAGGAACATCGCTGCTGTACCCACGGCAGCAGGTTGACGGCTTTGGCTTCAACTACA
ACACGCCGTACAACCCGGCAGTTCATCCTGGCTACCCATTTGGGGACCCTGGCTACCCAGGACAGATGTACTTCGGTCACGTGGCGGCCCAGCAGGCATTCCAGGCCCCGGGGAGGTCGGGTCAGAGTG
AAGTCCAGACTCCCATGAGTGGTGATCGTTCCAGCGCTGGTCAGGAGCTAAACCAGCGCAACCGCTCCCGTGACAGCCTCCAGTTCCCTGGGCCCCAGAGCACTTACGGTTACCCTCAGCCCCTTAGCA
ACAACACTAGTGACGTGCCAGCAACTAGGCTGTCGCCTCCCTGCACCAGTTACGCTGGAGCCTCAGAGCAGCAGCAGCAGGGTCCTGTTACCACCTCAGCAGCCAACTTCTGGCAGCCCAGCAACCTAA
GGCCCGAGCCTGGCACATCCTCACTTGGAAGCTCGAGCTTCGCCACTGCTGCCAATGGCGGATCTTTCTGTGAGAACCAAACGGGAACTTCTTCGAGCGCCGACCAGCAGCGACCGATGTCCGTGGGCC
TCGCTTGAGGCCTTGCCCTCAGCTGCAGCATGGTAGTCCTGTACTGTGTCATGCTACACCGTACTCTGCAGTACTGTAACGCCAGTAATACTGCGCGTCGTAGCGCTGTAACTTACAGAACTGTGTTAC
ATAGTACCATAGCCTGCACAAGTTGGACCCGTTGTTGTAGACCCCAGCTCATTGTACAGGAGTGTACTGTAGTCTTGCTGAAGATTCTCAGAAAAATTTTTGCTTCCTCATGCAAGGTGAAGTGGTTAA
ATTGGTGGTGGTGCTCTCTAGCTGTGGACGGCGAAGATTGCTCTGCAGGGTGCTTCTATTTATGTTGTAAGCTTAGTTTATAAGTGAAATCCACTTACCATGGTTGGGTCAATGGTCCAGGAAAAGAAA
AAAAAAAAAAAAA


>Ph_pby1B_ref [bp] 'Ph_pby1_ORF01 w/o primers, extended 5' by Ph_pby1_5R0101 (129 bp); extended 3' by Ph_pby1_3Rf316 (530
bp),433C>T, 1273T>C
GGATTTTACAAAGAACCCCTTTACCAACTCTTAACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCAATGGAGTGCATGTATAGCTTTAGGCCTCAG
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAGATGGTCCTGGATTCCTCTTTCCAATGCAAGGACAAGCTGCTGTGCTGAACATGTCTTTCTCGGATGGAGACGAACCTGAGAATTATAGCGGGAATCCCCCGGAAGCTTCTTATTAG
TTTTGTTTTACACGTCAACTTCCTCGCCACTTGGTGTTTCATATTGTAATGGCTCTAAATAAGGGAGTTGTTTTATGTTCTTCTAGCATTTCACTAAATTAAGCATGATTTGCTCCCCATAGAAGTAGC
AGTTTTCATATGATGTTTAGAGCGTGGGAACTGGGAACTACTGCCCCAGTAGTGCCAAGTCAGAAGTTGAATGGCTGAGCGCCACAATACGATTTAATAGATTCAGATGAACCCTTAGCGAACATCATT
GGTTGTATTTTATCGACTTCACTCAATCACTTTATAATATGATGCTACTTTTCTAAGCTGCCTAGTTTTAAACTTTTGACTACATCACTTCGTACTTACGCATTAACTTAACTTTGAGGACAGCCAGAG
TATCCTTCGTTTAAATTTTGAATATTATGGACTACTGTACAAGTGTGTTAATAGTACTGTAAATAAAGTAGTGGGAGCAATGACGAAAAAAAAAAAAAAAAAAAAA


>Ph_pby1B_ORF01 [bp] 'w/o 1614, 1615
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGCCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
```

```
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAGATGGTCCTGGATTCCTCTTTCCAATGCAAGGACAAGCTGCTGTGCTGAACATGTCTTTCTCGGATGGAGACGAACCTGAGAATTATA

>Ph_pby1B_ORF03 [bp] 'w/o 1614, 1615
CCCTTTGGGCCAATGGGTGTGTCTGCTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAGACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCATGCAATATTTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAACCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAATAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCGGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTCACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGTGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTGGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCACAGCGCCCTGCCAACAGCAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCCAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATGCCAAGATGGTCCTGGATTCCTCTTTCCAATGCAAGGACAAGCTGCTGTGCTGAACATGTCTTTCTCGGATGGAGACGAACCTGAGAATTATA

>Ph_pby1B_ORF05 [bp] 'w/o 1614, 1615
CCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACCGCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGC
CGCGTCAACCAGCTGGGGGGCAAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGAATCTGTCCCTGCAATATCTCCAAAGCCCTCAAAGTGTCC
CACGGCTGCGTCTCCAAGATACTTAACCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGGGTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCA
TTCAGTAATTGCCATGGCGTCATGACTCTGGAACAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTACAGGGAC
CAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAGAGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTG
GAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCT
AGCAACAGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTACGGA
ATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCG
CCTGCATACCAAGATGGTCCTGGATTCCTCTTTCCAATGCAAGGACAAGCTGCTGTGCTGAACATGTCTTTCTCGGATGGAGACGAACCTGAGAATTATA

>Ph_pby1B_3R1413 [bp] 'w/o 1602
GGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCTAGCAACAGCCCGGAAGAGGCCGCACAGCGCCCTGCCAACAACAGAAACGGCGGATCCTCAGTAGCTG
CCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGAATGTTCAACAGATACGGTCTGCAGGCTCCTGCCAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGC
ATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCGCCTGCATACCAAGATGGTCCTGGATTCCTCTTTCCAATGCAAGGACAAGCTGCTGTGCTGAACATGT
CTTTCTCGGATGGAGACGAACCTGAGAATTATAGCGGGAATCCCCCGGAAGCTTCTTATTAGTTTTGTTTTACACGTCAACTTCCTCGCCACTTGGTGTTTCATATTGTAATGGCTCTAAATAAGGGAG
TTGTTTTATGTTCTTCTAGCATTTCACTAAATTAAGCATGATTTGCTCCCCATAGAAGTAGCAGTTTTCATATGATGTTTAGAGCGTGGGAACTGGGAACTACTGCCCCAGTAGTGCCAAGTCAGAAGT
TGAATGGCTGAGCGCCACAATACGATTTAATAGATTCAGATGAACCCTTAGCGAACATCATTGGTTGTATTTTATCGACTTCACTCAATCACTTTATAATATGATGCTACTTTTCTAAGCTGCCTAGTT
TTAAACTTTTGACTACATCACTTCGTACTTACGCATTAACTTAACTTTGAGGACAGCCAGAGTATCCTTCGTTTAAATTTTGAATATTATGGACTACTGTACAGTATGTTAATAGTACTGTAAATAAAG
TAGTGGGAGCAATGACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_pby1B_3Rf21 [bp] 'w/o 1602
GGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCTAGCAACAGCCCGGAAGAGGCCGCACAGCGCCCTGCCAACAACAGAAACGGCGGATCCTCAGTAGCTG
CCATCAGTCCATTCGACCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGAATGTTCAACAGATACGGTCTGCAGGCTCCTGCCAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGC
ATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCGCTTGCATACCAAGATGGTCCTGGATTCCTCTTTCCAATGCAAGGACAAGCTGCTGTGCTGAACATGT
CTTTCTCGGATGGAGACGAACCTGAGAATTATAGCGGGAATCCCCCGGAAGCTTCTTATTAGTTTTGTTTTACACGTCAACTTCCTCGCCACTTGGTGTTTCATATTGTAATGGCTCTAAATAAGGGAG
TTGTTTTATGTTCTTCTAGCATTTCACTAAATTAAGCATGATTTGCTCCCCATAGAAGTAGCAGTTTTCATATGATGTTTAGAGCGTGGGAACTGGGAACTACTGCCCCGGTAGTGCCAAGTCAGAAGT
TGAATGGCTGAGCGCCACAATACGATTTAATAGATTCAGATGAACCCTTAGCGAACATCATTGGTTGTATTTTATCGACTTCACTCAATCACTTTATAATATGATGCTACTTTTCTAAGCTGCCTAGTT
TTAAACTTTTGACTACATCACTGCGTACTTACGCATTAACTTAACTTTGAGGACAGCCAGAGTATCCTTCGTTTAAATTTTGAATATTATGGACTACTGTACAAGTGTGTTAATAGTACTGTAAATAAA
GTAGTGGGAGCAAAAAAAAAAAAAA

>Ph_pby1B_3Rf316 [bp] 'w/o 1602
GGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCTAGCAACAGCCCGGAAGAGGCCGCACAGCGCCCTGCCAACAACAGAAACGGCGGATCCTCAGTAGCTG
CCATCAGTCCATTCGGCCCAGCAGTACCTGGACCCAGTCACGGTCTGGACGGCGGCTATGGAATGTTCAACAGATACGGTCTGCAGGCTCCTGCCAGCGATGCTGTTGTCCCCCCTGTGAAGATCGAGC
ATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCGCCTGCATACCAAGATGGTCCTGGATTCCTCTTTCCAATGCAAGGACAAGCTGCTGTGCTGAACATGT
CTTTCTCGGATGGAGACGAACCTGAGAATTATAGCGGGAATCCCTCGGAAGCTTCTTATTAGTTTTGTTTTACACGTCAACTTCCTCGCCACTTGGTGTTTCATATTGTAATGGCTCTAAATAAGGGAG
TTGTTTTATGTTCTTCTAGCATTTCACTAAATTAAGCATGATTTGCTCCCCATAGAAGTAGCAGTTTTCATATGATGTTTAGAGCGTGGGAACTGGGAACTACTGCCCCAGTAGTGCCAAGTCAGAAGT
TGAATGGCTGAGCGCCACAATACGATTTAATAGATTCAGATGAACCCTTAGCGAACATCATTGGTTGTATTTTATCGACTTCACTCAATCACTTTATAATATGATGCTACTTTTCTAAGCTGCCTAGTT
TTAAACTTTTGACTACATCACTTCGTACTTACGCATTAACTTAACTTTGAGGACAGCCAGAGTATCCTTCGTTTAAATTTTGAATATTATGGACTACTGTACAAGTGTGTTAATAGTACTGTAAATAAA
GTAGTGGGAGCAATGACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA



>Ph_pby1A_ref [aa]
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRL
DERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWF
SNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPG
YPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQSEVQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQGPVTTSAANFWQPSNLKPEPGTSSLGS
SSFATAANGGSFCENQTGTSSSADQQRPMSVGLA&


>Ph_pby1A_ORF09 [aa] 'w/o 1614, 1616
PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWS
FNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKA
SNSPEEAARRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYSG
QMYLGHVAAQQAFQAPGRSGHSEVQTPRSGDRPSVGQELNQRNRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQGPVTTSAANFWQPSNLKPGPGTSSLGSSSFATAANGG
SFCENQTGTSSSPDQQR

>Ph_pby1A_ORF13 [aa] 'w/o 1614, 1616
PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWS
FNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKA
SNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSNPNQPSYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYSG
QMYLGHVAAQQAFQAPGRSGHSEVQTPMSGDRPSVGQELNQRNRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQGPVTTSAANFWQPSNLKPEPGTSSLGSSSFATAANGG
SFCENQTGTSSSADQQR

>Ph_pby1A_ORF16 [aa] 'w/o 1614, 1616
PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWS
FNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKA
SNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYPG
QMYLGHVAAQQAFQAPGRSGQSEVQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQGPVTTSAANFRQPSNLKPEPGTSSLGSSSFATAANGG
SFCENQTGTSSSADQQR
```

```
>Ph_pby1_5R0101 [aa] 'w/o 1607
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRL
DERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQ

>Ph_pby1_5n54 [aa] 'w/o 1607
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRL
DERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQ

>Ph_pby1_5n55 [aa] 'w/o 1607
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRL
DERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQ

>Ph_pby1_vsdP_PhQW11 [aa] 'w/o fw_QRR, re_QVW
AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV

>Ph_pby1_vsdP_Phad02 [aa] 'w/o fw_QRR, re_QVW
AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV

>Ph_pby1_vsdP_Phad06 [aa] 'w/o fw_QRR, re_QVW
AIQLEILEQEFCSNQYPDITTREHLAERTGLTKARV

>Ph_pby1_vsdP_Phad08 [aa] 'w/o fw_QRR, re_QVW
AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV

>Ph_pby1_vsdP_Phad14 [aa] 'w/o fw_QRR, re_QVW
AIQLEILEQEFCSNQYPDITTREHLAERTGLTEVRV

>Ph_pby1A_3R2219 [aa] 'w/o 1602
LTEARVQVWFSNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQGTSLLYPRQQVDGFGFNYN
TPYNPAVHPGYPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQNEVQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTCDVTATRLSPPCTSYAGASEQQQQGPVTTSAANFWQPSNLK
PEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA&

>Ph_pby1A_3Rf27 [aa] 'w/o 1602
LTEARVQVWFSNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDTVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQGTSLLYPRQQVDGFGFNYN
TPYNPAVHPGYPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQSEVQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQGPVTTSAANFWQPSNLK
PEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA&

>Ph_pby1A_3Rf29 [aa] 'w/o 1602
LTEARVQVRFSNRRARQRKASNSPEEAAQCPTNNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQGTSLLYPRQQVDGFGFNYN
TPYNPAVHPGYPFGDPGYPGQMYFGHVAAQQAFQAPGRSGQSEVQTPMSGDRSSAGQELNQRNRSRDSLQFPGPQSTYGYPQPLSNNTSDVPATRLSPPCTSYAGASEQQQQGPVTTSAANFWQPSNLR
PEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA&


>Ph_pby1B_ref [aa]
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRL
DERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWF
SNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENY
SGNPPEASY@

>Ph_pby1B_ORF01 [aa] 'w/o 1614, 1615
PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGPIKPGVLAGVARSLPPRLDERIVQRIWS
FNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKA
SNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENY

>Ph_pby1B_ORF03 [aa] 'w/o 1614, 1615
PFGPMGVSAPFPSASAFPPTAPLPSTANLDHENNDFDNAHADSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWS
FNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEAGVQVWFSNRRARQRKA
SNSPEEAAQRPANSRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPACQDGPGFLFPMQGQAAVLNMSFSDGDEPENY

>Ph_pby1B_ORF05 [aa] 'w/o 1614, 1615
PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGICPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWS
FSNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKA
SNSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENY

>Ph_pby1B_3R1413 [aa] 'w/o 1602
LTEARVQVWFSNRRARQRKASNSPEEAAQRPANNRNGGSSVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQDGPGFLFPMQGQAAVLNMS
FSDGDEPENYSGNPPEASY@

>Ph_pby1B_3Rf21 [aa] 'w/o 1602
LTEARVQVWFSNRRARQRKASNSPEEAAQRPANNRNGGSSVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPLAYQDGPGFLFPMQGQAAVLNMS
FSDGDEPENYSGNPPEASY@

>Ph_pby1B_3Rf316 [aa] 'w/o 1602
LTEARVQVWFSNRRARQRKASNSPEEAAQRPANNRNGGSSVAAISPFGPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQDGPGFLFPMQGQAAVLNMS
FSDGDEPENYSGNPSEASY@
```

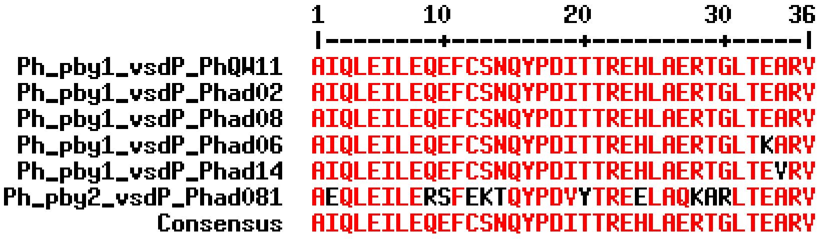
[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.6_Ph_pby1_pby2]
[PhD_A2.6.1_Ph_pby1_pby2_sequences]

## “Ph_pby2_vsdPCR.txt”

```
>Ph_pby2_vsdP_Phad081 [bp] 'w/o fw_QRR, re_QVW
TGCAGAGCAGCTGGAAATCCTTGAGAGATCCTTCGAGAAGACGCAATATCCTGACGTCTACACTAGGGAGGAGCTCGCACAGAAGGCGCGCCTGACGGAGGCTAGGGTG

>Ph_pby2_vsdP_Phad081 [aa] 'w/o fw_QRR, re_QVW
AEQLEILERSFEKTQYPDVYTREELAQKARLTEARV
```

[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.6_Ph_pby1_pby2]
[PhD_A2.6.2_Ph_pby1_pby2_alignments]

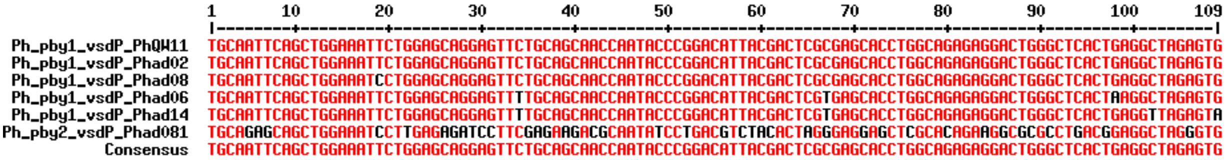
**“alignment_Ph_pby1_pby2_vsdPCR_aa_sequences.gif”**



```
                        1         10            20            30     36
                        |---------+---------+---------+-----|
  Ph_pby1_vsdP_PhQW11   AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
  Ph_pby1_vsdP_Phad02   AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
  Ph_pby1_vsdP_Phad08   AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
  Ph_pby1_vsdP_Phad06   AIQLEILEQEFCSNQYPDITTREHLAERTGLTKARV
  Ph_pby1_vsdP_Phad14   AIQLEILEQEFCSNQYPDITTREHLAERTGLTEVRV
 Ph_pby2_vsdP_Phad081   AEQLEILERSFEKTQYPDVYTREELAQKARLTEARV
            Consensus   AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
```

[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.6_Ph_pby1_pby2]
[PhD_A2.6.2_Ph_pby1_pby2_alignments]

**“alignment_Ph_pby1_pby2_vsdPCR_sequences.gif”**



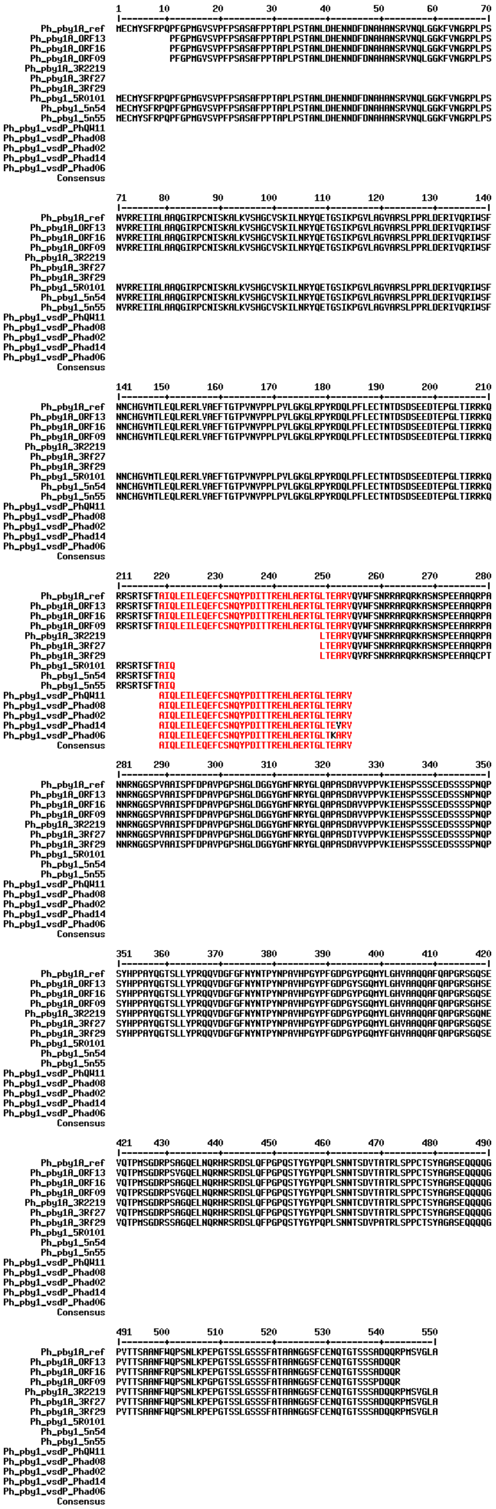
```
                       1        10        20        30        40        50        60        70        80        90        100      109
                       |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+--------|
  Ph_pby1_vsdP_PhQW11  TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTG
  Ph_pby1_vsdP_Phad02  TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTG
  Ph_pby1_vsdP_Phad08  TGCAATTCAGCTGGAAATCCTGGAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTG
  Ph_pby1_vsdP_Phad06  TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGTGAGCACCTGGCAGAGAGGACTGGGCTCACTAAGGCTAGAGTG
  Ph_pby1_vsdP_Phad14  TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGTGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGTTAGAGTA
 Ph_pby2_vsdP_Phad081  TGCAGAGCAGCTGGAAATCCTTGAGAGATCCTTCGAGAAGACGCAATATCCTGACGTCTACACTAGGGAGGAGCTCGCACAGAAGGCGCGCCTGACGGAGGCTAGGGTG
            Consensus  TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTG
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.6_Ph_pby1_pby2]

[PhD_A2.6.2_Ph_pby1_pby2_alignments]

## “alignment_Ph_pby1A_all_aa_sequences.gif”



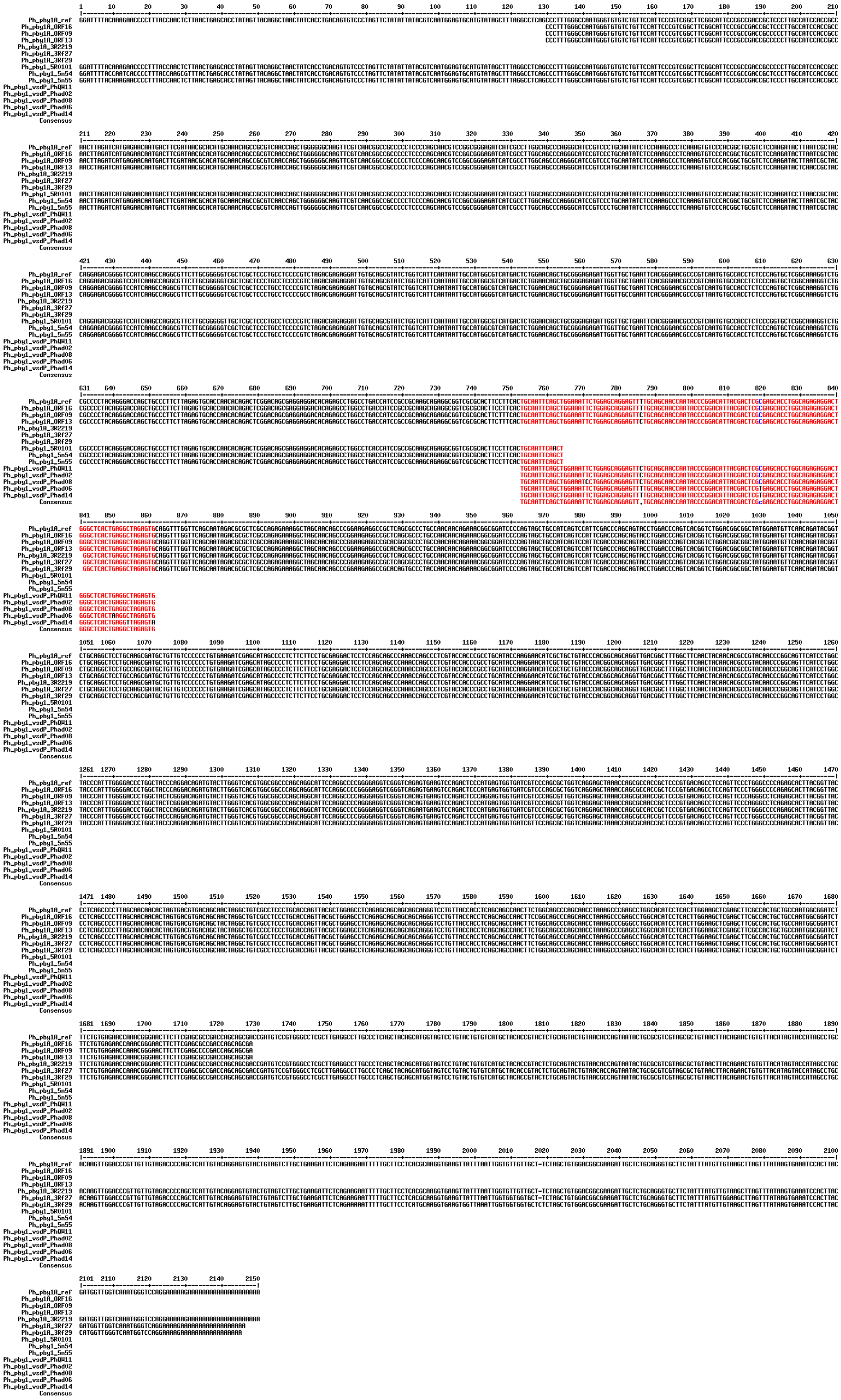
```
                     1       10        20        30        40        50        60        70
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref  MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
     Ph_pby1A_ORF13            PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
     Ph_pby1A_ORF16            PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
     Ph_pby1A_ORF09            PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
    Ph_pby1A_3R2219
     Ph_pby1A_3Rf27
     Ph_pby1A_3Rf29
     Ph_pby1_5R0101  MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
       Ph_pby1_5n54  MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
       Ph_pby1_5n55  MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     71      80        90       100       110       120       130       140
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
     Ph_pby1A_ORF13  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
     Ph_pby1A_ORF16  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
     Ph_pby1A_ORF09  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
    Ph_pby1A_3R2219
     Ph_pby1A_3Rf27
     Ph_pby1A_3Rf29
     Ph_pby1_5R0101  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
       Ph_pby1_5n54  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
       Ph_pby1_5n55  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     141    150       160       170       180       190       200       210
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
     Ph_pby1A_ORF13  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
     Ph_pby1A_ORF16  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
     Ph_pby1A_ORF09  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
    Ph_pby1A_3R2219
     Ph_pby1A_3Rf27
     Ph_pby1A_3Rf29
     Ph_pby1_5R0101  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
       Ph_pby1_5n54  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
       Ph_pby1_5n55  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     211    220       230       240       250       260       270       280
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref  RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1A_ORF13  RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1A_ORF16  RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1A_ORF09  RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAARRPA
    Ph_pby1A_3R2219                                        LTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1A_3Rf27                                        LTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1A_3Rf29                                        LTEARVQVRFSNRRARQRKASNSPEEAAQCPT
     Ph_pby1_5R0101  RRSRTSFTAIQ
       Ph_pby1_5n54  RRSRTSFTAIQ
       Ph_pby1_5n55  RRSRTSFTAIQ
Ph_pby1_vsdP_PhQW11          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
Ph_pby1_vsdP_Phad08          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
Ph_pby1_vsdP_Phad02          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
Ph_pby1_vsdP_Phad14          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEVRV
Ph_pby1_vsdP_Phad06          AIQLEILEQEFCSNQYPDITTREHLAERTGLTKARV
          Consensus          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV

                     281    290       300       310       320       330       340       350
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1A_ORF13  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSNPNQP
     Ph_pby1A_ORF16  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1A_ORF09  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
    Ph_pby1A_3R2219  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1A_3Rf27  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDTVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1A_3Rf29  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1_5R0101
       Ph_pby1_5n54
       Ph_pby1_5n55
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     351    360       370       380       390       400       410       420
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref  SYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQSE
     Ph_pby1A_ORF13  SYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYSGQMYLGHVAAQQAFQAPGRSGHSE
     Ph_pby1A_ORF16  SYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQSE
     Ph_pby1A_ORF09  SYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYSGQMYLGHVAAQQAFQAPGRSGHSE
    Ph_pby1A_3R2219  SYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQNE
     Ph_pby1A_3Rf27  SYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQSE
     Ph_pby1A_3Rf29  SYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYPGQMYFGHVAAQQAFQAPGRSGQSE
     Ph_pby1_5R0101
       Ph_pby1_5n54
       Ph_pby1_5n55
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     421    430       440       450       460       470       480       490
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref  VQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQG
     Ph_pby1A_ORF13  VQTPMSGDRPSVGQELNQRNRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQG
     Ph_pby1A_ORF16  VQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQG
     Ph_pby1A_ORF09  VQTPRSGDRPSVGQELNQRNRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQG
    Ph_pby1A_3R2219  VQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTCDVTATRLSPPCTSYAGASEQQQQG
     Ph_pby1A_3Rf27  VQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQG
     Ph_pby1A_3Rf29  VQTPMSGDRSSAGQELNQRNRSRDSLQFPGPQSTYGYPQPLSNNTSDVPATRLSPPCTSYAGASEQQQQG
     Ph_pby1_5R0101
       Ph_pby1_5n54
       Ph_pby1_5n55
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     491    500       510       520       530       540       550
                     |--------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref  PVTTSAANFWQPSNLKPEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA
     Ph_pby1A_ORF13  PVTTSAANFWQPSNLKPEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQR
     Ph_pby1A_ORF16  PVTTSAANFRQPSNLKPEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQR
     Ph_pby1A_ORF09  PVTTSAANFWQPSNLKPGPGTSSLGSSSFATAANGGSFCENQTGTSSSPDQQR
    Ph_pby1A_3R2219  PVTTSAANFWQPSNLKPEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA
     Ph_pby1A_3Rf27  PVTTSAANFWQPSNLKPEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA
     Ph_pby1A_3Rf29  PVTTSAANFWQPSNLRPEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA
     Ph_pby1_5R0101
       Ph_pby1_5n54
       Ph_pby1_5n55
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.6_Ph_pby1_pby2]

[PhD_A2.6.2_Ph_pby1_pby2_alignments]

## “alignment_Ph_pby1A_all_bp_sequences.gif”

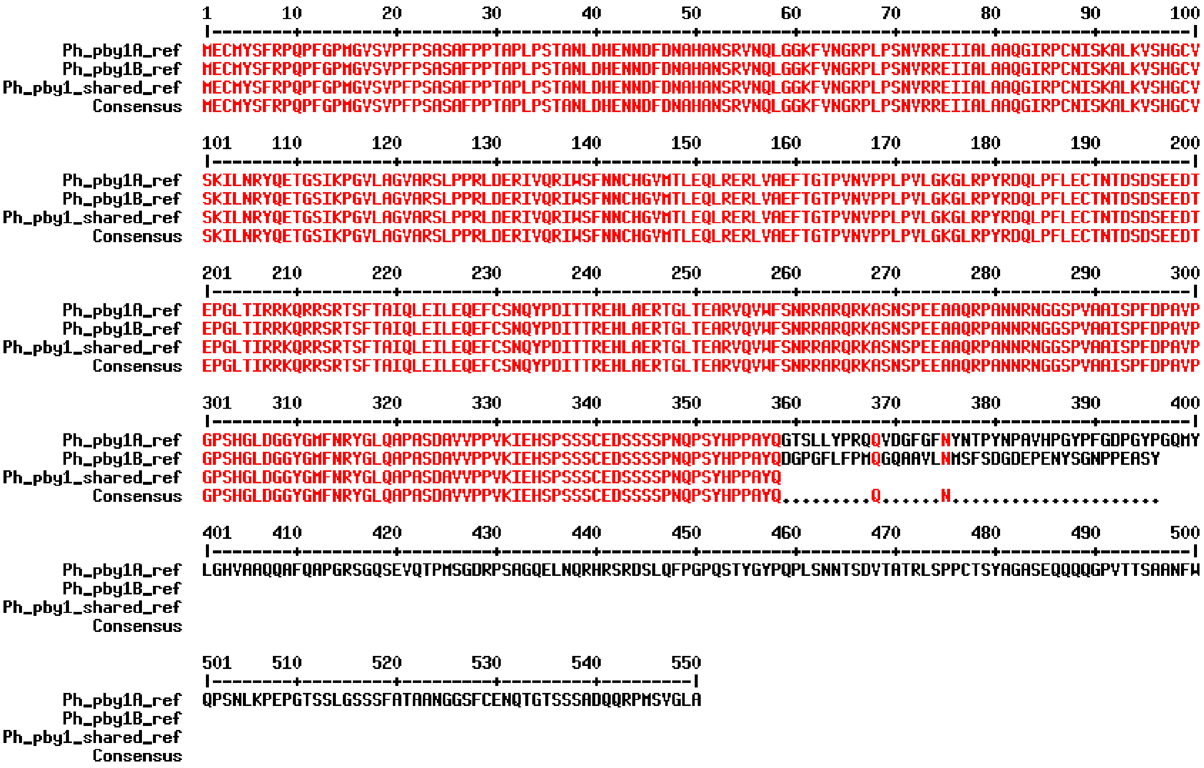


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.6_Ph_pby1_pby2]

[PhD_A2.6.2_Ph_pby1_pby2_alignments]

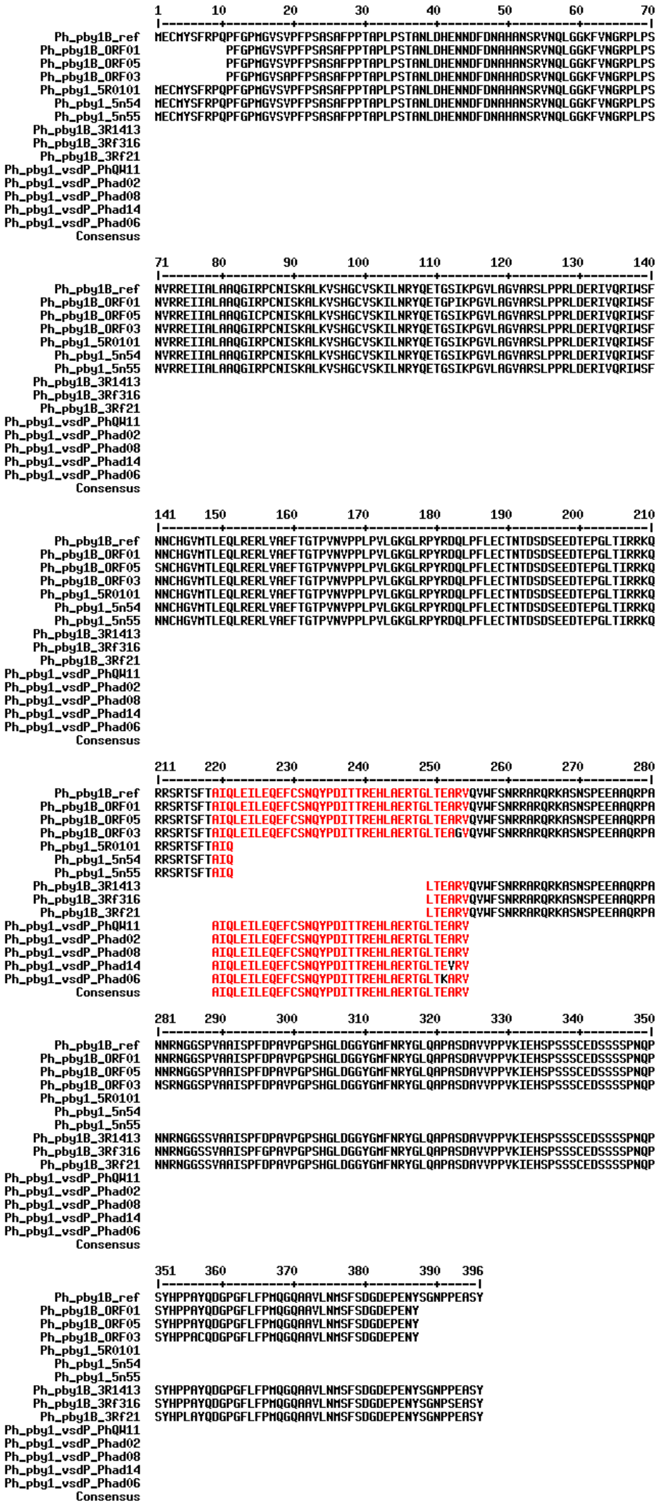
## “alignment_Ph_pby1A_pby1B_pbyshared_aa.gif”



```
                   1       10        20        30        40        50        60        70        80        90        100
                   |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
      Ph_pby1A_ref MECMYSFRPQPFGPMGVSYPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNYRREIIALAAQGIRPCNISKALKVSHGCV
      Ph_pby1B_ref MECMYSFRPQPFGPMGVSYPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNYRREIIALAAQGIRPCNISKALKVSHGCV
 Ph_pby1_shared_ref MECMYSFRPQPFGPMGVSYPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNYRREIIALAAQGIRPCNISKALKVSHGCV
         Consensus MECMYSFRPQPFGPMGVSYPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNYRREIIALAAQGIRPCNISKALKVSHGCV

                   101     110       120       130       140       150       160       170       180       190       200
                   |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
      Ph_pby1A_ref SKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDT
      Ph_pby1B_ref SKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDT
 Ph_pby1_shared_ref SKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDT
         Consensus SKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDT

                   201     210       220       230       240       250       260       270       280       290       300
                   |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
      Ph_pby1A_ref EPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVP
      Ph_pby1B_ref EPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVP
 Ph_pby1_shared_ref EPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVP
         Consensus EPGLTIRRKQRRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPANNRNGGSPVAAISPFDPAVP

                   301     310       320       330       340       350       360       370       380       390       400
                   |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
      Ph_pby1A_ref GPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYPGQMY
      Ph_pby1B_ref GPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENYSGNPPEASY
 Ph_pby1_shared_ref GPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQ
         Consensus GPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQPSYHPPAYQ.........Q......N.....................

                   401     410       420       430       440       450       460       470       480       490       500
                   |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
      Ph_pby1A_ref LGHVAAQQAFQAPGRSGQSEVQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQGPVTTSAANFW
      Ph_pby1B_ref
 Ph_pby1_shared_ref
         Consensus

                   501     510       520       530       540       550
                   |--------+---------+---------+---------+---------|
      Ph_pby1A_ref QPSNLKPEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA
      Ph_pby1B_ref
 Ph_pby1_shared_ref
         Consensus
```

[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.6_Ph_pby1_pby2]
[PhD_A2.6.2_Ph_pby1_pby2_alignments]

## “alignment_Ph_pby1B_all_aa_sequences.gif”



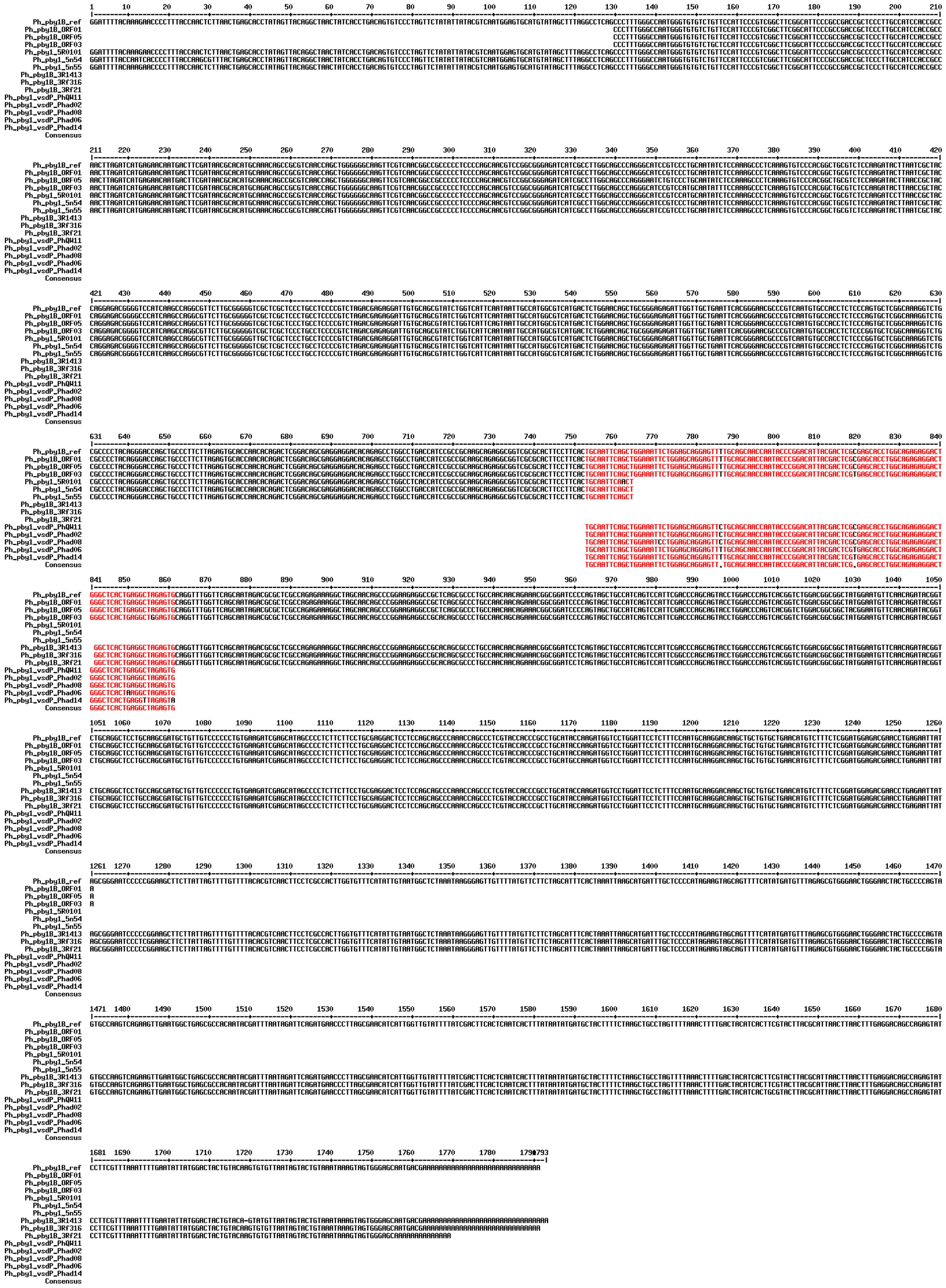
```
                     1       10        20        30        40        50        60        70
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
     Ph_pby1B_ORF01            PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
     Ph_pby1B_ORF05            PFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
     Ph_pby1B_ORF03            PFGPMGVSAPFPSASAFPPTAPLPSTANLDHENNDFDNAHADSRVNQLGGKFVNGRPLPS
     Ph_pby1_5R0101  MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
       Ph_pby1_5n54  MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
       Ph_pby1_5n55  MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
    Ph_pby1B_3R1413
    Ph_pby1B_3Rf316
     Ph_pby1B_3Rf21
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     71      80        90       100       110       120       130       140
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
     Ph_pby1B_ORF01  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGPIKPGVLAGVARSLPPRLDERIVQRIWSF
     Ph_pby1B_ORF05  NVRREIIALAAQGICPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
     Ph_pby1B_ORF03  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
     Ph_pby1_5R0101  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
       Ph_pby1_5n54  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
       Ph_pby1_5n55  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
    Ph_pby1B_3R1413
    Ph_pby1B_3Rf316
     Ph_pby1B_3Rf21
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     141    150       160       170       180       190       200       210
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
     Ph_pby1B_ORF01  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
     Ph_pby1B_ORF05  SNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
     Ph_pby1B_ORF03  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
     Ph_pby1_5R0101  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
       Ph_pby1_5n54  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
       Ph_pby1_5n55  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
    Ph_pby1B_3R1413
    Ph_pby1B_3Rf316
     Ph_pby1B_3Rf21
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     211    220       230       240       250       260       270       280
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1B_ORF01  RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1B_ORF05  RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1B_ORF03  RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEAGVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1_5R0101  RRSRTSFTAIQ
       Ph_pby1_5n54  RRSRTSFTAIQ
       Ph_pby1_5n55  RRSRTSFTAIQ
    Ph_pby1B_3R1413                                        LTEARVQVWFSNRRARQRKASNSPEEAAQRPA
    Ph_pby1B_3Rf316                                        LTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1B_3Rf21                                        LTEARVQVWFSNRRARQRKASNSPEEAAQRPA
Ph_pby1_vsdP_PhQW11          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
Ph_pby1_vsdP_Phad02          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
Ph_pby1_vsdP_Phad08          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
Ph_pby1_vsdP_Phad14          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEVRV
Ph_pby1_vsdP_Phad06          AIQLEILEQEFCSNQYPDITTREHLAERTGLTKARV
          Consensus          AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV

                     281    290       300       310       320       330       340       350
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1B_ORF01  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1B_ORF05  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1B_ORF03  NSRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1_5R0101
       Ph_pby1_5n54
       Ph_pby1_5n55
    Ph_pby1B_3R1413  NNRNGGSSVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
    Ph_pby1B_3Rf316  NNRNGGSSVAAISPFGPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1B_3Rf21  NNRNGGSSVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     351    360       370       380       390   396
                     |--------+---------+---------+---------+-----|
       Ph_pby1B_ref  SYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENYSGNPPEASY
     Ph_pby1B_ORF01  SYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENY
     Ph_pby1B_ORF05  SYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENY
     Ph_pby1B_ORF03  SYHPPACQDGPGFLFPMQGQAAVLNMSFSDGDEPENY
     Ph_pby1_5R0101
       Ph_pby1_5n54
       Ph_pby1_5n55
    Ph_pby1B_3R1413  SYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENYSGNPPEASY
    Ph_pby1B_3Rf316  SYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENYSGNPSEASY
     Ph_pby1B_3Rf21  SYHPLAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENYSGNPPEASY
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.6_Ph_pby1_pby2]

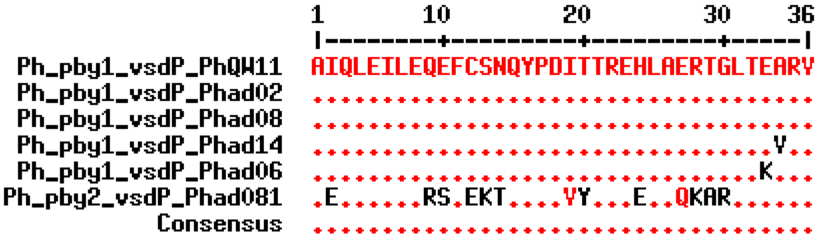
[PhD_A2.6.2_Ph_pby1_pby2_alignments]

## “alignment_Ph_pby1B_all_bp_sequences.gif”



[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.6_Ph_pby1_pby2]
[PhD_A2.6.2_Ph_pby1_pby2_alignments]

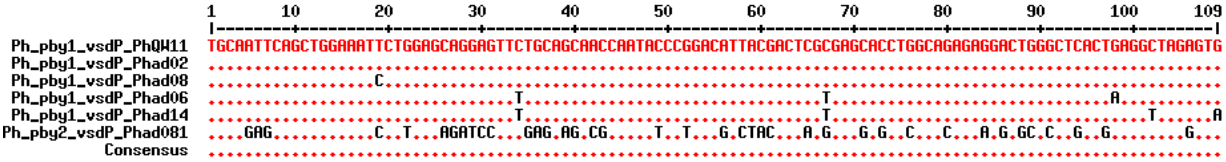
### “difference_alignment_Ph_pby1_pby2_vsdPCR_aa_sequences.gif”



```
                     1       10        20        30    36
                     |--------+---------+---------+-----|
Ph_pby1_vsdP_PhQW11  AIQLEILEQEFCSNQYPDITTREHLAERTGLTEARV
Ph_pby1_vsdP_Phad02  ....................................
Ph_pby1_vsdP_Phad08  ....................................
Ph_pby1_vsdP_Phad14  .................................V..
Ph_pby1_vsdP_Phad06  ................................K...
Ph_pby2_vsdP_Phad081 .E......RS.EKT....VY...E..QKAR......
           Consensus ....................................
```

[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.6_Ph_pby1_pby2]
[PhD_A2.6.2_Ph_pby1_pby2_alignments]

### “difference_alignment_Ph_pby1_pby2_vsdPCR_sequences.gif”



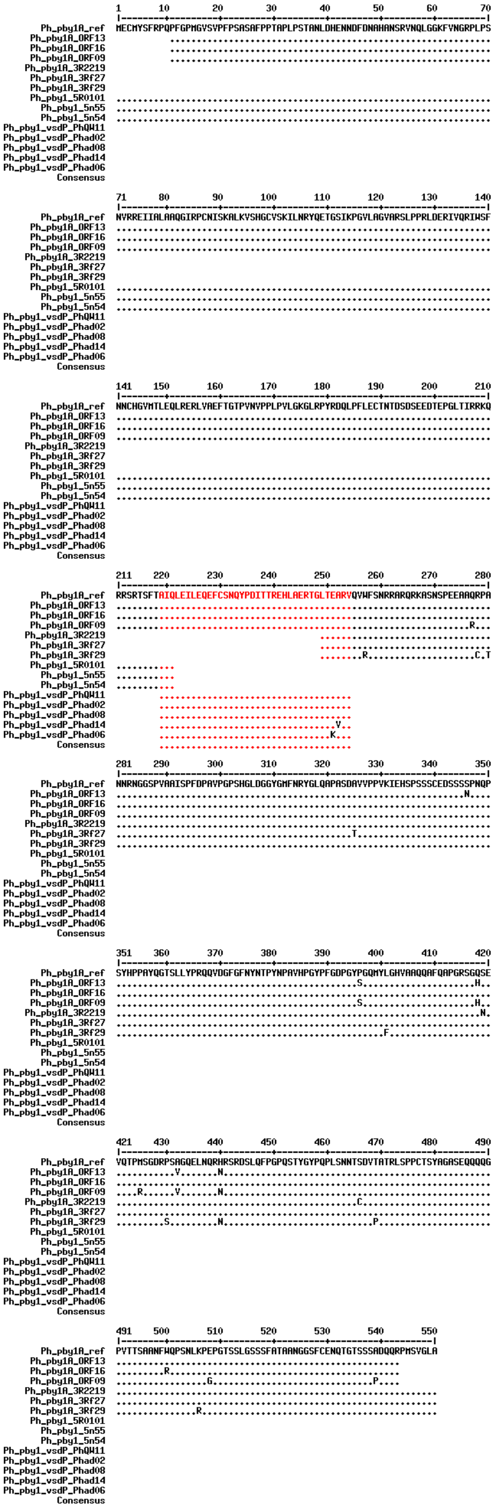
```
                     1       10        20        30        40        50        60        70        80        90       100      109
                     |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+--------|
Ph_pby1_vsdP_PhQW11  TGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTCTGCAGCAACCAATACCCGGACATTACGACTCGCGAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTG
Ph_pby1_vsdP_Phad02  .............................................................................................................
Ph_pby1_vsdP_Phad08  ..................C..........................................................................................
Ph_pby1_vsdP_Phad06  .................................T................................T..............................A...........
Ph_pby1_vsdP_Phad14  .................................T................................T..................................T......A
Ph_pby2_vsdP_Phad081 ....GAG...........C..T...AGATCC...GAG.AG.CG.....T..T...G.CTAC...A.G...G.G..C...C...A.G.GC.C..G..G........G...
           Consensus .............................................................................................................
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.6_Ph_pby1_pby2]

[PhD_A2.6.2_Ph_pby1_pby2_alignments]

## **“difference_alignment_Ph_pby1A_all_aa_sequences.gif”**



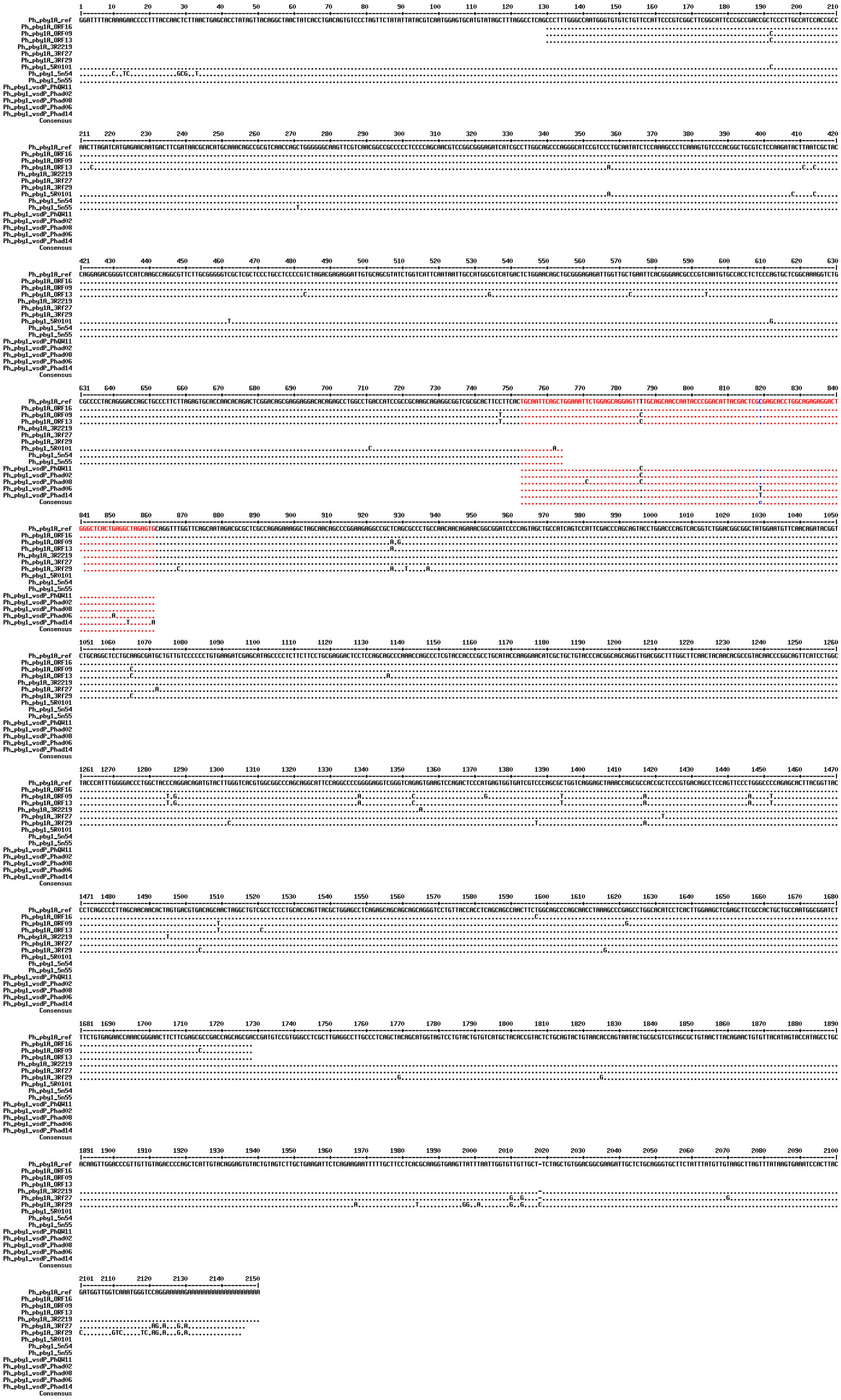
```
                    1       10        20        30        40        50        60        70
                    |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
     Ph_pby1A_ORF13           ............................................................
     Ph_pby1A_ORF16           ............................................................
     Ph_pby1A_ORF09           ............................................................
    Ph_pby1A_3R2219
     Ph_pby1A_3Rf27
     Ph_pby1A_3Rf29
     Ph_pby1_5R0101 ......................................................................
       Ph_pby1_5n55 ......................................................................
       Ph_pby1_5n54 ......................................................................
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                    71      80        90       100       110       120       130       140
                    |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
     Ph_pby1A_ORF13 ......................................................................
     Ph_pby1A_ORF16 ......................................................................
     Ph_pby1A_ORF09 ......................................................................
    Ph_pby1A_3R2219
     Ph_pby1A_3Rf27
     Ph_pby1A_3Rf29
     Ph_pby1_5R0101 ......................................................................
       Ph_pby1_5n55 ......................................................................
       Ph_pby1_5n54 ......................................................................
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                    141    150       160       170       180       190       200       210
                    |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
     Ph_pby1A_ORF13 ......................................................................
     Ph_pby1A_ORF16 ......................................................................
     Ph_pby1A_ORF09 ......................................................................
    Ph_pby1A_3R2219
     Ph_pby1A_3Rf27
     Ph_pby1A_3Rf29
     Ph_pby1_5R0101 ......................................................................
       Ph_pby1_5n55 ......................................................................
       Ph_pby1_5n54 ......................................................................
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                    211    220       230       240       250       260       270       280
                    |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1A_ORF13 ......................................................................
     Ph_pby1A_ORF16 ......................................................................
     Ph_pby1A_ORF09 ..................................................................R...
    Ph_pby1A_3R2219                                       ................................
     Ph_pby1A_3Rf27                                       ................................
     Ph_pby1A_3Rf29                                       ........R....................C.T
     Ph_pby1_5R0101 ...........
       Ph_pby1_5n55 ...........
       Ph_pby1_5n54 ...........
Ph_pby1_vsdP_PhQW11         ....................................
Ph_pby1_vsdP_Phad02         ....................................
Ph_pby1_vsdP_Phad08         ....................................
Ph_pby1_vsdP_Phad14         .................................V..
Ph_pby1_vsdP_Phad06         ................................K...
          Consensus         ....................................

                    281    290       300       310       320       330       340       350
                    |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1A_ORF13 .................................................................N....
     Ph_pby1A_ORF16 ......................................................................
     Ph_pby1A_ORF09 ......................................................................
    Ph_pby1A_3R2219 ......................................................................
     Ph_pby1A_3Rf27 ............................................T.........................
     Ph_pby1A_3Rf29 ......................................................................
     Ph_pby1_5R0101
       Ph_pby1_5n55
       Ph_pby1_5n54
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                    351    360       370       380       390       400       410       420
                    |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref SYHPPAYQGTSLLYPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQSE
     Ph_pby1A_ORF13 .............................................S.....................H..
     Ph_pby1A_ORF16 ......................................................................
     Ph_pby1A_ORF09 .............................................S.....................H..
    Ph_pby1A_3R2219 ....................................................................N.
     Ph_pby1A_3Rf27 ......................................................................
     Ph_pby1A_3Rf29 ..................................................F...................
     Ph_pby1_5R0101
       Ph_pby1_5n55
       Ph_pby1_5n54
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                    421    430       440       450       460       470       480       490
                    |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref VQTPMSGDRPSAGQELNQRHRSRDSLQFPGPQSTYGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQG
     Ph_pby1A_ORF13 ...........V.......N..................................................
     Ph_pby1A_ORF16 ......................................................................
     Ph_pby1A_ORF09 ....R......V.......N..................................................
    Ph_pby1A_3R2219 .............................................C........................
     Ph_pby1A_3Rf27 ......................................................................
     Ph_pby1A_3Rf29 .........S.........N............................P.....................
     Ph_pby1_5R0101
       Ph_pby1_5n55
       Ph_pby1_5n54
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                    491    500       510       520       530       540       550
                    |--------+---------+---------+---------+---------+---------|
       Ph_pby1A_ref PVTTSAANFWQPSNLKPEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA
     Ph_pby1A_ORF13 .....................................................
     Ph_pby1A_ORF16 .........R...........................................
     Ph_pby1A_ORF09 .................G..............................P....
    Ph_pby1A_3R2219 ............................................................
     Ph_pby1A_3Rf27 ............................................................
     Ph_pby1A_3Rf29 ...............R............................................
     Ph_pby1_5R0101
       Ph_pby1_5n55
       Ph_pby1_5n54
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.6_Ph_pby1_pby2]

[PhD_A2.6.2_Ph_pby1_pby2_alignments]

## “difference_alignment_Ph_pby1A_all_bp_sequences.gif”

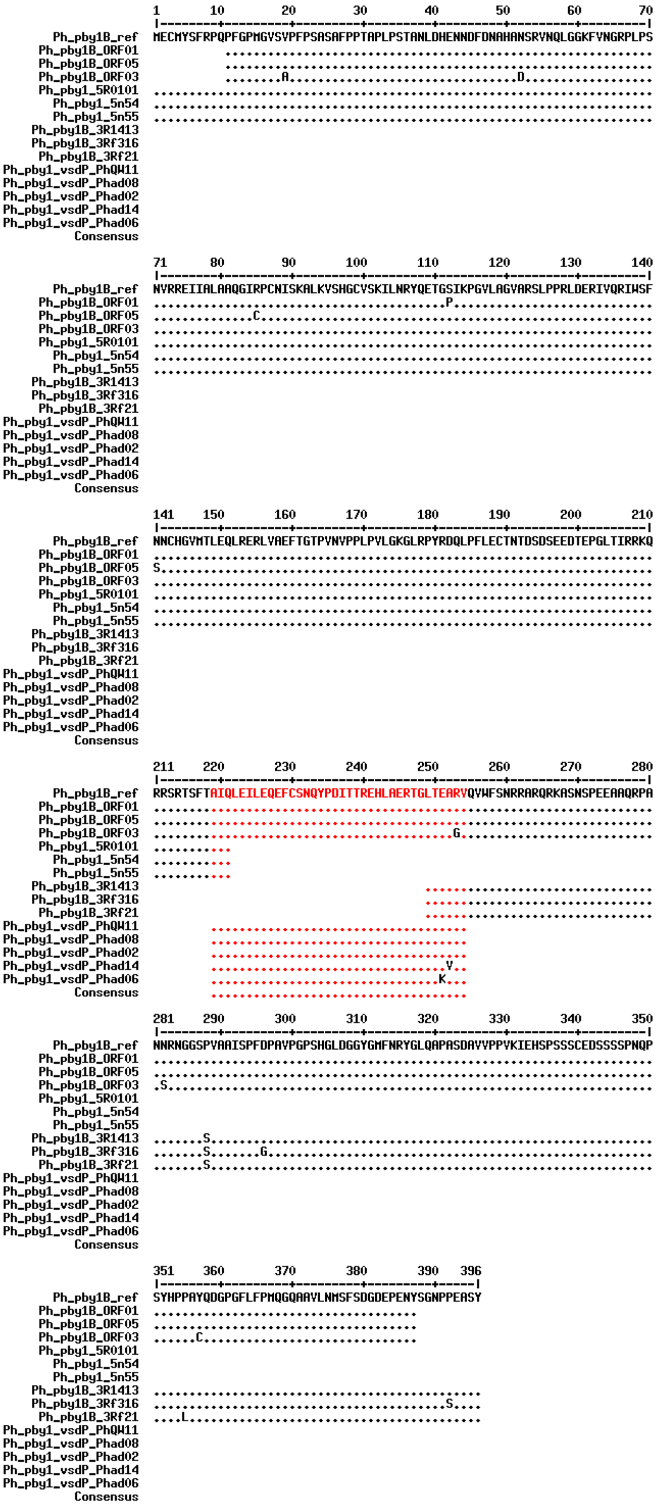


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.6_Ph_pby1_pby2]

[PhD_A2.6.2_Ph_pby1_pby2_alignments]

## “difference_alignment_Ph_pby1B_all_aa_sequences.gif”



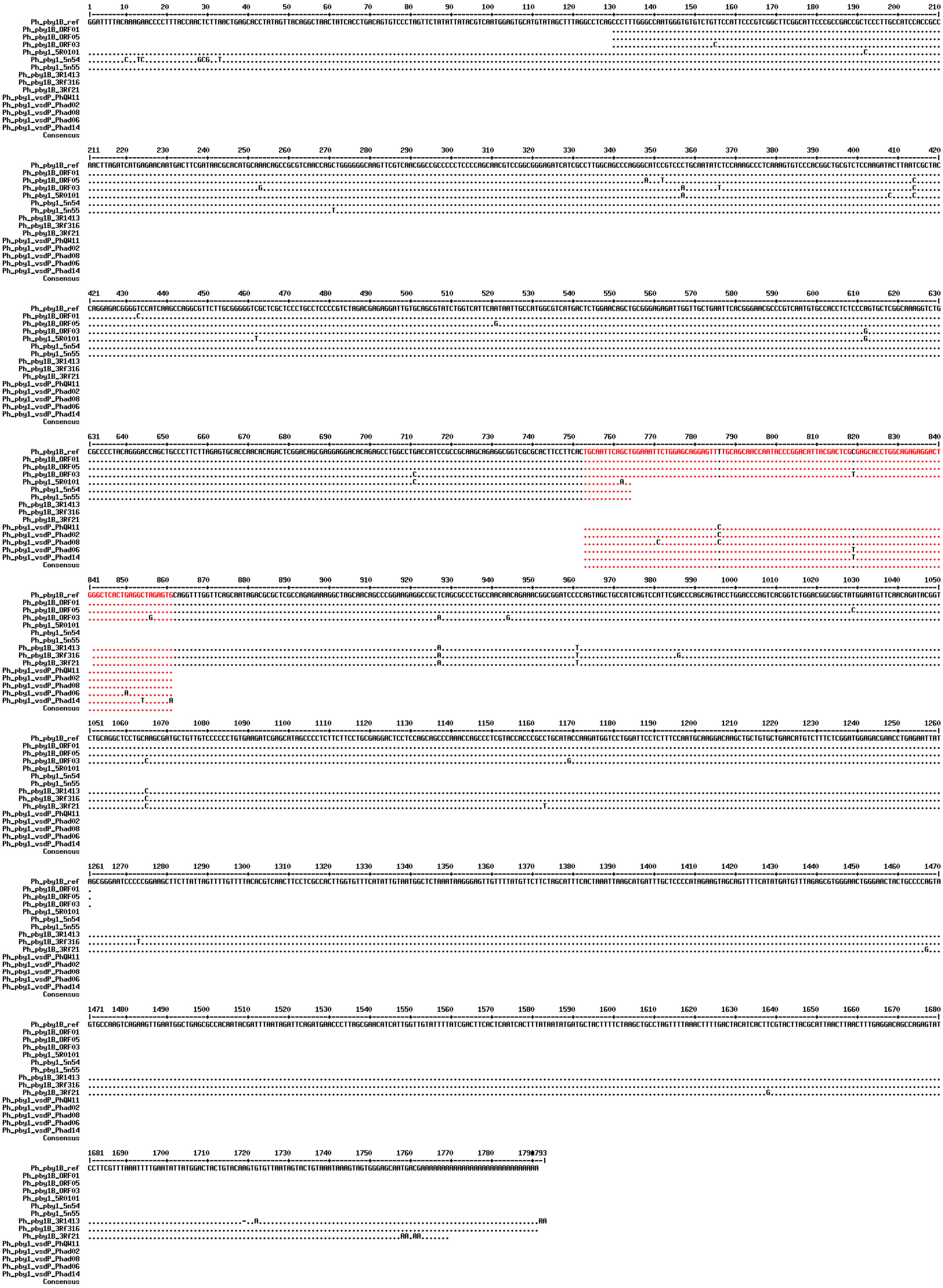
```
                     1       10        20        30        40        50        60        70
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPSTANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPS
     Ph_pby1B_ORF01            ............................................................
     Ph_pby1B_ORF05            ............................................................
     Ph_pby1B_ORF03            ........A................................D..................
     Ph_pby1_5R0101  ......................................................................
       Ph_pby1_5n54  ......................................................................
       Ph_pby1_5n55  ......................................................................
    Ph_pby1B_3R1413
    Ph_pby1B_3Rf316
     Ph_pby1B_3Rf21
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     71      80        90       100       110       120       130       140
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  NVRREIIALAAQGIRPCNISKALKVSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSF
     Ph_pby1B_ORF01  .........................................P............................
     Ph_pby1B_ORF05  ..............C.......................................................
     Ph_pby1B_ORF03  ......................................................................
     Ph_pby1_5R0101  ......................................................................
       Ph_pby1_5n54  ......................................................................
       Ph_pby1_5n55  ......................................................................
    Ph_pby1B_3R1413
    Ph_pby1B_3Rf316
     Ph_pby1B_3Rf21
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     141    150       160       170       180       190       200       210
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  NNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTNTDSDSEEDTEPGLTIRRKQ
     Ph_pby1B_ORF01  ......................................................................
     Ph_pby1B_ORF05  S.....................................................................
     Ph_pby1B_ORF03  ......................................................................
     Ph_pby1_5R0101  ......................................................................
       Ph_pby1_5n54  ......................................................................
       Ph_pby1_5n55  ......................................................................
    Ph_pby1B_3R1413
    Ph_pby1B_3Rf316
     Ph_pby1B_3Rf21
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     211    220       230       240       250       260       270       280
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  RRSRTSFTAIQLEILEQEFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKASNSPEEAAQRPA
     Ph_pby1B_ORF01  ......................................................................
     Ph_pby1B_ORF05  ......................................................................
     Ph_pby1B_ORF03  ..........................................G...........................
     Ph_pby1_5R0101  ...........
       Ph_pby1_5n54  ...........
       Ph_pby1_5n55  ...........
    Ph_pby1B_3R1413                                        ................................
    Ph_pby1B_3Rf316                                        ................................
     Ph_pby1B_3Rf21                                        ................................
Ph_pby1_vsdP_PhQW11          ....................................
Ph_pby1_vsdP_Phad08          ....................................
Ph_pby1_vsdP_Phad02          ....................................
Ph_pby1_vsdP_Phad14          .................................V..
Ph_pby1_vsdP_Phad06          ................................K...
          Consensus          ....................................

                     281    290       300       310       320       330       340       350
                     |--------+---------+---------+---------+---------+---------+---------|
       Ph_pby1B_ref  NNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFNRYGLQAPASDAVVPPVKIEHSPSSSCEDSSSSPNQP
     Ph_pby1B_ORF01  ......................................................................
     Ph_pby1B_ORF05  ......................................................................
     Ph_pby1B_ORF03  .S....................................................................
     Ph_pby1_5R0101
       Ph_pby1_5n54
       Ph_pby1_5n55
    Ph_pby1B_3R1413  .......S..............................................................
    Ph_pby1B_3Rf316  .......S.......G......................................................
     Ph_pby1B_3Rf21  .......S..............................................................
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus

                     351    360       370       380       390   396
                     |--------+---------+---------+---------+-----|
       Ph_pby1B_ref  SYHPPAYQDGPGFLFPMQGQAAVLNMSFSDGDEPENYSGNPPEASY
     Ph_pby1B_ORF01  .....................................
     Ph_pby1B_ORF05  .....................................
     Ph_pby1B_ORF03  ......C..............................
     Ph_pby1_5R0101
       Ph_pby1_5n54
       Ph_pby1_5n55
    Ph_pby1B_3R1413  ..............................................
    Ph_pby1B_3Rf316  .........................................S....
     Ph_pby1B_3Rf21  ....L.........................................
Ph_pby1_vsdP_PhQW11
Ph_pby1_vsdP_Phad08
Ph_pby1_vsdP_Phad02
Ph_pby1_vsdP_Phad14
Ph_pby1_vsdP_Phad06
          Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.6_Ph_pby1_pby2]

[PhD_A2.6.2_Ph_pby1_pby2_alignments]

## “difference_alignment_Ph_pby1B_all_bp_sequences.gif”



# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.7_Ph_awh]

### [PhD_A2.7.1_Ph_awh_sequences]

## “Ph_awh_general_read_me.txt”

```
General aspects:

(i)       All sequences are in FASTA format.

(ii)      [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)     Comments are preceded by an apostrophe (')

(iv)      All sequences without olignonucleotide sequence are indicated (e.g. "w/o 1414/1415"). In these cases, only novel
Parhyale sequence is presented.


Parhyale hawaiensis awh terminology:


(i)       In general, sequence denotation begins with: ">Ph_awh_"

(ii)      The above (i) is followed by: "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                        "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                        "ORF"               for coherent ORF sequences
                                        "cDNA"              for complete cDNA sequences

                                        "ref"               if reference sequence is presented
                                        "HD"                if homeobox/homeodomain fragment is presented
                                        "LIM1", "LIM2"      if LIM (encoding) fragment is presented

(iii)     The above (ii) is followed by clone id.


Parhyale hawaiensis awh synopsis:

cDNA:               3567 bp
5'UTR:              1181 bp (1-1181)
ORF:                1119 bp (1182-2300), 372 aa
3'UTR:              1267 bp (2301-3567)
LIM1:                159 bp (1530-1688), 53 aa
LIM2:                162 bp (1713-1874), 54 aa
HD:                  180 bp (1947-2126), 60 aa


>Ph_awh_LIM1 [bp]
TGTGCCGGCTGTGCCAGGATCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCAGAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATA
CGCGACCAAAACATTTACTGCAAGTCCGAC

>Ph_awh_LIM2 [bp]
TGCGCCAAATGCTGCCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGGGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCTGTAAGAGACAACTGAGCACAGGAGAGGAGTTCGCC
TTGCATGAGAGCAGAGTATTGTGTAAACAGCAT

>Ph_awh_HD [bp]
AGTAAGCGGGTCCGGACGACCTTCACGGACGAGCAGCTGCAGGTGCTCCAGGCTAACTTCCAGATCGACTCGAACCCTGACAGCCAGGACCTGGAGCGCATCGCACAGCTCACGGGGCTGAGCAAGCGG
GTGACGCAGGTGTGGTTCCAAAACAGCAGAGCGCGGCAGAAGAAGCACATG

>Ph_awh_ref [bp]
GAATTTTCACACTCCCCTTTACCGAGTTTTACTGTTACTCATGAATTATTTCTCAGAGCACTTAAAAGAAAATGACATGACTTCAGAATTTAATGCATGTGGAGGGTGCATTTTCCAGAAAATATTTGC
TGATGTGTGAGAAAATATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGCTGGGAAGAAGCAACAAACAAAAACAAGCAACAAGCTTGTGACGGAACTGCTGTTAGTATTGTAGGTGAT
ATTGAGTTCCTCGCCGCTTTTTGACCAGAATATTAATTTGTTAATCAGATAGAAAATATCACTGCAATAGTTGTTACAGAAGTTTACTGAAATGGACTGTGCAAACTTGAAGTAAACTTCGGCGGCGTG
TGTTGTATGGACGTGCGGACCTCGTTCTGAACTGTAACTGTTGTACATAGAAACTAACTAGTTGTAGGACTGGCTGTGTGGGGGGTAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAAAGAGGTTA
ATACGGTAAGTAGCGTTTCAAGCTACGGATATGATAGAACACAACTTCTGGATAAAGAACTACCCGCGAAATTCGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGACAGTGGAATATTG
CTCCAATAAGTTTAGTTTCTGCCAGTAGAATTGCAGAAATTGATTATATTTTAATTATCAGCAGAACTGTGCACTACTGTAGCTGAACTAATAGAGATTTTCATATCCTTATGTAAGTGTTCTTAAGAA
CAATAGAGAACAGAAGAACCACTGGAAAAACTAATATATATAACTCAGTGAACAAGGATACCGGCAGCACCCGTTATTAAGGACTATTTGTGAAATAAGACAAGGTAAGGTTCTTACTAATAGCAAGCC
GAACCAGAACCAGTAACGGAAGAAAATTAGTTAAAGTAGAACAGGTTCTCAGCGCTAATATACCACGGAACAACAAACGAAAGAGGAATCTGTGAATAGAAATGAAAACTGATATTACGGACTGTCAAG
GAAATGATCTAAAGTTCTTAGAAGACAACTAAACAAAAAGAGAAACTGCGAGTGAACCAGTTAGGCAAAACGTCAGAACCAATAATTAGAACTATTAGTGAAGTACTGCCAGGTTCTAAAGCGCAACAG
ACAACAGCAGAACTAACAGGATGATGGCCAGCGCAACAGTTCTGTTGCCAATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGGTGATGGGGACC
CCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAGCGGCCTACAACACACGGCGAGAGAGCGGAGGCTGGGGTGTTTGCCTCTGAGCGTAAACAACAACAACAACAGTTCCACCA
CAAACAACAACAATAGCAACAGTGCCAGTGCTGCCCCACACGGGGTGGAGGGAGGTGGGGGTGTGCGGGGGCTCTCTGGAGGAGGGTGCGGCGCCCCCACCACAGAGAGCTGTGCCGGCTGTGCCAGGA
TCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCAGAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATTTACT
GCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTGCCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGGGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCT
GTAAGAGACAACTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAGAGTATTGTGTAAACAGCATTACTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCG
GCCAGCGCAAGAGTAAGCGGGTCCGGACGACCTTCACGGACGAGCAGCTGCAGGTGCTCCAGGCTAACTTCCAGATCGACTCGAACCCTGACAGCCAGGACCTGGAGCGCATCGCACAGCTCACGGGGC
TGAGCAAGCGGGTGACGCAGGTGTGGTTCCAAAACAGCAGAGCGCGGCAGAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATCACAGGCAGCTTAGCCCGTA
CAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCCAGAGGACCTGTCGTCTCCCGAGAACTCGCTTATGACGCTCATGCCTGGAGAAGCATGACGGTACGACCCCTCCACCTCCC
CTCCACCTCCTCCTCCACCCACCCTCCACTGACGGCGCTGTGTCACCCGTTGTGGCACTCCTGGCGAGAGTCGCACTTTGTGACACCAACAGACTGCTCTGTGTCACCAGTTGTGACACTAACCGGCAC
ACGCAATGAAGCAAACATTTCCAGAGAAGTTGACCTTAGAGCAGAATGGTATTAACGAGTTGGGAATGTAACGTCAAGCTGGCATTGATTTTCAATTTGAAGAAATGAAAATATAACATACTTGGAAGT
CATATATCTGAAACTAAACAAATGACAGCAATCTTGACTACCAGTCGGACAATTACACGAGACAATCTCAGCTACCAGGCTTACGATTACACGAGTCATCTCAAGGGAAGCAGGATTTTTTAGAGGTCG
GATGAGATGGGACTCCGTGTGTCGCTACTGAAGTTGGTTAAATTCTTCATGAGACCCAGTGGATCCGTAAGGTCAGGAGTCAAATGGATCTGAGATCCAGTGGATCAGTTACACAGAGTTCAATGACAG
```

```
CGACTAGCAATAACTAGTTTCCAGACTAGTTACTTCTGCTAGTTCCAAACATATTATGGGTAGTGACAAGAATTTTCGAATAATCATTGCGGCGAATTCGGCCCACTCGTGGCTGTCGAGGAAAGCCAG
TCAGCTTGGCTTCTCCCGGTGGCAGAGAGCAGAGACTCTGGCACCGTCCGAGTGCCAGTAAGTCAAGCTTGAAGCTCAATCTAAAGCGCGACTCGAACGCCGGCGATGGAAACTGCGGAAACGTTAAAG
CTAAGTGGGATTAATTTAATCGATGACAGTAGGCGAATTGCCCACGAGTGTGTCTTATTCAGCATACTATAGATAAGCATGCGTATGATCTAACGCAGCACTCTAGGGAAGGTTGCGTATGGTCAAACA
GGGCACCCTAGGGAACGTTGCGTATGGTTGCAGCACACTTCTGCGTATGTTCTGAGTCGCGCGGCCGCAGATAAGTGCAGGGCGGAACCACGCAGCTTTGCTCAAGTTGGACAGATTCAGCTCTGTAGG
AATGGAAGAATTACACAACAGTACGCGGGCGATGAGTTGCAGCAAATACGACACCGTGGTCATCTGGAAATGTTCTGCACAGCTTCTACATCAAGAGTACCCTAGGTGGGGGCACGGACCGTCAGAGAG
CGCGTGATACCATTGTAGGGGATTCATGATAAATATAATTGGTACACTCAGTACCGGTAAAAAAAAAAAAAAAAAAAAAAAAAAA


>Ph_awh_LIM1 [aa]
CAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSD

>Ph_awh_LIM2 [aa]
CAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQH

>Ph_awh_HD [aa]
SKRVRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVTQVWFQNSRARQKKHM

>Ph_awh_ref [aa]
MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFL
LRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKR
VRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA&
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.7_Ph_awh]

### [PhD_A2.7.1_Ph_awh_sequences]

### “Ph_awh_cDNA_sequences.txt”

```
>Ph_awh_ref [bp] 'Ph_awh_ORF21 w/o primers, extended 5' by Ph_awh_5RCR02 (1210 bp); extended 3' by Ph_awh_3Rn21 (1298 bp)
GAATTTTCACACTCCCCTTTACCGAGTTTTACTGTTACTCATGAATTATTTCTCAGAGCACTTAAAAGAAAATGACATGACTTCAGAATTTAATGCATGTGGAGGGTGCATTTTCCAGAAAATATTTGC
TGATGTGTGAGAAAATATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGCTGGGAAGAAGCAACAAACAAAAACAAGCAACAAGCTTGTGACGGAACTGCTGTTAGTATTGTAGGTGAT
ATTGAGTTCCTCGCCGCTTTTTGACCAGAATATTAATTTGTTAATCAGATAGAAAATATCACTGCAATAGTTGTTACAGAAGTTTACTGAAATGGACTGTGCAAACTTGAAGTAAACTTCGGCGGCGTG
TGTTGTATGGACGTGCGGACCTCGTTCTGAACTGTAACTGTTGTACATAGAAACTAACTAGTTGTAGGACTGGCTGTGTGGGGGGTAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAAAGAGGTTA
ATACGGTAAGTAGCGTTTCAAGCTACGGATATGATAGAACACAACTTCTGGATAAAGAACTACCCGCGAAATTCGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGACAGTGGAATATTG
CTCCAATAAGTTTAGTTTCTGCCAGTAGAATTGCAGAAATTGATTATATTTTAATTATCAGCAGAACTGTGCACTACTGTAGCTGAACTAATAGAGATTTTCATATCCTTATGTAAGTGTTCTTAAGAA
CAATAGAGAACAGAAGAACCACTGGAAAAACTAATATATATAACTCAGTGAACAAGGATACCGGCAGCACCCGTTATTAAGGACTATTTGTGAAATAAGACAAGGTAAGGTTCTTACTAATAGCAAGCC
GAACCAGAACCAGTAACGGAAGAAAATTAGTTAAAGTAGAACAGGTTCTCAGCGCTAATATACCACGGAACAACAAACGAAAGAGGAATCTGTGAATAGAAATGAAAACTGATATTACGGACTGTCAAG
GAAATGATCTAAAGTTCTTAGAAGACAACTAAACAAAAAGAGAAACTGCGAGTGAACCAGTTAGGCAAAACGTCAGAACCAATAATTAGAACTATTAGTGAAGTACTGCCAGGTTCTAAAGCGCAACAG
ACAACAGCAGAACTAACAGGATGATGGCCAGCGCAACAGTTCTGTTGCCAATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGGTGATGGGGACC
CCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAGCGGCCTACAACACACGGCGAGAGAGCGGAGGCTGGGGTGTTTGCCTCTGAGCGTAAACAACAACAACAACAGTTCCACCA
CAAACAACAACAATAGCAACAGTGCCAGTGCTGCCCCACACGGGGTGGAGGGAGGTGGGGGTGTGCGGGGGCTCTCTGGAGGAGGGTGCGGCGCCCCCACCACAGAGAGCTGTGCCGGCTGTGCCAGGA
TCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCAGAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATTTACT
GCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTGCCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGGGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCT
GTAAGAGACAACTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAGAGTATTGTGTAAACAGCATTACTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCG
GCCAGCGCAAGAGTAAGCGGGTCCGGACGACCTTCACGGACGAGCAGCTGCAGGTGCTCCAGGCTAACTTCCAGATCGACTCGAACCCTGACAGCCAGGACCTGGAGCGCATCGCACAGCTCACGGGGC
TGAGCAAGCGGGTGACGCAGGTGTGGTTCCAAAACAGCAGAGCGCGGCAGAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATCACAGGCAGCTTAGCCCGTA
CAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCCAGAGGACCTGTCGTCTCCCGAGAACTCGCTTATGACGCTCATGCCTGGAGAAGCATGACGGTACGACCCCTCCACCTCCC
CTCCACCTCCTCCTCCACCCACCCTCCACTGACGGCGCTGTGTCACCCGTTGTGGCACTCCTGGCGAGAGTCGCACTTTGTGACACCAACAGACTGCTCTGTGTCACCAGTTGTGACACTAACCGGCAC
ACGCAATGAAGCAAACATTTCCAGAGAAGTTGACCTTAGAGCAGAATGGTATTAACGAGTTGGGAATGTAACGTCAAGCTGGCATTGATTTTCAATTTGAAGAAATGAAAATATAACATACTTGGAAGT
CATATATCTGAAACTAAACAAATGACAGCAATCTTGACTACCAGTCGGACAATTACACGAGACAATCTCAGCTACCAGGCTTACGATTACACGAGTCATCTCAAGGGAAGCAGGATTTTTTAGAGGTCG
GATGAGATGGGACTCCGTGTGTCGCTACTGAAGTTGGTTAAATTCTTCATGAGACCCAGTGGATCCGTAAGGTCAGGAGTCAAATGGATCTGAGATCCAGTGGATCAGTTACACAGAGTTCAATGACAG
CGACTAGCAATAACTAGTTTCCAGACTAGTTACTTCTGCTAGTTCCAAACATATTATGGGTAGTGACAAGAATTTTCGAATAATCATTGCGGCGAATTCGGCCCACTCGTGGCTGTCGAGGAAAGCCAG
TCAGCTTGGCTTCTCCCGGTGGCAGAGAGCAGAGACTCTGGCACCGTCCGAGTGCCAGTAAGTCAAGCTTGAAGCTCAATCTAAAGCGCGACTCGAACGCCGGCGATGGAAACTGCGGAAACGTTAAAG
CTAAGTGGGATTAATTTAATCGATGACAGTAGGCGAATTGCCCACGAGTGTGTCTTATTCAGCATACTATAGATAAGCATGCGTATGATCTAACGCAGCACTCTAGGGAAGGTTGCGTATGGTCAAACA
GGGCACCCTAGGGAACGTTGCGTATGGTTGCAGCACACTTCTGCGTATGTTCTGAGTCGCGCGGCCGCAGATAAGTGCAGGGCGGAACCACGCAGCTTTGCTCAAGTTGGACAGATTCAGCTCTGTAGG
AATGGAAGAATTACACAACAGTACGCGGGCGATGAGTTGCAGCAAATACGACACCGTGGTCATCTGGAAATGTTCTGCACAGCTTCTACATCAAGAGTACCCTAGGTGGGGGCACGGACCGTCAGAGAG
CGCGTGATACCATTGTAGGGGATTCATGATAAATATAATTGGTACACTCAGTACCGGTAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_awh_ORF20 [bp] 'w/o 1414/1415
AATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGGTGATGGGGACCCCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAG
CGGCCTACAACACACGGCGAGAGAGCGGAGGCTGGGGTGTTTGCCTCTGAGCGTAAACAACAACAACAACAGTTCCACCACAAACAACAACAATAGCAACAGTGCCAGTGCTGCCCCACACGGGGTGGA
GGGAGGTGGGGGTGTGCGGGGGCTCTCTGGAGGAGGGTGCGGCGCCCCCACCACAGAGAGCTGTGCCGGCTGTGCCAGGATCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCA
GAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACTAAAACATTTACTGCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTG
CCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGGGATCGGGTCTACCATCTCGCCCGTTTCGCCTGCGACGCCTGTAAGAGACAACTGAGCACAGGAGAGGAGCTCGCCTTGCATGAGAGCAG
AGTATTGTGTAAACAGCATTACTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCGGCCAGCGCAAGAGTAAGCGGGTCCGGACGACCTTCACGGACGAGCAGCT
GCAGGTGCTCCAGGCTAACTTCCAGATCGACTCGAACCCTGACAGCCAGGACCTGGAGCGCATCGCACAGCTCACGGGGCTGAGCAAGCGGGTGACGCAGGTGTGGTTCCAAAACAGCAGAGCGCGGCA
GAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATTACAGGCAGCTTAGCCCGTACAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGTAGAGCC
AGAGGACCTGTCGTCTCCCGAGAACT

>Ph_awh_ORF21 [bp] 'w/o 1414/1415
AATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGGTGATGGGGACCCCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAG
CGGCCTACAACACACGGCGAGAGAGCGGAGGCTGGGGTGTTTGCCTCTGAGCGTAAACAACAACAACAACAGTTCCACCACAAACAACAACAATAGCAACAGTGCCAGTGCTGCCCCACACGGGGTGGA
GGGAGGTGGGGGTGTGCGGGGGCTCTCTGGAGGAGGGTGCGGCGCCCCCACCACAGAGAGCTGTGCCGGCTGTGCCAGGATCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCA
GAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATTTACTGCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTG
CCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGGGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCTGTAAGAGACAACTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAG
AGTATTGTGTAAACAGCATTACTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCGGCCAGCGCAAGAGTAAGCGGGTCCGGACGACCTTCACGGACGAGCAGCT
GCAGGTGCTCCAGGCTAACTTCCAGATCGACTCGAACCCTGACAGCCAGGACCTGGAGCGCATCGCACAGCTCACGGGGCTGAGCAAGCGGGTGACGCAGGTGTGGTTCCAAAACAGCAGAGCGCGGCA
GAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATCACAGGCAGCTTAGCCCGTACAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCC
AGAGGACCTGTCGTCTCCCGAGAACT

>Ph_awh_ORF22 [bp] 'w/o 1414/1415
AATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGGTGATGGGGACCCCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAG
CGGCCTACAACACACGGCGAGAGAGCGGAGGCTGGGGTGTTTGCCTCTGAGCGTAAACAACAACAACAACAGTTCCACCACAAACAACAACAATAGCAACAGTGCCAGTGCTGCCCCACACGGGGTGGA
GGGAGGTGGGGGTGTGCGGGGGCTCTCTGGAGGAGGGTGCGGCGCCCCCACCACAGAGAGCTGTGCCGGCTGTGCCAGGATCATCACTGACAGGTTCCTGCTGCGCGTCAACTGCCTCTCCTGGCACCA
GAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATTTACTGCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTG
CCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGGGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCTGTAAGAGACAACTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAG
AGTATTGTGTAAACAGCATTACTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCGGCCAGCGCAAGAGTAAGCGGGTCCGGACGACCTTCACGGACGAGCAGCT
GCAGGTGCTCCAGGCTAACTTCCAGATCGACTCGAACCCTGACAGCCAGGACCTGGAGCGCATCGCACAGCTCACGGGGCTGAGCAAGCGGGTGACGCAGGTGTGGTTCCAAAACAGCAGAGCGCGGCA
GAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATCACAGGCAGCTTAGCCCGTACAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCC
AGAGGACCTGTCGTCTCCCGAGAACT

>Ph_awh5RCR1 [bp] 'w/o 1406
GGATTTTACCAATCACCCCTTTACCAAGCGTTTACTGGCATACCAGCCGTAAATGCAAAATTCAGTGTTAGAGTTTTAGTAATGATGTTGATGTTTGTTACAGTTACTCATGAATTATTTCTCAGAGCA
CTTAAAAGAAAATGACATGACTTCAGAATTTAATGCATGTGGAGGGTGCATTTTCCAGAAAATATTTGCTGATGTGTGAGAAAATATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGC
TGGGAAGAAGCAACAAACAAAAACAAGCAACAAGCTTGTGACGGAACTGCTGTTAGTATTGTAGGTGATATTGAGTTCCTCGCCGCTTTTTGACCAGAATATTAATTTGTTAATCAGATAGAAAATATC
ACTGCAATAGTTGTTACAGAAGTTTACTGAAATGGACTGTGCAAACTTGAAGTAAACTTCGGCGGCGTGTGTTGTATGGACGTGCGGACCTCGTTCTGAACTGTAATTGTTGTACATAGAAACTAACTA
GTTGTAGGACTGGCTGTGTGGGGGGTAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAAAGAGGTTAATACGGTAAGTAGCGTTTCAAGCTACGGATATGATAGAACACAACTTCTGGATAAAGAAC
TACCCGCGAAATTCGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGACAGTGGAATATTGCTCCAATAAGTTTAGTTTCTGCCAGTAGAATTGCAGAAATTGATTATATTTTAATTATCA
GCAGAACTGTGCACTACTGTAGCTGAACTAATAGAGATTTTCATATCCTTATGTAAGTGTTCTTAAGAACAATAGAGAACAGAAGAACCACTGGAAAAACTAATATATATAACTCAGTGAACAAGGATA
CCGGCAGCACCCGTTATTAAGGACTATTTGTGAAATAAGACAAGGTAAGGTTCTTACTAATAGCAAGCCGAACCAGAACCAGTAACGGAAGAAAATTAGTTAAAGTAGAACAGGTTCTCAGCGCTAATA
TACCACGGAACAACAAACGAAAGAGGAATCTGTGAATAGAAATGAAAACTGATATTACGGACTGTCAAGGAAATGATCTAAAGTTCTTAGAAGACAACTAAACAAAAAGAGAAACTGCGAGTGAACCAG
TTAGGCAAAACGTCAGAACCAATAATTAGAACTATTAGTGAAGTACTACCAGGTTCTAAAGCGCAACAGACAACAGCAGAACTAACAGGATGATGGCCAGCGCAACAGTTCTGTTGCCAATGAACAGAA
TCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGGTGATGGGGACCCCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAGCGGCCTACAAC
```

```
ACACGGCGAGAGAGCGGAGGCTGGGGTGCTTGCCTCTGAGTGTAAACAACAACAACAACAGTTCCACCACAAACAACAACAATAGCAACGGTGCCAGTGCTGCCCCACACGGGGTAGAGGGAGGTGGGG
GTGTGCGGGGGCTCTCTGGAGGAGGGTGCGGTGCCCCCACCACAGAGAGCTGTGCTGGCTGTGCCAGGATCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCAAAGCTGCCTGC
GCTGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATCTACTGCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTGCCGGGCCATCG
GGTCGACGGACTGGGTCCGGAGGGCGAGAGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCTGTAAGAGACAGCTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAGAGTATTGTGTA
AACAGCATTATTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCGGCCAGCGCAAGAGCAAGCGGGTCCGAACGACCTTCACGGACGAACAGCTGCAGGTGCTCC
AGGCCAACTTCCAGATCGACTCGAA

>Ph_awh5RCR2 [bp] 'w/o 1406
GAATTTTCACACTCCCCTTTACCGAGTTTTACTGTTACTCATGAATTATTTCTCAGAGCACTTAAAAGAAAATGACATGACTTCAGAATTTAATGCATGTGGAGGGTGCATTTTCCAGAAAATATTTGC
TGATGTGTGAGAAAATATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGCTGGGAAGAAGCAACAAACAAAAACAAGCAACAAGCTTGTGACGGAACTGCTGTTAGTATTGTAGGTGAT
ATTGAGTTCCTCGCCGCTTTTTGACCAGAATATTAATTTGTTAATCAGATAGAAAATATCACTGCAATAGTTGTTACAGAAGTTTACTGAAATGGACTGTGCAAACTTGAAGTAAACTTCGGCGGCGTG
TGTTGTATGGACGTGCGGACCTCGTTCTGAACTGTAACTGTTGTACATAGAAACTAACTAGTTGTAGGACTGGCTGTGTGGGGGGTAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAAAGAGGTTA
ATACGGTAAGTAGCGTTTCAAGCTACGGATATGATAGAACACAACTTCTGGATAAAGAACTACCCGCGAAATTCGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGACAGTGGAATATTG
CTCCAATAAGTTTAGTTTCTGCCAGTAGAATTGCAGAAATTGATTATATTTTAATTATCAGCAGAACTGTGCACTACTGTAGCTGAACTAATAGAGATTTTCATATCCTTATGTAAGTGTTCTTAAGAA
CAATAGAGAACAGAAGAACCACTGGAAAAACTAATATATATAACTCAGTGAACAAGGATACCGGCAGCACCCGTTATTAAGGACTATTTGTGAAATAAGACAAGGTAAGGTTCTTACTAATAGCAAGCC
GAACCAGAACCAGTAACGGAAGAAAATTAGTTAAAGTAGAACAGGTTCTCAGCGCTAATATACCACGGAACAACAAACGAAAGAGGAATCTGTGAATAGAAATGAAAACTGATATTACGGACTGTCAAG
GAAATGATCTAAAGTTCTTAGAAGACAACTAAACAAAAAGAGAAACTGCGAGTGAACCAGTTAGGCAAAACGTCAGAACCAATAATTAGAACTATTAGTGAAGTACTGCCAGGTTCTAAAGCGCAACAG
ACAACAGCAGAACTAACAGGATGATGGCCAGCGCAACAGTTCTGTTGCCAATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGGTGATGGGGACC
CCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAGCGGCCTACAACACACGGCGAGAGAGCGGAGGCTGGGGTGCTTGCCTCTGAGTGTAAACAACAACAACAACAGTTCCACCA
CAAACAACAACAATAGCAACAGTGCCAGTGCTGCCCCACACGGGGTAGAGGGAGGTGGGGGTGTGCGGGGGCTCTCTGGAGGAGGGTGCGGTGCCCCCACCACAGAGAGCTGTGCTGGCTGTGCCAGGA
TCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCAAAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATCTACT
GCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTGCCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGAGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCT
GTAAGAGACAGCTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAGAGTATTGTGTAAACAGCATTATTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCG
GCCAGCGCAAGAGCAAGCGGGTCCGAACGACCTTCACGGACGAACAGCTGCAGGTGCTCCAGGCCAACTTCCAGATCGACTCGAA

>Ph_awh5RCR5 [bp] 'w/o 1406
GAATTTTCACACTCCCCTTTACCGAGTTTTACTGTTACTCATGAATTATTTCTCAGAGCACTTAAAAGAAAATGACATGACTTCAGAATTTAATGCATGTGGAGGGTGCATTTTCCAGAAAATATTTGC
TGATGTGTGAGAAAATATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGCTGGGAAGAAGCAACAAACAAAAACAAGCAACAAGCTTGTGACGGAACTGCTGTTAGTATTGTAGGTGAT
ATTGAGTTCCTCGCCGCTTTTTGACCAGAATATTAATTTGTTAATCAGATAGAAAATATCACTGCAATAGTTGTTACAGAAGTTTACTGAAATGGACTGTGCAAACTTGAAGTAAACTTCGGCGGCGTG
TGTTGTATGGACGTGCGGACCTCGTTCTGAACTGTAACTGTTGTACATAGAAACTAACTAGTTGTAGGACTGGCTGTGTGGGGGGTAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAAAGAGGTTA
ATACGGTAAGTAGCGTTTCAAGCTACGGATATGATAGAACACAACTTCTGGATAAAGAACTACCCGCGAAATTCGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGACAGTGGAATATTG
CTCCAATAAGTTTAGTTTCTGCCAGTAGAATTGCAGAAATTGATTATATTTTAATTATCAGCAGAACTGTGCACTACTGTAGCTGAACTAATAGAGATTTTCATATCCTTATGTAAGTGTTCTTAAGAA
CAATAGAGAACAGAAGAACCACTGGAAAAACTAATATATATAACTCAGTGAACAAGGATACCGGCAGCACCCGTTATTAAGGACTATTTGTGAAATAGGACAAGGTAAGGTTCTTACTAATAGCAAGCC
GAACCAGAACCAGTAACGAAAGAAAATTAGTTAAAGTAGAACAGGTTCTCAGCGCTAATATACCACGGAACAACAAACGAAAGAGGAATCTGTGAATAGAAATGAAAACTGATATTACGGACTGTCAAG
GAAATGATCTAAAGTTCTTAGAAGACAACTAAACAAAAAGAGAAACTGCGAGTGAACCAGTTAGGCAAAACGTCAGAACCAATAATTAGAACTATTAGTGAAGTACTGCCAGGTTCTAAAGCGCAACAG
ACAACAGCAGAACTAACAGGATGATGGCCAGCGCAACAGTTCTGTTGCCAATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGGTGATGGGGACC
CCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAGCGGCCTACAACACACGGCGAGAGAGCGGAGGCTGGGGTGCTTGCCTCTGAGTGTAAACAACAACAACAACAGTTCCACCA
CAAACAACAACAATAGCAACAGTGCCAGTGCTGCCCCACACGGGGTAGAGGGAGGTGGGGGTGTGCGGGGGCTCTCTGGAGGAGGGTGCGGTGCCCCCACCACAGAGAGCTGTGCTGGCTGTGCCAGGA
TCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCAAAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATCTACT
GCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTGCCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGAGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCT
GTAAGAGACAGCTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAGAGTATTGTGTAAACAGCATTATTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCG
GCCAGCGCAAGAGCAAGCGGGTCCGAACGACCTTCACGGACGAACAGCTGCAGGTGCTCCAGGCCAACTTCCAGATCGACTCGAA

>Ph_awh5RCR6 [bp] 'w/o 1406
GAATTTTCACTGTTCCCTTTACCACGTTTTACTGGCATACCAGCTGTAAATGCAAAATTCAGTGTTAGAGTTTTAGTAATGATGTTGATGTTTGTTACAGTTACTCATGAATTATTTCTCAGAGCACTT
AAAAGAAAATGACATGACTTCAGAATTTAATGCATGTGGAGGGTGCATTTTCCAGAAAATATTTGCTGATGTGTGAGAAAATATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGCTGG
GAAGAAGCAACAAACAAAAACAAGCAACAAGCTTGTGACGGAACTGCTGTTAGTATTGTAGGTGATATTGAGTTCCTCGCCGCTTTTTGACCAGAATATTAATTTGTTAATCAGATAGAAAATATCACT
GCAATAGTTGTTACAGAAGTTTACTGAAATAGACTGTGCAAACTTGAAGTAAACTTCGGCGGCGTGTGTTGTATGGACGTGCGGACCTCGTTCTGAACTGTAACTGTTGTACATAGAAACTAACTAGTT
GTAGGACTGGCTGTGTGGGGGGTAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAAAGAGGTTAATACGGTAAGTAGCGTTTCAAGCTACGGATATGATAGAACACAACTTCTGGATAAAGAACTAC
CCGCGAAAATTCGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGACAGTGGAATATTGCTCCAATAAGTTTAGTTTCTGCCAGTAGAATTGCAGAAATTGATTATATTTTAATTATCAGC
AGAACTGTGCACTACTGTAGCTGAACTAATAGAGATTTTCATATCCTTATGTAAGTGTTCTTAAGAACAATAGAGAACAGAAGAACCACTGGAAAAACTAATATATATAACTCAGTGAACAAGGATACC
GGCAGCACCCGTTATTAAGGACTATTTGTGAAATAAGACAAGGTAAGGTTCTTACTAATAGCAAGCCGAACCAGAACCAGTAACGGAAGAAAATTAGTTAAAGTAGAACAGGTTCTCAGCGCTAATATA
CCACGGAACAACAAACGAAAGAGGAATCTGTGAATAGAAATGAAAACTGATATTACGGACTGTCAAGGAAATGATCTAAAGTTCTTAGAAGACAACTAAACAAAAAGAGAAACTGCGAGTGAACCAGTT
AGGCAAAACGTCAGAACCAATAATTAGAACTATTAGTGAAGTACTGCCAGGTTCTAAAGCGCAACAGACAACAGCAGAACTAACAGGATGATGGCCAGCGCAACAGTTCTGTTGCCAATGAACAGAATC
AAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGGTGATGGGGACCCCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAGCGGCCTACAACAC
ACGGCGAGAGAGCGGAGGCTGGGGTGCTTGCCTCTGAGTGTAAACAACAACAACAACAGTTCCACCACAAACAACAACAATAGCAACAGTGCCAGTGCTGCCCCACACGGGGTAGAGGGAGGTGGGGGT
GTGCGGGGGCTCTCTGGAGGAGGGTGCGGTGCCCCCACCACAGAGAGCTGTGCTGGCTGTGCCAGGATCATCACTGACAGGTTCCTACTGCGTGTCAACTGCCTCTCCTGGCACCAAAGCTGCCTGCGC
TGCTGCGTGTGCCAGCTGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATCTACTGCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTGCCGGGCCATCGGG
TCGACGGACTGGGTCCGGAGGGCGAGAGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCTGTAAGAGACAGCTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAGAGTATTGTGTAAA
CAGCATTATTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCGGCCAGCGCAAGAGCAAGCGGGTCCGAACGACCTTCACGGACGAACAGCTGCAGGTGCTCCAG
GCCAACTTCCAGATCGACTCGAA

>Ph_awh_5Rn121 [bp] 'w/o 1406
TATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGCTGGGAAGAAGCAACAAGAGCTTGTGACGGAGCTGTTAGTATTGTAGGTGATATTGAGTCCCTCGCCGCTTTTTGACCAGAATAT
TAATTTGTTAAACAGATGGAAAATATCACTGCAGTAGCTGTTACAGAAGTTCACTGAAATGGACTGTGCAGACTTGAAGTAAACTTCGGCGGCGTGTGTTGTATGGAAGTGCGGACTTCGTTCTGAACT
GTAAATGTCGTACATAGAAACTAACTCGTTGTAGGACTGGCTGTGTAGGGGTAAAGTTATGAGCAGTTCGGAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAATAAGGTTAATACGGTAAGTAGCG
TTTCAAGCTACGGATATGATAGAACACGACTTCTGGATAAAGAACTACCCGCGAAATACGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGAAAGTGGAATATTGCTCCAATAAGTTTAG
TTTCTGCCAGTAGAATTGCAGAAAAGATTATATTTTAATTATCAGCAGAACTGTGCACTACTGTAGCTGAACTAATAGAGATTTTCCCATCCTTATGTACGTGTTCTTAAGAACAATAGAGAACAGAAG
AACCACTGAAAAAACTAATATATATAACTCAGTGAACAAGGATACCGACAGCACCCGTTATTAAGGACTATTTGTGAAATAAGACAAGGTAAGGTTCTTACTATTAGCAAGCCGGACCAGAACCAGTAA
CGGAAGAAAATTAATTAAGGTAGAACAGGTTCTCAGCGCTAATATACCACGGAACAACAAACGAAAGATGAATCTGTGAATAAAAATAAAAACTGTCAAGGAAATGATCTAAAGTTCTTGGAAGACAAC
TAAACAAAAAGAGAAACTGCGAGTGAACCAGTTAGGCAAAACGTCAGAACCAATAATTAGAACTTTAGTGAAGTACTGCCAGGCTCTGAAGCGCAACAGACAACAGCAGAACTAACAGGATGATGGCCA
GCGCAACAGTTCTGTTGCCAATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGTTGATGGGGACCCCTCGGGACCCCCCTGCTGTATGGGCGGCA
GCGTTAAGCTGACTGTAAGCGGCCTACAACACACGGCGAGAGAGCGGAGGCTGGGGTGTTTGCCTCTGAGCGTAAACAACAACAACAACAGTTCCACCACAAACAACAACAATAGCAACAGTGCCAGTG
CTACCCCACACGGGGTGGAGGGAGGTGGGAATGTGCGGGGGCTCTCTGGAGGAGGGTGCGGCGCCCCCACCACAGAGAGCTGTGCCGGCTGTGCCAGGATCATCACTGACAGGTTCCTGCTGCGTGTCA
ACTGCCTCTCCTGGCACCAGAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCCCTGGACCGTCAGCCGTCCTGCTTCATACGCGACCAAAACATCTACTGCAAGTCCGACTACTCTCGAAATTTTGGGG
CCCGCTGCGCCAAATGCTGCCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGAGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCTGTAAGAGACAGCTGAGCACAGGAGAGGAGT
TCGCCTTGCATGAGAGCAGAGTATTGTGTAAACAGCATTATTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCGGCCAGCGCAAGAGCAAGCGGGTCCGAACGA
CCTTCACGGACGAACAGCTGCAGGTGCTCCAGGCCAACTTCCAGATCGACTCGAA

>Ph_awh_5Rn122 [bp] 'w/o 1406
GAATTTTCACTGTTCCCTTTACCACGTTTTACTGTTACTCTTGAATTATTTCCCAGAGCACTTGAAAGAAAATAACATGACTTCAGAATTTATTGCATGGGGAGGGTGCATTTTTCAGAAAATACTTGC
TGATGTATGAGAAAATATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGCTGGGAAGAAGCAACAAGAGCTTGTGACGGAGCTGTTAGTATTGTAGGTGATATTGAGTCCCTCGCCGCT
TTTTGACCAGAATATTAATTTGTTAAACAGATGGAAAATATCACTGCAGTAGCTGTTACAGAAGTTCACTGAAATGGACTGTGCAGACTTGAAGTAAACTTCGGCGGCGTGTGTTGTATGGAAGTGCGG
ACTTCGTTCTGAACTGTAAATGTCGTACATAGAAACTAACTCGTTGTAGGACTGGCTGTGTAGGGGTAAAGTTATGAGCAGTTCGGAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAATAAGGTTA
ATACGGTAAGTAGCGTTTCAAGCTACGGATATGATAGAACACGACTTCTGGATAAAGAACTACCCGCGAAATACGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGAAAGTGGAATATTG
CTCCAATAAGTTTAGTTTCTGCCAGTAGAATTGCAGAAAAGATTATATTTTAATTATCAGCAGAACTGTGCACTACTGTAGCTGAACTAATAGAGATTTTCCCATCCTTATGTACGTGTTCTTAAGAAC
AATAGAGAACAGAAGAACCACTGAAAAAACTAATATATATAACTCAGTGAACAAGGATACCGACAGCACCCGTTATTAAGGACTATTTGTGAAATAAGACAAGGTAAGGTTCTTACTATTAGCAAGCCG
GACCAGAACCAGTAACGGAAGAAAATTAATTAAGGTAGAACAGGTTCTCAGCGCTAATATACCACGGAACAACAAACGAAAGATGAATCTGTGAATAAAAATAAAAACTGTCAAGGAAATGATCTAAAG
TTCTTGGAAGACAACTAAACAAAAAGAGAAACTGCGAGTGAACCAGTTAGGCAAAACGTCAGAACCAATAATTAGAACTTTAGTGAAGTACTGCCAGGCTCTGAAGCGCAACAGACAACAGCAGAACTA
ACAGGATGATGGCCAGCGCAACAGTTCTGTTGCCAATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGTTGATGGGGACCCCTCGGGACCCCCCT
```

```
GCTGTATGGGCGGCAGCGTTAAGTGTTTGCCTCTGAGCGTAAACAACAACAACAACAGTTCCACCACAAACAACAACAATAGCAACAGTGCCAGTGCTACCCCACACGGGGTGGAGGGAGGTGGGAATG
TGCGGGGGCTCTCTGGAGGAGGGTGCGGCGCCCCCACCACAGAGAGCTGTGCCGGCTGTGCCAGGATCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCAGAGCTGCCTGCGCT
GCTGCGTGTGCCAGCTGGCCCTGGACCGTCAGCCGTCCTGCTTCATACGCGACCAAAATATTTACTGCAAGTCTGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTGCCGGGCCATCGGGT
CGACGGACTGGGTCCGGAGGGCGAGGGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCTGTAAGAGACAACTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAGAGTATTGTGTAAAC
AGCATTACTTGGAGGCCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCGGCCAGCGCAAGAGTAAGCGGGTCCGGACGACCTTCACGGACGAGCAGCTGCAGGTGCTCCAGG
CTAACTTCCAGATCGACTCGAA

>Ph_awh_5Rn150 [bp] 'w/o 1406
TATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGCTGGGAAGAAGCAACAAGAGCTTGTGACGGAGCTGTTAGTATTGTAGGTGATATTGAGTCCCTCGCCGCTTTTTGACCAGAATAT
TAATTTGTTAAACAGATGGAAAATATCACTGCAGTAGCTGTTGCAGAAGTTCACTGAAATGGACTGTGCAGACTTGAAGTAAACTTCGGCGGCGTGTGTTGTATGGAAGTGCGGACTTCGTTCTGAACT
GTAAATGTCGTACATAGAAACTAACTCGTTGTAGGACTGGCTGTGTAGGGGTAAAGTTATGAGCAGTTCGGAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAATAAGGTTAATACGGTAAGTAGCG
TTTCAAGCTACGGATATGATAGAACACGACTTCTGGATAAAGAACTACCCGCGAAATACGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGAAAGTGGAATATTGCTCCAATAAGTTTAG
TTTCTGCCAGTAGAATTGCAGAAAAGATTATATTTTAATTATCAGCAGAACTGTGCACTACTGTAGCTGAACTAATAGAGATTTTCCCATCCTTATGTACGTGTTCTTAAGAACAATAGAGAACAGAAG
AACCACTGAAAAAATTAATATATATAACTCAGTGAACAAGGATACCGACAGCACCCGTTATTAAGGACTATTTGTGAAATAAGACAAGGTAAGGTTCTTACTATTAGCAAGCCGGACCAGAACCAGTAA
CGGAAGAAAATTAATTAAGGTAGAACAGGTTCTCAGCGCTAATATACCACGGAACAACAAACGAAAGATGAATCTGTGAATAAAAATAAAAACTGTCAAGGAAATGATCTAAAGTTCTTGGAAGACAAC
TAAACAAAAAGAGAAACTGCGAGTGAACCAGTTAGGCAAAACGTCAGAACCAATAATTAGAACTTTAGTGAAGTACTGCCAGGCTCTGAAGCGCAACAGACAACAGCAGAACTAACAGGATGATGGCCA
GCGCAACAGTTCTGTTGCCAATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGTAACACAGTTGATGGGGACCCCTCGGGACCCCCCTGCTGTATGGGCGGCA
GCGTTAAGTGCTTGCCTCTGAGTGTAAACAACAACAACAACAGTTCCACCACAAACAACAACAATAGCAACAGTGCCAGTGCTGCCCCACACGGGGTAGAGGGAGGTGGGGGTGTGCGGGGGCTCTCTG
GAGGAGGGTGCGGTGCCCCCACCACAGAGAGCTGTGCTGGCTGTGCCAGGATCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCAGAGCTGCCTGCGCTGCTGCGTGTGCCAGC
TGGCCCTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATCTACTGCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAATGCTGCCGGGCCATCGGGTCGACGGACTGGGTCC
GGAGGGCGAGAGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCTGTAAGAGACAGCTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGTAGAGTATTGTGTAAACAGCATTATTTGGAGG
CCGTGGAGGGAGGAGCAACTAGTAATGATGAGAACACGGACGGTGAGGGCGGCCAGCGCAAGAGCAAGCGGGTCCGAACGACCTTCACGGACGAACAGCTGCAGGTGCTCCAGGCCAACTTCCAGATCG
ACTCGAA

>Ph_awh_vsdP10 [bp] 'w/o deg oligos, identical to >Ph_awh_vsdP19 [bp]
GGACGAGCAGCTGCAGGTGCTCCAGGCTAACTTCCAGATCGACTCGAACCCTGACAGCCAGGACCTGGAGCGCATCGCACAGCTCACGGGGCTGAGCAAGCGGGTGACG

>Ph_awh_vsdP15 [bp] 'w/o deg oligos
GGACGAGCAGCTGCAGGTGCTCCAGGCTAACTTCCAGATTGACTCGAACCCTGACAGCCAGGACCTGGAGCGCATCGCACAGCTCACGGGGCTGAGCAAGCGGGTGACG

>Ph_awh_3Rn21 [bp] 'w/o 1401
ATCGCACAGCTCACGGGGCTGAGCAAGCGGGTGACGCAGGTGTGGTTCCAAAACAGCAGAGCGCGGCAGAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATC
ACAGGCAGCTTAGCCCGTACAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCCAGAGGACCTGTCGTCTCCCGAGAACTCGCTTATGACGCTCATGCCTGGAGAAGCATGACGG
TACGACCCCTCCACCTCCCCTCCACCTCCTCCTCCACCCACCCTCCACTGACGGCGCTGTGTCACCCGTTGTGGCACTCCTGGCGAGAGTCGCACTTTGTGACACCAACAGACTGCTCTGTGTCACCAG
TTGTGACACTAACCGGCACACGCAATGAAGCAAACATTTCCAGAGAAGTTGACCTTAGAGCAGAATGGTATTAACGAGTTGGGAATGTAACGTCAAGCTGGCATTGATTTTCAATTTGAAGAAATGAAA
ATATAACATACTTGGAAGTCATATATCTGAAACTAAACAAATGACAGCAATCTTGACTACCAGTCGGACAATTACACGAGACAATCTCAGCTACCAGGCTTACGATTACACGAGTCATCTCAAGGGAAG
CAGGATTTTTTAGAGGTCGGATGAGATGGGACTCCGTGTGTCGCTACTGAAGTTGGTTAAATTCTTCATGAGACCCAGTGGATCCGTAAGGTCAGGAGTCAAATGGATCTGAGATCCAGTGGATCAGTT
ACACAGAGTTCAATGACAGCGACTAGCAATAACTAGTTTCCAGACTAGTTACTTCTGCTAGTTCCAAACATATTATGGGTAGTGACAAGAATTTTCGAATAATCATTGCGGCGAATTCGGCCCACTCGT
GGCTGTCGAGGAAAGCCAGTCAGCTTGGCTTCTCCCGGTGGCAGAGAGCAGAGACTCTGGCACCGTCCGAGTGCCAGTAAGTCAAGCTTGAAGCTCAATCTAAAGCGCGACTCGAACGCCGGCGATGGA
AACTGCGGAAACGTTAAAGCTAAGTGGGATTAATTTAATCGATGACAGTAGGCGAATTGCCCACGAGTGTGTCTTATTCAGCATACTATAGATAAGCATGCGTATGATCTAACGCAGCACTCTAGGGAA
GGTTGCGTATGGTCAAACAGGGCACCCTAGGGAACGTTGCGTATGGTTGCAGCACACTTCTGCGTATGTTCTGAGTCGCGCGGCCGCAGATAAGTGCAGGGCGGAACCACGCAGCTTTGCTCAAGTTGG
ACAGATTCAGCTCTGTAGGAATGGAAGAATTACACAACAGTACGCGGGCGATGAGTTGCAGCAAATACGACACCGTGGTCATCTGGAAATGTTCTGCACAGCTTCTACATCAAGAGTACCCTAGGTGGG
GGCACGGACCGTCAGAGAGCGCGTGATACCATTGTAGGGGATTCATGATAAATATAATTGGTACACTCAGTACCGGTAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_awh_3Rn22 [bp] 'w/o 1401, identical to >Ph_awh_3Rn75 [bp] 'w/o 1401
ATCGCGCAGCTCACGGGGCTGAGCAAGCGGGTGACGCAGGTCTGGTTCCAGAACAGCAGAGCGCGGCAGAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATC
ACAGGCAGCTTAGCCCGTACAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCCAGAGGACCTGTCGTCTCCCGAGAACTCGCTTATGACGCTCATGCCTGGAGAAGCATGACGG
TACGACCCCTCCACCTCCCCTCCACCTCCTCCTCCACCCACCCTCCACTGACGGCGCTGTGTCACCCGTTGTGGCACTCCTGGCGAGAGTCGCACTTTGTGACACCAACAGACTGCTCTGTGTCACCAG
TTGTGACACTAACCGGCACACGCAATGAAGCAAACATTTCCAGAGAAGTTGACCTTAGAGCAGAATGGTATTAACGAGTTGGGAATGTAACGTCAAGCTGGCATTGATTTTCAATTTGAAGAAATGAAA
ATATAACATACTTGGAAGTCATATATCTGAAACTAAACAAATGACAGCAATCTTGACTACCAGTCGGACAATTACACGAGACAATCTCAGCTACCAGGCTTACGATTACACGAGTCATCTCAAGGGAAG
CAGGATTTTTTAGAGGTCGGATGAGATGGGACTCCGTGTGTCGCTACTGAAGTTGGTTAAATTCTTCATGAGACCCAGTGGATCCGTAAGGTCAGGAGTCAAATGGATCTGAGATCCAGTGGATCAGTT
ACACAGAGTTCAATGACAGCGACTAGCAATAACTAGTTTCCAGACTAGTTACTTCTGCTAGTTCCAAACATATTATGGGTAGTGACAAGAATTTTCGAATAATCATTGCGGTGAATTCGGGCCGCTCGT
GGCTGTCGAGGGAAGCCAGTCAGCTTGGCTTCTCCCGGTGGCAGAGAGCAGAGACTCTGGCACCGTCCGAGTGCCAGTAAGTCAAGCTTGAAGCTCAATCTAAAGCGCGACTCGAACGCCGGCGATGGA
AACTGCGGAAACGTTAAAGCTAAGTGGAATTAATTTAATCGATGACAGTAGGCGAATTGCCCACGAGTGTGTCTTATTCAGCATACTATAGATAAGCATGCGTATGATCTAACGCAGCACTCTAGGGAA
GGTTGCGTATGGTCAAACAGGGCACCCTAGGGAACGTTGCGTATGGTTGCAGCACACTTCTGCGTATGTTCTGAGTCGCGCGGCCGCAGATAAGTGCAGGGCGGAACCACGCAGCTTTGCTCAAGTTGG
ACAGATTCAGCTCTGTAGGAATGGAAGAATTACACAACAGTACGCGGGCGATGAGTTGCAGAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_awh_3Rn78 [bp] 'w/o 1401
ATCGCACAGCTCACGGGGCTGAGCAAGCGGGTGACGCAGGTGTGGTTCCAAAACAGCAGAGCGCGGCAGAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATC
ACAGGCAGCTTAGCCCGTACAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCCAGAGGACCTGTCGTCTCCCGAGAACTCGCTTATGACGCTCATGCCTGGAGAAGCATGACGG
TACGACCCCTCCACCTCCCCTCCACCTCCTCCTCCACCCACCCTCCACTGACGGCGCTGTGTCACCCGTTGTGGCACTCCTGGCGAGAGTCGCACTTTGTGACACCAACAGACTGCTCTGTGTCGCCAG
TTGTGACACTAACCGGCACACGCAATGAAGCAAACATTTCCAGAGAAGTTGACCTTAGAGCAGAATGGTATTAACGAGTTGGGAATGTAACGTCAAGCTGGCATTGATTTTCAATTTGGAGAAATGAAA
ATATAACATACTTGGAAGTCATATATCTGAAACTAAACAAATGACAGCAATCTTGACTACCAGTCGGACAATTACACGAGACAATCTCAGCTACCAGGCTTACGATTACACGAGTCATCTCAAGGGAAG
CAGGATTTTTTAGAGGTCGGATGAGATGGGACTCCGTGTGTCGCTACTGAAGTTGGTTAAATTCTTCATGAGACCCAGTGGATCCGTAAGGTCAGGAGTCAAATGGATCTGAGATCCAGTGGATCAGTT
ACACAGAGTTCAATGACAGCGACTAGCAATAACTAGTTTCCAGACTAGTTACTTCTGCTAGTTCCAAACATATTATGGGTAGTGACAAGAATTTTCGAATAATCATTGCGGCGAATTCGGCCCACTCGT
GGCTGTCGAGGGAAGCCAGTCAGCTTGGCTTCTCCCGGTGGCAGAGAGCAGAGACTCTGGCACCGTCCGAGTGCCAGTAAGTCAAGCTTGAAGCTCAATCTAAAGCGCGACTCGAACGCCGGCGATGGA
AACTGCGGAAACGTTAAAGCTAAGTGGGATTAATTTAATCGATGACAGTAGGCGAATTGCCCACGAGTGTGTCTTATTCAGCATACTATAGATAAGCATGCGTATGATCTAACGCAGCACTCTAGGGAA
GGTTGCGTATGGTCAAACAGGGCACCCTAGGGAACGTTGCGTATGGTTGCAGCACACTTCTGCGTATGTTCTGAGTCGCGCGGCCGCAGATAAGTGCAGGGCGGAACCACGCAGCTTTGCTCAAGTTGG
ACAGATTCAGCTCTGTAGGAATGGAAGAATTACACAACAGTACGCGGGCGATGAGTTGCAGCAAATACGACACCGTGGTCATCTGGAAATGTTCTGCACAGCTTCTACATCAAGAGTACCCTAGGTGGG
GGCACGGACCGTCAGAGAGCGCGTGATACCATTGTAGGGGATTCATGATAAATATAATTGGTACACTCAGTACCGGTAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_awh_3RCR07 [bp] 'w/o 1401
ATCGCGCAGCTCACGGGGCTGAGCAAGCGGGTGACGCAGGTCTGGTTCCAGAACAGCAGAGCGCGGCAGAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATC
ACAGGCAGCTTAGCCCGCACAATCGACCTACACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCCAGAGGACCTGTCGTCTCCCGAGAACTCGCTTATGACGCTCATGCCTGGAGAAGCATGACGG
TACGACCCCTCCACCTCCCCTCCACCTCCTCCTCCACCCACCCTCCACTGACGGCGCTGTGTCACCCGTTGTGGCACTCCTGGCGAGAGTCGCACTTTGTGACACCAACAGACTGCTCTGTGTCACCAG
TTGTGACACTAACCGGCACACGCAATGAAGCAAACATTTCCAGAGAAGTTGACCTTAGAGCAGAATGGTATTAACGAGTTGGGAATGTAACGTCAAGCTGGCATTGATTTTCAATTTGAAGAAATGAAA
ATATAACATACTTGGAAGTCATATATCTTAAACTAAACAAATGGCAGCAATCTTGACTACCAGTCGGACAATTACACGAGACAATCTCAGCTACCAGGCTTACGATTACACGAGTCATCTCAAGGGAAG
CAGGATTTTTTAGAGGTCGGATGAGATGGGACTCCGTGTGTCGCTACTGAAGTTGGTTAAATTCTTCATGAGACCCAGTGGATCCGTAAGGTCAGGAGTCAAATGGATCTGAGATCCAGTGGATCAGTT
ACACAGAGTTCAATGACAGCGACTAGCAATAACTAGTTTCCAGACTAGTTACTTCTGCTAGTTCCAAACATATTATGGGTAGTGACAAGAATTTTCGAATAATCATTGCGGCGAATTCGGGCCGCTCGT
GGCTGTCGAGGAAAGCCAGTCAGCTTGGCTTCTCCCGGTGGCAGAGAGCAGAGACTCTGGCACCGTCCGAGTGCCAGTAAGTCAAGCTTGAAGCTCAATCTAAAGCGCGACTCGAACGCCGGCGATGGA
AACTGCGGAAACGTTAAAGCTAAGTGGGATTAATTTAATCGATGACAGTAGGCGAATTGTCCACAAGTGTGTCTTATTCAGCATACTATAGGTAAACATGCGTATGGTCAAACGCAGCACTCTAGGGAA
GGTTGCGTACGATCAAACAGAGCACCTTAGGGAAGGTTGCGTATGGTCAAACAGAGCACCCTAGGGAAGGTTGCGTATGGTTGCAGCACACTTCTGCGTATGTTCTGAGTCGCACGGCCGCAGATAAGT
TCAGGGCGGAAACACGCAGCTCTGGCTCAAGTTTGACAGATTCAGCTCTGTAGGAATGGAGGAATTACACAACAGTACGCGGGCGACGAGTTGCAGCAAATGCGACACCGTGGTCATCTGGAAATGTTG
TGCACAGCTTCTACATCAAGAGTACCCTAGGTGGGGGCACGGACCCTCAGAGAGCGCGTGATACCATTGTAGGGGATTCATGATAAATATAATTGGGTACACTCAGTACCGGTAAATGTAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAA

>Ph_awh_3RCR09 [bp] 'w/o 1401
ATCGCGCAGCTCACGGGGCTGAGCAAGCGGGTGACGCAGGTCTGGTTCCAGAACAGCAGAGCGCGGCAGAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATC
ACAGGCAGCTTAGCCCGCACAATCGACCTACACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCCAGAAGACCTGTCGTCTCCCGAGAACTCGCTTATGACGCTCATGCCTGGAGAAGCATGACGG
TACGACCCCTCCACCTCCCCTCCACCTCCTCCTCCACCCACCCTCCACTGACGGCGCTGTGTCACCCGTTGTGGCACTCCTGGCGAGAGTCGCACTTTGTGACACCAACAGACTGCTCTGTGTCACCAG
```

```
TTGTGACACTAACCGGCACACGCAATGAAGCAAGCATTTCCAGAGAAGTTGACCTTAGAGCAGAATGGTATTAGCGAGTTGGGAATGTAACGTCAAGCTGGCATTGATTTTCAATTTGAAGAAATGAAA
ATATAACATACTTGGAAGTCATATATCTTAAACTAAACAAATGACAGCAATCTTGGCTACCAGTCGGACAATTACACGAAACAATCTCAGCTACCAGGCTTACGATTACACGAGTCATCTCAAGGGAAG
CAGGATTTTTTAGAGGTCGGATGAGATGGGACTCCGTGTGTCGCTACTGAAGTTGGTTAAATTCTTCATGAGACCCAGTGGATCCGTAAGGTCAGGAGTCAAATGGATCTGAGATCCAGTGGATCAGTT
ACACAGAGTTCAATGACAGCGACTAGCAATAACTAGTTTCCAGACTAGTTACTTCTGCTAGTTCCAAACATATTATGGGTAGTGACAAGAATTTTCGAATAATCATTGCGGTGAATTCGGGCCGCTCGT
GGCTGTCGAGGAAAGCCAGTCAGCTTGGCTTCTCCCGGTGGCAGAGAGCAGAGATTCTGGCACCGTCCGAGTGCCAGTAAGTCAAGCTTGAAGCTCAATCTAAAGCGCGACTCGAACGCCGGCGATGGA
AACTGCGGAAACGTTAAAGCTAAGTGGAATTAATTTAATCGATGACAGTAGGCGAATTGTCCACAAGTGTGTCTTATTCAGCATACTATAGGTAAGCATGCGTATGGTCTAAGGCAGCACTCTAGGGAA
GGTTACGTATGGTCAAACAGAGCACCCTAGGGAACGTTGCGTATGGTTGCAGCACACTTCTGCGTATGTTCTGAGTCGCGCGGCCGCAGATAAGTGCAGGGCGGAACCACGCAGCTCTGGCTCAAGTTT
GACAGATTCAGCTCTGTAGGAATGGAGGAATTACACAACAGTACGCGGGCGACGAGTTGCAGCAAATGCGACACCGTGGTCATCTGGAAATGTTCTGCGCAGCTTCTACATCAAGAGTACCCTAGGTGG
GGGCACGGACCCTCAGAGAGCGCGTGATACCATTGTAGGGGATTCATGATAAATATAATTGGGTACACTCAGTACCGGTAAATGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_awh_3RCR10 [bp] 'w/o 1401
ACCGCACAGCTCACGGGGCTGAGCAAGCGGGTGACGCAGGTGTGGTTCCAAAACAGCAGAGCGCGGCAGAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGACATC
ACAGGCAGCTTAGCCCGTACAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCCAGAGGACCTGTCGTCTCCCGAGAACTCGCTTATGACGCTCATGCCTGGAGAAGCATGACGG
TACGACCCCTCCACCTCCCCTCCACCTCCTCCTCCACCCACCCTCCACTGACGGCGCTGTGTCACCCGTCGTGGCACTCCTGGCGAGAGTCGCACTTTGTGACACCAACAGACTGCTCTGTGTCACCAG
TTGTGACACTCACCGGCACACGCAATGAAGCAAACATTTCCAGAGAAGTTGACCTTAGAGCAGAATGGTATTAGCGAGTTGGGAATGTAACGTCAAGCTGGCATTGATTTTCGATTTGAAGAAATGAAA
ATATAACATACTTGGAAGTCATATATCTTAAACTAAACAAATGACAGCAATCTTGGCTACCAGTCGGACAATTACACGAAACAATCTCAGCTACCAGGCTTACGATTACACGAGTCATCTCAAGGGAAG
CAGGATTTTTTAGAGGTCGGATGAGATGGGACTCCGTGTGTCGCTACTGAAGTTGGTTAAATTCTTCATGAGACCCAGTGGATCCGTAAGGTCAGGAGTCAAATGGATCTGAGATCCAGTGGATCAGTT
ACACAGAGTTCAATGACAGCGACTAGCAATAACTAGTTTCCAGACTAGTTACTTCTGCTAGTTCCAAACATATTATGGGTAGTGACAAGAATTTTCGAATAATCATTGCGGTGAATTCGGGCCGCTCGT
GGCTGTCGAGGAAAGCCAGTCAGCTTGGCTTCTCCCGGTGGCAGAGAGCAGAGATTCTGGCACCGTCCGAGTGCCAGTAAGTCAAGCTTGAAGCTCAATCTAAAGCGCGACTCGAACGCCGGCGATGGA
AACTGCGGAAACGTTAAAGCTAAGTGGAATTAATTTAATCGATGACAGTAGGCGAATTGTCCACAAGTGTGTCTTATTCAGCATACTATAGGTAAGCATGCGTATGGTCTAAGGCAGCACTCTAGGGAA
GGTTACGTATGGTCAAACAGAGCACCCTAGGGAACGTTGCGTATGGTTGCAGCACACTTCTGCGTATGTTCTGAGTCGCGCGGCCGCAGATAAGTGCAGGGCGGAACCACGCAGCTTCGCTCAAGTTGG
ACAGATTCAGCTCTGTAGGAATGGAAGAATTACACAACAGTACGCGGGCGATGAGTTGCAGCAAATACGACACCGTGGTCATCTGGAAATGTTCTGCACAGCTTCTACATCAAGAGTACCCTAGGTGGG
GACACGGACCGTCAGAGAGCGCGTGATACCATTGTAGGGGATTCATGATAAATATAATTGGTACACTCAGTACCGGTAAATGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA



>Ph_awh_ref [aa] 'Ph_awh_ORF21 w/o primers, extended 5' by Ph_awh_5RCR02 (1210 bp); extended 3' by Ph_awh_3Rn21 (1298 bp)
MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNNSSTTNNNNNSNSASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFL
LRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKR
VRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA&

>Ph_awh_ORF20 [bp] 'w/o 1414/1415
MNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNNSSTTNNNNNSNSASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQ
SCLRCCVCQLALDRQPSCFIRD#NIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLARFACDACKRQLSTGEELALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQL
QVLQANFQIDSNPDSQDLERIAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSVEPEDLSSPEN

>Ph_awh_ORF21 [bp] 'w/o 1414/1415
MNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNNSSTTNNNNNSNSASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQ
SCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQL
QVLQANFQIDSNPDSQDLERIAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPEN

>Ph_awh_ORF22 [bp] 'w/o 1414/1415
MNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNNSSTTNNNNNSNSASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQ
SCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQL
QVLQANFQIDSNPDSQDLERIAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPEN

>Ph_awh5RCR1 [bp] 'w/o 1406
MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNNSSTTNNNNNSNGASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFL
LRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKR
VRTTFTDEQLQVLQANFQIDS

>Ph_awh5RCR2 [bp] 'w/o 1406
MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNNSSTTNNNNNSNSASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFL
LRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKR
VRTTFTDEQLQVLQANFQIDS

>Ph_awh5RCR5 [bp] 'w/o 1406
MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNNSSTTNNNNNSNSASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFL
LRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKR
VRTTFTDEQLQVLQANFQIDS

>Ph_awh5RCR6 [bp] 'w/o 1406
MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNNSSTTNNNNNSNSASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFL
LRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKR
VRTTFTDEQLQVLQANFQIDS

>Ph_awh_5Rn121 [bp] 'w/o 1406
MMASATVLLPMNRIKAERNSPVSSPEASSGNTVDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNNSSTTNNNNNSNSASATPHGVEGGGNVRGLSGGGCGAPTTESCAGCARIITDRFL
LRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKR
VRTTFTDEQLQVLQANFQIDS

>Ph_awh_5Rn122 [bp] 'w/o 1406
MMASATVLLPMNRIKAERNSPVSSPEASSGNTVDGDPSGPPCCMGGSVKCLPLSVNNNNNNSSTTNNNNNSNSASATPHGVEGGGNVRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCC
VCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQAN
FQIDS

>Ph_awh_5Rn150 [bp] 'w/o 1406
MMASATVLLPMNRIKAERNSPVSSPEASSGNTVDGDPSGPPCCMGGSVKCLPLSVNNNNNNSSTTNNNNNSNSASAAPHGVEGGGGVRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCC
VCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQAN
FQIDS

>Ph_awh_vsdP10 [bp] 'w/o deg oligos
DEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVT

>Ph_awh_vsdP15 [bp] 'w/o deg oligos
DEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVT

>Ph_awh_3Rn21 [bp] 'w/o 1401
IAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA&
```

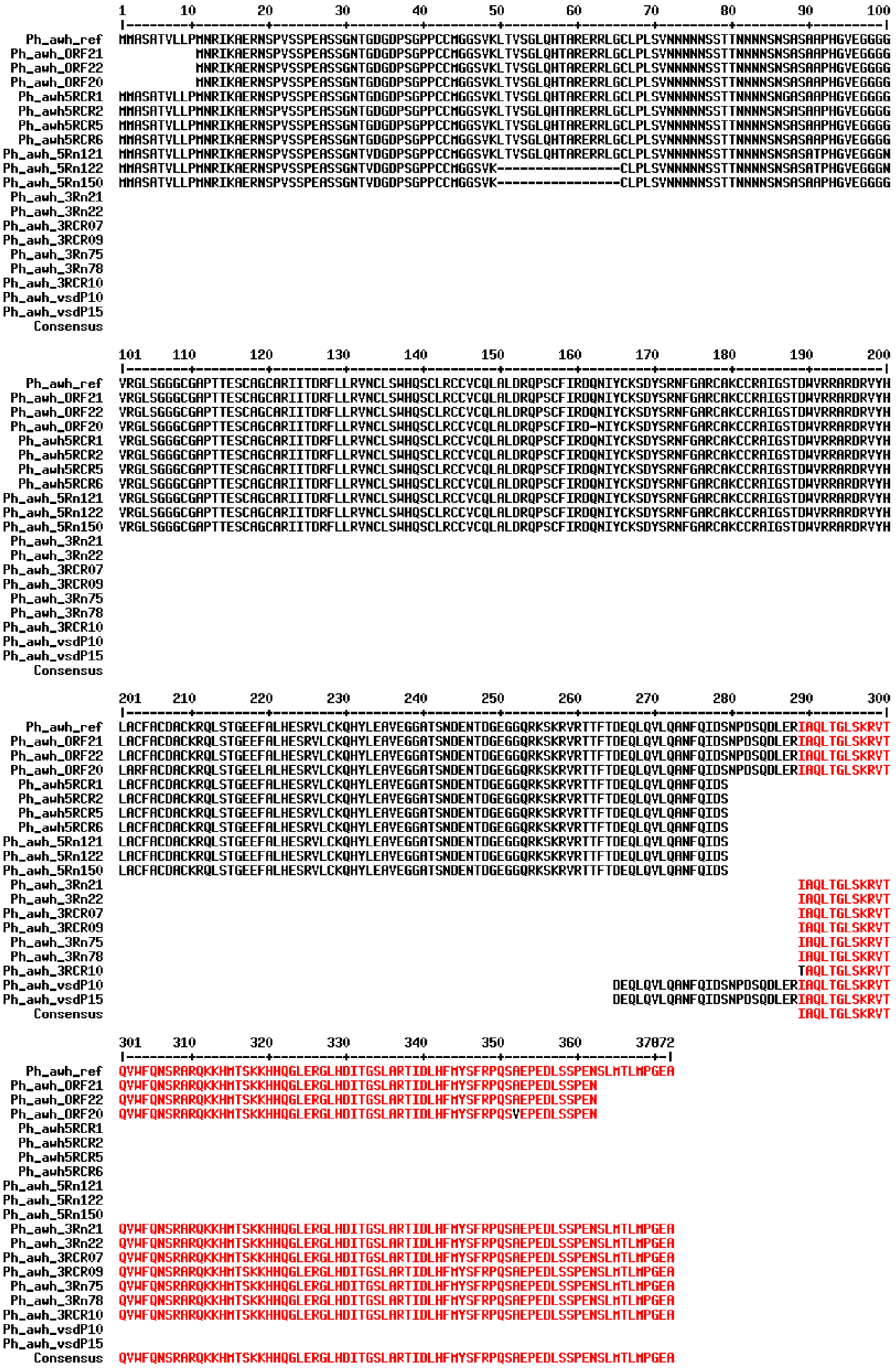
```
>Ph_awh_3Rn22 [bp] 'w/o 1401
IAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA&

>Ph_awh_3Rn75 [bp] 'w/o 1401
IAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA&

>Ph_awh_3Rn78 [bp] 'w/o 1401
IAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA&

>Ph_awh_3RCR07 [bp] 'w/o 1401
IAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA&

>Ph_awh_3RCR09 [bp] 'w/o 1401
IAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA&

>Ph_awh_3RCR10 [bp] 'w/o 1401
TAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA&
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.7_Ph_awh]

### [PhD_A2.7.2_Ph_awh_alignments]

**“alignment_Ph_awh_all_aa_sequences.gif”**



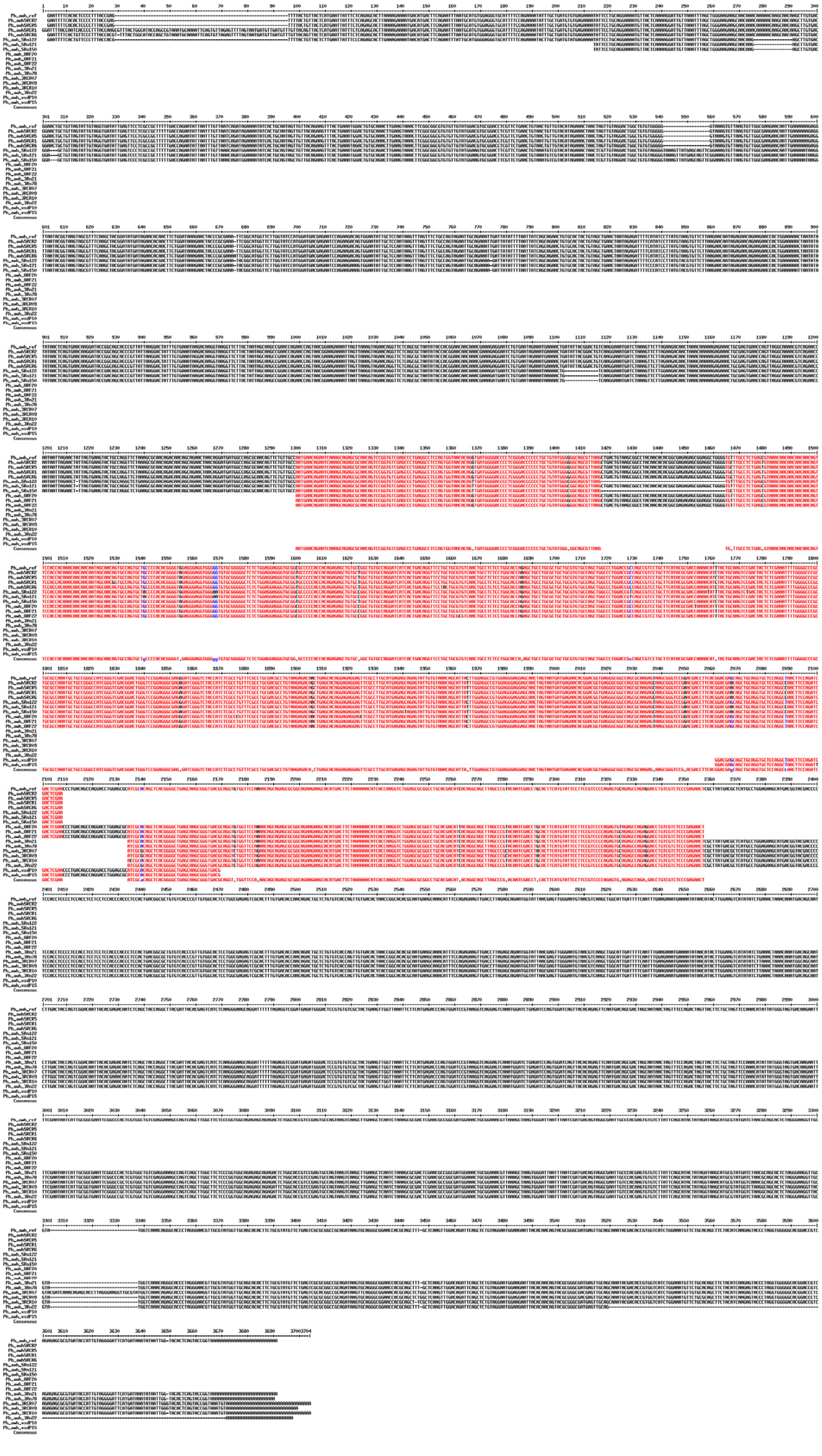
```
              1       10        20        30        40        50        60        70        80        90       100
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_awh_ref MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGG
 Ph_awh_ORF21           MNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGG
 Ph_awh_ORF22           MNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGG
 Ph_awh_ORF20           MNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGG
  Ph_awh5RCR1 MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNGASAAPHGVEGGGG
  Ph_awh5RCR2 MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGG
  Ph_awh5RCR5 MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGG
  Ph_awh5RCR6 MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGG
Ph_awh_5Rn121 MMASATVLLPMNRIKAERNSPVSSPEASSGNTVDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASATPHGVEGGGN
Ph_awh_5Rn122 MMASATVLLPMNRIKAERNSPVSSPEASSGNTVDGDPSGPPCCMGGSVK----------------CLPLSVNNNNNSSTTNNNNSNSASATPHGVEGGGN
Ph_awh_5Rn150 MMASATVLLPMNRIKAERNSPVSSPEASSGNTVDGDPSGPPCCMGGSVK----------------CLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGG
 Ph_awh_3Rn21
 Ph_awh_3Rn22
Ph_awh_3RCR07
Ph_awh_3RCR09
 Ph_awh_3Rn75
 Ph_awh_3Rn78
Ph_awh_3RCR10
Ph_awh_vsdP10
Ph_awh_vsdP15
    Consensus

              101    110       120       130       140       150       160       170       180       190       200
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_awh_ref VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
 Ph_awh_ORF21 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
 Ph_awh_ORF22 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
 Ph_awh_ORF20 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRD-NIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
  Ph_awh5RCR1 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
  Ph_awh5RCR2 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
  Ph_awh5RCR5 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
  Ph_awh5RCR6 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
Ph_awh_5Rn121 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
Ph_awh_5Rn122 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
Ph_awh_5Rn150 VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWYRRARDRVYH
 Ph_awh_3Rn21
 Ph_awh_3Rn22
Ph_awh_3RCR07
Ph_awh_3RCR09
 Ph_awh_3Rn75
 Ph_awh_3Rn78
Ph_awh_3RCR10
Ph_awh_vsdP10
Ph_awh_vsdP15
    Consensus

              201    210       220       230       240       250       260       270       280       290       300
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_awh_ref LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVT
 Ph_awh_ORF21 LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVT
 Ph_awh_ORF22 LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVT
 Ph_awh_ORF20 LARFACDACKRQLSTGEELALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVT
  Ph_awh5RCR1 LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDS
  Ph_awh5RCR2 LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDS
  Ph_awh5RCR5 LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDS
  Ph_awh5RCR6 LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDS
Ph_awh_5Rn121 LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDS
Ph_awh_5Rn122 LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDS
Ph_awh_5Rn150 LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDS
 Ph_awh_3Rn21                                                                                         IAQLTGLSKRVT
 Ph_awh_3Rn22                                                                                         IAQLTGLSKRVT
Ph_awh_3RCR07                                                                                         IAQLTGLSKRVT
Ph_awh_3RCR09                                                                                         IAQLTGLSKRVT
 Ph_awh_3Rn75                                                                                         IAQLTGLSKRVT
 Ph_awh_3Rn78                                                                                         IAQLTGLSKRVT
Ph_awh_3RCR10                                                                                         TAQLTGLSKRVT
Ph_awh_vsdP10                                                                 DEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVT
Ph_awh_vsdP15                                                                 DEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVT
    Consensus                                                                                         IAQLTGLSKRVT

              301    310       320       330       340       350       360       37072
              |--------+---------+---------+---------+---------+---------+---------+-|
   Ph_awh_ref QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
 Ph_awh_ORF21 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPEN
 Ph_awh_ORF22 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPEN
 Ph_awh_ORF20 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSVEPEDLSSPEN
  Ph_awh5RCR1
  Ph_awh5RCR2
  Ph_awh5RCR5
  Ph_awh5RCR6
Ph_awh_5Rn121
Ph_awh_5Rn122
Ph_awh_5Rn150
 Ph_awh_3Rn21 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
 Ph_awh_3Rn22 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
Ph_awh_3RCR07 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
Ph_awh_3RCR09 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
 Ph_awh_3Rn75 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
 Ph_awh_3Rn78 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
Ph_awh_3RCR10 QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
Ph_awh_vsdP10
Ph_awh_vsdP15
    Consensus QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.7_Ph_awh]

[PhD_A2.7.2_Ph_awh_alignments]

**“alignment_Ph_awh_all_bp_sequences.gif”**

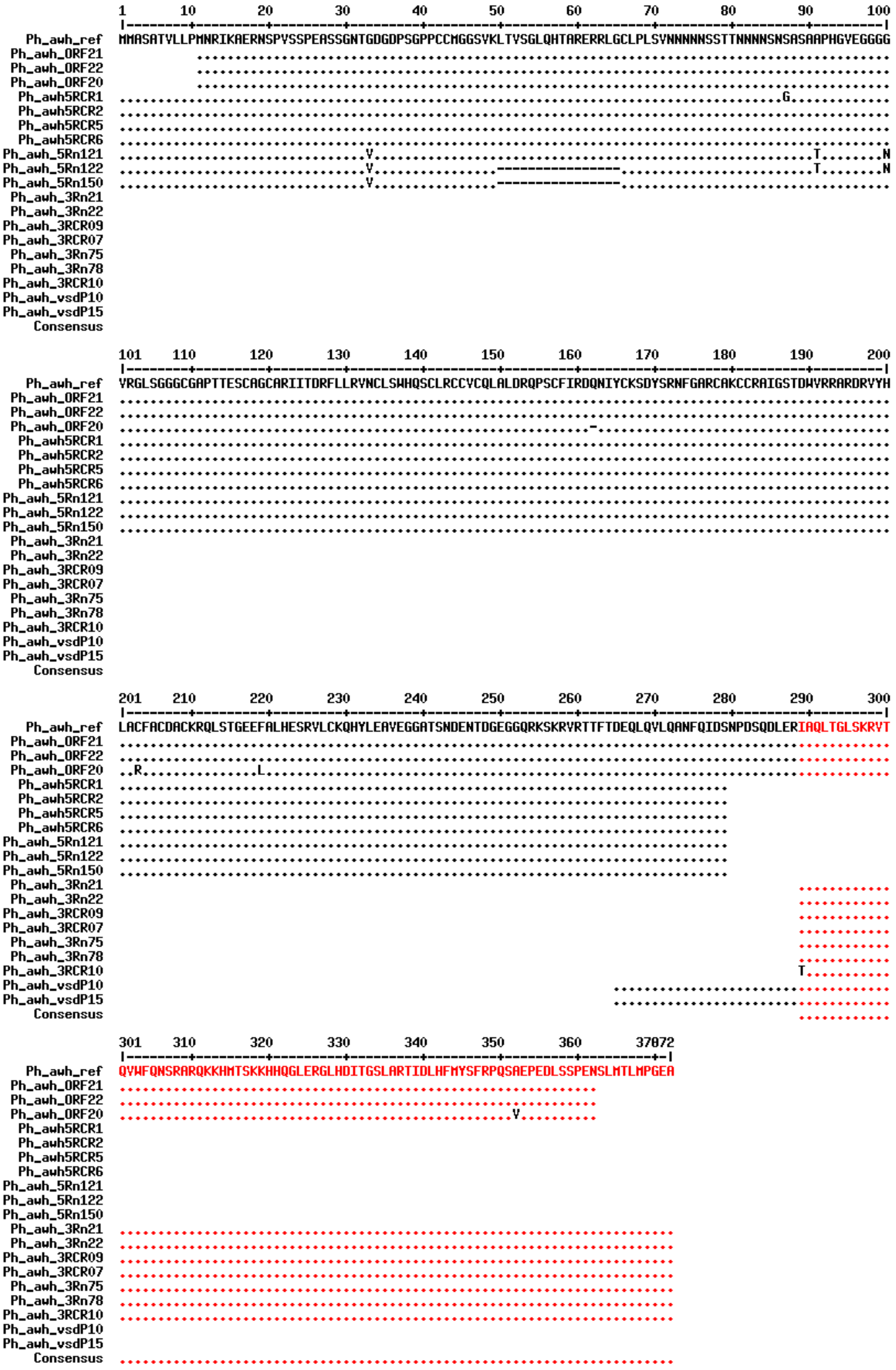


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.7_Ph_awh]

[PhD_A2.7.2_Ph_awh_alignments]

**“difference_alignment_Ph_awh_all_aa_sequences.gif”**



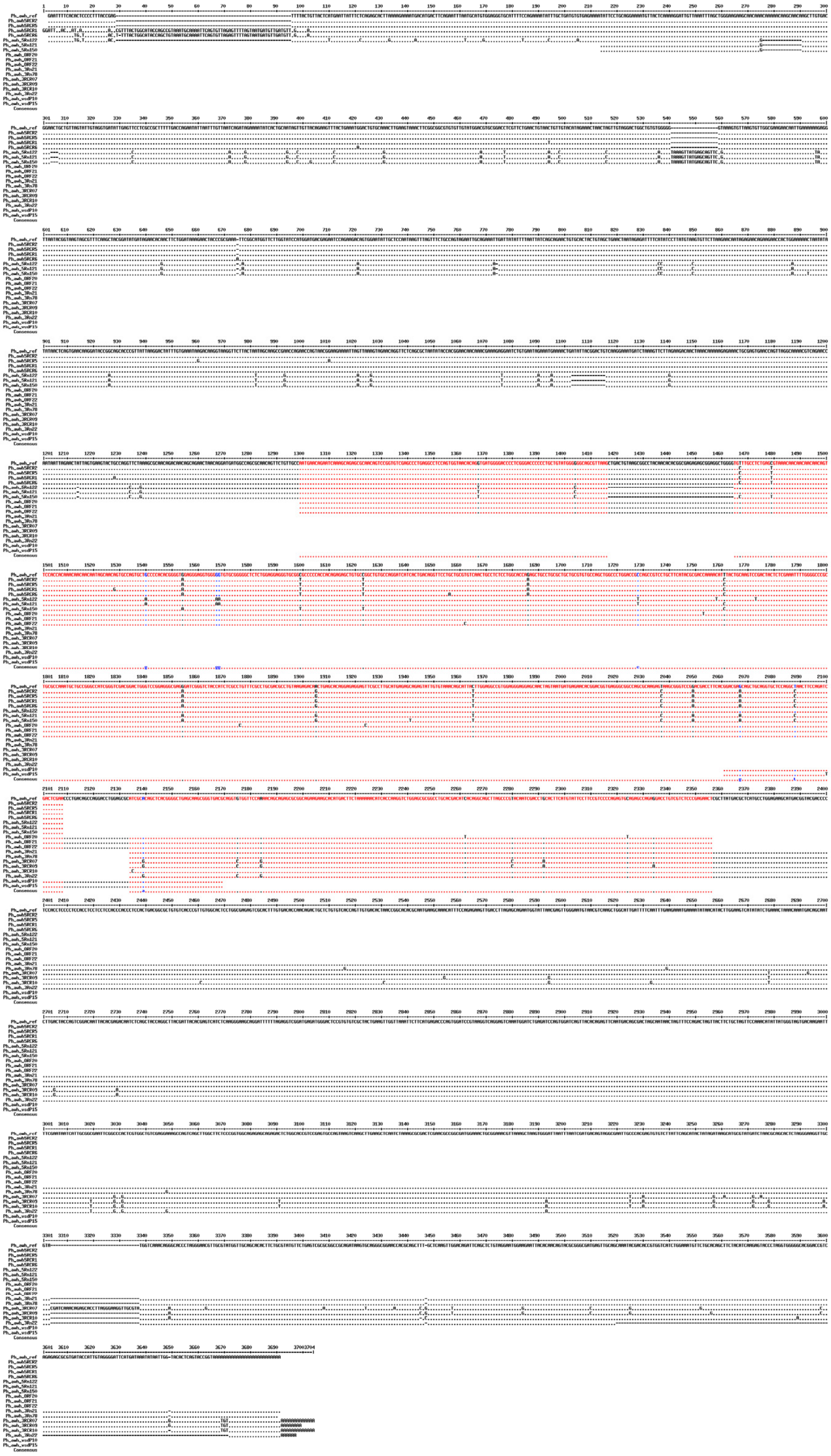
```
              1       10        20        30        40        50        60        70        80        90       100
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_awh_ref MMASATVLLPMNRIKAERNSPVSSPEASSGNTGDGDPSGPPCCMGGSVKLTVSGLQHTARERRLGCLPLSVNNNNNSSTTNNNNSNSASAAPHGVEGGGG
 Ph_awh_ORF21           ..........................................................................................
 Ph_awh_ORF22           ..........................................................................................
 Ph_awh_ORF20           ..........................................................................................
  Ph_awh5RCR1 ......................................................................................G.............
  Ph_awh5RCR2 ....................................................................................................
  Ph_awh5RCR5 ....................................................................................................
  Ph_awh5RCR6 ....................................................................................................
Ph_awh_5Rn121 ................................V.........................................................T........N
Ph_awh_5Rn122 ................................V................----------------.........................T........N
Ph_awh_5Rn150 ................................V................----------------...................................
 Ph_awh_3Rn21
 Ph_awh_3Rn22
Ph_awh_3RCR09
Ph_awh_3RCR07
 Ph_awh_3Rn75
 Ph_awh_3Rn78
Ph_awh_3RCR10
Ph_awh_vsdP10
Ph_awh_vsdP15
    Consensus

              101    110       120       130       140       150       160       170       180       190       200
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_awh_ref VRGLSGGGCGAPTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGARCAKCCRAIGSTDWVRRARDRVYH
 Ph_awh_ORF21 ....................................................................................................
 Ph_awh_ORF22 ....................................................................................................
 Ph_awh_ORF20 .............................................................-......................................
  Ph_awh5RCR1 ....................................................................................................
  Ph_awh5RCR2 ....................................................................................................
  Ph_awh5RCR5 ....................................................................................................
  Ph_awh5RCR6 ....................................................................................................
Ph_awh_5Rn121 ....................................................................................................
Ph_awh_5Rn122 ....................................................................................................
Ph_awh_5Rn150 ....................................................................................................
 Ph_awh_3Rn21
 Ph_awh_3Rn22
Ph_awh_3RCR09
Ph_awh_3RCR07
 Ph_awh_3Rn75
 Ph_awh_3Rn78
Ph_awh_3RCR10
Ph_awh_vsdP10
Ph_awh_vsdP15
    Consensus

              201    210       220       230       240       250       260       270       280       290       300
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_awh_ref LACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRKSKRVRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVT
 Ph_awh_ORF21 ....................................................................................................
 Ph_awh_ORF22 ....................................................................................................
 Ph_awh_ORF20 ..R...............L.................................................................................
  Ph_awh5RCR1 ...............................................................................
  Ph_awh5RCR2 ...............................................................................
  Ph_awh5RCR5 ...............................................................................
  Ph_awh5RCR6 ...............................................................................
Ph_awh_5Rn121 ...............................................................................
Ph_awh_5Rn122 ...............................................................................
Ph_awh_5Rn150 ...............................................................................
 Ph_awh_3Rn21                                                                                         ............
 Ph_awh_3Rn22                                                                                         ............
Ph_awh_3RCR09                                                                                         ............
Ph_awh_3RCR07                                                                                         ............
 Ph_awh_3Rn75                                                                                         ............
 Ph_awh_3Rn78                                                                                         ............
Ph_awh_3RCR10                                                                                         T...........
Ph_awh_vsdP10                                                                 ....................................
Ph_awh_vsdP15                                                                 ....................................
    Consensus                                                                                         ............

              301    310       320       330       340       350       360       37072
              |--------+---------+---------+---------+---------+---------+---------+-|
   Ph_awh_ref QVWFQNSRARQKKHMTSKKHHQGLERGLHDITGSLARTIDLHFMYSFRPQSAEPEDLSSPENSLMTLMPGEA
 Ph_awh_ORF21 ..............................................................
 Ph_awh_ORF22 ..............................................................
 Ph_awh_ORF20 ...................................................V..........
  Ph_awh5RCR1
  Ph_awh5RCR2
  Ph_awh5RCR5
  Ph_awh5RCR6
Ph_awh_5Rn121
Ph_awh_5Rn122
Ph_awh_5Rn150
 Ph_awh_3Rn21 ........................................................................
 Ph_awh_3Rn22 ........................................................................
Ph_awh_3RCR09 ........................................................................
Ph_awh_3RCR07 ........................................................................
 Ph_awh_3Rn75 ........................................................................
 Ph_awh_3Rn78 ........................................................................
Ph_awh_3RCR10 ........................................................................
Ph_awh_vsdP10
Ph_awh_vsdP15
    Consensus ........................................................................
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.7_Ph_awh]

[PhD_A2.7.2_Ph_awh_alignments]

**“difference_alignment_Ph_awh_all_bp_sequences.gif”**



# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.8_Ph_six3]

### [PhD_A2.8.1_Ph_six3_sequences]

### **“Ph_six3_general_read_me.txt”**

```
General aspects:

(i)       All sequences are in FASTA format.

(ii)      [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)     Comments are preceded by an apostrophe (')

(iv)      All sequences without olignonucleotide sequence are indicated (e.g. "w/o 2013, 2016"). In these cases, only novel
Parhyale sequence is presented.


Parhyale hawaiensis six3 terminology:


(i)       In general, sequence denotation begins with: ">Ph_six3_"

(ii)      The above (i) is followed by: "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                         "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                         "ORF"               for coherent ORF sequences
                                         "cDNA"              for complete cDNA sequences

                                         "ref"               if reference sequence is presented
                                         "HD"                if homeobox/homeodomain fragment is presented
                                         "SD"                if Six-domain (encoding) fragment is presented

(iii)     The above (ii) is followed by clone id.


Parhyale hawaiensis six3 synopsis:

cDNA:     3655 bp
5'UTR:    77 bp (1-77)
ORF:    1632 bp (78-1709), 343 aa
3'UTR:  1946 bp (1710-3655)
SD:      357 bp (195-551), 119 aa
HD:      180 bp (552-731), 60 aa

Parhyale six3 has been annotated as: "Parhyale hawaiensis optix; GenBank: EU908055.1"

>Ph_six3_SD [bp]
TTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGTTAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAA
GCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCTTAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCC
CACTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCGCGCACGATATGGGAT

>Ph_six3_HD [bp]
GGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCACGCGGTCCCTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCTCACCCCCACG
CAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGGGATAGAGCCGCCGCT

>Ph_six3_ref [bp]
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCG
CACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGT
TAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCT
TAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCA
GGAAAAAGTTCCCGTTTCCGCGCACGATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCACGCGGTCCCTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGC
GGGAACTGGCTCAGGCCACGACCCTCACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGGGATAGAGCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCT
CTGCCAAACTGGAGACGGGGCCGCGGTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGCTCTCCGACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGT
CTCCGCTCTCCAAACTGGGCCAGCCGCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAACACGGGGGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGA
AGTGCAGCTCTGCGGGGAGTTCGTTTTCCCTGAGTGAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTCAGCAGCCCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTT
CCTCACAAGAAATTTTTCTGCCCGCCTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCCGACGGGGACCACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTG
CATCAGCGCTGCCCCCCAGACTGCCTCTAGCGTTCCCTCCGCACCTCAGCACCTGGTCCCAATACGCTCCATTCCCCACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTC
TGTCCATCGACTCGCTCATTAGCCCGGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCACACCGTTACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCG
CGCAGGGTGACGGGCCCTCGCTGCGGTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCAGGCGACACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAAC
TAGAGTTCGCTCCCTCTACCGAGACCTCATGATTACATCTCAACATTTTTATTGTCTCCTATTAGCTCTCGATATTTATGATGTACATTAATTGTGTATACTTAAAGTTATAGAATTATTAGACTAGCA
GGAAATATTTCCTTAATTCTCTTGATCTTTATACATTTACAGTCATTATTTTTTTCCCTTGAATTTATTAGCGCAGGATGAAGAACTTGCTGTGGTATTTTTCTTAGTGAAATGTACAGGGAAGATAAA
AATGAACAAATTGTCTCGAGCCTAAATATGAGAGTCTACGGAGCATTTAGCTCCAGACAGCATTACCCATATCTAATTGCATCTACCTTATTTATCAAAGGCTTATATTTCACTTTTTCTTAGTAACTG
GAAACAATCTTCAGTAGGCAATCGAGAATATTTTTGTTAGTTAGTATTAAAACTGCAGGCTTCCATGTCTCTGAGAGGGTATGAAAGAACCTAAAGTATAGGGAATTTATTCCTGTGTGACAAAATTAT
TATCTCATGACATGAATCTTATGATCAGTTTATGTAACGCTTTGTGTACACACACGTGTTGTTATTCTCTCGGTTGAACCGCAGGTGGTGGGTGCATTATTTTCTTTTTTCCACACGACCGAAGTTTGA
GTCGGTTGCTTGGCCTCTCTGCCTCATGCAGAGGTCACAGCATTAAGTAGATTCAGCAGTGGGAGGTATGACAACTTAGGGCTCGTTCCATTTGCTGGTCGATGTAGAACTCAGCGAAAATCAATCGAG
CTTGACTTGTCTGGAACTGCAGATAGAATCTTTGGCTACGTCACTTTGAACCGCGTTGCAGGTGTTAAGCCCGGTGCCCACATAACTACTTTCGTAGCCGCTAGCAGGCTAGAGACAACTGAACTAATT
CGAGGGCGTAGTCAAGGTCGTGACTTCAGCATGAACTCTGTGAACATCAAGCTATAATTCTGAAAATATCAAGCTTTTCTTACTTAAACTGCTCATAATTGGGAGCGATTGTTGAAAGTAGCTACACCA
TTGCCGCTTGCAGTGGTGTAAAAATTGCAGTAGAGACTGAGATATTCGACAATTTTTTTAGACAGGTAAATTTTGCGCAAGCTGAAGCCGAGCCGCGGCTAAGTAAGCGGATTAGTCGAGTTGCTTCT
AAGCTGTTAGCAGCAACGAAACTTGTTTCGTGGGTACCGGGCTTTAGAAACAAGGCGATTATGTTGAGCTGACTGCAAACTCCATTTTGTTGAGTAAGATTAAGTTCTTGTGAAGTCTCAAAGGTAATC
GATTAAGGGTCAACACCTGAAATTTGAATTAACCTATTAAAGTCAGCCCACATGGCTGCAGTTTCGGACAAGTTAAGCCTCAAAAACCTATTTTGGCCCATCAAACAAACGCGCTCTTAGTGGACTACT
GAACTAATCCTAACTTGGCACGCCACGGGTACTGATATTGGTACAGTCGGCCTGATGAGGCCATAGCAAGCAGTCCGGCAGCAAATGCCACTGTCGCGCGGGGTGCATGTAGCCATGTGACGGTACCGA
```

TGTGGGGCTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGTTAGGATATCATGATACATATGTGTGTGCTAGGTCGTCTGA
TTTGATTGGGTAATTAACTGCCGCGTGATTGATTGATTGCTTGATTGCCGTGTCAAAGAGGGATAATAATAATGGTGCAAATAATTAATACTTTTTTGTGCAAAAATAGTTCGCCCACAGATAGCTGCT
CAGGTGGGCAAAGACTGGGTACGAATCAAGTCTCAACTCGGTTGACCAACTCTGAAGTTGTTGCCATGACCAGCACCGGACCTAAACTGACTCTAAAATTCTCTTCTAACCCCAGCCGAACGTTGCCAT
TCGGATCCCCCCCACCCAAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_six3_SD [aa]
FTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQEAERLRGRPLGPVDKYRVRKKFPFPRTIWD

>Ph_six3_HD [aa]
GEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAA

>Ph_six3_ref [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSPGADDDEDDLINPGSP
TPFEDDVSEDEVSLGQQSPLSKLGQPPLAAFDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDG
DHSALRLPLPGALFSPFASALPPRLPLAFPPHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGD
TSISDQQPRPHIFQPIKLEFAPSTETS&

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.8_Ph_six3]

[PhD_A2.8.1_Ph_six3_sequences]

**“Ph_six3_cDNA_sequences.txt”**

```
>Ph_six3_ref [bp] 'artificial consensus cDNA, based on Ph_six3_cDNA09; added 5' 25bp, from 5R02; altered 754G>A, 2080G>A,
2886G>A; added 3' 58bp from 3R08
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCG
CACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGT
TAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCT
TAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCA
GGAAAAAGTTCCCGTTTCCGCGCACGATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCACGCGGTCCCTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGC
GGGAACTGGCTCAGGCCACGACCCTCACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGGGATAGAGCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCT
CTGCCAAACTGGAGACGGGGCCGCGGTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGCTCTCCGACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGT
CTCCGCTCTCCAAACTGGGCCAGCCGCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAACACGGGGGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGA
AGTGCAGCTCTGCGGGGAGTTCGTTTTCCCTGAGTGAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTCAGCAGCCCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTT
CCTCACAAGAAATTTTTCTGCCCGCCTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCCGACGGGGACCACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTG
CATCAGCGCTGCCCCCCAGACTGCCTCTAGCGTTCCCTCCGCACCTCAGCACCTGGTCCCAATACGCTCCATTCCCCACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTC
TGTCCATCGACTCGCTCATTAGCCCGGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCACACCGTTACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCG
CGCAGGGTGACGGGCCCTCGCTGCGGTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCAGGCGACACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAAC
TAGAGTTCGCTCCCTCTACCGAGACCTCATGATTACATCTCAACATTTTTATTGTCTCCTATTAGCTCTCGATATTTATGATGTACATTAATTGTGTATACTTAAAGTTATAGAATTATTAGACTAGCA
GGAAATATTTCCTTAATTCTCTTGATCTTTATACATTTACAGTCATTATTTTTTTCCCTTGAATTTATTAGCGCAGGATGAAGAACTTGCTGTGGTATTTTTCTTAGTGAAATGTACAGGGAAGATAAA
AATGAACAAATTGTCTCGAGCCTAAATATGAGAGTCTACGGAGCATTTAGCTCCAGACAGCATTACCCATATCTAATTGCATCTACCTTATTTATCAAAGGCTTATATTTCACTTTTTCTTAGTAACTG
GAAACAATCTTCAGTAGGCAATCGAGAATATTTTTGTTAGTTAGTATTAAAACTGCAGGCTTCCATGTCTCTGAGAGGGTATGAAAGAACCTAAAGTATAGGGAATTTATTCCTGTGTGACAAAATTAT
TATCTCATGACATGAATCTTATGATCAGTTTATGTAACGCTTTGTGTACACACACGTGTTGTTATTCTCTCGGTTGAACCGCAGGTGGTGGGTGCATTATTTTCTTTTTTCCACACGACCGAAGTTTGA
GTCGGTTGCTTGGCCTCTCTGCCTCATGCAGAGGTCACAGCATTAAGTAGATTCAGCAGTGGGAGGTATGACAACTTAGGGCTCGTTCCATTTGCTGGTCGATGTAGAACTCAGCGAAAATCAATCGAG
CTTGACTTGTCTGGAACTGCAGATAGAATCTTTGGCTACGTCACTTTGAACCGCGTTGCAGGTGTTAAGCCCGGTGCCCACATAACTACTTTCGTAGCCGCTAGCAGGCTAGAGACAACTGAACTAATT
CGAGGGCGTAGTCAAGGTCGTGACTTCAGCATGAACTCTGTGAACATCAAGCTATAATTCTGAAAATATCAAGCTTTTCTTACTTAAACTGCTCATAATTGGGAGCGATTGTTGAAAGTAGCTACACCA
TTGCCGCTTGCAGTGGTGTAAAAATTGCAGTAGAGACTGAGATATTCGACAATTTTTTTAGACAGGTAAATTTTGCGCAAGCTGAAGCCGAGCCGCGGCTAAGTAAGCGGATTAGTCGAGTTGCTTCT
AAGCTGTTAGCAGCAACGAAACTTGTTTCGTGGGTACCGGGCTTTAGAAACAAGGCGATTATGTTGAGCTGACTGCAAACTCCATTTTGTTGAGTAAGATTAAGTTCTTGTGAAGTCTCAAAGGTAATC
GATTAAGGGTCAACACCTGAAATTTGAATTAACCTATTAAAGTCAGCCCACATGGCTGCAGTTTCGGACAAGTTAAGCCTCAAAAACCTATTTTGGCCCATCAAACAAACGCGCTCTTAGTGGACTACT
GAACTAATCCTAACTTGGCACGCCACGGGTACTGATATTGGTACAGTCGGCCTGATGAGGCCATAGCAAGCAGTCCGGCAGCAAATGCCACTGTCGCGCGGGGTGCATGTAGCCATGTGACGGTACCGA
TGTGGGGCTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGTTAGGATATCATGATACATATGTGTGTGCTAGGTCGTCTGA
TTTGATTGGGTAATTAACTGCCGCGTGATTGATTGATTGCTTGATTGCCGTGTCAAAGAGGGATAATAATAATGGTGCAAATAATTAATACTTTTTTGTGCAAAAATAGTTCGCCCACAGATAGCTGCT
CAGGTGGGCAAAGACTGGGTACGAATCAAGTCTCAACTCGGTTGACCAACTCTGAAGTTGTTGCCATGACCAGCACCGGACCTAAACTGACTCTAAAATTCTCTTCTAACCCCAGCCGAACGTTGCCAT
TCGGATCCCCCCCACCCAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_six3_cDNA09 [bp] 'w/o 2013, 2016
GCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCGCACCCCACGGAGGGGCCCCAGGGCC
CGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGTTAGGTAGATTTTTGTGGTCCCTCCC
AGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCTTAATTGAGTGCAACAGATTCACCCG
CGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCGCGCAC
GATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCACGCGGTCCCTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCT
CACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGGGATAGAGCCGCCGCTATCAAGAACAGACTTCACCAGCGGAGTTCGTCAATATCGTCCTCTGCCAAACTGGAGACGGGGCCGCG
GTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGCTCTCCGACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGTCTCCGCTCTCCAAACTGGGCCAGCC
GCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAACACGGGGGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGAAGTGCAGCTCTGCGGGGAGTTCGTT
TTCCCTGAGTGAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTCAGCAGCCCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTTCCTCACAAGAAATTTTTCTGCCCGC
CTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCCGACGGGGACCACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTGCATCAGCGCTGCCCCCCAGACTGCC
TCTAGCGTTCCCTCCGCACCTCAGCACCTGGTCCCAATACGCTCCATTCCCCACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTCTGTCCATCGACTCGCTCATTAGCCC
GGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCACACCGTTACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCGCGCAGGGTGACGGGCCCTCGCTGCG
GTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCAGGCGACACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAACTAGAGTTCGCTCCCTCTACCGAGAC
CTCATGATTACATCTCAACATTTTTATTGTCTCCTATTAGCTCTCGATATTTATGATGTACATTAATTGTGTATACTTAAAGTTATAGAATTATTAGACTAGCAGGAAATATTTCCTTAATTCTCTTGA
TCTTTATACATTTACAGTCATTATTTTTTTCCCTTGAATTTATTAGCGCAGGATGAAGAACTTGCTGTGGTATTTTTCTTAGTGAAATGTACAGGGAAGATAAAAATGAACAAATTGTCTCGAGCCTAA
ATATGAGAGTCTACGGAGCATTTAGCTCCAGACAGCATTACCCATATCTAATTGCATCTACCTTATTTATCAAAGGCTTATATTTCACTTTTTCTTAGTAACTGGAAACAATCTTCAGTGGGCAATCGA
GAATATTTTTGTTAGTTAGTATTAAAACTGCAGGCTTCCATGTCTCTGAGAGGGTATGAAAGAACCTAAAGTATAGGGAATTTATTCCTGTGTGACAAAATTATTATCTCATGACATGAATCTTATGAT
CAGTTTATGTAACGCTTTGTGTACACACACGTGTTGTTATTCTCTCGGTTGAACCGCAGGTGGTGGGTGCATTATTTTCTTTTTTCCACACGACCGAAGTTTGAGTCGGTTGCTTGGCCTCTCTGCCTC
ATGCAGAGGTCACAGCATTAAGTAGATTCAGCAGTGGGAGGTATGACAACTTAGGGCTCGTTCCATTTGCTGGTCGATGTAGAACTCAGCGAAAATCAATCGAGCTTGACTTGTCTGGAACTGCAGATA
GAATCTTTGGCTACGTCACTTTGAACCGCGTTGCAGGTGTTAAGCCCGGTGCCCACATAACTACTTTCGTAGCCGCTAGCAGGCTAGAGACAACTGAACTAATTCGAGGGCGTAGTCAAGGTCGTGACT
TCAGCATGAACTCTGTGAACATCAAGCTATAATTCTGAAAATATCAAGCTTTTCTTACTTAAACTGCTCATAATTGGGAGCGATTGTTGAAAGTAGCTACACCATTGCCGCTTGCAGTGGTGTAAAAAT
TGCAGTAGAGACTGAGATATTCGACAATTTTTTTAGACAGGTAAATTTTGCGCAAGCTGAAGCCGAGCCGCGGCTAAGTAAGCGGATTAGTCGAGTTGCTTCTAAGCTGTTAGCAGCAACGAAACTTG
TTTCGTGGGTACCGGGCTTTAGGAACAAGGCGATTATGTTGAGCTGACTGCAAACTCCATTTTGTTGAGTAAGATTAAGTTCTTGTGAAGTCTCAAAGGTAATCGATTAAGGGTCAACACCTGAAATTT
GAATTAACCTATTAAAGTCAGCCCACATGGCTGCAGTTTCGGACAAGTTAAGCCTCAAAAACCTATTTTGGCCCATCAAACAAACGCGCTCTTAGTGGACTACTGAACTAATCCTAACTTGGCACGCCA
CGGGTACTGATATTGGTACAGTCGGCCTGATGAGGCCATAGCAAGCAGTCCGGCAGCAAATGCCACTGTCGCGCGGGGTGCATGTAGCCATGTGACGGTACCGATGTGGGGCTAGCCATGTGGCTGCAC
CGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGTTAGGATATCATGATACATATGTGTGTGCTAGGTCGTCTGATTTGATTGGGTAATTAACTGCCGCG
TGATTGATTGATTGCTTGATTGCCGTGTCAAAGAGGGATAATAATAATGGTGCAAATAATTAATACTTTTTTGTGCAAAAATAGTTCGCCCACAGATAGCTGCTCAGGTGGGCAAAGACTGGGTACGAA
TCAAGTCTCAACTCGGTTGACCAACTCTGAAGTTGTTGCCATGACCAGCACCGGACCTAAACTGACTCTAAAATTCTCTTCTAACCCCA

>Ph_six3_cDNA13 [bp] 'w/o 2013, 2016
GCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCACTCTCCTGGGCCCCGCACCCCACGGAGGGGCCCCAGGGCC
CGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGTTAGGTAGATTTTTGTGGTCCCTCCC
AGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCTTAATTGAGTACAACAGATTCACCCG
CGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCGCGCAC
GATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCACGCGGTCCCTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCGGGAAGCGGGAACTGGCTCAGGCCACGACCCT
CACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGGGATAGAGCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCTCTGCCAAACTGGAGACGGGGCCGCG
GTCGTCTGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGCTCTCCGACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGTCTCCGCTCTCCAAACTGGGCCAGCC
ACCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAACACGGGGGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGAAGTGCAGCTCTGCGGGGAGTTCGTT
TTCCCTGAGTGAAGGTGCCTCGTCTGTCACTTCCAACAGCGCCTTCAGCAGCCCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTTCCTCACAAGAAATTTTTCTGCCCGC
CTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCCGACGGGGACCACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTGCATCAGCGCTGCCCCCCAGACTGCC
TCTAGCGTTCCCTCCGCACCTCAGCACCTGGTCCCAATACGCTCCATTCCCCACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTCTGTCCATCGACTCGCTCATTAGCCC
GGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCACACCGTTACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCGCGCAGGGTGACGGGCCCTCGCTGCG
```

```
GTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCAGGCGACACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAACTAGAGTTCGCTCCCTCTACCGAGAC
CTCATGATTACATCTCAACATTTTTATTGTCTCCTATTAGCTCTCGATATTTATGATGTACATTAATTGTGTATACTTAAAGTTATAGAATTATTAGACTAGCAGGAAATATTTCCTTAATTCTCTTGA
TCTTTATACATTTACAGTCATTATTTTTTTCCCTTGAATTTATTAGCGCAGGATGAAGAACTTGCTGTGGTATTTTTCTTAGTGAAATGTACAGGGAAGATAAAAATGAACAAATTGTCTCGAGCCTAA
ATATGAGAGTCTACGGAGCATTTAGCTCCAGACAGCATTACCCATATCTAATTGCATCTACCTTATTTATCAAAGGCTTATATTTCACTTTTTCTTAGTAACTGGAAACAATCTTCAGTAGGCAATCGA
GAATATTTTTGTTAGTTAGTATTAAAACTGCAGGCTTCCATGTCTCTGAGAGGGTATGAAAGAACCTAAAGTATAGGGAATTTATTCCTGTGTGACAAAATTATTATCTCATGACATGAATCTTATGAT
CAGTTTATGTAACGCTTTGTGTACACACACGTGTTGTTATTCTCTCGGTTGAACCGCAGGTGGTGGGTGCATTATTTTCTTTTTTCCACACGACCGAAGTTTGAGTCGGTTGCTTGGCCTCTCTGCCTC
ATGCAGAGGTCACAGCATTAAGTAGATTCAGCAGTGGGAGGTATGACAACTTAGGGCTCGTTCCATTTGCTGGTCGATGTAGAACTCAGCGAAAATCAATCGAGCTTGACTTGTCTGGAACTGCAGATA
GAATCTTTGGCTACGTCACTTTGAACCGCGTTGCAGGTGTTAAGCCCGGTGCCCACATAACTACTTTCGTAGCCGCTAGCAGGCTAGAGACAACTGAACTAATTCGAGGGCGTAGTTAAGGTCGTGACT
TCAGCATGAACTCTGTGAACATCAAGCTATAATTCTGAAAATATCAAGCTTTTCTTACTTAAACTGCTCATAATTGGGAGCGATTGTTGAAAGTAGCTACACCATTGCCGCTTGCAGTGGTGTAAAAAT
TGCAGTAGAGACTGAGATATTCGACAATTTTTTTAGACAGGTAAACTTTGCGCAAGCTGAAGCCGAGCCGCGGCTAAGTAAGCGGATTAGTCGAGTTGCTTCTAAGCTGTTAGCAGCAACGAAACTTG
TTTCGTGGGTACCGGGCTTTAGAAACAAGGCGATTATGTTGAGCTGACTGCAAACTCCATTTTGTTGAGTAAGATTAAGTTCTTGTGAAGTCTCAAAGGTAATCGATTAAGGGTCAACACCTGAAATTT
GAATTAACCTATTAAAGTCAGCCCACATGGCTGCAGTTTCGGACAAGTTAAGCCTCAAAAACCTATTTTGGCCCATCAAACGAACGCGCTCTTAGTGGACTACTGAACTAATCCTAACTTGGCACGCCA
CGGGTACTGATATTGGTACAGTCGGCCTGATGAGGCCATAGCAAGCAGTCCGGCAGCAAATGCCACTGTCGCGCGGGGTGCATGTAGCCATGTGACGGTACCGATGTGGGGCTAGCCATGTGGCTGCAC
CGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGTTAGGATATCATGATACATATGTGTGTGCTAGGTCGTCTGATTTGATTGGGTAATTAACTGCCGCG
TGATTGATTGATTGCTTGATTGCCGTGTCAAAGAGGGATAATAATAATGGTGCAAATAATTAATACTTTTTTGTGCAAAAATAGTTCGCCCACAGATAGCTGCTCAGGTGGGCAAAGACTGGGTACGAA
TCAAGTCTCAACTCGGTTGACCAACTCTGAAGTTGTTGCCATGACCAGCACCGGACCTAAACTGACTCTAAAATTCTCTTCTAACCCCA

>Ph_six3_ORF10 [bp] 'w/o 2014, 2015
ATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCGCACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACA
CAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGTTAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGT
GCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCTTAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAG
GCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCGCGCACGATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCACGCGGTCC
CTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCTCACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGGGATAGA
GCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCTCTGCCAAACTGGAGACGGGGCCGCGGTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGCTCTCCG
ACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGTCTCCGCTCTCCAAACTGGGCCAGCCGCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAACACGGG
GGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGAAGTGCAGCTCTGCGGGGAGTTCGTTTTCCCTGAGTGAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTCAGCAGC
CCTTACCCTCCCACCAGCCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTTCCTCACAAGAAATTTTTCTGCCCGCCTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGGATAAACCCCGACGGG
GACCACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTGCATCAGCGCTGCCCCCCAGACTGCCTCTAGCGTTCCCTCCGCACCTCAGCACCTGGTCCCAATACGCTCCATTCCCC
ACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTCTGTCCATCGACTCGCTCATTAGCCCGGCACGAAATAAGAGAGAAGCTCCTCCGCGCTCTCCTTCCCCCCACACCGTT
ACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCGCGCAGGGTGACGGGCCCTCGCTGCGGTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCAGGCGAC
ACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAACTAGAGTTCGCTCCCTCTACCGAGACCTCATGATTACAT

>Ph_six3_ORF11 [bp] 'w/o 2014, 2015
ATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCGCACCCCACGGAGGGGCCCCAGGGTCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACA
CAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGTTAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGT
GCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCTTAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAG
GCGGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCGCGCACGATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCACGCGGTCC
CTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCTCACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGGGATAGA
GCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCTCTGCCAAACTGGAGACGGGGCCGCGGTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGCTCTCCG
ACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGTCTCCGCTCTCCAAACTGGGCCAGCCGCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAACACGGG
GGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGAAGTGCAGCTCTGCGGGGAGTTCGTTCTCCCTGAGTGAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTCAGCAGC
CCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTTCCTCACAAGAAATTTTTCTGCCCGCCTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCCGACGGG
GACCACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTGCATCAGCGCTGCCCCCCAGACTGCCTCTAGCGTTCCCTCCGCACCTCAGCGCCTGGTCCCAATACGCTCCATTCCCC
ACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTCTGTCCATCGACTCGCTCATTAGCCCGGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCACACCGTT
ACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCGCGCAGGGTGACGGGCCCTCGCTGCGGTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCAGGCGAC
ACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAACTAGAGTTCGCTCCCTCTACCGAGACCTCATGATTACAT

>Ph_six3_ORF12 [bp] 'w/o 2014, 2015
ATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCGCACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACA
CAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGTTAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTTATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGT
GCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCTTAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAG
GCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCGCGCACGATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCACGCGGTCC
CTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCTCACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGGGATAGA
GCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCTCTGCCAAACTGGAGACGGGGCCGCGGTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGCTCTCCG
ACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGTCTCCGCTCTCCAAACTGGGCCAGCCGCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAACACGGG
GGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGAAGTGCAGCTCTGCGGGGAGTTCGTTTTCCCTGAGTGAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTCAGCAGC
CCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTTCCTCACAAGAAATTTTTCTGCCCGCCTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCCGACGGG
GACCACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTGCATCAGCGCTGCCCCCCAGACTGCCTCTAGCGTTCCCTCCGCACCTCAGCACCTGGTCCCAATACGCTCCATTCCCC
ACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTCTGTCCATCGACTCGCTCATTAGCCCGGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCACACCGTT
ACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCGCGCAGGGTGACGGGCCCTCGCTGCGGTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCAGGCGAC
ACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAACTAGAGTTCGCTCCCTCTACCGAGACCTCATGATTACAT

>Ph_six3_5R01 [bp] 'w/o 2006
TGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCACACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCT
TCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGTTAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAG
CAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCTTAATTGAGTGCAACAGATTCACCCGCGCGTCGTGCCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCC
ACTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCACGCACGATATGGGACGGGGAGCAGAA

>Ph_six3_5R02 [bp] 'w/o 2006
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCG
CACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTGTCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGT
TAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATGCGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCTT
TAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGATTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGCGGCCTTTGGGACCCGTCGACAAGTACCGCGTCA
GGAAAAAGTTCCCGTTTCCTCGCACGATATGGGATGGGGAGCAGAA

>Ph_six3_5R04 [bp] 'w/o 2006
GCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCGCACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTT
CACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGTTAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGC
AGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCTTAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCA
CTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCGCGCACGATATGGGATGGGGAGCAGAA

>Ph_six3_5R05 [bp] 'w/o 2006
GACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCGCACCCCACGGAGGGGCCCCAGG
GCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTGTCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGTTAGGTAGATTTTTGTGGTCCCT
CCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCTTTAATTGAGTGCAACAGATTCAC
CCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGCGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCTCG
CACGATATGGGGTGGGGAGCAGAA

>Ph_six3_5R06 [bp] 'w/o 2006
```

```
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCG
CACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGT
TAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCT
TAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCA
GGAAAAAGTTCCCGTTTCCGCGCACGATATGGGATGGGGAGCAGAA

>Ph_six3_5R16 [bp] 'w/o 2007
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCG
CACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGT
TAGATAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCT
TAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGTAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTT

>Ph_six3_5R17 [bp] 'w/o 2007
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCG
CACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTGTCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAGGCGGTGACATTGAAAGGT
TAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCTT
TAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGCGGCCTTT

>Ph_six3_5R18 [bp] 'w/o 2007
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAACATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCG
CACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCCTCAAACGTGTCTTTACCTTCCCTGTCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGGT
TAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTTGTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCTT
TAATTGAGTGCAACAGATTCACCCACGCGTCGTACCCTCGGTTGCAGGCCCTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGCGGCCTTT

>Ph_six3_5R19 [bp] 'w/o 2008
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGTGTGCCCTGCAGCCCAA

>Ph_six3_5R23 [bp] 'w/o 2008
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAA

>Ph_six3_5R24 [bp] 'w/o 2008, removed 5' G
GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAA

>Ph_six3_vsdP13 [bp] 'w/o deg primer
CCAGGAGGCAGAGCGCCTCAGAGGGCGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCTCGCACGATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCAC
GCGGTCCCTCCTGAGGGAGGCCTACCTGCAGGACCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCGCAGGCCACGACCCTTACC

>Ph_six3_vsdP17 [bp] 'w/o deg primer
CCAGGAGGCAGAGCGCCTCAGAGGGCGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAGAGTTCCCGTTTCCTCGCACGATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCAC
GCGGTCCCTCCTGAGGGAGGCCTACCTGCAGGACCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCGCAGGCCACGACCCTTACC

>Ph_six3_vsdP19 [bp] 'w/o deg primer
CCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTCCCGTTTCCGCGCACGATATGGGACGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCAC
GCGGTCCCTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCTCACC

>Ph_six3_3R08 [bp] 'w/o 2001
CGGTCCCTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCTCACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGG
GATAGAGCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCTCTGCCAAACTGGAGACGGGGCCGCGGTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGC
TCTCCGACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGTCTCCGCTCTCCAAACTGGGCCAGCCGCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCCCAA
CACGGGGGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGAAGTGCAGCTCTGCGGGGAGTTCGTTTTCCCTGAGTGAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTC
AGCAGCCCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTTCCTCACAAGAAATTTTTCTGCCCGCCTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCC
GACGGGGACCACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTGCATCAGCGCTGCCCCCCAGACTGCCTCTAGCGTTCCCTCCGCACCTCAGCACCTGGTCCCAATACGCTCCA
TTCCCCACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTCTGTCCATCGACTCGCTCATTAGCCCGGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCAC
ACCGTTACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCGCGCAGGGTGACGGGCCCTCGCTGCGGTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCA
GGCGACACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAACTAGAGTTCGCTCCCTCTACCGAGACCTCATGATTACATCTCAACATTTTTATTGTCTCCTATTAGCTCTCG
ATATTTATGATGTACATTAATTGTGTATACTTAAAGTTATAGAATTATTAGACTAGCAGGAAATATTTCCTTAATTCTCTTGATCTTTATACATTTACAGTCATTATTTTTTTCCCTTGAATTTATTAG
CGCAGGATGAAGAACTTGCTGTGGTATTTTTCTTAGTGAAATGTACAGGGAAGATAAAAATGAACAAATTGTCTCGAGCCTAAATATGAGAGTCTACGGAGCATTTAGCTCCAGACAGCATTACCCATA
TCTAATTGCATCTACCTTATTTATCAAAGGCTTATATTTCACTTTTTCTTAGTAACTGGAAACAATCTTCAGTAGGCAATCGAGAATATTTTTGTTAGTTAGTATTAAAACTGCAGGCTTCCATGTCTC
TGAGAGGGTATGAAAGAACCTAAAGTATAGGGAATTTATTCCTGTGTGACAAAATTATTATCTCATGACATGAATCTTATGATCAGTTTATGTAACGCTTTGTGTACACACACGTGTTGTTATTCTCTC
GGTTGAACCGCAGGTGGTGGGTGCATTATTTCTTTTTTCCACACGACCGAAGTTTGAGTCGGTTGCTTGGCCTCTCTGCCTCATGCAGAGGTCACAGCATTAAGTAGATTCAGCAGTGGGAGGTATGA
CAACTTAGGGCTCGTTCCATTTGCTGGTCGATGTAGAACTCAGCGAAAATCAATCGAGCTTGACTTGTCTGGAACTGCAGATAGAATCTTTGGCTACGTCACTTTGAACCGCGTTGCAGGTGTTAAGCC
CGGTGCCCACATAACTACTTTCGTGGCCGCTAGCAGGCTAGAGACAACTGAACTAATTCGAGGGCGTAGTCAAGGTCGTGACTTCAGCATGAACTCTGTGAACATCAAGCTATAATTCTGAAAATATCA
AGCTTTTCTTACTTAAACTGCTCATAATTGGGAGCGATTGTTGAAAGTAGCTACACCATTGCCACTTGCAGTGGTGTAAAAATTGCAGTAGAGACTGAGATATTCGACAATTTTTTTAGACAGGTAAAT
TTTGCGCAAGCTGAAGCCGAGCCGCGGCTAAGTAAGCGGATTGGTCGAGTTGCTTCTGAGCTGTTAGCAGCAACGAAACTTGTTTCGTGGGTACCGGGCTTTAGAAACAAGGCGATTATGTTGAGCTGA
CTGCAAACTCCATTTTGTTGAGTAAGATTAAGTTCTTGTGAAGTCTCAAAGGTAATCGATTAAGGGTCAACACCTGAAATTTGAATTAACCTATTAAAGTCAGCCCACATGGCTGCAGTTTCGGACAAG
TTAAGCCTCAAAAACCTATTTTGGCCCATCAAACAAACGCGCTCTTAGTGGACTACTGAACTAATCCTAACTTGGCACGCCACGGGTACTGATATTGGTACAGTCGGCCTGATGAGGCCATAGCAAGCA
GTCCGGCAGCAAATGCCACTGTCGCGCGGGGTGCATGTAGCCATGTGACGGTACCGATGTGGGGCTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATG
TGGCTGCACCGATGTGTTAGGATATCATGATACATATGTGTGTGCTGGGTCGTCTAGTTTGGTTGGGTAATTAACTGCCGCGTGATTGATTGATTGCTTGGTTGCCGTGTCGAAGAGGAATAATAGTAA
TGATGCAAATGATTGATACTTTTTTGTGCAAAAATAGTTCGCCCACAGATAGCTGCTCGGGTGGACAAAGACTGGGTACGAATCAAGTCTCAACTCGGTTGACCAACTCTGAAGTTGTTGCCATGACCA
GCACCGGACCTAAACTGACTCTAAAATTCTCTTCTAACCCCAGCCGAACGTTGCCATTCGGATCCCCCCCACCCAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_six3_3R09 [bp] 'w/o 2001
CGGTCCCTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCTCACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGG
GATAGAGCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCTCTGCCAAACTGGAGACGGGGCCGCGGTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGC
TCTCCGACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGTCTCCGCTCTCCAAACTGGGCCAGCCGCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAA
CACGGGGGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGAAGTGCAGCTCTGTGGGGAGTTCGTTTTCCCTGAGTGAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTC
AGCAGCCCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTTCCTCACAAGAAATTTTTCTGCCCGCCTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCC
GACGGGGACCACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTGCATCAGCGCTGCCCCCCAGACTGCCTCTAGCGTTCCCTCCGCACCTCAGCACCTGGTCCCAATACGCTCCA
TTCCCCACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTCTGTCCATCGACTCGCTCATTAGCCCGGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCAC
ACCGTTACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCGCGCAGGGTGACGGGCCCTCGCTGCGGTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCA
GGCGACACCGGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAACTAGAGTTCGCTCCCTCTACCGAGACCTCATGATTACATCTCAACATTTTTATTGTCTCCTATTAGCTCTCG
ATATTTATGATGTACATTAATTGTGTATACTTAAAGTTATAGAATTATTAGACTAGCAGGAAATATTTCCTTAATTCTCTTGATCTTTATACATTTACAGTCATTATTTTTTTCCCTTGAATTTATTAG
CGCAGGATGAAGAACTTGCTGTGGTATTTTTCTTAGTGAAATGTACAGGGAAGATAAAAATGAACAAATTGTCTCGAGCCTAAATATGAGAGTCTACGGAGCATTTAGCTCCAGACAGCATTACCCATA
TCTAATTGCATCTACCTTATTTATCAAAGGCTTATATTTCACTTTTTCTTAGTAACTGGAAACAATCTTCAGTAGGCAATCGAGAATATTTTTGTTAGTTAGTATTAAAACTGCAGGCTTCCATGTCTC
TGAGAGGGTATGAAAGAACCTAAAGTATAGGGAATTTATTCCTGTGTGACAAAATTATTATCTCATGACATGAATCTTATGATCAGTTTATGTAACGCTTTGTGTACACACACGTGTTGTTATTCTCTC
GGTTGAACCGCAGGTGGTGGGTGCATTATTTCTTTTTTCCACACGACCGAAGTTTGAGTCGGTTGCTTGGCCTCTCTGCCTCATGCAGAGGTCACGGCATTAAGTAGATTCAGCAGTGGGAGGTATGA
CAACTTAGGGCTCGTTCCATTTGCTGGTCGATGTAGAACTCAGCGAAAATCAATCGAGCTTGACTTGTCTGGAACTGCAGATAGAATCTTTGGCTACGTCACTTTGAACCGCGTTGCAGGTGTTAAGCC
CGGTGCCCACATAACTACTTTCGTGGCCGCTAGCAGGCTAGAGACAACTGAACTAATTCGAGGGCGTAGTCAAGGTCGTGACTTCAGCATGAACTCTGTGAACATCAAGCTATAATTCTGAAAATATCA
AGCTTTTCTTACTTAAACTGCTCATAATTGGGAGCGATTGTTGAAAGTAGCTACACCATTGCCGCTTGCAGTGGTGTAAAAATTGCAGTAGAGACTGAGATATTCGACAATTTTTTTAGACAGGTAAAT
TTTGCGCAAGCTGAAGCCGAGCCGCGGCTAAGTAAGCGGATTGGTCGAGTTGCTTCTGAGCTGTTAGCAGCAACGAAACTTGTTTCGTGGGTACCGGGCTTTAGAAACAAGGCGATTATGTTGAGCTGA
```

```
CTGCAAACTCCATTTTGTTGAGTAAGATTAAGTTCTTGTGAAGTCTCAAAGGTAATCGATTAAGGGTCAACACCTGAAATTTGAATTAACCTATTAAAGTCAGCCCACATGGCTGCAGTTTCGGACAAG
TTAAGCCTCAAAAACCTATTTTGGCCCATCAAACAAACGCGCTCTTAGTGGACTACTGAACTAATCCTAACTTGGCACGCCACGGGTACTGATATTGGTACAGTCGGCCTGATGAGGCCATAGCAAGCA
GTCCGGCAGCAAATGCCACTGTCGCGCGGGGTGCATGTAGCCATGTGACGGTACCGATGTGGGGCTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATG
TGGCTGCACCGATGTGTTAGGATATCATGATACATATGTGTGTGCTGGGTCGTCTAGTTTGGTTGGGTAATTAACTGCCGCGTGATTGATTGATTGCTTGGTTGCCGTGTCGAAGAGGAATAATAGTAA
TGATGCAAATGATTGATACTTTTTTGTGCAAAAATAGTTCGCCCACAGATAGCTGCTCGGGTGGGCAAAGACTGGGTACGAATCAAGTCTCAACTCGGTTGACCAACTCTGAAGTTGTTGCCATGACCA
GCACCGGACCTAAACTGACTCTAAAATTCTCTTCTAACCCCAGCCGAACGTTGCCATTCGGATCCCCCCACCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_six3_3R10 [bp] 'w/o 2001
CGGTCCCTCCTGAGGGAGGCCTACCTGCAGGATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCTCACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGAAGACAGCGG
GATAGAGCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCTCTGCCAAACTGGAGACGGGGCCGCGGTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGC
TCTCCGACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTAAGTCTTGGGCAGCAGTCTCCGCTCTCCAAACTGGGCCAGCCGCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAA
CACGGGGGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGAAGTGCAGCTCTGCGGGGAGTTCGTTTTCCCTGAGTGAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTC
AGCAGCCCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAATCTTCCTCACAAGAAATTTTTCTGCCCGCCTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCC
GACGGGGACTACAGTGCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTGCATCAGCGCTGCCCCCCAGACTGCCTCTAGCGTTCCCTCCGCACCTCAGCACCTGGTCCCAATACGCTCCA
TTCCCCACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTCTGTCCATCGACTCGCTCATTAGCCCGGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCAC
ACCGTTACGTCCAAGACGGAACCTGAGGACCCTGGCATTGGTGCTGACCTGACCACCGCGCAGGGTGACGGGCCCTCGCTGCGGTCCCCCACACAGTCCCCAGTGGCCAAACGGCTCCACACAAGTCCA
GGCGACACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAACTAGAGTTCGCTCCCTCTACCGAGACCTCATGATTACATCTCAACATTTTTATTGTCTCCTATTAGCTCTCG
ATATTTATGATGTACATTAATTGTGTATACTTAAAGTTATAGAATTATTAGACTAGCAGGAAATATTTCCTTAATTCTCTTGATCTTTATACATTTACAGTCATTATTTTTTTCCCTTGAATTTATTAG
CGCAGGATGAAGAACTTGCTGTGGTATTTTTCTTAGTGAAATGTACAGGGAAGATAAAAATGAACAAATTGTCTCGAGCCTAAATATGAGAGTCTACGGAGCATTTAGCTCCAGACAGCATTACCCATA
TCTAATTGCATCTACCTTATTTATCAAAGGCTTATATTTCACTTTTTCTTAGTAACTGGAAACAATCTTCAGTAGGCAATCGAGAATATTTTTGTTAGTTAGTATTAAAACTGCAGGCTTCCATGTCTC
TGAGAGGGTATGAAAGAACCTAAAGTATAGGGAATTTATTCCTGTGTGACAAAATTATTATCTCATGACATGAATCTTATGATCAGTTTATGTAACGCTTTGTGTACACGCACGTGTTGTTATTCTCTC
GGTTGAACCGCAGGTGGTGGGTGCATTATTTCTTTTTTCCACACGACCGAAGTTTGAGTCGGTTGCTTGGCCTCTCTGCCTCATGCAGAGGTCACAGCATTAAGTAGATTCAGCAGTGGGAGGTATGA
CAACTTAGGGCTCGTTCCATTTGCTGGTCGATGTAGAACTCAGCGAAAATCAATCGAGCTTGACTTGTCTGGAACTGCAGATAGAATCTTTGGCTACGTCACTTTGAACCGCGTTGCAGGTGTTAAGCC
CGGTGCCCACATAACTACTTTCGTGGCCGCTAGCAGGCTAGAGACAACTGAACTAATTCGAGGGCGTAGTCAAGGTCGTGACTTCAGCATGAACTCTGTGAACATCAAGCTATAATTCTGAAAATATCA
AGCTTTTCTTACTTAAACTGCTCATAATTGGGAGCGATTGTTGAAAGTAGCTACACCATTGCCGCTTGCAGTGGTGTAAAAATTGCAGTAGAGACTGAGATATTCGACAATTTTTTTAGACAGGTAAAT
TTTGCGCAAGCTGAAGCCGAGCCGCGGCTAAGTAAGCGGATTGGTCGAGTTGCTTCTGAGCTGTTAGCAGCAACGAAACTTGTTTCGTGGGTACCGGGCTTTAGAAACAAGGCGATTATGTTGAGCTGA
CTGCAAACTCCATTTTGTTGAGTAAGATTAAGTTCTTGTGAAGTCTCAAAGGTAATCGATTAAGGGTCAACACCTGAAATTTGAATTAACCTATTAAAGTCAGCCCACATGGCTGCAGTTTCGGACAAG
TTAAGCCTCAAAAACCTATTTTGGCCCATCAAACAAACGCGCTCTTAGTGGACTACTGAACTAATCCTAACTTGGCACGCCACGGGTACTGATATTGGTACAGTCGGCCTGATGAGGCCATAGCAAGCA
GTCCGGCAGCAAATGCCACTGTCGCGCGGGGTGCATGTAGCCATGTGACGGTACCGATGTGGGGCTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATG
TGGCTGCACCGATGTGTTAGGATATCATGATACATATGTGTGTGCTGGGTCGTCTAGTTTGGTTGGGTAATTAACTGCCGCGTGATTGATTGATTGCTTGGTTGCCGTGTCGAAGAGGAATAATAGTAA
TGATGCAAATGATTGATACTTTTTTGTGCAAAAATAGTTCGCCCACAGATAGCTGCTCGGGTGGGCAAAGACTGGGTACGAATCAAGTCTCAACTCGGTTGACCAACTCTGAAGTTGTTGCCATGACCA
GCACCGGACCTAAACTGACTCTAAAATTCTCTTCTAACCCCAGCCGAACGTTGCCATTCGGATCCCCCCCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA


>Ph_six3_ref [aa] 'artificial consensus cDNA
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSPGADDDEDDLINPGSP
TPFEDDVSEDEVSLGQQSPLSKLGQPPLAAFDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDG
DHSALRLPLPGALFSPFASALPPRLPLAFPPHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGD
TSISDQQPRPHIFQPIKLEFAPSTETS&

>Ph_six3_cDNA09 [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQRSSSISSSAKLETGPRSPGADDDEDDLINPGSP
TPFEDDVSEDEVSLGQQSPLSKLGQPPLAAFDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDG
DHSALRLPLPGALFSPFASALPPRLPLAFPPHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGD
TSISDQQPRPHIFQPIKLEFAPSTETS&

>Ph_six3_cDNA13 [aa]
MALLAGLSAYSATLLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIEYNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSGKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSSGADDDEDDLINPGSP
TPFEDDVSEDEVSLGQQSPLSKLGQPPLAAFDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSVTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDG
DHSALRLPLPGALFSPFASALPPRLPLAFPPHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGD
TSISDQQPRPHIFQPIKLEFAPSTETS&

>Ph_six3_ORF10 [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSPGADDDEDDLINPGSP
TPFEDDVSEDEVSLGQQSPLSKLGQPPLAAFDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDG
DHSALRLPLPGALFSPFASALPPRLPLAFPPHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGD
TSISDQQPRPHIFQPIKLEFAPSTETS&

>Ph_six3_ORF11 [aa]
MALLAGLSAYSAALLGPAPHGGAPGSVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSPGADDDEDDLINPGSP
TPFEDDVSEDEVSLGQQSPLSKLGQPPLAAFDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDG
DHSALRLPLPGALFSPFASALPPRLPLAFPPHLSAWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGD
TSISDQQPRPHIFQPIKLEFAPSTETS&

>Ph_six3_ORF12 [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPYLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSPGADDDEDDLINPGSP
TPFEDDVSEDEVSLGQQSPLSKLGQPPLAAFDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDG
DHSALRLPLPGALFSPFASALPPRLPLAFPPHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGD
TSISDQQPRPHIFQPIKLEFAPSTETS&

>Ph_six3_5R01 [aa]
MALLAGLSAYSAALLGPTPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASCPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRS

>Ph_six3_5R02 [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVMRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRS

>Ph_six3_5R04 [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRS

>Ph_six3_5R05 [aa]
```

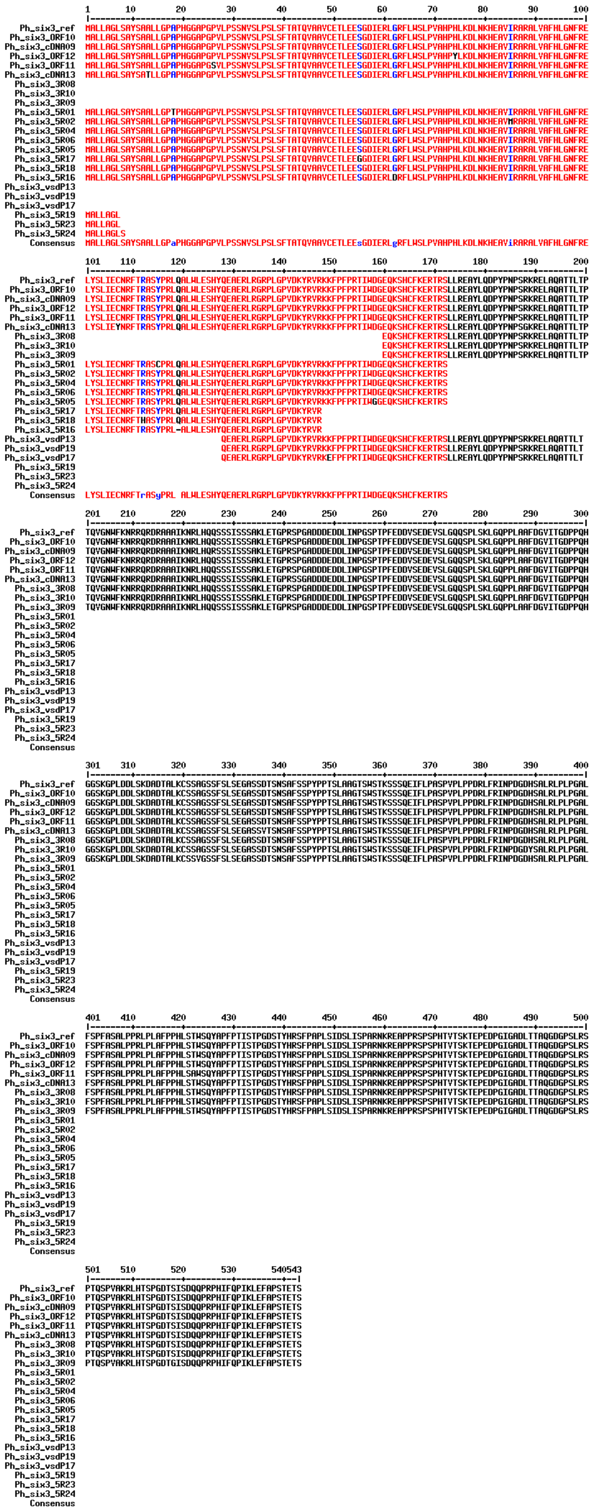
```
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWGGEQKSHCFKERTRS

>Ph_six3_5R06 [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRS

>Ph_six3_5R16 [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLDRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRL@ALWLESHYQE
AERLRGRPLGPVDKYRVR

>Ph_six3_5R17 [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEEGGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVR

>Ph_six3_5R18 [aa]
MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTHASYPRLQALWLESHYQE
AERLRGRPLGPVDKYRVR

>Ph_six3_5R19 [aa]
MALLAGL

>Ph_six3_5R23 [aa]
MALLAGL

>Ph_six3_5R24 [aa]
MALLAGLS

>Ph_six3_3R09 [aa]
EQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSPGADDDEDDLINPGSPTPFEDDVSEDEVSLGQQSPLSKLGQPPLAA
FDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSVGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDGDHSALRLPLPGALFSPFASALPPRLPLAFP
PHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGDTGISDQQPRPHIFQPIKLEFAPSTETS&

>Ph_six3_3R08 [aa]
EQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSPGADDDEDDLINPGSPTPFEDDVSEDEVSLGQQSPLSKLGQPPLAA
FDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDGDHSALRLPLPGALFSPFASALPPRLPLAFP
PHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGDTSISDQQPRPHIFQPIKLEFAPSTETS&

>Ph_six3_3R10 [aa]
EQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSPGADDDEDDLINPGSPTPFEDDVSEDEVSLGQQSPLSKLGQPPLAA
FDGVITGDPPQHGGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDGDYSALRLPLPGALFSPFASALPPRLPLAFP
PHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRSPTQSPVAKRLHTSPGDTSISDQQPRPHIFQPIKLEFAPSTETS&

>Ph_six3_vsdP13 [aa] 'w/o deg primer
QEAERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLT

>Ph_six3_vsdP17 [aa] 'w/o deg primer
QEAERLRGRPLGPVDKYRVRKEFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLT

>Ph_six3_vsdP19 [aa] 'w/o deg primer
QEAERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLT
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.8_Ph_six3]

[PhD_A2.8.2_Ph_six3_alignments]

## “alignment_Ph_six3_all_aa_sequences.gif”

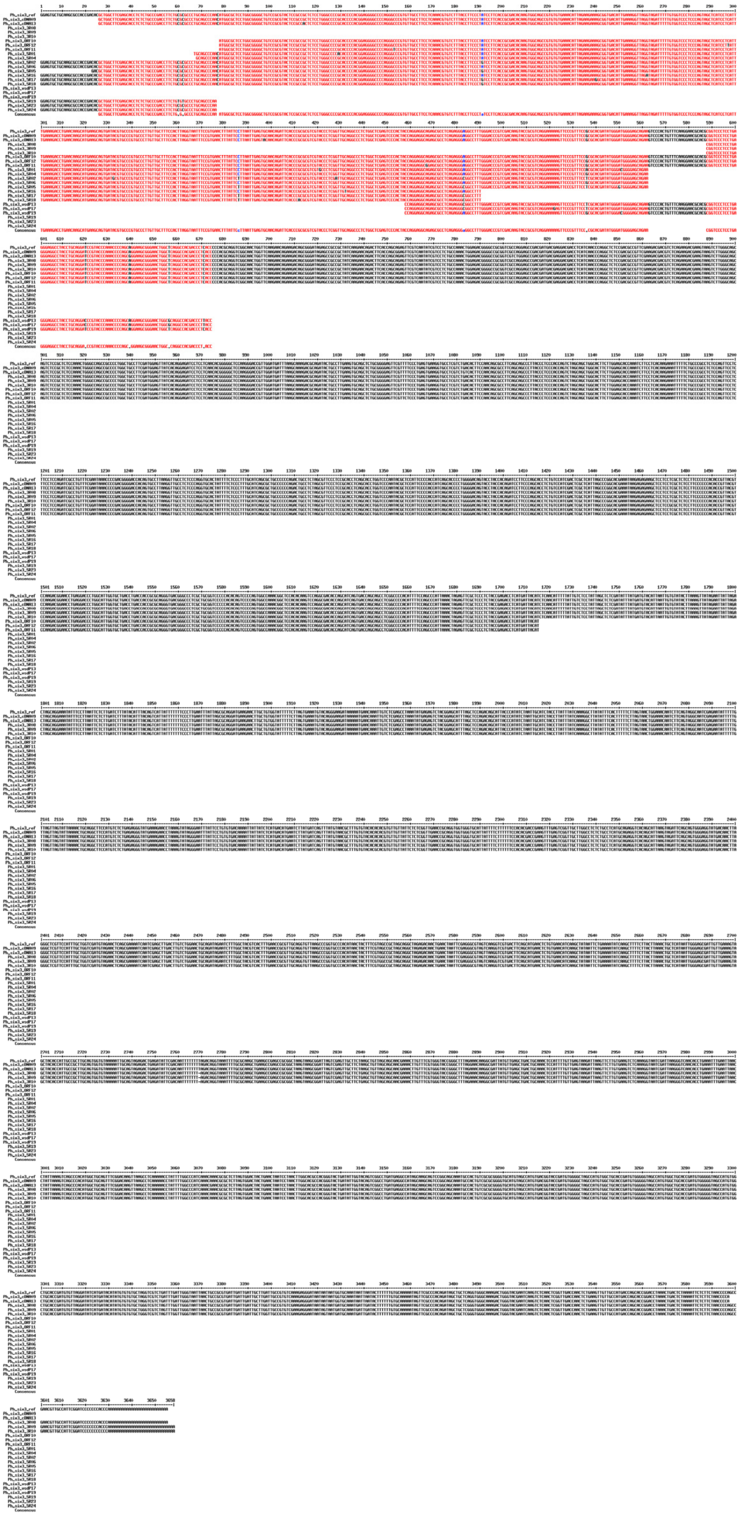


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.8_Ph_six3]

[PhD_A2.8.2_Ph_six3_alignments]

**“alignment_Ph_six3_all_bp_sequences.gif”**

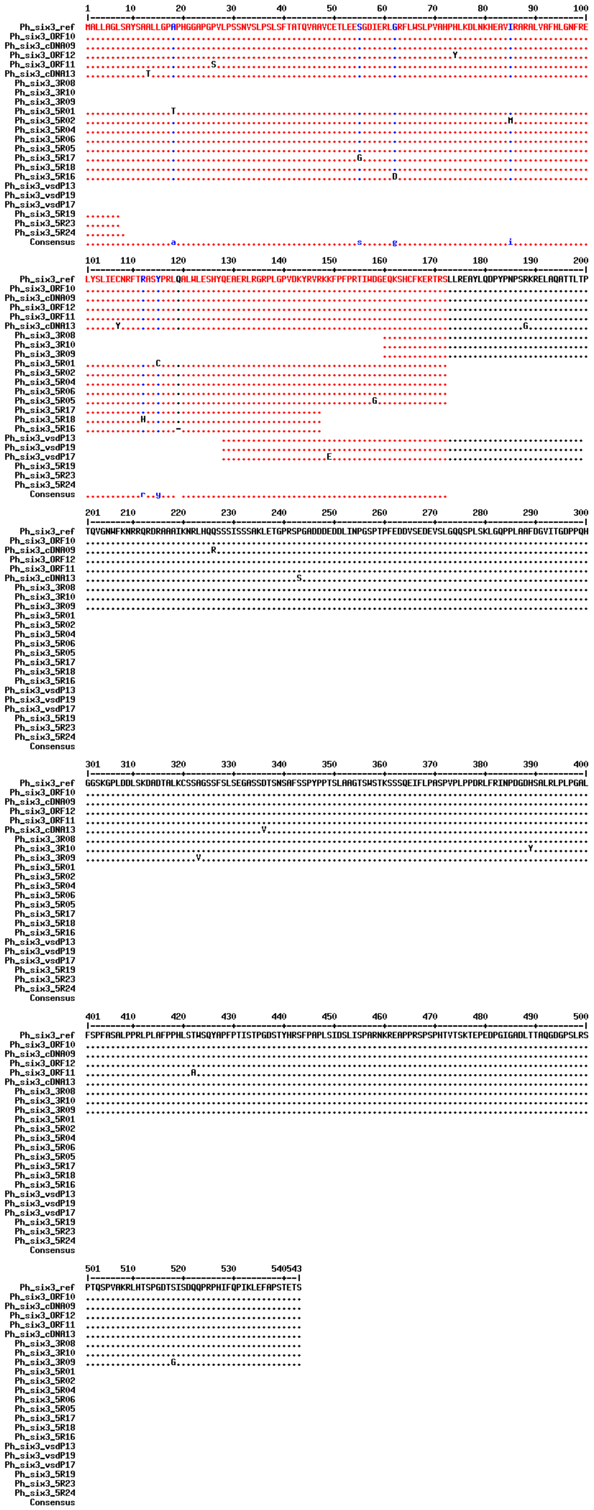


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.8_Ph_six3]

[PhD_A2.8.2_Ph_six3_alignments]

**“difference_alignment_Ph_six3_all_aa_sequences.gif”**



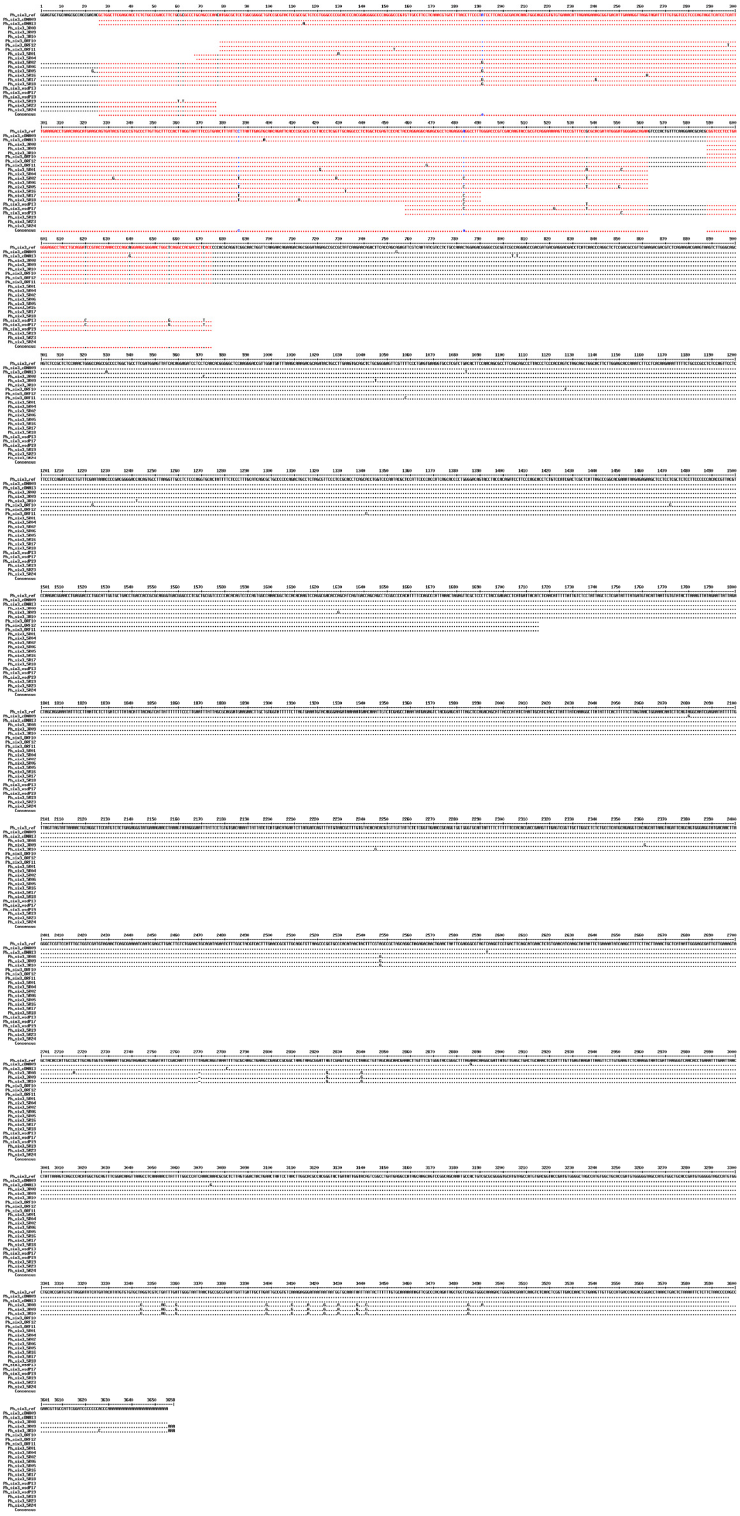
```
               1       10        20        30        40        50        60        70        80        90       100
               |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six3_ref MALLAGLSAYSAALLGPAPHGGAPGPVLPSSNVSLPSLSFTATQVAAVCETLEESGDIERLGRFLWSLPVAHPHLKDLNKHEAVIRARALVAFHLGNFRE
 Ph_six3_ORF10 ....................................................................................................
Ph_six3_cDNA09 ....................................................................................................
 Ph_six3_ORF12 .........................................................................Y..........................
 Ph_six3_ORF11 .........................S..........................................................................
Ph_six3_cDNA13 ............T.......................................................................................
  Ph_six3_3R08
  Ph_six3_3R10
  Ph_six3_3R09
  Ph_six3_5R01 .................T..................................................................................
  Ph_six3_5R02 ....................................................................................M...............
  Ph_six3_5R04 ....................................................................................................
  Ph_six3_5R06 ....................................................................................................
  Ph_six3_5R05 ....................................................................................................
  Ph_six3_5R17 ......................................................G.............................................
  Ph_six3_5R18 ....................................................................................................
  Ph_six3_5R16 .............................................................D......................................
Ph_six3_vsdP13
Ph_six3_vsdP19
Ph_six3_vsdP17
  Ph_six3_5R19 .......
  Ph_six3_5R23 .......
  Ph_six3_5R24 ........
     Consensus .................a....................................s......g......................i...............

               101    110       120       130       140       150       160       170       180       190       200
               |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six3_ref LYSLIECNRFTRASYPRLQALWLESHYQEAERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTP
 Ph_six3_ORF10 ....................................................................................................
Ph_six3_cDNA09 ....................................................................................................
 Ph_six3_ORF12 ....................................................................................................
 Ph_six3_ORF11 ....................................................................................................
Ph_six3_cDNA13 ......Y................................................................................G............
  Ph_six3_3R08                                                            .........................................
  Ph_six3_3R10                                                            .........................................
  Ph_six3_3R09                                                            .........................................
  Ph_six3_5R01 ..............C.........................................................
  Ph_six3_5R02 ........................................................................
  Ph_six3_5R04 ........................................................................
  Ph_six3_5R06 ........................................................................
  Ph_six3_5R05 .........................................................G..............
  Ph_six3_5R17 ...............................................
  Ph_six3_5R18 ...........H...................................
  Ph_six3_5R16 ..................-............................
Ph_six3_vsdP13                            ........................................................................
Ph_six3_vsdP19                            ........................................................................
Ph_six3_vsdP17                            .....................E..................................................
  Ph_six3_5R19
  Ph_six3_5R23
  Ph_six3_5R24
     Consensus ...........r..y... .....................................................

               201    210       220       230       240       250       260       270       280       290       300
               |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six3_ref TQVGNWFKNRRQRDRAAAIKNRLHQQSSSISSSAKLETGPRSPGADDDEDDLINPGSPTPFEDDVSEDEVSLGQQSPLSKLGQPPLAAFDGVITGDPPQH
 Ph_six3_ORF10 ....................................................................................................
Ph_six3_cDNA09 .........................R..........................................................................
 Ph_six3_ORF12 ....................................................................................................
 Ph_six3_ORF11 ....................................................................................................
Ph_six3_cDNA13 ..........................................S.........................................................
  Ph_six3_3R08 ....................................................................................................
  Ph_six3_3R10 ....................................................................................................
  Ph_six3_3R09 ....................................................................................................
  Ph_six3_5R01
  Ph_six3_5R02
  Ph_six3_5R04
  Ph_six3_5R06
  Ph_six3_5R05
  Ph_six3_5R17
  Ph_six3_5R18
  Ph_six3_5R16
Ph_six3_vsdP13
Ph_six3_vsdP19
Ph_six3_vsdP17
  Ph_six3_5R19
  Ph_six3_5R23
  Ph_six3_5R24
     Consensus

               301    310       320       330       340       350       360       370       380       390       400
               |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six3_ref GGSKGPLDDLSKDADTALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQEIFLPASPVPLPPDRLFRINPDGDHSALRLPLPGAL
 Ph_six3_ORF10 ....................................................................................................
Ph_six3_cDNA09 ....................................................................................................
 Ph_six3_ORF12 ....................................................................................................
 Ph_six3_ORF11 ....................................................................................................
Ph_six3_cDNA13 ...................................V................................................................
  Ph_six3_3R08 ....................................................................................................
  Ph_six3_3R10 ........................................................................................Y...........
  Ph_six3_3R09 ......................V.............................................................................
  Ph_six3_5R01
  Ph_six3_5R02
  Ph_six3_5R04
  Ph_six3_5R06
  Ph_six3_5R05
  Ph_six3_5R17
  Ph_six3_5R18
  Ph_six3_5R16
Ph_six3_vsdP13
Ph_six3_vsdP19
Ph_six3_vsdP17
  Ph_six3_5R19
  Ph_six3_5R23
  Ph_six3_5R24
     Consensus

               401    410       420       430       440       450       460       470       480       490       500
               |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six3_ref FSPFASALPPRLPLAFPPHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADLTTAQGDGPSLRS
 Ph_six3_ORF10 ....................................................................................................
Ph_six3_cDNA09 ....................................................................................................
 Ph_six3_ORF12 ....................................................................................................
 Ph_six3_ORF11 .....................A..............................................................................
Ph_six3_cDNA13 ....................................................................................................
  Ph_six3_3R08 ....................................................................................................
  Ph_six3_3R10 ....................................................................................................
  Ph_six3_3R09 ....................................................................................................
  Ph_six3_5R01
  Ph_six3_5R02
  Ph_six3_5R04
  Ph_six3_5R06
  Ph_six3_5R05
  Ph_six3_5R17
  Ph_six3_5R18
  Ph_six3_5R16
Ph_six3_vsdP13
Ph_six3_vsdP19
Ph_six3_vsdP17
  Ph_six3_5R19
  Ph_six3_5R23
  Ph_six3_5R24
     Consensus

               501    510       520       530       540543
               |--------+---------+---------+---------+--|
   Ph_six3_ref PTQSPVAKRLHTSPGDTSISDQQPRPHIFQPIKLEFAPSTETS
 Ph_six3_ORF10 ...........................................
Ph_six3_cDNA09 ...........................................
 Ph_six3_ORF12 ...........................................
 Ph_six3_ORF11 ...........................................
Ph_six3_cDNA13 ...........................................
  Ph_six3_3R08 ...........................................
  Ph_six3_3R10 ...........................................
  Ph_six3_3R09 .................G.........................
  Ph_six3_5R01
  Ph_six3_5R02
  Ph_six3_5R04
  Ph_six3_5R06
  Ph_six3_5R05
  Ph_six3_5R17
  Ph_six3_5R18
  Ph_six3_5R16
Ph_six3_vsdP13
Ph_six3_vsdP19
Ph_six3_vsdP17
  Ph_six3_5R19
  Ph_six3_5R23
  Ph_six3_5R24
     Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.8_Ph_six3]

[PhD_A2.8.2_Ph_six3_alignments]

**“difference_alignment_Ph_six3_all_bp_sequences.gif”**



# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.9_Ph_six4]

### [PhD_A2.9.1_Ph_six4_sequences]

**“Ph_six4_general_read_me.txt”**

```
General aspects:

(i)     All sequences are in FASTA format.

(ii)    [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)   Comments are preceded by an apostrophe (')

(iv)    All sequences without olignonucleotide sequence are indicated (e.g. "w/o 2114, 2115"). In these cases, only novel Parhyale
sequence is presented.


Parhyale hawaiensis six4 terminology:


(i)     In general, sequence denotation begins with: ">Ph_six4_"

(ii)    The above (i) is followed by:   "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                        "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                        "ORF"               for coherent ORF sequences
                                        "cDNA"              for complete cDNA sequences

                                        "ref"               if reference sequence is presented
                                        "HD"                if homeobox/homeodomain fragment is presented
                                        "SD"                if Six-domain (encoding) fragment is presented

(iii)   The above (ii) is followed by clone id.


Parhyale hawaiensis six4 synopsis:

cDNA: 2711 bp
5'UTR: 590 bp (1-590)
ORF:  1515 bp (591-2105), 504 aa
3'UTR: 606 bp (2106-2711)
SD:    351 bp (1227-1577), 117 aa
HD:    180 bp (1578-1757), 60 aa


>Ph_six4_SD [bp]
TTCAACAAGGAACAAATAGACTGCATATGCGAGGTTCTCATACAGGCGAGAGATATGGACAAACTCTCTAAATTTTTAAACAATTTACCGGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATA
CTAAGGGCCAAAGTTGAAGTTGCTTTTGCGAAGGGCAATTACAAGGAAGTATATAATATCTTAGAGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAGGGCATGTGGTACAAGGCGCACTAT
AAAGAAGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGAAATACCCGCTGCCGCGCACCATCTGGGAC

>Ph_six4_HD [bp]
GGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGATGGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTTTGTCCCGCCAGACCGGCCTCACACTCACC
CAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCGAAATCCTGCG

>Ph_six4_ref [bp]
ATTCCCCGGCATGGACGATGCGAGGAAAGTACGCTAGACACCGTAAACTGACTCGGCCTGTATTGAATTTGATCTGCTAGTTGATTCGAAAGTGATATGATCCTGTTGCGCTTCTTGTCTGTCTTGCAT
GAGGTCTTCATAGATAAAGTGGCGGAAAATGTGTTGTTGTAATAATGACTCATAGTTTTTGTGGAGTTCAGGGCTTCCGAAGTGTCGTGTTAATTTAGAAGTGTTTGTGTTGTGCGAGTTTTGAACTTT
TGGTGTTACCGTTGGTGTTTGGTGCGAAAGCTATTCGAAGAAAGTGGGCTATGGAAACTGACTTTACTGCTATTCAAAAATAGTGCTTATAATAAGCTGAATCCATTTACAACTGTAGATTTTGTTTTG
TGGATTTTACGAAAAGTGTTTTTATGAACTCAATCGAGACTCGGTAATTAAAATTACTTCGTATTTTATTTGTGTGTGTAGAAATATAATCTGAAACTACCATCTCCGAAGAGAATTTGAAAACTTTTT
AAGATTTGTCTTTCTATGAATAATTTTTACTGAGTTCATGATTGAACTTTGTTTTTTAAATTGTATCAAGAAAAATGAAAACGGAACATCCGCGGCCGTCGTCGGCGGAATCGTACGGAAGCTTCTCAA
TTCCTTCCCCCGACGGGCTTAGCACTCCTAACCCCCGACTACCGCACTACAGCAGCATGAATTCTAGTTCCTCCAGCGTAGCTAATTTTTCAAACCTAAGCAATTTTCAAAGCACTGCGTCTAACCTTT
GCAGTCGCTCCTTCAAAAACGATACCTACACTCCACCACCTCATATCGGCATGGGTATGTCGCCAATGTCTTCTCAATCTTCTGTAACAACGCCGCCTTTATCTTCGTCTCCTTTAAGCCCATCGGAAC
CGCTAAGAAATATAACAAATAACCCGCATTTGATTCACGGGTCCGCCGCCTCGCCATTCCACTACGGGGCTTTAGCAGCCAACACGTCCGGCAGTACAGCCCCCTGCGCCGTCAGCCAATCCACTTCGT
CCAGTTGCACCACTAATATTTCCAATTTATCCAAGTCGCCGGCTATAAAACCACAGCCAACGTCCCCGCATATTGCTACCGTGACACCTGAACCCCATCACCGTGAATCACCGTCCCTATCAACCCCGA
AGAAAACCCCCTCCTCCTCAGCTTCGCCAAGCTCCCCATCATCGTCCCCAGTAAAAACGTCCCAGTTCAACAAGGAACAAATAGACTGCATATGCGAGGTTCTCATACAGGCGAGAGATATGGACAAAC
TCTCTAAATTTTTAAACAATTTACCGGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATACTAAGGGCCAAAGTTGAAGTTGCTTTTGCGAAGGGCAATTACAAGGAAGTATATAATATCTTAG
AGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAGGGCATGTGGTACAAGGCGCACTATAAAGAAGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGA
AATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGATGGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTT
TGTCCCGCCAGACCGGCCTCACACTCACCCAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCGAAATCCTGCGCCTCGACCGGACCTTCTTTTGGGAGCGCACATGGATCACGGCGGTCTGG
CCATGGACCATCACGTCCATCCTGACATGAGAGCCATGTACCACCACGCCAAGATGGCCGCCCATGCGTACGATCCGTGCAAGATGGGCACCGGCGTCCATCCAGCTTATGACCCCAACAACGTCATGT
GCATGAGCATGAGAGGTCATCGAGCTGATGCCATTCTACACCACCTCCCTAATGTGGCCATGGGTAGCTACCCCTATAGTAGCTACGACCTACAGAGCGTGGCGGCGACACACGGCAACTCTCACCAAG
ACATGCGGGAAGACAAGTGGACGGCCGCGGGAGTGCAATAATTGTTCGCCTCTCAAATCTTACGAGCTTCTGACATTTCACGTACCATCGATTGGGACAGTTTCATGGAGACGAACGGATATGAGCTGA
AACTTGTTCAGACTTTCACTTTTAACCATGCAATTCGCTACTGTGGATCGAGCTTCAAGGGTTAATTCCAAATAGATACCCGCAAAGTTGAGGTTAAAGCCGTCTCATGAATAATTTTTGTTTTGAAAT
ATGTCTCTGAAACTCTATCTAAATAATTTTAATGAACGCTTAGTTATTGCTGAAGTTCTGGGTACGTTAAGGCAAATGAGAGTAGTTTGCCTATTGACACTGTCGCAATCGATATTACTTTCAATTGTG
TATATGTTTTATAATGTACTTGTATAGTCTACCTAGCTATAATTGTCATAGGCAGCAGCTCCTTAGCATAGGATATCGACGTACAACCGAACGTGAATGTGAAAAAGTAGGTGGCCACGTCACAAGCCT
TTTCATTGTGTACTGAGTTTTGTTAGTAATACCTTGCCGTCCATTGTATGCTGAAGAAAATTTAGGGTTTGTTTACGGCTCGATCACATAGATCTACATGTATCAAAAAAAAAAAAAAAAAAAAAAAAA
AA


>Ph_six4_SD [aa]
FNKEQIDCICEVLIQARDMDKLSKFLNNLPASHFATDAVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRTIWD
```

```
>Ph_six4_HD [aa]
GEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPA

>Ph_six4_ref [aa]
MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMGMSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGAL
AANTSGSTAPCAVSQSTSSSCTTNISNLSKSPAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSSASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLNNLPASHFATDAVSEVILRA
KVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRN
PAPRPDLLLGAHMDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDLQSVAATHGNSHQDMREDKWTAAGVQ#
```

[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.9_Ph_six4]
[PhD_A2.9.1_Ph_six4_sequences]

## “Ph_six4_cDNA_sequences.txt”

```
>Ph_six4_ref [bp] '5R13/586bp-ORF15-3R24/569bp; 1884C>T, 2685G>A
ATTCCCCGGCATGGACGATGCGAGGAAAGTACGCTAGACACCGTAAACTGACTCGGCCTGTATTGAATTTGATCTGCTAGTTGATTCGAAAGTGATATGATCCTGTTGCGCTTCTTGTCTGTCTTGCAT
GAGGTCTTCATAGATAAAGTGGCGGAAAATGTGTTGTTGTAATAATGACTCATAGTTTTTGTGGAGTTCAGGGCTTCCGAAGTGTCGTGTTAATTTAGAAGTGTTTGTGTTGTGCGAGTTTTGAACTTT
TGGTGTTACCGTTGGTGTTTGGTGCGAAAGCTATTCGAAGAAAGTGGGCTATGGAAACTGACTTTACTGCTATTCAAAAATAGTGCTTATAATAAGCTGAATCCATTTACAACTGTAGATTTTGTTTTG
TGGATTTTACGAAAAGTGTTTTTATGAACTCAATCGAGACTCGGTAATTAAAATTACTTCGTATTTTATTTGTGTGTGTAGAAATATAATCTGAAACTACCATCTCCGAAGAGAATTTGAAAACTTTTT
AAGATTTGTCTTTCTATGAATAATTTTTACTGAGTTCATGATTGAACTTTGTTTTTTAAATTGTATCAAGAAAAATGAAAACGGAACATCCGCGGCCGTCGTCGGCGGAATCGTACGGAAGCTTCTCAA
TTCCTTCCCCCGACGGGCTTAGCACTCCTAACCCCCGACTACCGCACTACAGCAGCATGAATTCTAGTTCCTCCAGCGTAGCTAATTTTTCAAACCTAAGCAATTTTCAAAGCACTGCGTCTAACCTTT
GCAGTCGCTCCTTCAAAAACGATACCTACACTCCACCACCTCATATCGGCATGGGTATGTCGCCAATGTCTTCTCAATCTTCTGTAACAACGCCGCCTTTATCTTCGTCTCCTTTAAGCCCATCGGAAC
CGCTAAGAAATATAACAAATAACCCGCATTTGATTCACGGGTCCGCCGCCTCGCCATTCCACTACGGGGCTTTAGCAGCCAACACGTCCGGCAGTACAGCCCCCTGCGCCGTCAGCCAATCCACTTCGT
CCAGTTGCACCACTAATATTTCCAATTTATCCAAGTCGCCGGCTATAAAACCACAGCCAACGTCCCCGCATATTGCTACCGTGACACCTGAACCCCATCACCGTGAATCACCGTCCCTATCAACCCCGA
AGAAAACCCCCTCCTCCTCAGCTTCGCCAAGCTCCCCATCATCGTCCCCAGTAAAAACGTCCCAGTTCAACAAGGAACAAATAGACTGCATATGCGAGGTTCTCATACAGGCGAGAGATATGGACAAAC
TCTCTAAATTTTTAAACAATTTACCGGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATACTAAGGGCCAAAGTTGAAGTTGCTTTTGCGAAGGGCAATTACAAGGAAGTATATAATATCTTAG
AGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAGGGCATGTGGTACAAGGCGCACTATAAAGAAGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGA
AATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGATGGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTT
TGTCCCGCCAGACCGGCCTCACACTCACCCAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCGAAATCCTGCGCCTCGACCGGACCTTCTTTTGGGAGCGCACATGGATCACGGCGGTCTGG
CCATGGACCATCACGTCCATCCTGACATGAGAGCCATGTACCACCACGCCAAGATGGCCGCCCATGCGTACGATCCGTGCAAGATGGGCACCGGCGTCCATCCAGCTTATGACCCCAACAACGTCATGT
GCATGAGCATGAGAGGTCATCGAGCTGATGCCATTCTACACCACCTCCCTAATGTGGCCATGGGTAGCTACCCCTATAGTAGCTACGACCTACAGAGCGTGGCGGCGACACACGGCAACTCTCACCAAG
ACATGCGGGAAGACAAGTGGACGGCCGCGGGAGTGCAATAATTGTTCGCCTCTCAAATCTTACGAGCTTCTGACATTTCACGTACCATCGATTGGGACAGTTTCATGGAGACGAACGGATATGAGCTGA
AACTTGTTCAGACTTTCACTTTTAACCATGCAATTCGCTACTGTGGATCGAGCTTCAAGGGTTAATTCCAAATAGATACCCGCAAAGTTGAGGTTAAAGCCGTCTCATGAATAATTTTTGTTTTGAAAT
ATGTCTCTGAAACTCTATCTAAATAATTTTAATGAACGCTTAGTTATTGCTGAAGTTCTGGGTACGTTAAGGCAAATGAGAGTAGTTTGCCTATTGACACTGTCGCAATCGATATTACTTTCAATTGTG
TATATGTTTTATAATGTACTTGTATAGTCTACCTAGCTATAATTGTCATAGGCAGCAGCTCCTTAGCATAGGATATCGACGTACAACCGAACGTGAATGTGAAAAAGTAGGTGGCCACGTCACAAGCCT
TTTCATTGTGTACTGAGTTTTGTTAGTAATACCTTGCCGTCCATTGTATGCTGAAGAAAATTTAGGGTTTGTTTACGGCTCGATCACATAGATCTACATGTATCAAAAAAAAAAAAAAAAAAAAAAAAA
AA

>Ph_six4_ORF14 [bp] 'w/o 2114, 2115
AAAAATGAAAACGGAACATCCGCGGCCGTCGTCGGCGGAATCGTACGGAAGCTTCTCAATTCCTTCCCCCGACGGGCTTAGCACTCCTAACCCCCGACTACCGCACTACAGCAGCATGAATTCTAGTTC
CTCCAGCGTAGCTAATTTTTCAAACCTAAGCAATTTTCAAAGCACTGCGTCTAACCTTTGCAGTCGCTCCTTCAAAAACGATACCTACACTCCACCACCTCATATCGGCATGGGTATGTCGCCAATGTC
TTCTCAATCTTCTGTAACAACGCCGCCTTTATCTTCGTCTCCTTTAAGCCCATCGGAACCGCTAAGAAATATAACAAATAACCCGCATTTGATTCACGGGTCCGCCGCCTCGCCATTCCACTACGGGAC
TTTAGCAGCCAACACGTCCGGCAGTACAGCCCCCTGCGCCGTCAGCCAATCCACTTCGTCCAGTTGCACCACTAATATTTCCAATTTATCCAAGTCGCCGGCTATAAAACCACAGCCAACGTCCCCGCA
TATTGCTACCGTGACACCTGAACCCCATCACCGTGAATCACCGTCCCTATCAACCCCGAAGAAAACCCCCTCCTCCTCAGCTTCGCCAAGCTCCCCATCATCGTCCCCAGTAAAAACGTCCCAGTTCAA
CAAGGAACAAATAGACTGCATATGCGAGGTTCTCATACAGGCGAGAGATATGGACAAACTCTCTAAATTTTTAAACAATTTACCGGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATACTAAG
GGCCAAAGTTGAAGTTGCTTTTGCGAAGGGCAATTACAAGGAAGTATATAATATCTTAGAGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAGGGCATGTGGTACAAGGCGCACTATAAAGA
AGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGAAATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGAT
GGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTTTGTCCCGCCAGACCGGCCTCACACTCACCCAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCG
AAATCCTGCGCCTCGACCGGACCTTCTTTTGGGAGCGCACATGGATCACGGCGGTCTGGCCATGGACCATCACGTCCATCCTGACATGAGAGCCATGTACCACCACGCCAAGATGGCCGCCCATGCGTA
CGATCCGTGCAAGATGGGCACCGGCGTCCATCCAGCTTATGACCCCAACAACGTCATGTGCATGAGCATGAGAGGTCATCGAGCTGATGCCATTCTACACCACCTCCCTAACGTGGCCATGGGTAGCTA
CCCCTATAGTAGCTACGACCTACAGAGCGTGGCGGCGACACACGGCAACTCTCACCAAGACATGCGGGAAGACAAGTGGACGGCCGCGGGAGTGCAATAATTGTTCGCCTCTCAAATCTTACGAGCTTC
TGACATTT

>Ph_six4_ORF15 [bp] 'w/o 2114, 2115
AAAAATGAAAACGGAACATCCGCGGCCGTCGTCGGCGGAATCGTACGGAAGCTTCTCAATTCCTTCCCCCGACGGGCTTAGCACTCCTAACCCCCGACTACCGCACTACAGCAGCATGAATTCTAGTTC
CTCCAGCGTAGCTAATTTTTCAAACCTAAGCAATTTTCAAAGCACTGCGTCTAACCTTTGCAGTCGCTCCTTCAAAAACGATACCTACACTCCACCACCTCATATCGGCATGGGTATGTCGCCAATGTC
TTCTCAATCTTCTGTAACAACGCCGCCTTTATCTTCGTCTCCTTTAAGCCCATCGGAACCGCTAAGAAATATAACAAATAACCCGCATTTGATTCACGGGTCCGCCGCCTCGCCATTCCACTACGGGGC
TTTAGCAGCCAACACGTCCGGCAGTACAGCCCCCTGCGCCGTCAGCCAATCCACTTCGTCCAGTTGCACCACTAATATTTCCAATTTATCCAAGTCGCCGGCTATAAAACCACAGCCAACGTCCCCGCA
TATTGCTACCGTGACACCTGAACCCCATCACCGTGAATCACCGTCCCTATCAACCCCGAAGAAAACCCCCTCCTCCTCAGCTTCGCCAAGCTCCCCATCATCGTCCCCAGTAAAAACGTCCCAGTTCAA
CAAGGAACAAATAGACTGCATATGCGAGGTTCTCATACAGGCGAGAGATATGGACAAACTCTCTAAATTTTTAAACAATTTACCGGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATACTAAG
GGCCAAAGTTGAAGTTGCTTTTGCGAAGGGCAATTACAAGGAAGTATATAATATCTTAGAGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAGGGCATGTGGTACAAGGCGCACTATAAAGA
AGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGAAATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGAT
GGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTTTGTCCCGCCAGACCGGCCTCACACTCACCCAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCG
AAATCCTGCGCCTCGACCGGACCTTCTTTTGGGAGCGCACATGGATCACGGCGGTCTGGCCATGGACCATCACGTCCATCCTGACATGAGAGCCATGTACCACCACGCCAAGATGGCCGCCCATGCGTA
CGATCCGCGCAAGATGGGCACCGGCGTCCATCCAGCTTATGACCCCAACAACGTCATGTGCATGAGCATGAGAGGTCATCGAGCTGATGCCATTCTACACCACCTCCCTAATGTGGCCATGGGTAGCTA
CCCCTATAGTAGCTACGACCTACAGAGCGTGGCGGCGACACACGGCAACTCTCACCAAGACATGCGGGAAGACAAGTGGACGGCCGCGGGAGTGCAATAATTGTTCGCCTCTCAAATCTTACGAGCTTC
TGACATTT

>Ph_six4_ORF16 [bp] 'w/o 2114, 2115
AAAAATGAAAACGGAACATCCGCGGCCGTCGTCGGCGGAATCGTACGGAAGCTTCTCAATTCCTTCCCCCGACGGGCTTAGCACTCCTAACCCCCGACTACCGCACTACAGCAGCATGAATTCTAGTTC
CTCCAGCGTAGCTAATTTTTCAAACCTAAGCAATTTTCAAAGCACTGCGTCTAACCTTTGCAGTCGCTCCTTCAAAAACGATACCTACACTCCACCACCTCATATCGGCATGGGTATGTCGCCAATGTC
TTCTCAATCTTCTGTAACAACGCCGCCTTTATCTTCGTCTCCTTTAAGCCCATCGGAACCGCTAAGAAATATAACAAATAACCCGCATTTGATTCACGGGTCCGCCGCCTCGCCATTCCACTACGGGGC
TTTAGCAGCCAACACGTCCGGCAGTACAGCCCCCTGCGCCGTCAGCCAATCCACTTCGTCCAGTTGCACCACTAATATTTCCAATTTATCCAAGTCGCCGGCTATAAAACCACAGCCAACGTCCCCGCA
TATTGCTACCGTGACACCTGAACCCCATCACCGTGAATCACCGTCCCTATCAACCCCGAAGAAAACCCCCTCCTCCTCAGCTTCGCCAAGCTCCCCATCATCGTCCCCAGTAAAAACGTCCCAGTTCAA
CAAGGAACAAATAGACTGCATATGCGAGGTTCTCGTACAGGCGAGAGATATGGACAAACTCTCTAAATTTTTAAACAATTTACCGGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATACTAAG
GGCCAAAGTTGAAGTTGCTTTTGCGAAGGGCAATTACAAGGAAGTATATAATATCTTAGAGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAGGGCATGTGGTACAAGGCGCACTATAAAGA
AGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGAAATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGAT
GGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTTTGTCCCGCCAGACCGGCCTCACACTCACCCAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCG
AAATCCTGCGCCTCGACCGGACCTTCTTTTGGGAGCGCACATGGATCACGGCGGTCTGGCCATGGACCATCACGTCCATCCTGACATGAGAGCCATGTACCACCACGCCAAGATGGCCGCCCATGCGTA
CGATCCGTGCAAGATGGGCACCGGCGTCCATCCAGCTTATGACCCCAACAACGTCATGTGCATGAGCATGAGAGGTCATCGAGCTGATGCCATTCTACACCACCTCCCTAATGTGGCCATGGGTAGCTA
CCCCTATAGTAGCTACGACCTACAGAGCGTGGCGGCGACACACGGCAACTCTCACCAAGACATGCGGGAAGACAAGTGGACGGCCGCGGGAGTGCAATAATTGTTCGCCTCTCAAATCTTACGAGCTTC
TGACATTT

>Ph_six4_5R13 [bp] 'w/o 2106
ATTCCCCGGCATGGACGATGCGAGGAAAGTACGCTAGACACCGTAAACTGACTCGGCCTGTATTGAATTTGATCTGCTAGTTGATTCGAAAGTGATATGATCCTGTTGCGCTTCTTGTCTGTCTTGCAT
GAGGTCTTCATAGATAAAGTGGCGGAAAATGTGTTGTTGTAATAATGACTCATAGTTTTTGTGGAGTTCAGGGCTTCCGAAGTGTCGTGTTAATTTAGAAGTGTTTGTGTTGTGCGAGTTTTGAACTTT
TGGTGTTACCGTTGGTGTTTGGTGCGAAAGCTATTCGAAGAAAGTGGGCTATGGAAACTGACTTTACTGCTATTCAAAAATAGTGCTTATAATAAGCTGAATCCATTTACAACTGTAGATTTTGTTTTG
TGGATTTTACGAAAAGTGTTTTTATGAACTCAATCGAGACTCGGTAATTAAAATTACTTCGTATTTTATTTGTGTGTGTAGAAATATAATCTGAAACTACCATCTCCGAAGAGAATTTGAAAACTTTTT
AAGATTTGTCTTTCTATGAATAATTTTTACTGAGTTCATGATTGAACTTTGTTTTTTAAATTGTATCAAGAAAAATGAAAACGGAACATCCGCGGCCGTCGTCGGCGGAATCGTACGGAAGCTTCTCAA
```

```
TTCCTTCCCCCGACGGGCTTAGCACTCCTAACCCCCGACTACCGCACTACAGCAGCATGAATTCTAGTTCCTCCAGCGTAGCTAATTTTTCAAACCTAAGCAATTTTCAAAGCACTGCGTCTAACCTTT
GCAGTCGCTCCTTCAAAAACGATACCTACACTCCACCACCTCATATCGGCATGGGTATGTCGCCAATGTCTTCTCAATCTTCTGTAACAACGCCGCCTTTATCTTCGTCTCCTTTAAGCCCATCGGAAC
CGCTAAGAAATATAACAAATAACCCGCATTTGATTCACGGGTCCGCCGCCTCGCCATTCCACTACGGGGCTTTAGCAGCCAACACGTCCGGCAGTACAGCCCCCTGCGCCGTCAGCCAATCCACTTCGT
CCAGTTGCACCACTAATATTTCCAATTTATCCAAGTCGCCGGCTATAAAACCACAGCCAACGTCCCCGCATATTGCTACCGTGACACCTGAACCCCATCACCGTGAATCACCGTCCCTATCAACCCCGA
AGAAAACCCCCTCCTCCTCAGCTTCGCCAAGCTCCCCATCATCGTCCCCAGTAAAAACGTCCCAGTTCAACAAGGAACAAATAGACTGCATATGCGAGGTTCTCATACAGGCGAGAGATATGGACAAAC
TCTCTAAATTTTTAAACAATTTACCGGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATACTAAGGGCCAAAGTTGAAGTTGCTTTTGCGAAGGGCAATTACAAGGAAGTATATAATATCTTAG
AGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAGGGCATGTGGTACAAGGCGCACTATAAAGAAGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGA
AATACCCGCTGCCGCGCGCCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGATGGCGTTAAAAGACTGCTACCGACA

>Ph_six4_5R14 [bp] 'w/o 2106
AGTTGATTCGAAGGTGATATGATCCTGTTGCGCTTCTTGTCTGTCTTGCATGAGGTCTTCATAGATAAAGTGGCGGAAAATGTGTTGTTGTAATAATGACTCATAGTTTTTGTGGAGTTCAGGGCTTCC
GAAGTGTCGTGTTAATTTAGAAGTGTTTGTGTTGTGCGAGTTTTGAACTTTTGGTGTTACCGTTGGTGTTTGGTGCGAAAGCTATTCGAAGAAAGTGGGCTATGGAAACTGACTTTACTGCTATTCAAA
AATAGTGCTTATAATAAGCTGAATCCATTTACAACTGTAGATTTTGTTTTGTGGATTTTACGAAAAGTGTTTTTATGAACTCAATCGAGACTCGGTAATTAAAATTACTTCGTATTTTATTTGTGTGTG
TAGAAATATAATCTGAAACTACCATCTCCGAAGAGAATTTGAAAACTTTTTAAGATTTGTCTTTCTATGAATAATTTTTACTGAGTTCATGATTGAACTTTGTTTTTTAAATTGTATCAAGAAAAAATGA
AAACGGAACATCCGCGGCCGTCGTCGACGGAATCGTACGGAAGCTTCTTAATTCCTTCCCCCGACGGGCTTAGCACTCCTAACCCCCGACTACCGCACTACAGCAGCATGAATTCTAGTTCCTCCAGCG
TAGCTAATTTTTCAAACCTAAGCAATTTTCAAAGCACTGCGTCTAACCTTTGCAGTCGCTCCTTCAAAAACGATACCTACACTCCACCACCTCATATCGGCATGGGTATGTCGCCAATGTCTTCTCAAT
CTTCTGTAACAACGCCGCCTTTATCTTCGTCTCCTTTAAGCCCATCGGAACCGCTAAGAAATATAACAAATAACCCGCATTTGATTCGCGGGTCCGCCGCCTCGCCATTCCACTACGGGGCTTTAGCAG
CCAACACGTCCGGCAGTACAGCCCCCCGCGCCGTCAGCCAATCCACTTCGTCCAGTTGCACCACTAATATTTCCAATTTATCCAAGTCGCCGGCTATAAAACCACAGCCAACGTCCCCGCATATTGCTA
CCGTGACACCTGAACCCCATCACCGTGAATCACCGTCCCTATCAACCCCGAAGAAAACCCCCTCCTCCTCAGCTTCGCCAAGCTCCCCATCATCGTCCCCAGTAAAAACGTCCCAGTTCAACAAGGAAC
AAATAGACTGCATATGCGAGGTTCTCATACAGGCGAGAGATATGGACAAACTCTCTAAATTTTTAAACAATTTACCGGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATACTAAGGGCCAAAG
TTGAAGTTGCTTTTGCGAAGGGCAATTACAAGGAAGTATATAATATCTTAGAGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAGGGCATGTGGTACAAGGCGCACTATAAAGAAGCCGAGA
GAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGAAATACCCACTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGATGGCGTTAA
AAGACTGCTACCGACA

>Ph_six4_5R15 [bp] 'w/o 2106
ATTCCCCGGCATGGACGATGCGAGGAAAGTACGCTAGACGCCGTAAACTGACTCGGCCTGTATTGAATTTGATCTGCTAGTTGATTCGAAAGTGATATGATCCTGTTGCGCTTCTTGTCTGTCTTGCAT
GAGGTCTTCATAGATAAAGTGGCGCAAAATTTGTTGTTGTAATAATGACTCATAGTTTTTGTGGAGTTCAGGTCTTCCGAAGTGTCGTGTTAATTTAGAAGTGTTTGTGTTGTGCGAGTTTCGAACTTT
TGGTGTTACCGTTGGTGTTTGGTGCGAAATCTATACGAAGAAAGTGGGCTATGGAAACTGACTTTACTGCTATTCAAATATAGTGCTTATAATAAGCGGAATCCATTTACAACTGTAGATTTTGTTTTG
TGAATTTTATGAAAAGTGTTTTTATGAACTCAATCAAGACTCGGTAATTAAAATTACTTCGTATTTTATTTGTGTGTTTAGAAATATAATCTGAAAGTACCATCTCCGAAGAGAGTTTGAAAACTTTTT
TAAGGATTTGTCTTATTATGAATAATTTTTACTTAGTTCATGATTGAACTTTGTTTTTTAAATTGTATCAAGAAAGATGAAAACGGAACATCCGCGGCCGTCGTCGGCGGAATCGTACGGAAGCTTCTC
AATTCCTTCCCCCGACGGGCTAAGCACTCCTAACCCCCGACTGCCGCTCTACAGCAGCATGAATTCCAGTTCATCCAGCGCAGCTAATTTTTCAAACCTAAGCAATTTTCAAAGCACTGCGTCTAACCT
TTGCAGTCGCTCCTTCAAAAACGATACCTACACTCCGCCACCTCATATCGGCATGGGTATGTCGCCAATGTCTTCACAATCTTCCGTAACAACGCCGCCTTTATCTTCGTCTCCTTTAAGCCCATCGGA
ACCGCTAAGAAATATAACAAATAACCCGCATTTAATTCACGGGTCCGCCGCCTCGCCATTCCACTACGGGGCTTTAGCAGCCCACACGTCGGGCAGCACAGCCCCCTGTGCCGTCAGCCAATCCACTTC
GTCCAACTGCACCGCTAATATTTCCAATTTATCCAAGTCGCCGGCTATAAAACCGCAGCCAACGTCCCCGCTTGTTGCTACCGTGACACCTGAACCCCATCACCGTGAATCACCGTCCCTATCAACCCC
GAAGAAAACCCCCTCTTCCTCCTCAGCTTCGCCAAGCTCCCCATCATCGTCCCAAGTAAAAACGTCCCAGTTCAACAAGGAACAAATAGACTGCATATGCGAGGTTCTCATACAGGCGAGAGATATGGA
CAAACTCTCTAAGTTTTTAAACAATTTACCTGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATACTAAGGGCCAAAGTTGAAGTTGCTTTTGCGAAGGGCAATTACAAGGAAGTATATAATAT
CTTAGAGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAAGGCATGTGGTACAAGGCACACTATAAAGAAGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACG
CAAGAAATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGTTTCAAAGAGAAGTCCAGGATGGCGTTAAAAGACTGCTACCGACA

>Ph_six4_vsdP14 [bp] 'w/o deg primer
TAAAGAAGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGAAATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTC
CAGGATGGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTTTGTCCCGCCAGACCGGCCTCACA

>Ph_six4_vsdP16 [bp] 'w/o deg primer
TAAAGAAGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGAAATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTC
CAGGATGGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTTTGTCCCGCCAGACCGGCCTCACA

>Ph_six4_vsdP21 [bp] 'w/o deg primer
TAAAGAAGCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGAAATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTC
CAGGATGGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTTTGTCCCGCCAGACCGGCCTCACA

>Ph_six4_3R19 [bp] 'w/o 2100
AAATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGATGGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACT
TTGTCCCGCCAGACCGGCCTCACACTCACCCAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCGAAATCCTGCGCCTCGACCGGACCTTCTTTTGGGAGCGCACATGGATCACGGCGGTCTG
GCCATGGACCATCACGTCCATCCTGACATGAGAGCCATGTACCACCACGCCAAGATGGCCGCCCATGCGTACGATCCGTGCAAGATGGGCACCGGCGTCCATCCAGCTTATGACCCCAACAACGTCATG
TGCATGAGCATGAGAGGTCATCGAGCTGATGCCATTCTACACCACCTCCCTAATGTGGCCATGGGTAGCTACCCCTATAGTAGCTACGACCTACAGAGCGTGGCGGCGACACACGGCAACTCTCACCAA
GACATGCGGGAAGACAAGTGGACGGCCGCGGGAGTGCAATAATTGTTCGCCTCTCAAATCTTACGAGCTTCTGACATTTCACGTACCATCGATTGGGACAGTTTCATGGAGACGAACGGATATGAGCTG
AAACTTGTTCAGACTTTCACTTTTAACCATGCAATTCGCTACTGTGGATCGAGCTTCAAGGGTTAATTCCAAATAGATACCCGCAAAGTTGAGGTTAAAGCCGTCTCATGAATAATTTTTGTTTTGAAA
TATGTCTCTGAAACTCTATCTAAATAATTTTAATGAACGCTTAGTTATTGCTGAAGTTCTGGGTACGTTAAGGCAAATGAGAGTAGTTTGCCTATTGACACTGTCGCAATCGATATTACTTTCAATTGT
GTATATGTTTTATAATGTACTTGTATAGTCTACCTAGCTATAATTGTCATAGGCAGCAGCTCCTTAGCATAGGATATCGACGTACAACCGAACGTGAATGTGAAAAAGTAGGTGGCCACGTCCCAAGCC
TTTTCATTGTGTACTGAGTTTTGTTAGTAATACCTTGCCGTCCATTGTATGCTGAAGAAAATTTAGGGTTTGTTTACGGCTCGATCACATAGATCTCCATGTATCAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAA

>Ph_six4_3R20 [bp] 'w/o 2100
AAATACCCGCTGCCGCGCACCATCTGGGACGGGGAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGATGGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGCAGAAGACT
TTGTCCCGCCAGACCGGCCTCACACTCACCCAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCGAAATCCTGCGCCTCGACCGGACCTTCTTTTGGGAGCGCACATGGATCACGGCGGTCTG
GCCATGGACCATCACGTCCATCCTGACATGAGAGCCATGTACCACCACGCCAAGATGGCCGCCCATGCGTACGATCCGTGCAAGATGGGCACCGGCGTCCATCCAGCTTATGACCCCAACAACGTCATG
TGCATGAGCATGAGAGGTCATCGAGCTGATGCCATTCTACACCACCTCCCTAATGTGGCCATGGGTAGCTACCCCTATAGTAGCTACGACCTACAGAGCGTGGCGGCGACACACGGCAACTCTCACCAA
GACATGCGGGAAGACAAGTGGACGGCCGCGGGAGTGCAATAATTGTTCGCCTCTCAAATCTTACGAGCTTCTGACATTTCACGTACCATCGATTGGGACAGTTTCATGGAGACGAACGGATATGAGCTG
AAACTTGTTCAGACTTTCACTTTTAACCATGCAATTCGCTACTGTGGATCGAGCTTCAAGGGTTAATTCCAAATAGATACCCGCAAAGTTGAGGTTAAAGCCGTCTCATGAATAATTTTTGTTTTGAAA
TATGTCTCTGAAACTCTATCTAAATAATTTTAATGAACGCTTAGTTATTGCTGAAGTTCTGGGTACGTTAAGGCAAATGAGAGTAGTTTGCCTATTGACACTGTCGCAATCGATATTACTTTCAATTGT
GTATATGTTTTATAATGTACTTGTATAGTCTACCTAGCTATAATTGTCATAGGCAGCAGCTCCTTAGCATAGGATATCGACGTACAACCGAACGTGAATGTGAAAAAGTAGGTGGCCACGTCACAAGCC
TTTTCATTGTGTACTGAGTTTTGTTAGTAATACCTTGCCGTCCATTGTATGCTGAAGAAAATTTAGGGTTTGTTTACGGCTCGATCACATAGATCTACATGTACAAAAAAAAAAAAAAAAAAAAAAAA
AAA

>Ph_six4_3R24 [bp] 'w/o 2101
GCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAGAAGACTTTGTCCCGCCAGACCGGCCTCACACTCACCCAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCGA
AATCCTGCGCCTCGACCGGACCTTCTTTTGGGAGCGCACATGGATCACGGCGGTCTGGCCATGGACCATCACGTCCATCCTGACATGAGAGCCATGTACCACCACGCCAAGATGGCCGCCCATGCGTAC
GATCCGTGCAAGATGGGCACCGGCGTCCATCCAGCTTATGACCCCAACAACGTCATGTGCATGAGCATGAGAGGTCATCGAGCTGATGCCATTCTACACCACCTCCCTAATGTGGCCATAGGTAGCTAC
CCCTATAGTAGCTACGACCTACAGAGCGTGGCGGCGACACACGGCAACTCTCACCAAGACATGCGGGAAGACAAGTGGACGGCCGCGGGAGTGCAATAATTGTTCGCCTCTCAAATCTTACGAGCTTCT
GACATTTCACGTACCATCGATTGGGACAGTTTCATGGAGACGAACGGATATGAGCTGAAACTTGTTCAGACTTTCACTTTTAACCATGCAATTCGCTACTGTGGATCGAGCTTCAAGGGTTAATTCCAA
ATAGATACCCGCAAAGTTGAGGTTAAAGCCGTCTCATGAATAATTTTTGTTTTGAAATATGTCTCTGAAACTCTATCTAAATAATTTTAATGAACGCTTAGTTATTGCTGAAGTTCTGGGTACGTTAAG
GCAAATGAGAGTAGTTTGCCTATTGACACTGTCGCAATCGATATTACTTTCAATTGTGTATATGTTTTATAATGTACTTGTATAGTCTACCTAGCTATAATTGTCATAGGCAGCAGCTCCTTAGCATAG
GATATCGACGTACAACCGAACGTGAATGTGAAAAAGTAGGTGGCCACGTCACAAGCCTTTTCATTGTGTACTGAGTTTTGTTAGTAATACCTTGCCGTCCATTGTATGCTGAAGAAAATTTAGGGTTTG
TTTACGGCTCGATCACATAGATCTACATGTATCGAAAAAAAAAAAAAAAAAAAAAAAAAA


>Ph_six4_ref [aa]
MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMGMSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGAL
AANTSGSTAPCAVSQSTSSSCTTNISNLSKSPAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSSASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLNNLPASHFATDAVSEVILRA
```

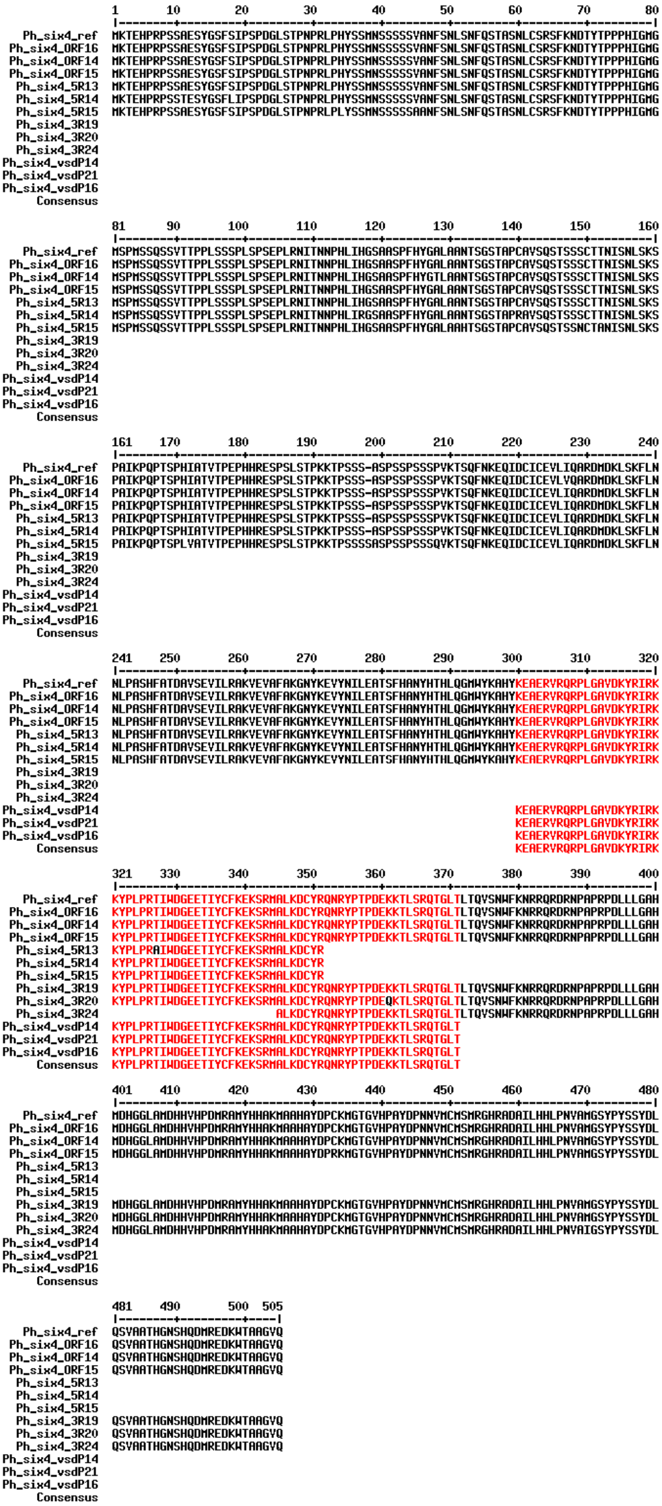
```
KVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRN
PAPRPDLLLGAHMDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDLQSVAATHGNSHQDMREDKWTAAGVQ#

>Ph_six4_ORF14 [aa]
MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMGMSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGTL
AANTSGSTAPCAVSQSTSSSCTTNISNLSKSPAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSSASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLNNLPASHFATDAVSEVILRA
KVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRN
PAPRPDLLLGAHMDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDLQSVAATHGNSHQDMREDKWTAAGVQ#

>Ph_six4_ORF15 [aa]
MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMGMSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGAL
AANTSGSTAPCAVSQSTSSSCTTNISNLSKSPAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSSASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLNNLPASHFATDAVSEVILRA
KVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRN
PAPRPDLLLGAHMDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPRKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDLQSVAATHGNSHQDMREDKWTAAGVQ#

>Ph_six4_ORF16 [aa]
MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMGMSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGAL
AANTSGSTAPCAVSQSTSSSCTTNISNLSKSPAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSSASPSSPSSSPVKTSQFNKEQIDCICEVLVQARDMDKLSKFLNNLPASHFATDAVSEVILRA
KVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRN
PAPRPDLLLGAHMDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDLQSVAATHGNSHQDMREDKWTAAGVQ#

>Ph_six4_5R13 [aa]
MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMGMSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGAL
AANTSGSTAPCAVSQSTSSSCTTNISNLSKSPAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSSASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLNNLPASHFATDAVSEVILRA
KVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRAIWDGEETIYCFKEKSRMALKDCYR

>Ph_six4_5R14 [aa]
MKTEHPRPSSTESYGSFLIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMGMSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIRGSAASPFHYGAL
AANTSGSTAPRAVSQSTSSSCTTNISNLSKSPAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSSASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLNNLPASHFATDAVSEVILRA
KVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYR

>Ph_six4_5R15 [aa]
MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPLYSSMNSSSSSAANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMGMSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGAL
AAHTSGSTAPCAVSQSTSSNCTANISNLSKSPAIKPQPTSPLVATVTPEPHHRESPSLSTPKKTPSSSSASPSSPSSSQVKTSQFNKEQIDCICEVLIQARDMDKLSKFLNNLPASHFATDAVSEVILR
AKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYR

>Ph_six4_vsdP14 [aa] 'w/o deg primer
KEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLT

>Ph_six4_vsdP16 [aa] 'w/o deg primer
KEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLT

>Ph_six4_vsdP21 [aa] 'w/o deg primer
KEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLT

>Ph_six4_3R19 [aa] 'w/o 2100
KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAHMDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVM
CMSMRGHRADAILHHLPNVAMGSYPYSSYDLQSVAATHGNSHQDMREDKWTAAGVQ#

>Ph_six4_3R20 [aa] 'w/o 2100
KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEQKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAHMDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVM
CMSMRGHRADAILHHLPNVAMGSYPYSSYDLQSVAATHGNSHQDMREDKWTAAGVQ#

>Ph_six4_3R24 [aa] 'w/o 2101
ALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAHMDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAIGSY
PYSSYDLQSVAATHGNSHQDMREDKWTAAGVQ#
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.9_Ph_six4]

### [PhD_A2.9.2_Ph_six4_alignments]

#### “alignment_Ph_six4_all_aa_sequences.gif”



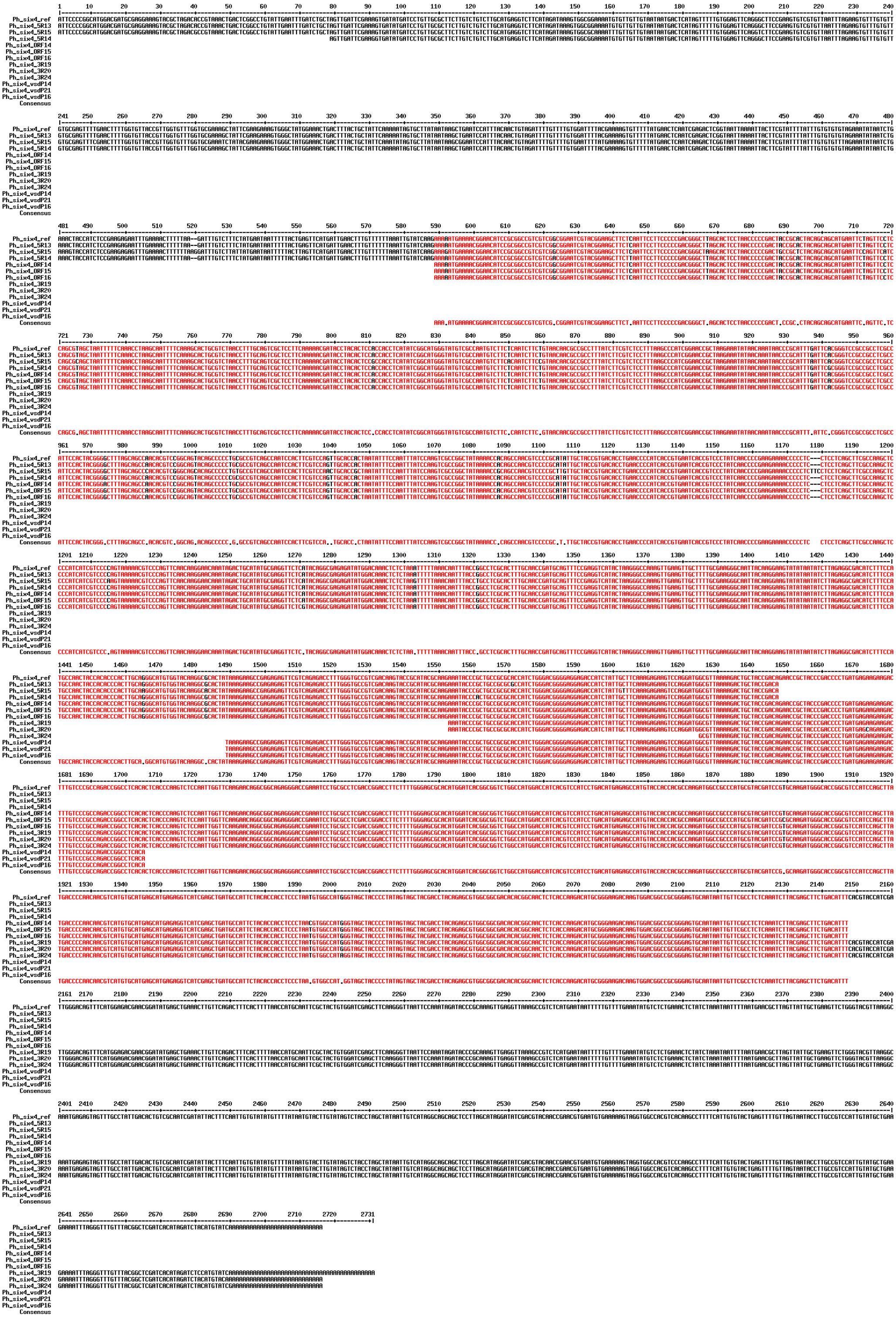
```
                1       10        20        30        40        50        60        70        80
                |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref  MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMG
 Ph_six4_ORF16  MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMG
 Ph_six4_ORF14  MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMG
 Ph_six4_ORF15  MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMG
  Ph_six4_5R13  MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMG
  Ph_six4_5R14  MKTEHPRPSSTESYGSFLIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMG
  Ph_six4_5R15  MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPLYSSMNSSSSSAANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMG
  Ph_six4_3R19
  Ph_six4_3R20
  Ph_six4_3R24
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus

                81      90       100       110       120       130       140       150       160
                |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref  MSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGALAANTSGSTAPCAVSQSTSSSCTTNISNLSKS
 Ph_six4_ORF16  MSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGALAANTSGSTAPCAVSQSTSSSCTTNISNLSKS
 Ph_six4_ORF14  MSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGTLAANTSGSTAPCAVSQSTSSSCTTNISNLSKS
 Ph_six4_ORF15  MSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGALAANTSGSTAPCAVSQSTSSSCTTNISNLSKS
  Ph_six4_5R13  MSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGALAANTSGSTAPCAVSQSTSSSCTTNISNLSKS
  Ph_six4_5R14  MSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIRGSAASPFHYGALAANTSGSTAPRAVSQSTSSSCTTNISNLSKS
  Ph_six4_5R15  MSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGALAAHTSGSTAPCAVSQSTSSNCTANISNLSKS
  Ph_six4_3R19
  Ph_six4_3R20
  Ph_six4_3R24
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus

                161    170       180       190       200       210       220       230       240
                |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref  PAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSS-ASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLN
 Ph_six4_ORF16  PAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSS-ASPSSPSSSPVKTSQFNKEQIDCICEVLVQARDMDKLSKFLN
 Ph_six4_ORF14  PAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSS-ASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLN
 Ph_six4_ORF15  PAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSS-ASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLN
  Ph_six4_5R13  PAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSS-ASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLN
  Ph_six4_5R14  PAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSS-ASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLN
  Ph_six4_5R15  PAIKPQPTSPLVATVTPEPHHRESPSLSTPKKTPSSSSASPSSPSSSQVKTSQFNKEQIDCICEVLIQARDMDKLSKFLN
  Ph_six4_3R19
  Ph_six4_3R20
  Ph_six4_3R24
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus

                241    250       260       270       280       290       300       310       320
                |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref  NLPASHFATDAVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRK
 Ph_six4_ORF16  NLPASHFATDAVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRK
 Ph_six4_ORF14  NLPASHFATDAVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRK
 Ph_six4_ORF15  NLPASHFATDAVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRK
  Ph_six4_5R13  NLPASHFATDAVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRK
  Ph_six4_5R14  NLPASHFATDAVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRK
  Ph_six4_5R15  NLPASHFATDAVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRK
  Ph_six4_3R19
  Ph_six4_3R20
  Ph_six4_3R24
Ph_six4_vsdP14                                                             KEAERVRQRPLGAVDKYRIRK
Ph_six4_vsdP21                                                             KEAERVRQRPLGAVDKYRIRK
Ph_six4_vsdP16                                                             KEAERVRQRPLGAVDKYRIRK
     Consensus                                                             KEAERVRQRPLGAVDKYRIRK

                321    330       340       350       360       370       380       390       400
                |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAH
 Ph_six4_ORF16  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAH
 Ph_six4_ORF14  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAH
 Ph_six4_ORF15  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAH
  Ph_six4_5R13  KYPLPRAIWDGEETIYCFKEKSRMALKDCYR
  Ph_six4_5R14  KYPLPRTIWDGEETIYCFKEKSRMALKDCYR
  Ph_six4_5R15  KYPLPRTIWDGEETIYCFKEKSRMALKDCYR
  Ph_six4_3R19  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAH
  Ph_six4_3R20  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEQKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAH
  Ph_six4_3R24                          ALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAH
Ph_six4_vsdP14  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLT
Ph_six4_vsdP21  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLT
Ph_six4_vsdP16  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLT
     Consensus  KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLT

                401    410       420       430       440       450       460       470       480
                |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref  MDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDL
 Ph_six4_ORF16  MDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDL
 Ph_six4_ORF14  MDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDL
 Ph_six4_ORF15  MDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPRKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDL
  Ph_six4_5R13
  Ph_six4_5R14
  Ph_six4_5R15
  Ph_six4_3R19  MDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDL
  Ph_six4_3R20  MDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDL
  Ph_six4_3R24  MDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAIGSYPYSSYDL
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus

                481    490       500  505
                |--------+---------+----|
   Ph_six4_ref  QSVAATHGNSHQDMREDKWTAAGVQ
 Ph_six4_ORF16  QSVAATHGNSHQDMREDKWTAAGVQ
 Ph_six4_ORF14  QSVAATHGNSHQDMREDKWTAAGVQ
 Ph_six4_ORF15  QSVAATHGNSHQDMREDKWTAAGVQ
  Ph_six4_5R13
  Ph_six4_5R14
  Ph_six4_5R15
  Ph_six4_3R19  QSVAATHGNSHQDMREDKWTAAGVQ
  Ph_six4_3R20  QSVAATHGNSHQDMREDKWTAAGVQ
  Ph_six4_3R24  QSVAATHGNSHQDMREDKWTAAGVQ
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.9_Ph_six4]

[PhD_A2.9.2_Ph_six4_alignments]

## “alignment_Ph_six4_all_bp_sequences.gif”

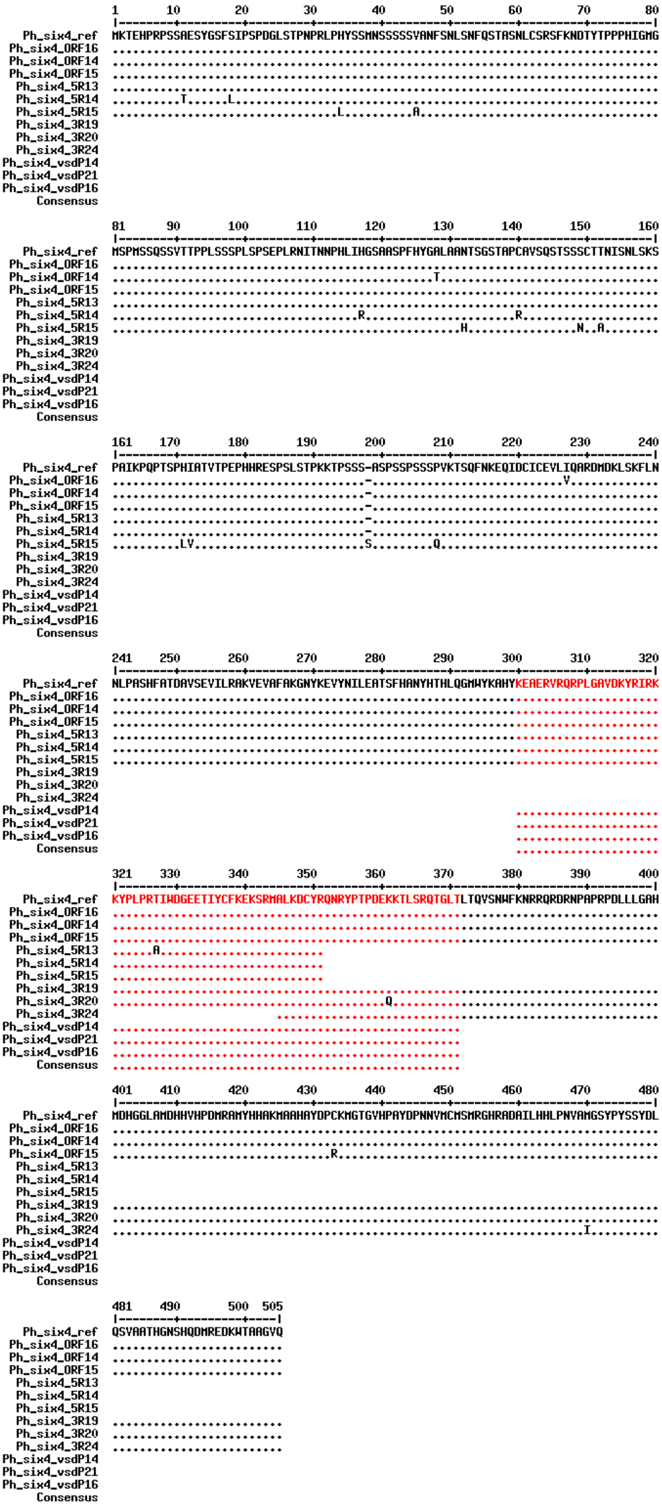


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.9_Ph_six4]

[PhD_A2.9.2_Ph_six4_alignments]

**“difference_alignment_Ph_six4_all_aa_sequences.gif”**



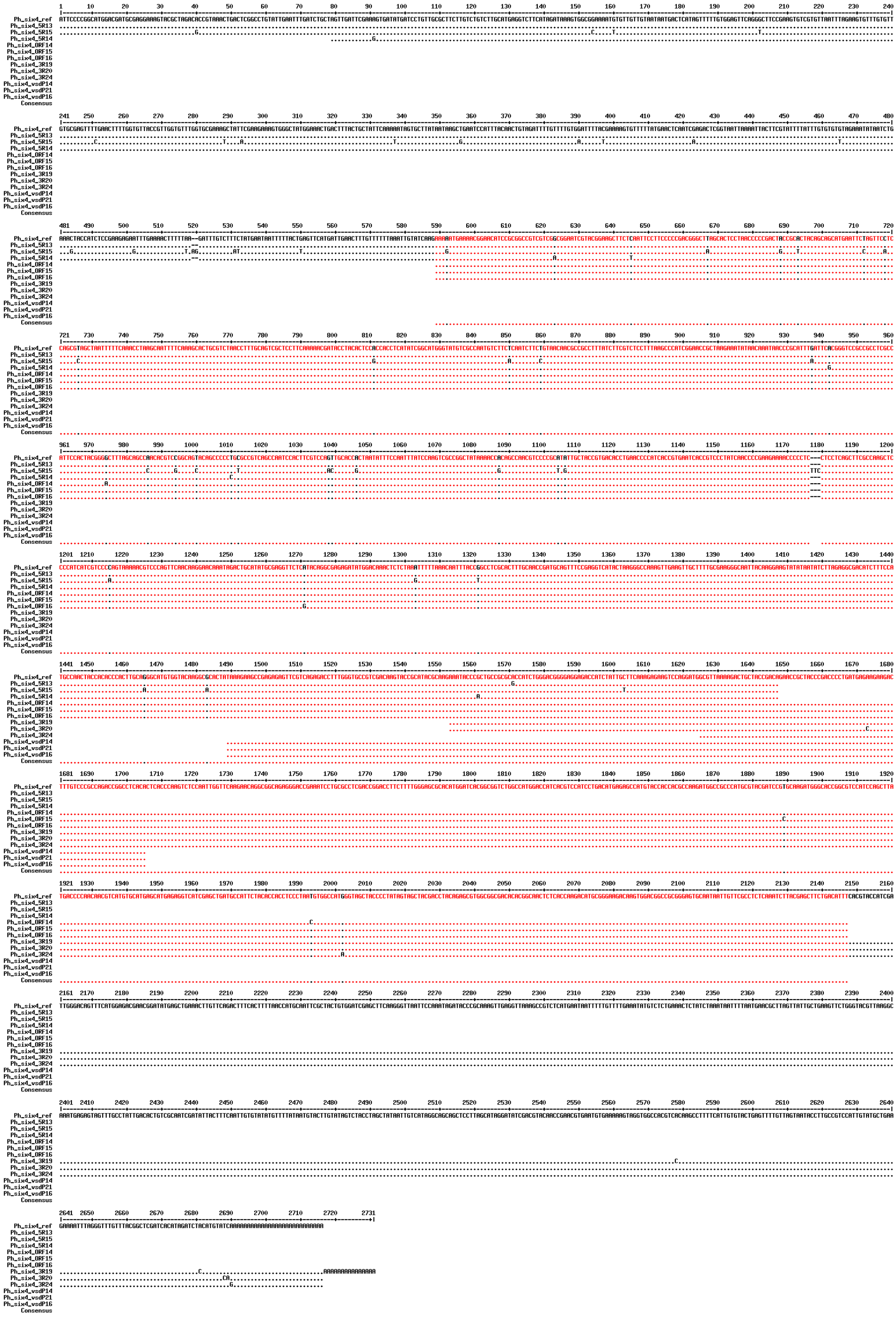
```
               1       10        20        30        40        50        60        70        80
               |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMG
 Ph_six4_ORF16 ................................................................................
 Ph_six4_ORF14 ................................................................................
 Ph_six4_ORF15 ................................................................................
  Ph_six4_5R13 ................................................................................
  Ph_six4_5R14 ..........T......L..............................................................
  Ph_six4_5R15 .................................L..........A...................................
  Ph_six4_3R19
  Ph_six4_3R20
  Ph_six4_3R24
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus

               81      90       100       110       120       130       140       150       160
               |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref MSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGALAANTSGSTAPCAVSQSTSSSCTTNISNLSKS
 Ph_six4_ORF16 ................................................................................
 Ph_six4_ORF14 ...............................................T................................
 Ph_six4_ORF15 ................................................................................
  Ph_six4_5R13 ................................................................................
  Ph_six4_5R14 ....................................R......................R....................
  Ph_six4_5R15 ...................................................H................N..A........
  Ph_six4_3R19
  Ph_six4_3R20
  Ph_six4_3R24
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus

               161     170       180       190       200       210       220       230       240
               |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref PAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSS-ASPSSPSSSPVKTSQFNKEQIDCICEVLIQARDMDKLSKFLN
 Ph_six4_ORF16 .....................................-............................V.............
 Ph_six4_ORF14 .....................................-..........................................
 Ph_six4_ORF15 .....................................-..........................................
  Ph_six4_5R13 .....................................-..........................................
  Ph_six4_5R14 .....................................-..........................................
  Ph_six4_5R15 ..........LV.........................S.........Q................................
  Ph_six4_3R19
  Ph_six4_3R20
  Ph_six4_3R24
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus

               241     250       260       270       280       290       300       310       320
               |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref NLPASHFATDAVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRK
 Ph_six4_ORF16 ................................................................................
 Ph_six4_ORF14 ................................................................................
 Ph_six4_ORF15 ................................................................................
  Ph_six4_5R13 ................................................................................
  Ph_six4_5R14 ................................................................................
  Ph_six4_5R15 ................................................................................
  Ph_six4_3R19
  Ph_six4_3R20
  Ph_six4_3R24
Ph_six4_vsdP14                                                            .....................
Ph_six4_vsdP21                                                            .....................
Ph_six4_vsdP16                                                            .....................
     Consensus                                                            .....................

               321     330       340       350       360       370       380       390       400
               |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref KYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWFKNRRQRDRNPAPRPDLLLGAH
 Ph_six4_ORF16 ................................................................................
 Ph_six4_ORF14 ................................................................................
 Ph_six4_ORF15 ................................................................................
  Ph_six4_5R13 ......A........................
  Ph_six4_5R14 ...............................
  Ph_six4_5R15 ...............................
  Ph_six4_3R19 ................................................................................
  Ph_six4_3R20 ........................................Q.......................................
  Ph_six4_3R24                         ........................................................
Ph_six4_vsdP14 ...................................................
Ph_six4_vsdP21 ...................................................
Ph_six4_vsdP16 ...................................................
     Consensus ...................................................

               401     410       420       430       440       450       460       470       480
               |--------+---------+---------+---------+---------+---------+---------+---------|
   Ph_six4_ref MDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGTGVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDL
 Ph_six4_ORF16 ................................................................................
 Ph_six4_ORF14 ................................................................................
 Ph_six4_ORF15 ................................R...............................................
  Ph_six4_5R13
  Ph_six4_5R14
  Ph_six4_5R15
  Ph_six4_3R19 ................................................................................
  Ph_six4_3R20 ................................................................................
  Ph_six4_3R24 .....................................................................I..........
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus

               481    490       500  505
               |--------+---------+----|
   Ph_six4_ref QSVAATHGNSHQDMREDKWTAAGVQ
 Ph_six4_ORF16 .........................
 Ph_six4_ORF14 .........................
 Ph_six4_ORF15 .........................
  Ph_six4_5R13
  Ph_six4_5R14
  Ph_six4_5R15
  Ph_six4_3R19 .........................
  Ph_six4_3R20 .........................
  Ph_six4_3R24 .........................
Ph_six4_vsdP14
Ph_six4_vsdP21
Ph_six4_vsdP16
     Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.9_Ph_six4]

[PhD_A2.9.2_Ph_six4_alignments]

## “difference_alignment_Ph_six4_all_bp_sequences.gif”



[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.10_Ph_gbx]

[PhD_A2.10.1_Ph_gbx_sequences]

## "Ph_gbx_general_read_me.txt"

```
General aspects:

(i)     All sequences are in FASTA format.

(ii)    [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)   Comments are preceded by an apostrophe (')

(iv)    All sequences without olignonucleotide sequence are indicated (e.g. "w/o 2314, 2315"). In these cases, only novel Parhyale
sequence is presented.


Parhyale hawaiensis gbx terminology:


(i)     In general, sequence denotation begins with: ">Ph_gbx_"

(ii)    The above (i) is followed by:  "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                       "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                       "ORF"               for coherent ORF sequences
                                       "cDNA"              for complete cDNA sequences

                                       "ref"               if reference sequence is presented
                                       "HD"                if homeobox/homeodomain fragment is presented
                                       "VPIPV"             if VPIPV (encoding) fragment is presented

(iii)   The above (ii) is followed by clone id.


Parhyale hawaiensis gbx synopsis:

cDNA    3456 bp
5'UTR:   505 bp (1-505)
ORF:    1665 bp (506-2170), 554 aa
3'UTR:  1286 bp (2171-3456)
HD:      180 bp (1346-1525), 60 aa
VPIPV     69 bp (1598-1666), 23 aa


>Ph_gbx_HD [bp]
AACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTC
CAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGAAG

>Ph_gbx_VPIPV [bp]
AAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGGCGCATCAGATAGAG

>Ph_gbx_ref [bp]
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGTGAAAGCTTCGTAAGTTTAGTTGAGCTTGTGAATGTAATGAAATATTTAGATATAAAAGACAATTGTACTGAACTGATACGTAATTTAGCCAGA
AGTGATGATTTACATGCAGGCATCCAATTTACTCCCAGTGAACTTCTATGAGTCAAAACTGTGCTATCAGAATGAAGTAAGTGTTATTGTGAGGCTACTATTTTCATATTTACGTAACGTACTATATTG
CATATCCCAGACGTACATCTACGTTTAGACAGAAGTACGATCGTGTGACTATACGAACTAATTATTTAACCAATAACTCTGCAATTTCTAGTTCTTAGTGAATTATTTAAAAAGTTGGTGTACGTCATG
CGAGATGAAACGTTTTGTGGTTATAATTGCTTGAACTTTCGCTATTTTTTACTGGACAACACAAAGTAAAAAAAAGTGCGCATTCGATCATTGTTCATCCCAGGGAATTATTTGAACGATGTTAGCGTC
TTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTCAAGACAAAAATGACTGCAATAGATC
AAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGTTGTCGCAAATAACCCAAATTTCTCA
GAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACTACCCGTACCCTGCCCCCTCCACCGT
ACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTCCTGGGGCTGGATCGAGATACTTCGG
GGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGACCAACGGAGATGAAAGCAGTCATAT
TTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACTCCGCGGGACTGGGGGGCGGTGGGGC
GCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAACGAAGAAATA
TCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGAAGGCTGGCCTTGTCGGCTCGAGCGG
CGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGGCGCATCAGATAGAGAAGTGTCGAGG
GTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAGAAGCGTGTACCTCCACCAAACACAG
AAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATCGCGCCCTACAGTCACGGCATTCTGC
GTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCCACTTACTCGGCCATACCAAGGAGGT
CACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGTAACCTTGATGTTTATAACTAAACAA
ACATATATCTATTATTTTTGGAACAGATTGGTGATAAAGAATATCATCGTGCACTATTACTTTGAAAATCAAGATGAGATGACGACCTAACCATAATTTACAAATAAACTTTCCTCTCAGGGAAACTTT
ACGGTATCTGATCTAGTGCAGATCGTACAACAATTTCATCGTCACAGCAATGCACTTTTCAAAATCACGGGGAATTTCTCTAGATAACCTTATCATTTATCATTAAAACATTGCTCACAAATAGCTTTA
TCATAATGGTTGCAGGCGGTCATCATCATCATTCCTGATTTAAGAATCAGGAAAGGATTTCTTCTTCGTTTGGAAACTCAGTAATAAAATCTCATTCACCAATATACCAAGATTTTCGTTGTTTTGTTG
TATCTTTAATAAGTCATCATCCTCGTCATTGCATTATTATTTCAAAATCGTGAAGGAAATCTTACTTAGCTTGAAATTCGCCCATAAAATTTCATTCACTAACAGCTTTGTGGCTTCCTAAGCTACGAC
AACCATATTATTAGTTTTTGTTTCCCCACTGCCATATTATTGGTATAAATATATCAGTATCTTACACCGGAAACATCTTTTCACAATTTTTAATGACACCAACTAACTGAATCGCAATAATTAAACTAC
ACTACTGTAACGAAAGTTAAAATCCTCTATGGGAAACCTCAGGAAGTAGAAGCTAAACTCTAGAATATACGGATCGTCACCATAATTTAGGCATTCTCGTTGGAATTATCTTAGTTGAAAGAAAAGAAA
GCAACTAATTTTCTATGTTTTATTTCGCATTAAGTGGAAGATTCTTTCAGAAACTTTACTATCAGTAGATCAAATTGTAACGCAAGTGGAAGGTACATAAATTTATCTTCAGTCAATGCTTCCCATATC
ACCACGTACGACGCAGCTAGTGCTTGAATATTAATAGGAAAAGGGAAAGGGCCTGCAGTAATTTAGGAAAAATGCATAATAGTTCTCAAGCACTGTAAATCACTTATAGGAATAGTAGTTCGGGAAAGA
AAACAAAAGAATTACAAAAATCATTAGTAAAAGTAAGAAAATAGTTAAGGAAACAAAAGCACAGAAAATAAAGATGTCCCGCTGTTCAGCTAGGACTACTGTTCTGTATGTACAGCTTGTAAATTATAT
GGACCACATTGCGTAGCTTGGGCGAACTTAACTAATCTGGCGGTACTTCCAAGCGATAGTTGTGTGTTTACTAACTGAAAAAAAAAAAAAAAAAAAAAAAAGT
```

```
>Ph_gbx_HD [aa]
NRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVK

>Ph_gbx_VPIPV [aa]
KIIVPIPVHVNRLSNLSQAHQIE

>Ph_gbx_ref [aa]
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#
```

### [PhD_A2_Parhyale_goi_sequences]

### [PhD_A2.10_Ph_gbx]

### [PhD_A2.10.1_Ph_gbx_sequences]

## “Ph_gbx_cDNA_sequences.txt”

```
>Ph_gbx_ref [bp] 'Ph_gbx_ORF58, extended 5' by Ph_gbx_5R05 (491bp), 3' by Ph_gbx_3R12 (1243bp)
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGTGAAAGCTTCGTAAGTTTAGTTGAGCTTGTGAATGTAATGAAATATTTAGATATAAAAGACAATTGTACTGAACTGATACGTAATTTAGCCAGA
AGTGATGATTTACATGCAGGCATCCAATTTACTCCCAGTGAACTTCTATGAGTCAAAACTGTGCTATCAGAATGAAGTAAGTGTTATTGTGAGGCTACTATTTTCATATTTACGTAACGTACTATATTG
CATATCCCAGACGTACATCTACGTTTAGACAGAAGTACGATCGTGTGACTATACGAACTAATTATTTAACCAATAACTCTGCAATTTCTAGTTCTTAGTGAATTATTTAAAAAGTTGGTGTACGTCATG
CGAGATGAAACGTTTTGTGGTTATAATTGCTTGAACTTTCGCTATTTTTTACTGGACAACACAAAGTAAAAAAAAGTGCGCATTCGATCATTGTTCATCCCAGGGAATTATTTGAACGATGTTAGCGTC
TTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTCAAGACAAAAATGACTGCAATAGATC
AAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGTTGTCGCAAATAACCCAAATTTCTCA
GAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACTACCCGTACCCTGCCCCCTCCACCGT
ACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTCCTGGGGCTGGATCGAGATACTTCGG
GGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGACCAACGGAGATGAAAGCAGTCATAT
TTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACTCCGCGGGACTGGGGGGCGGTGGGGC
GCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAACGAAGAAATA
TCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGAAGGCTGGCCTTGTCGGCTCGAGCGG
CGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGGCGCATCAGATAGAGAAGTGTCGAGG
GTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAGAAGCGTGTACCTCCACCAAACACAG
AAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATCGCGCCCTACAGTCACGGCATTCTGC
GTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCCACTTACTCGGCCATACCAAGGAGGT
CACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGTAACCTTGATGTTTATAACTAAACAA
ACATATATCTATTATTTTTGGAACAGATTGGTGATAAAGAATATCATCGTGCACTATTACTTTGAAAATCAAGATGAGATGACGACCTAACCATAATTTACAAATAAACTTTCCTCTCAGGGAAACTTT
ACGGTATCTGATCTAGTGCAGATCGTACAACAATTTCATCGTCACAGCAATGCACTTTTCAAAATCACGGGGAATTTCTCTAGATAACCTTATCATTTATCATTAAAACATTGCTCACAAATAGCTTTA
TCATAATGGTTGCAGGCGGTCATCATCATCATTCCTGATTTAAGAATCAGGAAAGGATTTCTTCTTCGTTTGGAAACTCAGTAATAAAATCTCATTCACCAATATACCAAGATTTTCGTTGTTTTGTTG
TATCTTTAATAAGTCATCATCCTCGTCATTGCATTATTATTTCAAAATCGTGAAGGAAATCTTACTTAGCTTGAAATTCGCCCATAAAATTTCATTCACTAACAGCTTTGTGGCTTCCTAAGCTACGAC
AACCATATTATTAGTTTTTGTTTCCCCACTGCCATATTATTGGTATAAATATATCAGTATCTTACACCGGAAACATCTTTTCACAATTTTTAATGACACCAACTAACTGAATCGCAATAATTAAACTAC
ACTACTGTAACGAAAGTTAAAATCCTCTATGGGAAACCTCAGGAAGTAGAAGCTAAACTCTAGAATATACGGATCGTCACCATAATTTAGGCATTCTCGTTGGAATTATCTTAGTTGAAAGAAAAGAAA
GCAACTAATTTTCTATGTTTTATTTCGCATTAAGTGGAAGATTCTTTCAGAAACTTTACTATCAGTAGATCAAATTGTAACGCAAGTGGAAGGTACATAAATTTATCTTCAGTCAATGCTTCCCATATC
ACCACGTACGACGCAGCTAGTGCTTGAATATTAATAGGAAAAGGGAAAGGGCCTGCAGTAATTTAGGAAAAATGCATAATAGTTCTCAAGCACTGTAAATCACTTATAGGAATAGTAGTTCGGGAAAGA
AAACAAAAGAATTACAAAAATCATTAGTAAAAGTAAGAAAATAGTTAAGGAAACAAAAGCACAGAAAATAAAGATGTCCCGCTGTTCAGCTAGGACTACTGTTCTGTATGTACAGCTTGTAAATTATAT
GGACCACATTGCGTAGCTTGGGCGAACTTAACTAATCTGGCGGTACTTCCAAGCGATAGTTGTGTGTTTACTAACTGAAAAAAAAAAAAAAAAAAAAAAAAGT
```

```
>Ph_gbx_ORF01 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGGACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGCGCCACTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGTGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAACCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
ACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG
```

```
>Ph_gbx_ORF02 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACGGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCCCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGTGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTCCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAACTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
ACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG
```

```
>Ph_gbx_ORF03 [bp] 'w/o 2314, 2315 'stranger seq!
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTATACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGAGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCTGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGGCCCCGTGGACAGCGATGACGAGAGTCATGTCAACGGTAACT
CAGCCGCGGGACTAGGCGGTGGGGCGCCGCTGGACAGCTCCACCACAACCGGCCCAGAGCACCGCGGCTCCGGCAGCAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCACACATTGCGCAGACGCTGCACCTCAGTGAGGTGCAAGTGAAAATTTGGTTCCAGAACCGGCGAGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGCGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCGCCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGGTCGCAGGTGTAACTCAACATAGCCAAGATCAACAGCAGCAGCAAC
ATCGCGCCCTACAGTCACGACATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGACAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAG
```

```
CCCACTTACTTGGCCATACCAAAGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGTCCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCC
CGTAGCCTTGATGTTTATAACTGAACAAACATGTATCTATTATTTTTG

>Ph_gbx_ORF04 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCGTGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATACCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
ACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG

>Ph_gbx_ORF05 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCATCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCAGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
ACTTGCTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG

>Ph_gbx_ORF07 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCGTCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAAGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
ACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAGCAAACATATATCTATTATTTTTG

>Ph_gbx_ORF52 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGCTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCGGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGTTCTGCCCTGGAAGCCC
ACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG

>Ph_gbx_ORF53 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCACGCACTTCCCCATCTGTACAACTAGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
ACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG

>Ph_gbx_ORF55 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCGGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
```

```
CCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACTGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGGACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACAGC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
ACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG

>Ph_gbx_ORF56 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGAGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
GCTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG

>Ph_gbx_ORF57 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACTCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGGCGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
ACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG

>Ph_gbx_ORF58 [bp] 'w/o 2314, 2315
GAATTATTTGAACGATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTC
AAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGT
TGTCGCAAATAACCCAAATTTCTCAGAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACT
ACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTC
CTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGA
CCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACT
CCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCG
AGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGA
AGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCCGGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGG
CGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAG
AAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATC
GCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCC
ACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGT
AACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTG

>Ph_gbx_5R02 [bp] 'w/o 2306
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGTGAAAGCTTCGTAAGTTTAGTTGAGCTTGTGAATGTAATGAAATATTTAGATATAAAAGACAATTGTACTGAACTGATACGTAATTTAGCCAGA
AGTGATGATTTACATGCAGGCATCCAATTTACTCCCAGTGAACTTCTATGAGTCAAAACTGTGCTATCAGAATGAAGTAAGTGTTATTGTGAGGCTACTATTTTCATATTTACGTAACGTACTATATTG
CATATCCCAGACGTACATCTACGTTTAGACAGAAGTACGATCGTGTGACTATACGAACTAATTATTTAACCAATAACTCTGCAATTTCTAGTTCTTAGTGAATTATTTAAAAAGTTGGTGTACGTCATG
CGAGATGAAACGTTTTGTGGTTATAATTGCTTGAACTTTCGCTATTTTTTACTGGACAACACAAGGTAAAAAAAAGTGCGCATTCGATCATTGTTCATCCCAGGGAATTATTTGAACGATGTTAGCGTC
TTCAGAAATGACAACTTGCGATTCACCACCTAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTCAAGACAAAAATGACTGCAATAGATCA
AAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGTTGTCGCAAATAACCCAAATTTCTCAG
AATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTAGCCCCGCACTACCCGTACCCTGCCCCCTCCACCGTA
CCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTCCTGGGGCTGGATCGAGATACTTCGGG
GGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGACCAACGGAGATGAAAGCAGTCATATT
TCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACTCCGCGGGACTGGGGGGCGGTGGGGCG
CCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAA

>Ph_gbx_5R03 [bp] 'w/o 2306
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGTGAAAGCTTCGTAAGTTTAGTTGAGCTTGTGAATGTAATGAAATATTTAGATATAAAAGACAATTGTACTGAACTGATACGTAATTTAGCCAGA
AGTGATGATTTACATGCAGGCATCCAATTTACTCCCAGTGAACTTCTATGAGTCAAAACTGTGCTATCAGAATGAAGTAAGTGTTATTGTGAGGCTACTATTTTCATATTTATGTAACGTACTATATTG
CATATCCCAGACGTACATCTACGTTTAGACAGAAGTACGATCGTGTGACTATACGAACTAATTATTTAACCAATAACTCTGCAATTTCTAGTTCTTAGTGAATTATTTAAAAAGTTGGTGTACGTCATG
CGAGATGAAACGTTTTGTGGTTATAATTGCTTGAACTTTCGCTATTTTTTACTGGACAACACAAAGTAAAAAAAAGTGCGCATTCGATCATTGTTCATCCCAGGGAATTATTTGAACGATGTTAGCGTC
TTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTCAAGACAAAAATGACTGCAATAGATC
AAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGTTGTCGCAAATAACCCAAATTTCTCA
GAATCTCGCACTTCCCCATCTGTACAACTCATGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACTACCCGTACCCTGCCCCCTCCACCGT
ACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAGGGATTTTGGTCCTGGGGCTGGATCGAGATACTTCGG
GGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGACCAACGGAGATGAAAGCAGTCATAT
TTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACTCCGCGGGACTGGGGGGCGGTGGGGC
GCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAA

>Ph_gbx_5R05 [bp] 'w/o 2306
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGTGAAAGCTTCGTAAGTTTAGTTGAGCTTGTGAATGTAATGAAATATTTAGATATAAAAGACAATTGTACTGAACTGATACGTAATTTAGCCAGA
AGTGATGATTTACATGCAGGCATCCAATTTACTCCCAGTGAACTTCTATGAGTCAAAACTGTGCTATCAGAATGAAGTAAGTGTTATTGTGAGGCTACTATTTTCATATTTACGTAACGTACTATATTG
CATATCCCAGACGTACATCTACGTTTAGACAGAAGTACGATCGTGTGACTATACGAACTAATTATTTAACCAATAACTCTGCAATTTCTAGTTCTTAGTGAATTATTTAAAAAGTTGGTGTACGTCATG
CGAGATGAAACGTTTTGTGGTTATAATTGCTTGAACTTTCGCTATTTTTTACTGGACAACACAAAGTAAAAAAAAGTGCGCATTCGATCATTGTTCATCCCAGGGAATTATTTGAACGATGTTAGCGTC
TTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATGGAATATCTCCTCACCATCTCCGGTCAAGACAAAAATGACTGCAATAGATC
AAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCCCGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGTTGTCGCAAATAACCCAAATTTCTCA
```

```
GAATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCGCTGGCCCCGCACTACCCGTACCCTGCCCCCTCCACCGT
ACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGTGGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTCCTGGGGCTGGATCGAGATACTTCGG
GGGATCTCCCACCGCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCGACCAACGGAGATGAAAGCAGTCATAT
TTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGACCCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACTCCGTGGGACTAGGGGGCGGTGGGGC
GCCGCTGGACAGCTCCACCACTACCGGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAA

>Ph_gbxvsdP105 [bp] 'w/o fw_AFT/re_KIW
TTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAG

>Ph_gbxvsdP106 [bp] 'w/o fw_AFT/re_KIW
TTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAG

>Ph_gbxvsdP108 [bp] 'w/o fw_AFT/re_KIW
TTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAG

>Ph_gbxvsdP109 [bp] 'w/o fw_AFT/re_KIW
TTCGGAGCAGCTCAAGGAACTCGAGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAG

>Ph_gbxvsdP110 [bp] 'w/o fw_AFT/re_KIW
TTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAG

>Ph_gbxvsdP111 [bp] 'w/o fw_AFT/re_KIW
TTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAG

>Ph_gbxvsdP112 [bp] 'w/o fw_AFT/re_KIW
TTCGGAGCAGCTCATGGAACTCGAGAGGGAATTCCAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAG

>Ph_gbx_3R03 [bp] 'w/o 2303
CGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGAAGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGA
GGTACAAACAAGGGGGGCGCCGGAGGAGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGGCGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGC
AACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAGAAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAAT
TTGCCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATCGCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCG
TCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCCACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAG
TCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGTAACCTTGATGTTTATAACTAAACAAACATATATCTATTATT
TTTGGAACAGATTGGTGATAAAGAATATCATCGTGCACTATTACTTTGAAAATCAAGATGAGATGACGACCTAACCATAATTTACAAATAAACTTTCCTCTCAGGGAAACTTTACGGTATCTGATCTAG
TGCAGATCGTACAACAATTTCATCGTCACAGCAATGCACTTTTCAAAATCACGGGGAATTTCTCTAGATAACCTTATCATTTATCATTAAAACATTGCTCACAAATAGCTTTATCATAATGGTTGCAGG
CGGTCATCATCATCATTCCTGATTTAAGAATCAGGAAAGGATTTCTTCTTCGTTTGGAAACTCAGTAATAAAATCTCATTCACCAATATACCAAGATTTTCGTTGTTTTGTTGTATCTTTAATAAGTCA
TCATCCTCGTCATTGCATTATTATTTCAAAATCGTGAAGGAAATCTTACTTAGCTTGAAATTCGCCCATAAAATTTCATTCACTAACAGCTTTGTGGCTTCCTAAGCTACGACAACCATATTATTAGTT
TTTGTTTCCCCACTGCCATATTATTGGTATAAATATATCAGTATCTTACACCGGAAACATCTTTTCACAATTTTTAATGACACCAATTAACTGAATCGCAATAATTAAACTACACTACTGTAACGAAAG
TTAAAATCCTCTATGGGAAACCTCAGGAAGTAGAAGCTAAACTCTAGAATATACGGATCGTCACCATAATTTAGGCATTCTCGTTGGAATTATCTTAGTTGAAAGAAAAGAAAGCAACTAATTTTCTAT
GTTTTATTTCGCATTAAGTGGAAGATTCTTTCAGAAACTTTACTATCAGTAGATCAAATTGTAACGCAAGTGGAAGGTACATAAATTTATCTTCAGTCAAAAAAAAAAAAAAAAAAGAAAAAAAGT

>Ph_gbx_3R11 [bp] 'w/o 2303
CGGTCGCACATTACTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGAAGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGA
GGTACAAACAAGGGGGGCGCCGGAGGAGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGGCGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGC
AACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAGAAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAAT
TTGTCGTCGAGCTTCTTGAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATCGCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCG
TCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCCACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAG
TCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGTAACCTTGATGTTTATAACTAAACAAACATATATCTATTATT
TTTGGAACAGATTGGTGATAAAGAATATCATCGTGCACTATTACTTTGAAAATCAAGATGAGATGACGACCTAACCATAATTTACAAATAAACTTTCCTCTCAGGGAAACTTTACGGTATCTGATCTAG
TGCAGATCGTACAACAATTTCATCGTCACAGCAATGCACTTTTCAAAATCACGGGGAATTTCTCTAGATAACCTTATCATTTATCATTAAAACATTGCTCACAAATAGCTTTATCATAATGGTTGCAGG
CGGTCATCATCATCATTCCTGATTTAAGAATCAGGAAAGGATTTCTTCTTCGTTTGGAAACTCAGTAATAAAATCTCACTCACCAATATACCAAGATTTTCGTTGTTTTGTTGTATCTTTAATAAGTCA
TCATCCTCGTCATTGCATTATTATTTCAAAATCGTGAAGGAAATCTTACTTAGCTTGAAATTCGCCCATAAAATTTCATTCACTAACAGCTTTGTGGCTTCCTAAGCTACGACAACCATATTATTAGTT
TTTGCTTCCCCACTGCCATATTATTGGTATAAATATATCAGTATCTTACACCGGAAACATCTTTTCACAATTTTTAATGACACCAATTAACTGAATCGCAATAATTAAACTACACTACTGTAACGAAAG
TTAAAATCCTCTATGGGAAACCTCAGGAAGTAGAAGCTAAACTCTAGAATATACGGATCGTCACCATAATTTAGGCATTCTCGCTGGAATTATCTTAGTTGAAAGAAAAGAAAGCAACTAATTTTCTAT
GTTTTATTTCGCATTAAGTGGAAGATTCTTTCAGAAACTTTACTATCAGTAGATCAAATTGTAACGCAAGTGGAAGGTACATAAATTTATCTTCAGTCAATGCTTCCTATATCACCACGTACGACGCAG
CTAGTGCTTGAATATTAATAGGAAAAGGGAAAGGGCCTGCAGTAATTTAGGAAAAATGCATAATAGTTCTCAAGCACTGTAAATCACTTATAGGAATAGTAGTTCGGGAAAGAAAACAAAAGAATTACA
AAAATCATTAGTAAAAGTAAGAAAATAGTTAAGGAAACAAAAGCACAGAAAATAAAGATGTCCCGCTGTTCGGCTAGGACTACTGTTCTGTATGTACAGCTTGTAAATTATATGGACCACATTGCGTAG
CTTGGGCGAACTTAACTAATCTGGCGGTACTTCCAAGCGATAGTTGTGTGTTTACTAACTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGT

>Ph_gbx_3R12 [bp] 'w/o 2303
CGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAGAACCGGCGGGCGAAATGGAAGCGCGTGAAGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGA
GGTACAAACAAGGGGGGCGCCGGAGGAGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGGCGCATCAGATAGAGAAGTGTCGAGGGTCGGCGGCTGGAAGC
AACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAACCAGAGGATGGCCACAGAAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAAT
TTGTCGTCGAGCTTCTTAAATTCAGAGACGTCGGAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATCGCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGGCTGTCG
TCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAAGCCCACTTACTCGGCTATACCAAGGAGGTCACGGAACAGTTGAAG
TCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACCTCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGTAACCTTGATGTTTATAACTAAACAAACATATATCTATTATT
TTTGGAACAGATTGGTGATAAAGAATATCATCGTGCACTATTACTTTGAAAATCAAGATGAGATGACGACCTAACCATAATTTACAAATAAACTTTCCTCTCAGGGAAACTTTACGGTATCTGATCTAG
TGCAGATCGTACAACAATTTCATCGTCACAGCAATGCACTTTTCAAAATCACGGGGAATTTCTCTAGATAACCTTATCATTTATCATTAAAACATTGCTCACAAATAGCTTTATCATAATGGTTGCAGG
CGGTCATCATCATCATTCCTGATTTAAGAATCAGGAAAGGATTTCTTCTTCGTTTGGAAACTCAGTAATAAAATCTCATTCACCAATATACCAAGATTTTCGTTGTTTTGTTGTATCTTTAATAAGTCA
TCATCCTCGTCATTGCATTATTATTTCAAAATCGTGAAGGAAATCTTACTTAGCTTGAAATTCGCCCATAAAATTTCATTCACTAACAGCTTTGTGGCTTCCTAAGCTACGACAACCATATTATTAGTT
TTTGTTTCCCCACTGCCATATTATTGGTATAAATATATCAGTATCTTACACCGGAAACATCTTTTCACAATTTTTAATGACACCAACTAACTGAATCGCAATAATTAAACTACACTACTGTAACGAAAG
TTAAAATCCTCTATGGGAAACCTCAGGAAGTAGAAGCTAAACTCTAGAATATACGGATCGTCACCATAATTTAGGCATTCTCGTTGGAATTATCTTAGTTGAAAGAAAAGAAAGCAACTAATTTTCTAT
GTTTTATTTCGCATTAAGTGGAAGATTCTTTCAGAAACTTTACTATCAGTAGATCAAATTGTAACGCAAGTGGAAGGTACATAAATTTATCTTCAGTCAATGCTTCCCATATCACCACGTACGACGCAG
CTAGTGCTTGAATATTAATAGGAAAAGGGAAAGGGCCTGCAGTAATTTAGGAAAAATGCATAATAGTTCTCAAGCACTGTAAATCACTTATAGGAATAGTAGTTCGGGAAAGAAAACAAAAGAATTACA
AAAATCATTAGTAAAAGTAAGAAAATAGTTAAGGAAACAAAAGCACAGAAAATAAAGATGTCCCGCTGTTCAGCTAGGACTACTGTTCTGTATGTACAGCTTGTAAATTATATGGACCACATTGCGTAG
CTTGGGCGAACTTAACTAATCTGGCGGTACTTCCAAGCGATAGTTGTGTGTTTACTAACTGAAAAAAAAAAAAAAAAAAAAAAAAAGT


>Ph_gbx_ref [aa] 'Ph_gbx2_ORF58 (repres. consenus of all ORFs!), extended 5' by Ph_gbx2_5R05 (498bp), 3' by Ph_gbx2_3R12 (1243bp)
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF01 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPDSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPTPVHVNRLSNLSQAHQIE
```

```
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF02 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGVPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESPIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF03 [aa] 'w/o 2314, 2315 'stranger seq!
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYEHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIGPVDSDDESHVNGNSAAGL
GGGAPLDSSTTTGPEHRGSGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSAKHRNKSANLSSSFLNSETSESSVAGVTQHSQDQQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELTNLGLSSGRNGSALEAHLLG
HTKEVTEQLKSMRNGGRTSTTSSSSSLQCYFQNHERISP@

>Ph_gbx_ORF04 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSVKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREYQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF05 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTARGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF07 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQARQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF52 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSGGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF53 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNHALPHLYN@CLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF55 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLRPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMGLEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQQRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF56 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEARLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF57 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNSPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSEASESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_ORF58 [aa] 'w/o 2314, 2315
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIE
KCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGH
TKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_5R02 [aa] 'w/o 2306, nonsense, premature stop
MLASSEMTTCDSPPNLPPQKWARTARVLPWNISSPSPVKTKMTAIDQNP&

>Ph_gbx_5R03 [aa] 'w/o 2306
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLRDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQ

>Ph_gbx_5R05 [aa] 'w/o 2306
```

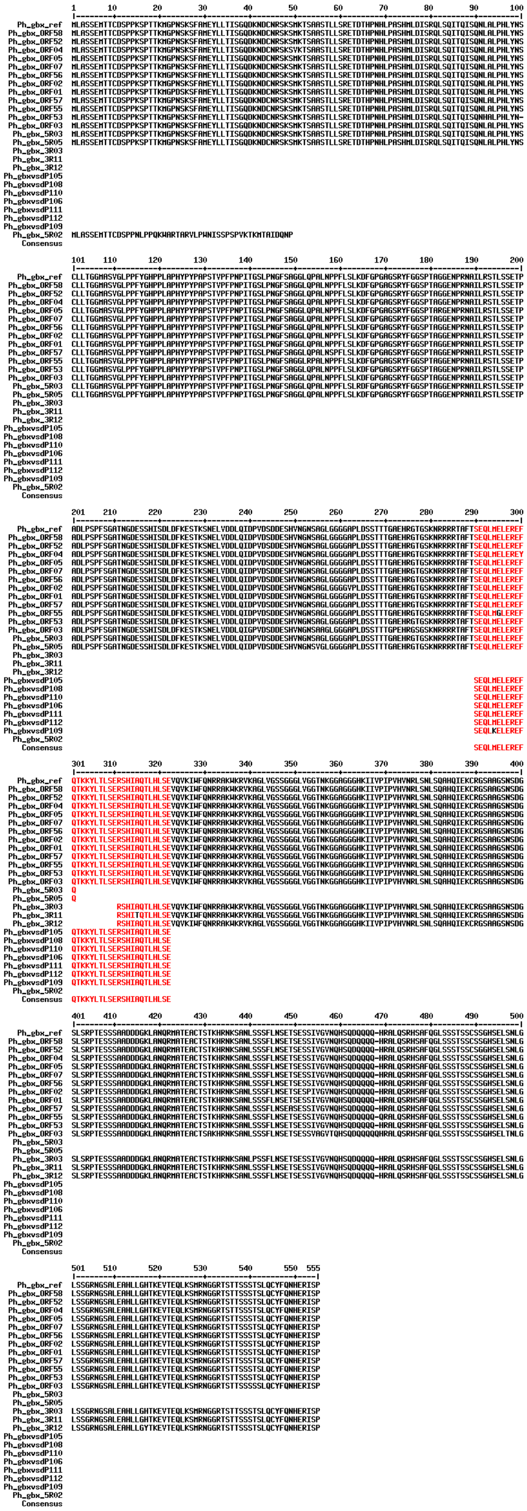
```
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGGMASVGLPPFYGHPPLAPHYPYPA
PSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSVGLG
GGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREFQ

>Ph_gbxvsdP105 [aa] 'w/o fw_AFT/re_KIW
SEQLMELEREFQTKKYLTLSERSHIAQTLHLSE

>Ph_gbxvsdP106 [aa] 'w/o fw_AFT/re_KIW
SEQLMELEREFQTKKYLTLSERSHIAQTLHLSE

>Ph_gbxvsdP108 [aa] 'w/o fw_AFT/re_KIW
SEQLMELEREFQTKKYLTLSERSHIAQTLHLSE

>Ph_gbxvsdP109 [aa] 'w/o fw_AFT/re_KIW
SEQLKELEREFQTKKYLTLSERSHIAQTLHLSE

>Ph_gbxvsdP110 [aa] 'w/o fw_AFT/re_KIW
SEQLMELEREFQTKKYLTLSERSHIAQTLHLSE

>Ph_gbxvsdP111 [aa] 'w/o fw_AFT/re_KIW
SEQLMELEREFQTKKYLTLSERSHIAQTLHLSE

>Ph_gbxvsdP112 [aa] 'w/o fw_AFT/re_KIW
SEQLMELEREFQTKKYLTLSERSHIAQTLHLSE

>Ph_gbx_3R03 [aa] 'w/o 2303
RSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSAN
LPSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_3R11 [aa] 'w/o 2303
RSHITQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSAN
LSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#

>Ph_gbx_3R12 [aa] 'w/o 2303
RSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSAN
LSSSFLNSETSESSIVGVNQHSQDQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGYTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP#
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.10_Ph_gbx]

[PhD_A2.10.2_Ph_gbx_alignments]

**“alignment_Ph_gbx_all_aa_sequences.gif”**



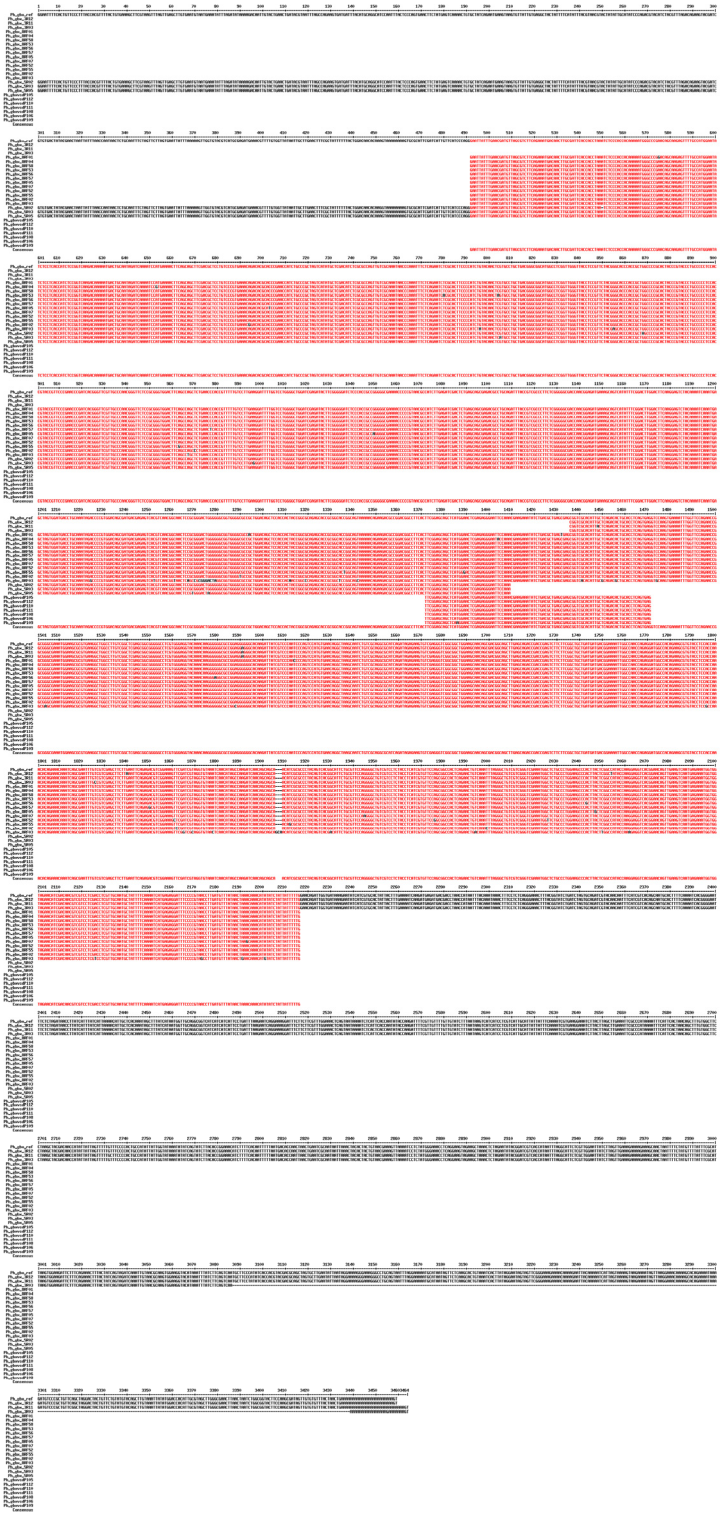
```
              1       10        20        30        40        50        60        70        80        90       100
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF58 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF52 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF04 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSVKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF05 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF07 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF56 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF02 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF01 MLASSEMTTCDSPPKSPTTKMGPDSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF57 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF55 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF53 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNHALPHLYN-
 Ph_gbx_ORF03 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
  Ph_gbx_5R03 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
  Ph_gbx_5R05 MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
  Ph_gbx_3R03
  Ph_gbx_3R11
  Ph_gbx_3R12
Ph_gbxvsdP105
Ph_gbxvsdP108
Ph_gbxvsdP110
Ph_gbxvsdP106
Ph_gbxvsdP111
Ph_gbxvsdP112
Ph_gbxvsdP109
  Ph_gbx_5R02 MLASSEMTTCDSPPNLPPQKWARTARVLPWNISSPSPVKTKMTAIDQNP
    Consensus

              101    110       120       130       140       150       160       170       180       190       200
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF58 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF52 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF04 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF05 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTARGENPRNAILRSTLSSETP
 Ph_gbx_ORF07 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF56 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF02 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF01 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF57 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNSPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF55 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLRPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF53 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF03 CLLTGGMASVGLPPFYEHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
  Ph_gbx_5R03 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLRDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
  Ph_gbx_5R05 CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
  Ph_gbx_3R03
  Ph_gbx_3R11
  Ph_gbx_3R12
Ph_gbxvsdP105
Ph_gbxvsdP108
Ph_gbxvsdP110
Ph_gbxvsdP106
Ph_gbxvsdP111
Ph_gbxvsdP112
Ph_gbxvsdP109
  Ph_gbx_5R02
    Consensus

              201    210       220       230       240       250       260       270       280       290       300
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF58 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF52 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF04 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREY
 Ph_gbx_ORF05 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF07 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF56 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF02 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGVPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF01 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF57 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF55 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMGLEREF
 Ph_gbx_ORF53 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF03 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIGPVDSDDESHVNGNSAAGLGGGAPLDSSTTTGPEHRGSGSKNRRRRTAFTSEQLMELEREF
  Ph_gbx_5R03 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
  Ph_gbx_5R05 ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSVGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
  Ph_gbx_3R03
  Ph_gbx_3R11
  Ph_gbx_3R12
Ph_gbxvsdP105                                                                                          SEQLMELEREF
Ph_gbxvsdP108                                                                                          SEQLMELEREF
Ph_gbxvsdP110                                                                                          SEQLMELEREF
Ph_gbxvsdP106                                                                                          SEQLMELEREF
Ph_gbxvsdP111                                                                                          SEQLMELEREF
Ph_gbxvsdP112                                                                                          SEQLMELEREF
Ph_gbxvsdP109                                                                                          SEQLKELEREF
  Ph_gbx_5R02
    Consensus                                                                                          SEQLMELEREF

              301    310       320       330       340       350       360       370       380       390       400
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF58 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF52 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF04 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF05 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF07 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQARQIEKCRGSAAGSNSDG
 Ph_gbx_ORF56 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF02 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF01 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPTPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF57 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF55 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF53 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF03 QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
  Ph_gbx_5R03 Q
  Ph_gbx_5R05 Q
  Ph_gbx_3R03           RSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
  Ph_gbx_3R11           RSHITQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
  Ph_gbx_3R12           RSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
Ph_gbxvsdP105 QTKKYLTLSERSHIAQTLHLSE
Ph_gbxvsdP108 QTKKYLTLSERSHIAQTLHLSE
Ph_gbxvsdP110 QTKKYLTLSERSHIAQTLHLSE
Ph_gbxvsdP106 QTKKYLTLSERSHIAQTLHLSE
Ph_gbxvsdP111 QTKKYLTLSERSHIAQTLHLSE
Ph_gbxvsdP112 QTKKYLTLSERSHIAQTLHLSE
Ph_gbxvsdP109 QTKKYLTLSERSHIAQTLHLSE
  Ph_gbx_5R02
    Consensus QTKKYLTLSERSHIAQTLHLSE

              401    410       420       430       440       450       460       470       480       490       500
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF58 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF52 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSGGHSELSNLG
 Ph_gbx_ORF04 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF05 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF07 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF56 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF02 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESPIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF01 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF57 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSEASESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF55 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-QRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF53 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF03 SLSRPTESSSAADDDGKLANQRMATEACTSAKHRNKSANLSSSFLNSETSESSVAGVTQHSQDQQQQQHRALQSRHSAFQGLSSSTSSCSSGHSELTNLG
  Ph_gbx_5R03
  Ph_gbx_5R05
  Ph_gbx_3R03 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLPSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
  Ph_gbx_3R11 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
  Ph_gbx_3R12 SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
Ph_gbxvsdP105
Ph_gbxvsdP108
Ph_gbxvsdP110
Ph_gbxvsdP106
Ph_gbxvsdP111
Ph_gbxvsdP112
Ph_gbxvsdP109
  Ph_gbx_5R02
    Consensus

              501    510       520       530       540       550  555
              |--------+---------+---------+---------+---------+----|
   Ph_gbx_ref LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF58 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF52 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF04 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF05 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF07 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF56 LSSGRNGSALEARLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF02 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF01 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF57 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF55 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF53 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF03 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSSSLQCYFQNHERISP
  Ph_gbx_5R03
  Ph_gbx_5R05
  Ph_gbx_3R03 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
  Ph_gbx_3R11 LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
  Ph_gbx_3R12 LSSGRNGSALEAHLLGYTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
Ph_gbxvsdP105
Ph_gbxvsdP108
Ph_gbxvsdP110
Ph_gbxvsdP106
Ph_gbxvsdP111
Ph_gbxvsdP112
Ph_gbxvsdP109
  Ph_gbx_5R02
    Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.10_Ph_gbx]

[PhD_A2.10.2_Ph_gbx_alignments]

## “alignment_Ph_gbx_all_bp_sequences.gif”

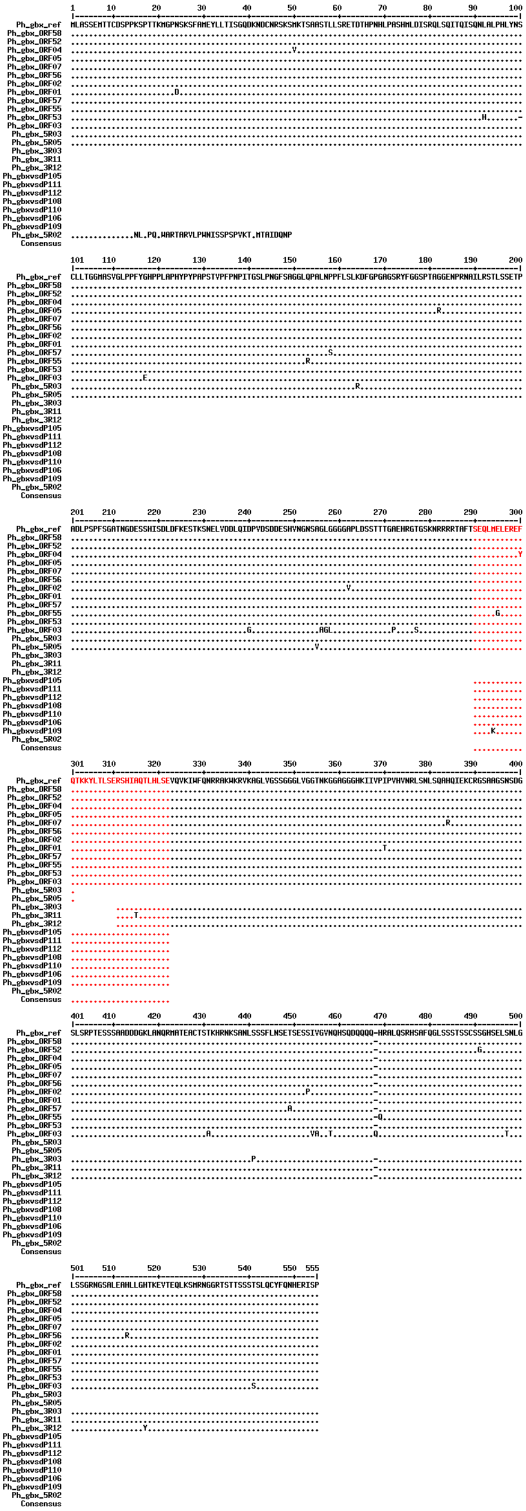


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.10_Ph_gbx]

[PhD_A2.10.2_Ph_gbx_alignments]

**“difference_alignment_Ph_gbx_all_aa_sequences.gif”**



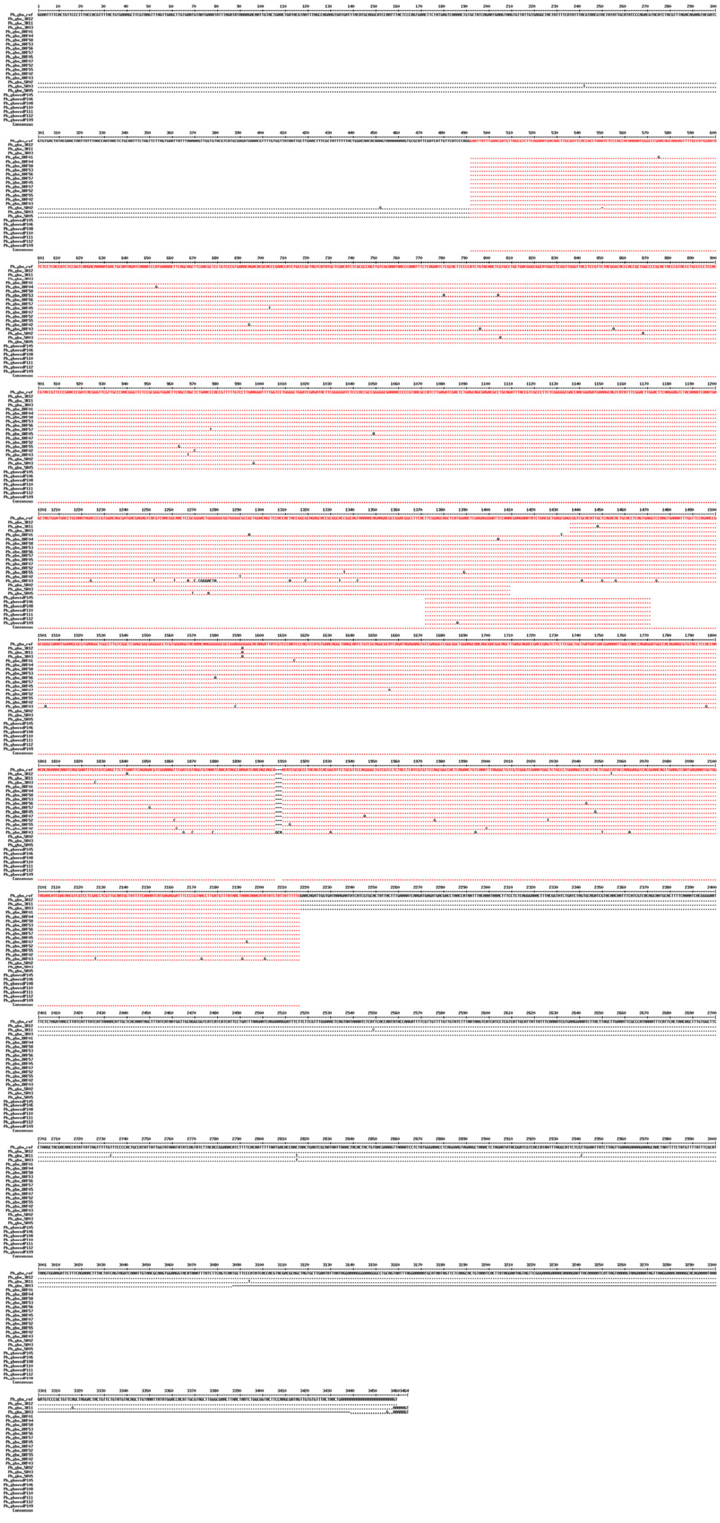
```
              1       10        20        30        40        50        60        70        80        90       100
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNS
 Ph_gbx_ORF58 ....................................................................................................
 Ph_gbx_ORF52 ....................................................................................................
 Ph_gbx_ORF04 .................................................V..................................................
 Ph_gbx_ORF05 ....................................................................................................
 Ph_gbx_ORF07 ....................................................................................................
 Ph_gbx_ORF56 ....................................................................................................
 Ph_gbx_ORF02 ....................................................................................................
 Ph_gbx_ORF01 .......................D............................................................................
 Ph_gbx_ORF57 ....................................................................................................
 Ph_gbx_ORF55 ....................................................................................................
 Ph_gbx_ORF53 ...........................................................................................H.......-
 Ph_gbx_ORF03 ....................................................................................................
  Ph_gbx_5R03 ....................................................................................................
  Ph_gbx_5R05 ....................................................................................................
  Ph_gbx_3R03
  Ph_gbx_3R11
  Ph_gbx_3R12
Ph_gbxvsdP105
Ph_gbxvsdP111
Ph_gbxvsdP112
Ph_gbxvsdP108
Ph_gbxvsdP110
Ph_gbxvsdP106
Ph_gbxvsdP109
  Ph_gbx_5R02 ..............NL.PQ.WARTARVLPWNISSPSPVKT.MTAIDQNP
    Consensus

              101     110       120       130       140       150       160       170       180       190      200
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref CLLTGGMASVGLPPFYGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQPALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPRNAILRSTLSSETP
 Ph_gbx_ORF58 ....................................................................................................
 Ph_gbx_ORF52 ....................................................................................................
 Ph_gbx_ORF04 ....................................................................................................
 Ph_gbx_ORF05 .................................................................................R..................
 Ph_gbx_ORF07 ....................................................................................................
 Ph_gbx_ORF56 ....................................................................................................
 Ph_gbx_ORF02 ....................................................................................................
 Ph_gbx_ORF01 ....................................................................................................
 Ph_gbx_ORF57 .........................................................S..........................................
 Ph_gbx_ORF55 ....................................................R...............................................
 Ph_gbx_ORF53 ....................................................................................................
 Ph_gbx_ORF03 ................E...................................................................................
  Ph_gbx_5R03 ...............................................................R....................................
  Ph_gbx_5R05 ....................................................................................................
  Ph_gbx_3R03
  Ph_gbx_3R11
  Ph_gbx_3R12
Ph_gbxvsdP105
Ph_gbxvsdP111
Ph_gbxvsdP112
Ph_gbxvsdP108
Ph_gbxvsdP110
Ph_gbxvsdP106
Ph_gbxvsdP109
  Ph_gbx_5R02
    Consensus

              201     210       220       230       240       250       260       270       280       290      300
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref ADLPSPFSGATNGDESSHISDLDFKESTKSNELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGAEHRGTGSKNRRRRTAFTSEQLMELEREF
 Ph_gbx_ORF58 ....................................................................................................
 Ph_gbx_ORF52 ....................................................................................................
 Ph_gbx_ORF04 ...................................................................................................Y
 Ph_gbx_ORF05 ....................................................................................................
 Ph_gbx_ORF07 ....................................................................................................
 Ph_gbx_ORF56 ....................................................................................................
 Ph_gbx_ORF02 .............................................................V......................................
 Ph_gbx_ORF01 ....................................................................................................
 Ph_gbx_ORF57 ....................................................................................................
 Ph_gbx_ORF55 ..............................................................................................G.....
 Ph_gbx_ORF53 ....................................................................................................
 Ph_gbx_ORF03 .......................................G...............AGL.............P....S.......................
  Ph_gbx_5R03 ....................................................................................................
  Ph_gbx_5R05 ......................................................V.............................................
  Ph_gbx_3R03
  Ph_gbx_3R11
  Ph_gbx_3R12
Ph_gbxvsdP105                                                                                          ...........
Ph_gbxvsdP111                                                                                          ...........
Ph_gbxvsdP112                                                                                          ...........
Ph_gbxvsdP108                                                                                          ...........
Ph_gbxvsdP110                                                                                          ...........
Ph_gbxvsdP106                                                                                          ...........
Ph_gbxvsdP109                                                                                          ....K......
  Ph_gbx_5R02
    Consensus                                                                                          ...........

              301     310       320       330       340       350       360       370       380       390      400
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref QTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDG
 Ph_gbx_ORF58 ....................................................................................................
 Ph_gbx_ORF52 ....................................................................................................
 Ph_gbx_ORF04 ....................................................................................................
 Ph_gbx_ORF05 ....................................................................................................
 Ph_gbx_ORF07 ...................................................................................R................
 Ph_gbx_ORF56 ....................................................................................................
 Ph_gbx_ORF02 ....................................................................................................
 Ph_gbx_ORF01 .....................................................................T..............................
 Ph_gbx_ORF57 ....................................................................................................
 Ph_gbx_ORF55 ....................................................................................................
 Ph_gbx_ORF53 ....................................................................................................
 Ph_gbx_ORF03 ....................................................................................................
  Ph_gbx_5R03 .
  Ph_gbx_5R05 .
  Ph_gbx_3R03           ..........................................................................................
  Ph_gbx_3R11           ....T.....................................................................................
  Ph_gbx_3R12           ..........................................................................................
Ph_gbxvsdP105 ......................
Ph_gbxvsdP111 ......................
Ph_gbxvsdP112 ......................
Ph_gbxvsdP108 ......................
Ph_gbxvsdP110 ......................
Ph_gbxvsdP106 ......................
Ph_gbxvsdP109 ......................
  Ph_gbx_5R02
    Consensus ......................

              401     410       420       430       440       450       460       470       480       490      500
              |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Ph_gbx_ref SLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQ-HRALQSRHSAFQGLSSSTSSCSSGHSELSNLG
 Ph_gbx_ORF58 ...................................................................-................................
 Ph_gbx_ORF52 ...................................................................-......................G.........
 Ph_gbx_ORF04 ...................................................................-................................
 Ph_gbx_ORF05 ...................................................................-................................
 Ph_gbx_ORF07 ...................................................................-................................
 Ph_gbx_ORF56 ...................................................................-................................
 Ph_gbx_ORF02 ....................................................P..............-................................
 Ph_gbx_ORF01 ...................................................................-................................
 Ph_gbx_ORF57 ................................................A..................-................................
 Ph_gbx_ORF55 ...................................................................-Q...............................
 Ph_gbx_ORF53 ...................................................................-................................
 Ph_gbx_ORF03 ..............................A......................VA..T.........Q............................T...
  Ph_gbx_5R03
  Ph_gbx_5R05
  Ph_gbx_3R03 ........................................P..........................-................................
  Ph_gbx_3R11 ...................................................................-................................
  Ph_gbx_3R12 ...................................................................-................................
Ph_gbxvsdP105
Ph_gbxvsdP111
Ph_gbxvsdP112
Ph_gbxvsdP108
Ph_gbxvsdP110
Ph_gbxvsdP106
Ph_gbxvsdP109
  Ph_gbx_5R02
    Consensus

              501     510       520       530       540       550  555
              |--------+---------+---------+---------+---------+----|
   Ph_gbx_ref LSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
 Ph_gbx_ORF58 .......................................................
 Ph_gbx_ORF52 .......................................................
 Ph_gbx_ORF04 .......................................................
 Ph_gbx_ORF05 .......................................................
 Ph_gbx_ORF07 .......................................................
 Ph_gbx_ORF56 ............R..........................................
 Ph_gbx_ORF02 .......................................................
 Ph_gbx_ORF01 .......................................................
 Ph_gbx_ORF57 .......................................................
 Ph_gbx_ORF55 .......................................................
 Ph_gbx_ORF53 .......................................................
 Ph_gbx_ORF03 ........................................S..............
  Ph_gbx_5R03
  Ph_gbx_5R05
  Ph_gbx_3R03 .......................................................
  Ph_gbx_3R11 .......................................................
  Ph_gbx_3R12 ................Y......................................
Ph_gbxvsdP105
Ph_gbxvsdP111
Ph_gbxvsdP112
Ph_gbxvsdP108
Ph_gbxvsdP110
Ph_gbxvsdP106
Ph_gbxvsdP109
  Ph_gbx_5R02
    Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.10_Ph_gbx]

[PhD_A2.10.2_Ph_gbx_alignments]

## “difference_alignment_Ph_gbx_all_bp_sequences.gif”



## [PhD_A2_Parhyale_goi_sequences]

### [PhD_A2.11_Ph_kni1]

### [PhD_A2.11.1_Ph_kni1_sequences]

## “Ph_kni1_general_read_me.txt”

```
General aspects:

(i)     All sequences are in FASTA format.

(ii)    [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)   Comments are preceded by an apostrophe (')

(iv)    All sequences without olignonucleotide sequence are indicated (e.g. "w/o 1913, 1919"). In these cases, only novel Parhyale
sequence is presented.


Parhyale hawaiensis kni1 terminology:


(i)     In general, sequence denotation begins with: ">Ph_kni1_"

(ii)    The above (i) is followed by:   "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                        "5R" or "3R",       in case sequence derives from 5' or 3' RACE, respectively
                                        "ORF"               for coherent ORF sequences
                                        "cDNA"              for complete cDNA sequences

                                        "ref"               if reference sequence is presented
                                        "ZNFD"              if Zinc finger domain (encoding) fragment is presented
                                        "KNID"              if knirps box/ knirps domain is presented
                                        "PIDLS"             if PIDLS motif (encoding fragment) is presented

(iii)   The above (ii) is followed by clone id.


Parhyale hawaiensis kni1 synopsis:


cDNA:   2228 bp
5'UTR:   698 bp (1-698)
ORF:    1194 bp (699-1892),  397 aa
3'UTR:   336 bp (1893-2228)
ZNFD:    198 bp (729-926),    66 aa
KNID:     57 bp (945-1001),   19 aa
PIDLS:    36 bp (1716-1751),  12 aa

>Ph_kni1_ZNFD [bp]
CAGTTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGA
GGCCAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTC

>Ph_kni1_KNID [bp]
GGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAG

>Ph_kni1_PIDLS [bp]
CCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAG

>Ph_kni1_ref [bp]
AGTTGTGCTCGTGCTGGCCGAAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTACACATTGATATCATTTTGAGGAAATTCA
CCGTTACCCTAATTTCATAATCTGGGAGCAGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAAA
TCGCCTAACTGTACTGAAGCCGCCATGAAGAGCAAATGTCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCGACTAGAATATCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTCT
GTTGAACTATTAACCAAGGCATCGCTAATATAAAGTGCCAGTGTAGGAAAAAAAGTTACTGTTGAACCAAACTTACAGTCTACAAAACTTTTATAGAAAGATAATTGTTTATAATTTTGTGTTGAATTG
AAAACAATAGGACAACATCCTTGAAATTCAGCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAATTCAAACTTTGAAATGCTGTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGCA
CAGTTTTTTGAAAATTAAGTCAAGTGTCACTAAGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTG
GAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTT
GTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGG
GTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGT
CGGACACAAGTGTAGATGTCGAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAA
ATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCAGTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGC
GTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAAATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTC
CTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTCTCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGC
CGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTTTCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTC
CGCTGCCTTTTAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTTCCAATGCTTGATAGTTTCGGAATCGGTGATGCAAGAGCGTATGTATGTCTGCTT
GACGCTTTTTCATTGCATTGGATGCATATCGCTAGCAAAAGTCAAGGTCTCTTGTCACCAAGATTAGGTGACTGAAATTACAGCACTGCTCACATCAAAGACTGAAACAACGTCGAGAAATTCGTGCTG
TAAATAGATGCTGTACTGTTACTGAGTAGCACAACCCGATAGCAAACTATTAGCCAGTTAAAAATTCTAAAATTGAGAATTCTACTAGTCCTAAGCTTGCACCCTACTGTTTTCAGAAGCCTCAGGGTG
CAAGAAGTTACCTTGTATACTTCTGGTCGTGCCAG


>Ph_kni1_ZNFD [aa]
QLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLF
```

```
>Ph_kni1_KNID [aa]
GSRYGRRSNWFKIHCALQE

>Ph_kni1_PIDLS [aa]
PEQDAPIDLSVK

>Ph_kni1_ref [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA&
```

[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.11_Ph_kni1]
[PhD_A2.11.1_Ph_kni1_sequences]

## **“Ph_kni1_cDNA_sequences.txt”**

```
>Ph_kni1_ref [bp] 'cDNAJZB5 w 1913, 1919
AGTTGTGCTCGTGCTGGCCGAAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTACACATTGATATCATTTTGAGGAAATTCA
CCGTTACCCTAATTTCATAATCTGGGAGCAGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAAA
TCGCCTAACTGTACTGAAGCCGCCATGAAGAGCAAATGTCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCGACTAGAATATCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTCT
GTTGAACTATTAACCAAGGCATCGCTAATATAAAGTGCCAGTGTAGGAAAAAAAGTTACTGTTGAACCAAACTTACAGTCTACAAAACTTTTATAGAAAGATAATTGTTTATAATTTTGTGTTGAATTG
AAAACAATAGGACAACATCCTTGAAATTCAGCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAATTCAAACTTTGAAATGCTGTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGCA
CAGTTTTTTGAAAATTAAGTCAAGTGTCACTAAGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTG
GAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTT
GTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGG
GTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGT
CGGACACAAGTGTAGATGTCGAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAA
ATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCAGTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGC
GTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAAATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTC
CTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTCTCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGC
CGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTTTCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTC
CGCTGCCTTTTAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTTCCAATGCTTGATAGTTTCGGAATCGGTGATGCAAGAGCGTATGTATGTCTGCTT
GACGCTTTTTCATTGCATTGGATGCATATCGCTAGCAAAAGTCAAGGTCTCTTGTCACCAAGATTAGGTGACTGAAATTACAGCACTGCTCACATCAAAGACTGAAACAACGTCGAGAAATTCGTGCTG
TAAATAGATGCTGTACTGTTACTGAGTAGCACAACCCGATAGCAAACTATTAGCCAGTTAAAAATTCTAAAATTGAGAATTCTACTAGTCCTAAGCTTGCACCCTACTGTTTTCAGAAGCCTCAGGGTG
CAAGAAGTTACCTTGTATACTTCTGGTCGTGCCAG

>Ph_kni1_cDNAJZB1 [bp] 'w/o 1913, 1919
AAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTACACATTGATATCATTTTGAGGAAATTCACCGTTACCCTAATTTCATAA
TCTGGGAGCAGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAAATCGCCTAACTGTACTGAAGC
CGCCATGAAGAGCAAGTGCCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCGACTAGAATGTCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTCTGTTGAACTATTAACCAAGGC
ATCGCTAATATAAAGTGCCTGTGTAGGAAAAAAAGTTACTGTTGAACTAACCTTACAGTCTACAAAACTTTTATAAAAAGATAATTGTTTATAATTTTGTGTTGAATTGAAAACAATAGGACGTCATCC
TTGAAATTCAGCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAATTCAAACTTTGAAATGCTGTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGCACAGTTTTTTGAAAATTAAGT
CAAGTGTCACTAAGCAGAAACCCTCGAGTCTACATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCCGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGC
TGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTT
TTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTG
CAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGATACAAGTGTAGATGTC
GAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCACAACTAACACTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGA
GGGAGAAGTCTGGAGCCACCATTTCCCGGCACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTG
TACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAAATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCC
CCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTCTCGGACTGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAA
CGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTTTCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTCAATTTTTTA
CATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTTCCAATGCTTGATAGTTTCGGAATCGGTGATGCAAGAGCGTATGTATGTCTGCTTGACGCTTTTTCGTTGCATTG
GATGCATATCGCTAGCAAAAGTCGAGGTCTTTTGTCACCAAGATTAGGTGACTGAAATTACAGCACTGCTCACATCAAAGACTGAAACAACGTCGAGAAATTCGTGCTGTAAATGGATGCTGTACTGTT
ACCGAGTAGCACAACCCGATAGCAAACTATTAGCCAATTAAAAATTCTAAAATTGAGAATTCTACTAGTCCTAAGCTTGCACCCTACTGTTTTCAGAAGCCTCAGGGTGCAAGAA

>Ph_kni1_cDNAJZB2 [bp] 'w/o 1913, 1919
AAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTACACATTGATATCATTTTGAGGAAATTCACCGATACCCTAATTTCATAA
TCTGGGAGCAGGAATTTATCGCCAGCTACAGAGTTGGCGTGTATGTCTTTCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGCAGCCGCAAAAATCGCCTAACTGTTCTGAACC
CGCCATGAAGAGCAAGTGACCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCATCATTCGACAAGAATGCCACTCTGTTGAATGTCAAGTTGTGCTGAGTGTCTGTTGAACTATTAACCAAGGC
ATCACTAATATAAAGTGCCAGTGTATTGAAAAAAGTTACTGTTGAACTAACCTTACAGTCTACAAAACTTGTATAAAAAGATAATTGTTTGTAATTTTGTGTTGATTTGAAAACCAAAGGACATCATCC
TTGAAATTCAGCAATTATTTTTTCACGCTGTTTAAGACTTATTACTGACAGTAACTCAAGCTTAGAAATGCTGTTGTCATATCTTGGGCAGCTGTAGTTATTTGAAGCACAGTTTTTGAAAATTAAATC
AAGTGTCACTAAGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCCGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCT
GCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCTCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTT
TCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGC
AGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGATACAAGTGTAGATGTCG
AACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAG
GGAGAAGTCTGGAGCCACCAGTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGT
ACCCAGGTATCCGGCCTGTAGTGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAAATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCC
CGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTCTCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAAC
GAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTTTCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTCAATTTTTTAC
ATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTTCCAATGCTTGATAGTTTCGGAATCGGTGATGCAAGAGCGTATGTATGTCTGCTTGACGCTTTTTCGTTGCATTGG
ATGCATATCGCTAGCAAAAGTCAAGGTCTTTTGTCACCAAGATTAGGTGACTGAAATTACAGCACTGCTCACATCAAAGACTAAAACAACGTCGAGAAATTCGTGCTGTAAATGGATGCTGTACTATTA
CTGAGTAGCACAACCCGATAGCAAACTATTAGCCAATTAAAAATTCTAAAATTGAGAATTCTACTAGTCCTAAGCTTGCACCCTACTGTTTTCAGAAGCCTCAGGGTGCAAGAA

>Ph_kni1_cDNAJZB5 [bp] 'w/o 1913, 1919
AAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTACACATTGATATCATTTTGAGGAAATTCACCGTTACCCTAATTTCATAA
TCTGGGAGCAGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAAATCGCCTAACTGTACTGAAGC
CGCCATGAAGAGCAAATGTCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCGACTAGAATATCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTCTGTTGAACTATTAACCAAGGC
ATCGCTAATATAAAGTGCCAGTGTAGGAAAAAAAGTTACTGTTGAACCAAACTTACAGTCTACAAAACTTTTATAGAAAGATAATTGTTTATAATTTTGTGTTGAATTGAAAACAATAGGACAACATCC
TTGAAATTCAGCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAATTCAAACTTTGAAATGCTGTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGCACAGTTTTTTGAAAATTAAGT
CAAGTGTCACTAAGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGC
TGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTT
TTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTG
CAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGACACAAGTGTAGATGTC
GAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGA
GGGAGAAGTCTGGAGCCACCAGTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTG
TACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAAATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCC
CCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTCTCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAA
CGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTTTCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTTAATTTTTTA
CATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTTCCAATGCTTGATAGTTTCGGAATCGGTGATGCAAGAGCGTATGTATGTCTGCTTGACGCTTTTTCATTGCATTG
```

```
GATGCATATCGCTAGCAAAAGTCAAGGTCTCTTGTCACCAAGATTAGGTGACTGAAATTACAGCACTGCTCACATCAAAGACTGAAACAACGTCGAGAAATTCGTGCTGTAAATAGATGCTGTACTGTT
ACTGAGTAGCACAACCCGATAGCAAACTATTAGCCAGTTAAAAATTCTAAAATTGAGAATTCTACTAGTCCTAAGCTTGCACCCTACTGTTTTCAGAAGCCTCAGGGTGCAAGAA

>Ph_kni1_ORF02 [bp] 'w/o 1914, 1915
TTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGCGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGC
CAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCCTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGC
GCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGC
AAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGACACAAGTGTAGATGTCGAACCTAGAAGAAATGTATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAA
CACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCAGTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATT
CACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAA
ATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTC
TCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTT
TCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTTAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACTTGTAGACTTGACTT

>Ph_kni1_ORF03 [bp] 'w/o 1914, 1915
TTGTGCAAGGTGTGCTCCGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGC
CAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGCAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGC
GCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGC
AGGCTTGGTTCCCCAGCACTTTCATCTCCTGGTAGCAACACGTCGGATACAAGTGTAGATGTCGAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAA
CACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCATTTCCCGGCACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATT
CACTCCGACTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAA
ATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAACCTTTACCTGCCCGTCATTTC
TCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTT
TCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTCAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTT

>Ph_kni1_ORF04 [bp] 'w/o 1914, 1915
TTGTGCAAGGTGTGCTCCGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAGCAACTTATCCCAAATTCATGAATGCAAGAATGGAGGC
CAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGC
GCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGC
AAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGATACAAGTGTAGATGTCGAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAA
CACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCATTTCCCGGCACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATT
CACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAA
ATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTC
TCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTT
TCGTCAACCACTCCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTCAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTT

>Ph_kni1_ORF07 [bp] 'w/o 1914, 1915
TTGTGCAAGGTGTGCTCCGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGAGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGC
CAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGC
GCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGC
AAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGATACAAGGGTAGATGTCGAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAA
CACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCATTTCCCGGCACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATT
CACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAA
ATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTC
TCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGGGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTT
TCGTCAACCCCTTCAGTTCCATCGCCTGATCTCACCCCTATTCCGCTGCCTTTCAATTTTTTACATTTAGCGGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTT

>Ph_kni1_ORF08 [bp] 'w/o 1914, 1915
TTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCTCAAATTCATGAATGCAAGAATGGAGGC
CAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGC
GCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGC
AAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGATACAAGTGTAGATGTCGAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAA
CACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCAGTTCCCGGCACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATT
CACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCTGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAA
ATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTC
TCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTT
TCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTCAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTT

>Ph_kni1_ORF11 [bp] 'w/o 1914, 1915
TTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGC
CAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGC
GCACTCCAAGAGCAAGCGGCCAACCTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGC
AAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGACACAAGTGTAGATGTCGAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGTCCAA
CACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCAGTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATT
CACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAA
ATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTC
TCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTT
TCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTTAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACTT

>Ph_kni1_ORF12 [bp] 'w/o 1914, 1915
TTGTGCAAGGTGTGCTCCGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCTCAAATTCATGAATGCAAGAATGGAGGC
CAGTGTGTGATCAACAAGCAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGC
GCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGC
AAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGATACAAGTGTAGATGTCGAACCTAGGAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAA
CACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCAGTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATT
CACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAA
ATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCGTCATTTC
TCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTT
TCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTCAATTTTTTACATTTAGCAGCGTCATCAACATCTTCAGCACCTTCGGTACCTGTAGACTTGACTT

>Ph_kni1_5R25 [bp] 'w/o 1906; all seq w/o SMART(like)seq 5' + w/o primer seq:
GAGTTGTGCTCGTGCTGGCCGAAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTACACATTGATATCATTTTGAGGAAATTC
ACCGTTACCCTAATTTCATAATCTGGGAGCAGGAATTTATCGCCCCCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAA
ATCGCCTAACTGTACCGAAGCCGCCATGAAGAGCAAGTGTCCAAGTACCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCGACTAGAATGTCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTC
TGTTGAACTATTAACCAAGGCATCGCTAATATAAAGTGCCTGTGTAGGAAAAAAAGTTACTGTTGAACTAACCTTACAGTCTACAAAACTTTTATAAAAAGATAATTGTTTATAATTTTGTGTTGAATT
GAAAACAATAGGACATCATCCTTGAAATTCAGCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAATTCAAACTTTGAAATGCTGTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGC
ACAGTTTTTTGAAAATTAAGTCAAGTGTCACTAGGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCCGAGCCGGCAGCAGGATTCCACTTT
GGAGCTTTAA

>Ph_kni1_5R26 [bp] 'w/o 1906
```

```
GAGTTGTGCTCGTGCTGGCCGAAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTACACATTGATATCATTTTGAGGAAATCC
ACCGTTACCCTAATTTCATAATCTGGGAGCAGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAA
ATCGCCTAACTGTACTGAAGCCGCCATGAAGAGCAAATGTCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCGACTAGAATATCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTC
TGTTGAACTATTAACCAAGGCATCGCTAATATAAAGTGCCAGTGTAGGAAAAAAAGTTACTGTTGAACCAAACTTACAGTCTACAAAACTTTTATAAAAAGATAATTGTTTATAATTTTGTGTTGAATT
GAAAACAATAGGACAACATCCTTGAAATTCAGCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAATTCAAACTTTGAAATGCTGTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGC
ACAGTTTTTTGAAAATTAAGTCAAGTGTCACTAGGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTT
GGAGCTTTTAC

>Ph_kni1_5R27 [bp] 'w/o 1906
GAGTTGTGCTCGTGCTGGCCGAAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTACACATTGATATCATTTTGAGGAAATTC
ACCGTTACCCTAATTTCATAATCTGGGAGCAGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAA
ATCGCCTAACTGTACTGAAGCCGCCATGAAGAGCAAGTGTCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCGACTAGAATGTCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTC
TGTTGAACTATTAACCAAGGCATCGCTAATATAAAGTGCCTGTGTAGGAAAAAAAGTTACTGTTGAACTAACCTTACAGTCTACAAAACTTTTATAAAAAGATAATTGTTTATAATTTTGTGTTGAATT
GAAAACAATAGGACATCATCCTTGAAATTCAGCAATTATTTTTTCACACAGTTAAAGACTCAATACTGACAGTAATTCAAACTTTGAAATGCTGTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGC
ACAGTTTTTTGAAAATTAAGTCAAGTGTCACTAAGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCCGAGCCGGCAGCAGGATTCCACTTT
GGAGCTTTTAC

>Ph_kni1_5R30 [bp] 'w/o 1906
ATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTGGACATTGATATCATTTTGAGGAAATTCACCGTTACCCTAATTTCATAATCTGGGAGC
AGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAAATCGCCTAACTGTACTGAAGCCGCCATGAA
GAGCAAGTGTCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCAACTAGAATGTCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTCTGTTGAACTATTAACCAAGGCATCGCTAAT
ATAAAGTGCCAGTGTAGGAAAAAAAGTTACTGTTGAACCAAACTTACAGTCTACAAAACTTTTATAAAAGATAATTGTTTATAATTTTGTGTTGAATTGAAAACAATAGGACATCATCCTTGAAATTCA
GCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAACTCAAGCTTAGAAATGCTGTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGCACAGTTTTTGAAAATTAAGTCAAGTGTCACT
AAGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTAC

>Ph_kni1_5R31 [bp] 'w/o 1906
CAGTTGTGCTCGTGCTGGCCGAAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTGGACATTGATATCATTTTGAGGAAACTC
ACCGTTACCCTAATTTGATAATCTGGGAGCAGGAATTTATCGCCAGCCACAGAGTTGGCGTGTATGTCTTTCCTTGATTGCCACTAAATGTAATTGAACGTATTTTTGACACGGATGGCAGCCGCAAAA
ATCGCCTAACTGTTCTGAACCCGCCATGAAGAGCAAGTGACCAAGTGCCCCAAGAAGCTTGCACCTTTTTGCTGGATCGTCATTCGACAAGAATGCCACTCTGTTGAATGTCAAGTTGTGCTGAGTGTC
TGTTGAACTATTAACCAAGGCATCACTAATATAAAGTGCCAGTGTATTGAAAAAAGTTACTGTTGAACTAACCTTACAGTCTACAAAACTTGTATAAAAAGATAATTGTTTGTAATTTTGTGTTGATTT
GAAAACCAAAGGACATCATCCTTGAAATTCAGCAATTATTTTTTCACGCTGTTTAAGACTTATTACTGACAGTAACTCAAGCTTAGAAATGCTGTTGTCATATCTTGGGCAGCTGTAGTTATTTGAAGC
ACAGTTTTTGAAAATTAAATCAAGTGTCACTAAGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCCGAGCCAGCAGCAGGATTCCACTTTG
GAGCTTTTAC

>Ph_kni1_5R33 [bp] 'w/o 1906
GAGTTGTGCTCGTGCTGGCCGAAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTGGACGTTGATATCATTTTGAGGAAATTC
ACCGTTACCCTAATTTCATAATCTGGGAGCAGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAA
ATCGCCTAACTGTACTGAAGCCGCCATGAAGAGCAAGTGTCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCAACTAGAATGTCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTC
TGTTGAACTATTAACCAAGGCATCGCTAATATAAAGTGCCAGTGTAGGAAAAAAAGTTACTGTTGAACCAAACTTACAGTCTACAAAACTTTTATAAAAGATAATTGTTTATAATTTTGTGTTGAATTG
AAAACAATAGGACATCATCCTTGAAATTCAGCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAACTCAAGCTTAGAAATGCTGTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGCA
CAGTTTTTGAAAATTAAGTCAAGTGTCACTAAGCAGAAATCCTCGAGTCTACATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTGG
AGCTTTTAC

>Ph_kni1_5R36 [bp] 'w/o 1906
GAGTTGTGCTCGTGCTGGCCGAAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCAGGATATCTACTGGATGCAATAGCATTGCAGTTGGACATTGATATCATTTTGAGGAAATTC
ACCGTTACCCTAATTTCATAATCTGGGAGCAGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTAACCGCAAAA
ATCGCCTAACTGTACTGAAGCCGCCATGAAGAGCAAGTGTCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTCATTCAACTAGAATGTCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTC
TGTTGAACTATTAACCAAGGCATCGCTAATATAAAGTGCCAGTGTAGGAAAAAAAGTTACTGTTGAACCAAACTTACAGTCTACAAAACTTTTATAAAAGATAATTGTTTATAATTTTGTGTTGAATTG
AAAACAATAGGACATCATCCTTGAAATTCAGCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAACTCAAGCTTAGAAATGCTGTCGTCATATCTTGAGCAGCTGTAGTTATTTGAAGCA
CAGTTTTTGAAAATTAAGTCAAGTGTCACTAAGCAGAAATCCTCGAGTCTGCATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTGG
AGCTTTTAC

>Ph_kni1_vsdP07 'w/o fw_CKV, re_LLQ
AGGATTCCACTTTGGAGCTTTTAAGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGTCAAAGTGTGATCAACAAGCAAAATCGAAC
TTTGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAAC

>Ph_kni1_vsdP09 'w/o fw_CKV, re_LLQ, ref seq for vsdPCR seqs kni1
AGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAAC
TTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAAC

>Ph_kni1_vsdP14 'w/o fw_CKV, re_LLQ
AGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCTCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAAC
TTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAAC

>Ph_kni1_vsdP16 'w/o fw_CKV, re_LLQ
AGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAAC
TTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGGAGATCAAAC

>Ph_kni1_vsdP18 'w/o fw_CKV, re_LLQ
AGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAAC
TTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAAC

>Ph_kni1_vsdP23 'w/o fw_CKV, re_LLQ
AGGATTCCACTTTGGAGCTTTTACGTGTGGAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAAC
TTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAAC

>Ph_kni1_vsdP24 'w/o fw_CKV, re_LLQ
AGGATTCCACTTTGGAGCTTTTACGTGTGAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAGCAAAATCGAAC
TTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCAAAC

>Ph_kni1_3R41 [bp] 'w/o 1901
GGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAA
AACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGATACAAGTGTAGATGTCGAACCTAGAAGAAATATATCT
GGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCA
GTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTA
ATGGATCCAATTGCACCACCTCAACAATCGCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_kni1_3R43 [bp] 'w/o 1901
GGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAA
AACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGATACAAGTGTAAATGTCGAACCTAGAAGGAATATATCT
```

```
GGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCA
GTCCCCGGCACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTA
ATGGACTCAATTGCACCACCTCAACAATCGCCATCACCTACTCAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_kni1_3R44 [bp] 'w/o 1901
GGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTGAGCGCAGCACGGGAA
AACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCGACACGTCGGATACAAGTGTAGATGTCGAACCTAGAAGAAATATATCT
GGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCA
GTTCCCGGCACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTA
ATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_kni1_3R45 [bp] 'w/o 1901
GGATCAAGATATGGGAGAAGATCAAACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGCGCTGGGAATATCAATCTGCAGCTGAGCGCAGCACGGGAA
AACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCTGATAGCAACACGTCGGATACAAGTGCAGATGTCGAACCTAGAAGAAATATATCT
GGTACTTCTATTCAGACTTCGCATTCGCAACTAACACTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCACCA
GTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAACTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAGCGTTTATACCCACCGTTTTTGTATCCAGGTATCCGGCCTGTA
ATGGACCCAATTGCACCACCTCAACAATCGCCCAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_kni1_3Rfwxn08 [bp] 'w/o 1904
TACTCAAATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCACCTTCACCTCAATCTTTACCTGCCCG
TCATTTCTCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATCGATTTAAGCGTGAAGCGACC
GCGCTTTTCGTCAACCACTTCAGTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTCAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACTTGACCTCCAA
TGCTTGATAGTTTCGGAATCGGTGATGCAAGAGCGTATGTATGTCTGCTTGACGCTTTTTCGTTGCATTGGATGCATATCGCTAGCAAAAGTCAAGGTCTTTTGTCACCAAGATTAGGTGACTGAAATT
ACAGCACTGCTCACATCAAAGACTAAAACAACGTCGAGAAATTCGTGCTGTAAATGGATGCTGTACTATTACTGAGTAGCACAACCCGATAGCAAACTATTAGCCAAAAAAAAAAAAAAAAAAAAAAAA
AAAAA

>Ph_kni1_3Rfwxn10 [bp] 'w/o 1904
TACTCAAATAAAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAACCCATCTTCACCTCAATCTTTACCTGCCCG
TCATTTCTCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATTTCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACC
GCGCTTTTCGTCAACCACTTCAGTTCCATTGCCTGACCTCACCCCTATTCCGCTGCCTTTTAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCTGTAGACCTGACTTCCAA
TGCTTGATAGTTTCGGAATCGGTGATGCAAGAGCGTATGTATGTCTGCTTGACGCTTTTTCATTGCATTGGATGCATATCGCTAGCAAAAGTCAAGGTCTTTTGTCACCAAGATTAGGTGACTGAAATT
ACAGCACTGCTCACATCAAAGACTGAAACAACGTCGAGAAATTCGTGCTGTAAATAGATGCTGTACTGTTACTGAGTAGCACAACCCGATAGCAAACTATTAGCCAATTAAAAATTCTAAAATTGAGAA
TTCTACTAGTCCTAAGCTTGCACCCTACTGTTTTCAGAAGCCTCAGGGTGCAAGAAGTTACCTTGTATACTTCTGGTCGTGCCAGTATTTTTTTCCAATTACACATAACCCATTCAGAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAA


>Ph_kni1_ref [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA&

>Ph_kni1_cDNAJZB1 [aa] 'pre-stop!
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPFPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSD&TAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA&

>Ph_kni1_cDNAJZB2 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVVDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA&

>Ph_kni1_cDNAJZB5 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA&

>Ph_kni1_ORF02 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCPFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNVSGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVLVDLTSNA&

>Ph_kni1_ORF03 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGRLGSPALSSPGSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPFPGTDFVAGLSLYGIHSDFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQPLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA

>Ph_kni1_ORF04 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYSNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPFPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTPVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA

>Ph_kni1_ORF07 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFERTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTRVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPFPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPGQDAPIDLSVKRPRFSSTPSVPSPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA

>Ph_kni1_ORF08 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIWPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA
```

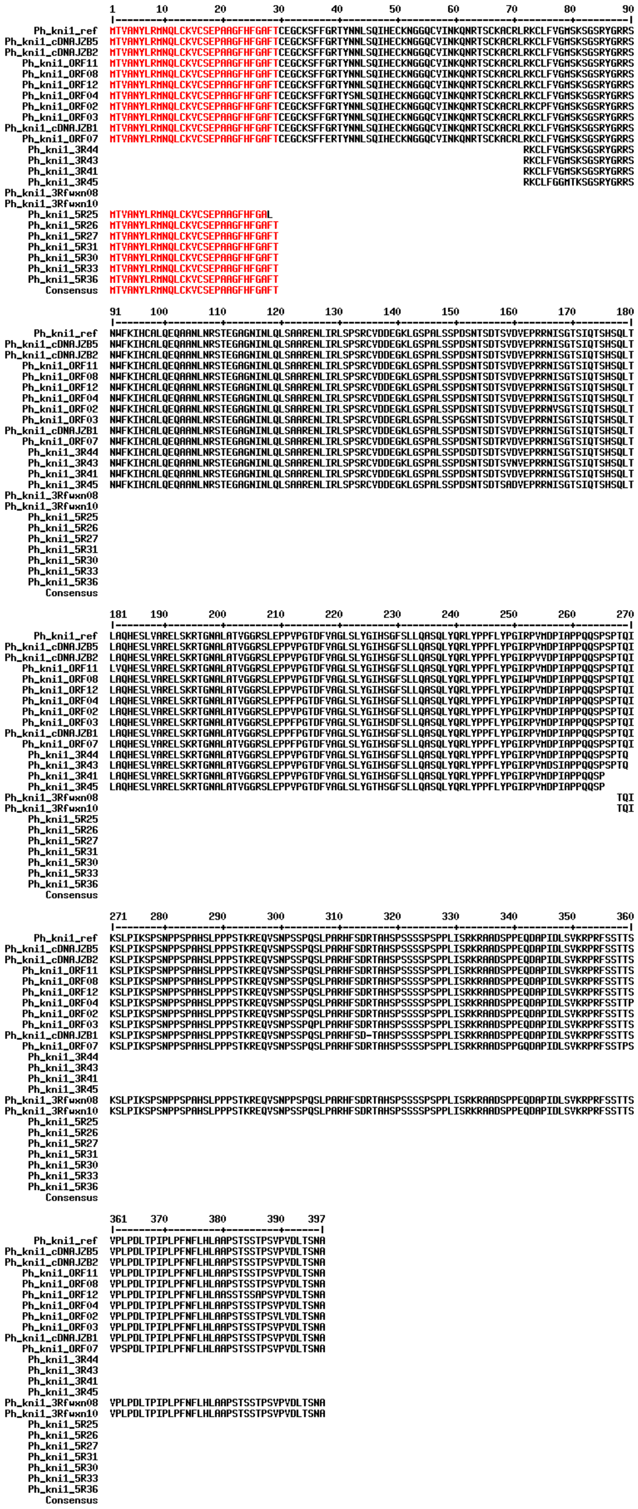
```
>Ph_kni1_ORF11 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLVQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTP
SVPVDLTSNA

>Ph_kni1_ORF12 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIR
LSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPI
APPQQSPSPTQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAASSTSSAP
SVPVDLTSNA

>Ph_kni1_5R25 [aa] 'all seq w/o primer seq:
MTVANYLRMNQLCKVCSEPAAGFHFGAL

>Ph_kni1_5R26 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFT

>Ph_kni1_5R27 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFT

>Ph_kni1_5R30 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFT

>Ph_kni1_5R31 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFT

>Ph_kni1_5R33 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFT

>Ph_kni1_5R36 [aa]
MTVANYLRMNQLCKVCSEPAAGFHFGAFT

>Ph_kni1_3R41 [aa]
RKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNAL
ATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSP

>Ph_kni1_3R43 [aa]
RKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVNVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNAL
ATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDSIAPPQQSPSPTQ

>Ph_kni1_3R44 [aa]
RKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSDTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNAL
ATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQ

>Ph_kni1_3R45 [aa]
RKCLFGGMTKSGSRYGRRSNWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSADVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNAL
ATVGGRSLEPPVPGTDFVAGLSLYGTHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSP

>Ph_kni1_3Rfwxn08 [aa]
TQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPPSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSN
A&

>Ph_kni1_3Rfwxn10 [aa]
TQIKSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTSVPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSN
A&
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.11_Ph_kni1]

### [PhD_A2.11.2_Ph_kni1_alignments]

#### **“alignment_Ph_kni1_all_aa_sequences.gif”**



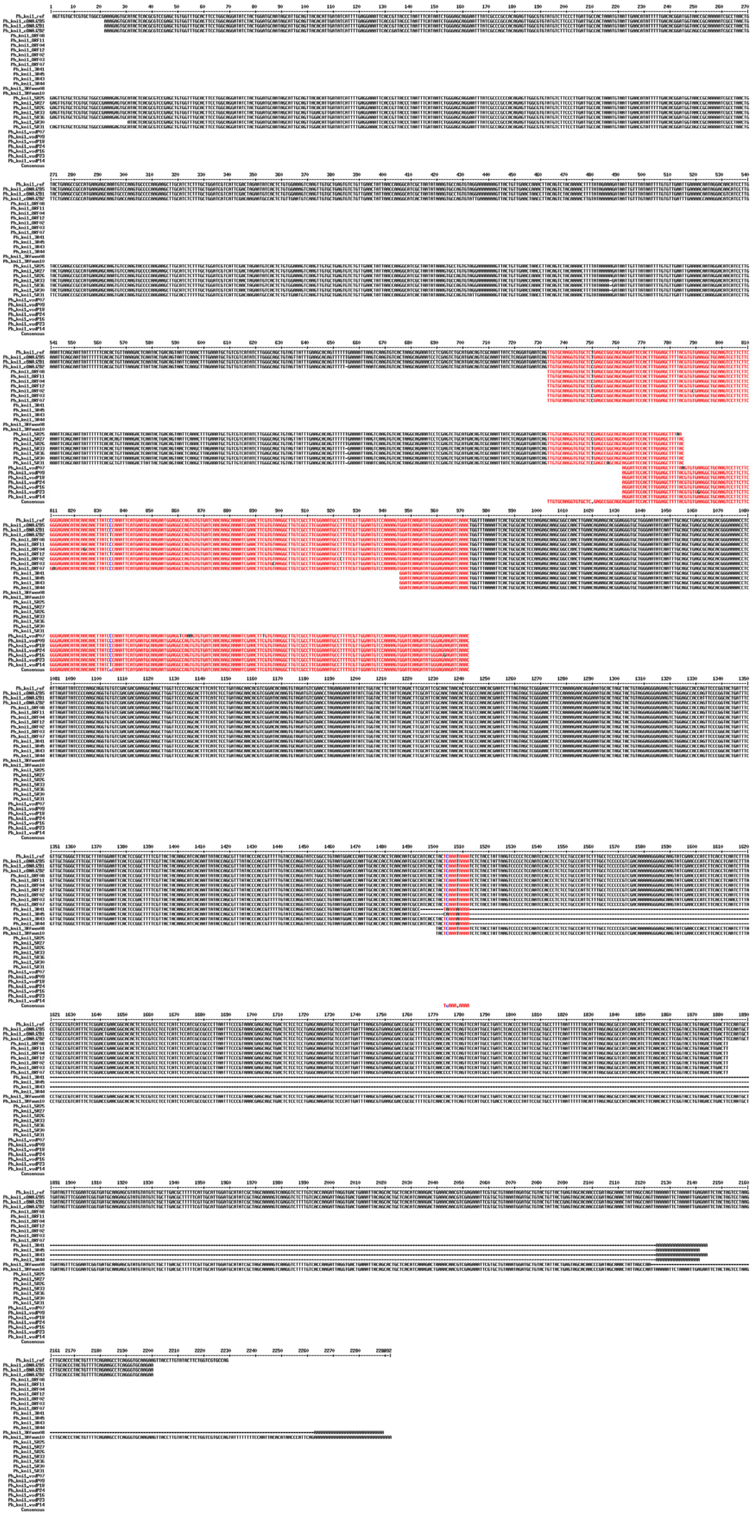
```
                 1       10        20        30        40        50        60        70        80        90
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------|
     Ph_kni1_ref MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
Ph_kni1_cDNAJZB5 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
Ph_kni1_cDNAJZB2 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
   Ph_kni1_ORF11 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
   Ph_kni1_ORF08 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
   Ph_kni1_ORF12 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
   Ph_kni1_ORF04 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYSNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
   Ph_kni1_ORF02 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCPFVGMSKSGSRYGRRS
   Ph_kni1_ORF03 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
Ph_kni1_cDNAJZB1 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
   Ph_kni1_ORF07 MTVANYLRMNQLCKVCSEPAAGFHFGAFTCEGCKSFFERTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
    Ph_kni1_3R44                                                                        RKCLFVGMSKSGSRYGRRS
    Ph_kni1_3R43                                                                        RKCLFVGMSKSGSRYGRRS
    Ph_kni1_3R41                                                                        RKCLFVGMSKSGSRYGRRS
    Ph_kni1_3R45                                                                        RKCLFGGMTKSGSRYGRRS
Ph_kni1_3Rfwxn08
Ph_kni1_3Rfwxn10
    Ph_kni1_5R25 MTVANYLRMNQLCKVCSEPAAGFHFGAL
    Ph_kni1_5R26 MTVANYLRMNQLCKVCSEPAAGFHFGAFT
    Ph_kni1_5R27 MTVANYLRMNQLCKVCSEPAAGFHFGAFT
    Ph_kni1_5R31 MTVANYLRMNQLCKVCSEPAAGFHFGAFT
    Ph_kni1_5R30 MTVANYLRMNQLCKVCSEPAAGFHFGAFT
    Ph_kni1_5R33 MTVANYLRMNQLCKVCSEPAAGFHFGAFT
    Ph_kni1_5R36 MTVANYLRMNQLCKVCSEPAAGFHFGAFT
       Consensus MTVANYLRMNQLCKVCSEPAAGFHFGAFT

                 91     100       110       120       130       140       150       160       170       180
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------|
     Ph_kni1_ref NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
Ph_kni1_cDNAJZB5 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
Ph_kni1_cDNAJZB2 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
   Ph_kni1_ORF11 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
   Ph_kni1_ORF08 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
   Ph_kni1_ORF12 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
   Ph_kni1_ORF04 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
   Ph_kni1_ORF02 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNVSGTSIQTSHSQLT
   Ph_kni1_ORF03 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGRLGSPALSSPGSNTSDTSVDVEPRRNISGTSIQTSHSQLT
Ph_kni1_cDNAJZB1 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
   Ph_kni1_ORF07 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTRVDVEPRRNISGTSIQTSHSQLT
    Ph_kni1_3R44 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSDTSDTSVDVEPRRNISGTSIQTSHSQLT
    Ph_kni1_3R43 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVNVEPRRNISGTSIQTSHSQLT
    Ph_kni1_3R41 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
    Ph_kni1_3R45 NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSADVEPRRNISGTSIQTSHSQLT
Ph_kni1_3Rfwxn08
Ph_kni1_3Rfwxn10
    Ph_kni1_5R25
    Ph_kni1_5R26
    Ph_kni1_5R27
    Ph_kni1_5R31
    Ph_kni1_5R30
    Ph_kni1_5R33
    Ph_kni1_5R36
       Consensus

                 181    190       200       210       220       230       240       250       260       270
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------|
     Ph_kni1_ref LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
Ph_kni1_cDNAJZB5 LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
Ph_kni1_cDNAJZB2 LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVVDPIAPPQQSPSPTQI
   Ph_kni1_ORF11 LVQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
   Ph_kni1_ORF08 LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIWPVMDPIAPPQQSPSPTQI
   Ph_kni1_ORF12 LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
   Ph_kni1_ORF04 LAQHESLVARELSKRTGNALATVGGRSLEPPFPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
   Ph_kni1_ORF02 LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
   Ph_kni1_ORF03 LAQHESLVARELSKRTGNALATVGGRSLEPPFPGTDFVAGLSLYGIHSDFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
Ph_kni1_cDNAJZB1 LAQHESLVARELSKRTGNALATVGGRSLEPPFPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
   Ph_kni1_ORF07 LAQHESLVARELSKRTGNALATVGGRSLEPPFPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
    Ph_kni1_3R44 LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQ
    Ph_kni1_3R43 LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDSIAPPQQSPSPTQ
    Ph_kni1_3R41 LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSP
    Ph_kni1_3R45 LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGTHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSP
Ph_kni1_3Rfwxn08                                                                                        TQI
Ph_kni1_3Rfwxn10                                                                                        TQI
    Ph_kni1_5R25
    Ph_kni1_5R26
    Ph_kni1_5R27
    Ph_kni1_5R31
    Ph_kni1_5R30
    Ph_kni1_5R33
    Ph_kni1_5R36
       Consensus

                 271    280       290       300       310       320       330       340       350       360
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------|
     Ph_kni1_ref KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
Ph_kni1_cDNAJZB5 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
Ph_kni1_cDNAJZB2 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
   Ph_kni1_ORF11 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
   Ph_kni1_ORF08 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
   Ph_kni1_ORF12 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
   Ph_kni1_ORF04 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTP
   Ph_kni1_ORF02 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
   Ph_kni1_ORF03 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQPLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
Ph_kni1_cDNAJZB1 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSD-TAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
   Ph_kni1_ORF07 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPGQDAPIDLSVKRPRFSSTPS
    Ph_kni1_3R44
    Ph_kni1_3R43
    Ph_kni1_3R41
    Ph_kni1_3R45
Ph_kni1_3Rfwxn08 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPPSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
Ph_kni1_3Rfwxn10 KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
    Ph_kni1_5R25
    Ph_kni1_5R26
    Ph_kni1_5R27
    Ph_kni1_5R31
    Ph_kni1_5R30
    Ph_kni1_5R33
    Ph_kni1_5R36
       Consensus

                 361    370       380       390    397
                 |--------+---------+---------+------|
     Ph_kni1_ref VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
Ph_kni1_cDNAJZB5 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
Ph_kni1_cDNAJZB2 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
   Ph_kni1_ORF11 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
   Ph_kni1_ORF08 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
   Ph_kni1_ORF12 VPLPDLTPIPLPFNFLHLAASSTSSAPSVPVDLTSNA
   Ph_kni1_ORF04 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
   Ph_kni1_ORF02 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVLVDLTSNA
   Ph_kni1_ORF03 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
Ph_kni1_cDNAJZB1 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
   Ph_kni1_ORF07 VPSPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
    Ph_kni1_3R44
    Ph_kni1_3R43
    Ph_kni1_3R41
    Ph_kni1_3R45
Ph_kni1_3Rfwxn08 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
Ph_kni1_3Rfwxn10 VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
    Ph_kni1_5R25
    Ph_kni1_5R26
    Ph_kni1_5R27
    Ph_kni1_5R31
    Ph_kni1_5R30
    Ph_kni1_5R33
    Ph_kni1_5R36
       Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.11_Ph_kni1]

[PhD_A2.11.2_Ph_kni1_alignments]

**“alignment_Ph_kni1_all_bp_sequences.gif”**

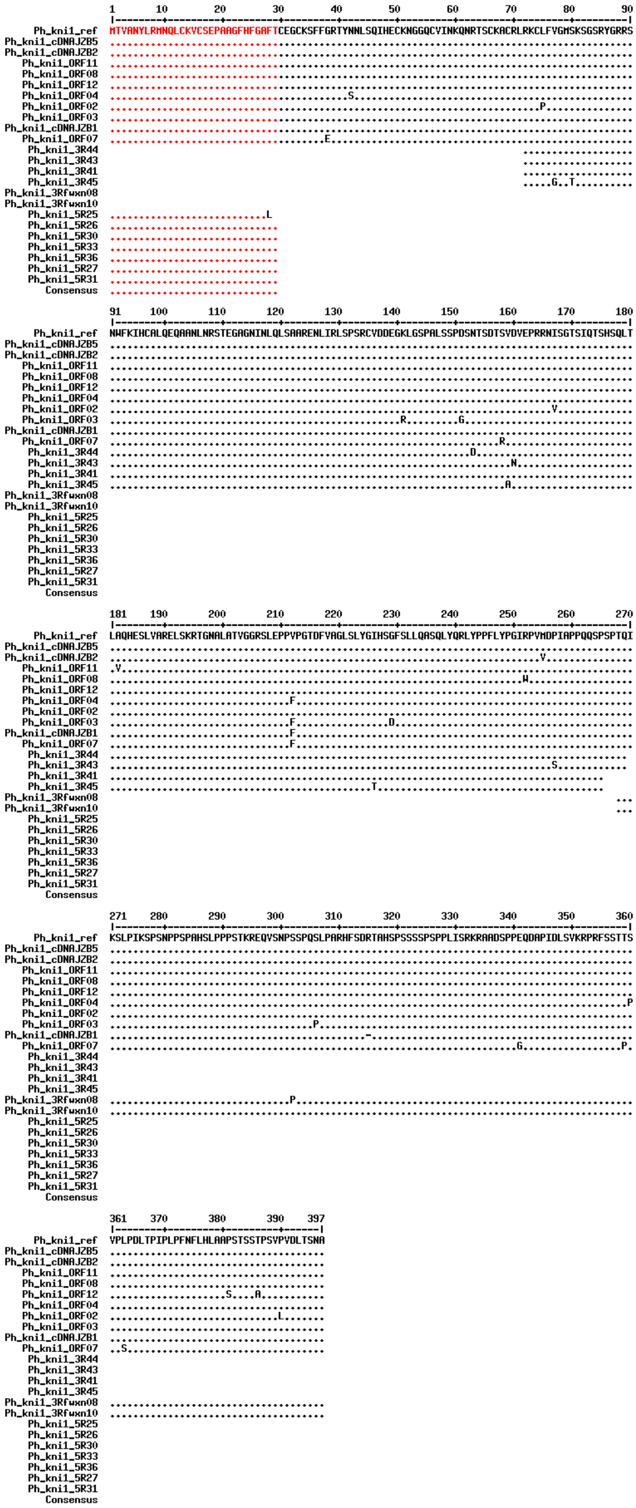


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.11_Ph_kni1]

[PhD_A2.11.2_Ph_kni1_alignments]

## **“difference_alignment_Ph_kni1_all_aa_sequences.gif”**



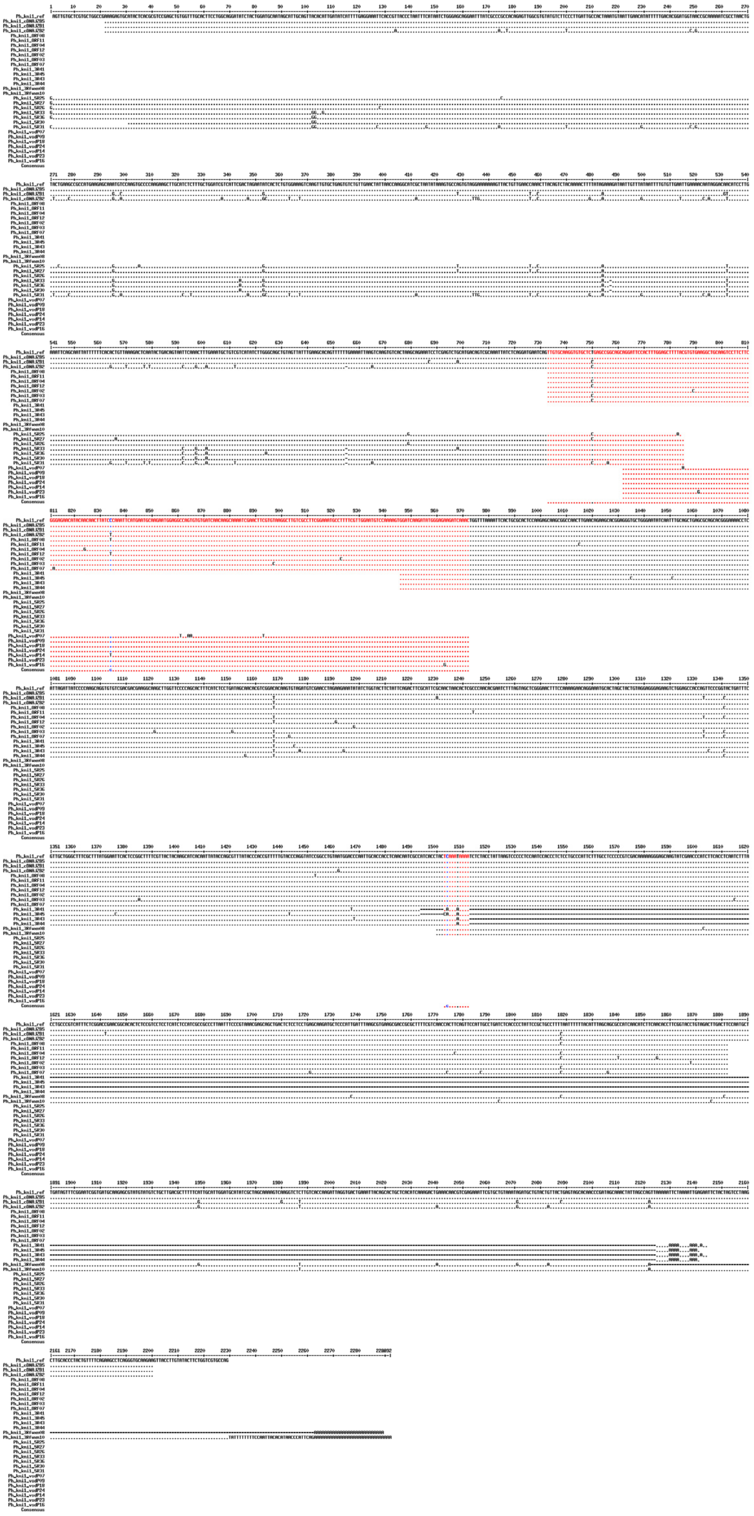
```
                 1        10        20        30        40        50        60        70        80        90
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------|
     Ph_kni1_ref MTVANYLRMNQLCKYCSEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRS
Ph_kni1_cDNAJZB5 ..........................................................................................
Ph_kni1_cDNAJZB2 ..........................................................................................
   Ph_kni1_ORF11 ..........................................................................................
   Ph_kni1_ORF08 ..........................................................................................
   Ph_kni1_ORF12 ..........................................................................................
   Ph_kni1_ORF04 .........................................S................................................
   Ph_kni1_ORF02 ..........................................................................P...............
   Ph_kni1_ORF03 ..........................................................................................
Ph_kni1_cDNAJZB1 ..........................................................................................
   Ph_kni1_ORF07 .....................................E....................................................
    Ph_kni1_3R44                                                                        ...................
    Ph_kni1_3R43                                                                        ...................
    Ph_kni1_3R41                                                                        ...................
    Ph_kni1_3R45                                                                        .....G..T..........
Ph_kni1_3Rfwxn08
Ph_kni1_3Rfwxn10
    Ph_kni1_5R25 ...........................L
    Ph_kni1_5R26 .............................
    Ph_kni1_5R30 .............................
    Ph_kni1_5R33 .............................
    Ph_kni1_5R36 .............................
    Ph_kni1_5R27 .............................
    Ph_kni1_5R31 .............................
       Consensus .............................

                 91     100       110       120       130       140       150       160       170       180
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------|
     Ph_kni1_ref NWFKIHCALQEQAANLNRSTEGAGNINLQLSAARENLIRLSPSRCVDDEGKLGSPALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLT
Ph_kni1_cDNAJZB5 ..........................................................................................
Ph_kni1_cDNAJZB2 ..........................................................................................
   Ph_kni1_ORF11 ..........................................................................................
   Ph_kni1_ORF08 ..........................................................................................
   Ph_kni1_ORF12 ..........................................................................................
   Ph_kni1_ORF04 ..........................................................................................
   Ph_kni1_ORF02 ............................................................................V.............
   Ph_kni1_ORF03 ..................................................R.........G.............................
Ph_kni1_cDNAJZB1 ..........................................................................................
   Ph_kni1_ORF07 ...................................................................R......................
    Ph_kni1_3R44 ..............................................................D...........................
    Ph_kni1_3R43 .....................................................................N....................
    Ph_kni1_3R41 ..........................................................................................
    Ph_kni1_3R45 ....................................................................A.....................
Ph_kni1_3Rfwxn08
Ph_kni1_3Rfwxn10
    Ph_kni1_5R25
    Ph_kni1_5R26
    Ph_kni1_5R30
    Ph_kni1_5R33
    Ph_kni1_5R36
    Ph_kni1_5R27
    Ph_kni1_5R31
       Consensus

                 181    190       200       210       220       230       240       250       260       270
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------|
     Ph_kni1_ref LAQHESLVARELSKRTGNALATVGGRSLEPPVPGTDFVAGLSLYGIHSGFSLLQASQLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQI
Ph_kni1_cDNAJZB5 ..........................................................................................
Ph_kni1_cDNAJZB2 ..........................................................................V...............
   Ph_kni1_ORF11 .V........................................................................................
   Ph_kni1_ORF08 .......................................................................W..................
   Ph_kni1_ORF12 ..........................................................................................
   Ph_kni1_ORF04 ...............................F..........................................................
   Ph_kni1_ORF02 ..........................................................................................
   Ph_kni1_ORF03 ...............................F................D.........................................
Ph_kni1_cDNAJZB1 ...............................F..........................................................
   Ph_kni1_ORF07 ...............................F..........................................................
    Ph_kni1_3R44 .........................................................................................
    Ph_kni1_3R43 ............................................................................S............
    Ph_kni1_3R41 .....................................................................................
    Ph_kni1_3R45 .............................................T.......................................
Ph_kni1_3Rfwxn08                                                                                        ...
Ph_kni1_3Rfwxn10                                                                                        ...
    Ph_kni1_5R25
    Ph_kni1_5R26
    Ph_kni1_5R30
    Ph_kni1_5R33
    Ph_kni1_5R36
    Ph_kni1_5R27
    Ph_kni1_5R31
       Consensus

                 271    280       290       300       310       320       330       340       350       360
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------|
     Ph_kni1_ref KSLPIKSPSNPPSPAHSLPPPSTKREQVSNPSSPQSLPARHFSDRTAHSPSSSSPSPPLISRKRAADSPPEQDAPIDLSVKRPRFSSTTS
Ph_kni1_cDNAJZB5 ..........................................................................................
Ph_kni1_cDNAJZB2 ..........................................................................................
   Ph_kni1_ORF11 ..........................................................................................
   Ph_kni1_ORF08 ..........................................................................................
   Ph_kni1_ORF12 ..........................................................................................
   Ph_kni1_ORF04 .........................................................................................P
   Ph_kni1_ORF02 ..........................................................................................
   Ph_kni1_ORF03 ...................................P......................................................
Ph_kni1_cDNAJZB1 ............................................-.............................................
   Ph_kni1_ORF07 ......................................................................G.................P.
    Ph_kni1_3R44
    Ph_kni1_3R43
    Ph_kni1_3R41
    Ph_kni1_3R45
Ph_kni1_3Rfwxn08 ...............................P..........................................................
Ph_kni1_3Rfwxn10 ..........................................................................................
    Ph_kni1_5R25
    Ph_kni1_5R26
    Ph_kni1_5R30
    Ph_kni1_5R33
    Ph_kni1_5R36
    Ph_kni1_5R27
    Ph_kni1_5R31
       Consensus

                 361    370       380       390    397
                 |--------+---------+---------+------|
     Ph_kni1_ref VPLPDLTPIPLPFNFLHLAAPSTSSTPSVPVDLTSNA
Ph_kni1_cDNAJZB5 .....................................
Ph_kni1_cDNAJZB2 .....................................
   Ph_kni1_ORF11 .....................................
   Ph_kni1_ORF08 .....................................
   Ph_kni1_ORF12 ....................S....A...........
   Ph_kni1_ORF04 .....................................
   Ph_kni1_ORF02 .............................L.......
   Ph_kni1_ORF03 .....................................
Ph_kni1_cDNAJZB1 .....................................
   Ph_kni1_ORF07 ..S..................................
    Ph_kni1_3R44
    Ph_kni1_3R43
    Ph_kni1_3R41
    Ph_kni1_3R45
Ph_kni1_3Rfwxn08 .....................................
Ph_kni1_3Rfwxn10 .....................................
    Ph_kni1_5R25
    Ph_kni1_5R26
    Ph_kni1_5R30
    Ph_kni1_5R33
    Ph_kni1_5R36
    Ph_kni1_5R27
    Ph_kni1_5R31
       Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.11_Ph_kni1]

[PhD_A2.11.2_Ph_kni1_alignments]

**“difference_alignment_Ph_kni1_all_bp_sequences.gif”**



[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.12_Ph_kni2]

[PhD_A2.12.1_Ph_kni2_sequences]

## “Ph_kni2_general_read_me.txt”

```
General aspects:

(i)     All sequences are in FASTA format.

(ii)    [bp] indicates nucleic acid sequences; [aa] indicates amino acid sequences

(iii)   Comments are preceded by an apostrophe (')

(iv)    All sequences without olignonucleotide sequence are indicated (e.g. "w/o 1813, 1816"). In these cases, only novel Parhyale
sequence is presented.


Parhyale hawaiensis kni2 terminology:


(i)     In general, sequence denotation begins with: ">Ph_kni2_"

(ii)    The above (i) is followed by:  "vsdP",   in case sequence originates from vsdPCR (5.2.3.1)
                                       "5R" or "3R",      in case sequence derives from 5' or 3' RACE, respectively
                                       "ORF"              for coherent ORF sequences
                                       "cDNA"             for complete cDNA sequences

                                       "ref"              if reference sequence is presented
                                       "ZNFD"             if Zinc finger domain (encoding) fragment is presented
                                       "KNID"             if knirps box/ knirps domain is presented
                                       "PIDLS"            if PIDLS motif (encoding fragment) is presented

(iii)   The above (ii) is followed by clone id.


Parhyale hawaiensis kni2 synopsis:

synopsis:
cDNA:    2656 bp
5'UTR:    244 bp (1-244)
ORF:     2079 bp (245-2323), 692 aa
3'UTR:    333 bp (2324-2656)
ZNFD:     198 bp (251-448),   66 aa
KNID:      57 bp (467-523),   19 aa
PIDLS:     36 bp (1967-2002), 12 aa

>Ph_kni2_ZNFD [bp]
CAACTGTGCAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGA
GGCCAGTGTGTGATCAACAAACAAAACAGGACTTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTT

>Ph_kni2_KNID [bp]
GGGTCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAA

>Ph_kni2_PIDLS [bp]
GCACCGGGTGAGCAGCCCATCGATCTCACGGTTCGG

>Ph_kni2_ref [bp]
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAACAAATTCGGTTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGAAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAG
ACAGACGAAAGCTTTGTATTCGGCCATCATTTTTCAAAATCAGGTTTTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGAAACAAGATAGCAAGCGAGATGAACCAACTGTG
CAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTG
TGTGATCAACAAACAAAACAGGACTTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACT
ACAAGAACAATGCGGGTACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGCTCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGA
TGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGCATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGT
CTCAGAACCGCACTTGTCGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAATGCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGAC
GTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAAACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGG
GGCTGAATACGAAAGAGTCCGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCCAGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATC
TGAATCTCCTCATAGTTCAAAAGTTGCTCTCAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAAATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAG
TACTCCAGAGGTTAAATCAGAAAGAAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCTAGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATT
ATCTTTCCTGAAGTCAGCCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGCATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCT
TTATGGTCCTTTGAACAACAACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTGGCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGG
AGCACCAATCTCACTTGATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTGCCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGA
TGCTATCCTTCAGGTTCAGAGAGAAAGTGCATGTGGACCTCGATCTCCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTCGCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCC
GATGGTGACAGCCACAACGACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCAAACGAAAGTGATCCTCAGGA
AGGCATGGAAGATGAGTATGAACAAGAACAGAAAGACGACGCAGATCCGAAGGAAGAATCTATCTTGAAAGACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGT
GAAAGGAACTTCAGAGATTGTCGATCTTAAAACGAGTTTGGAGTGTGCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTAGAAACTGCTGGTGATGGCCTCATCACCTG
ACGCCTGCTAAGTTAATTCACAGAGTCTCATGGGAGACTGTAATGTGCCATATTATGTAAGGTTTTGTTAAGTTACATTCTCAGATGAGTGTAGGGAGCGGCGGGTAAAGTATTACAGATAATGCAATG
AGAGCCATATGAACTGCATGAGATGTGTACAGTGTCTTGCGATCAATACACTGGCTCCTTACTGAGAATATAATTATGTTACAAATGCCATTATTCCAGTAAACTGCTCAGTAACTGTAAGGACTGAAT
TGAGTAGTTAAGTCTAACTGTTGTTTATTGTCTAAATTTCTATTGCTTAGCAGCTATTTCTTATAAATTCATTTCC
```

```
>Ph_kni2_ZNFD [aa]
QLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLV

>Ph_kni2_KNID [aa]
GSRYGRRSNWFKIHCLLQE

>Ph_kni2_PIDLS [aa]
APGEQPIDLTVR

>Ph_kni2_ref [aa]
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRD
ALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEP
VRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMT
EEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNK
KRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQPNESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVS
KEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT&
```

[PhD_A2_Parhyale_goi_sequences]
[PhD_A2.12_Ph_kni2]
[PhD_A2.12.1_Ph_kni2_sequences]

**“Ph_kni2_cDNA_sequences.txt”**

```
>Ph_kni2_ref [bp] 'Ph_kni2_cJZ1209 ref seq, first 50 bp from Ph_kni2_5R01; exchanged 546T>C, 837C>T, 1062T>C
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAACAAATTCGGTTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGAAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAG
ACAGACGAAAGCTTTGTATTCGGCCATCATTTTTCAAAATCAGGTTTTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGAAACAAGATAGCAAGCGAGATGAACCAACTGTG
CAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTG
TGTGATCAACAAACAAAACAGGACTTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACT
ACAAGAACAATGCGGGTACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGCTCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGA
TGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGCATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGT
CTCAGAACCGCACTTGTCGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAATGCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGAC
GTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAAACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGG
GGCTGAATACGAAAGAGTCCGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCCAGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATC
TGAATCTCCTCATAGTTCAAAAGTTGCTCTCAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAAATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAG
TACTCCAGAGGTTAAATCAGAAAGAAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCTAGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATT
ATCTTTCCTGAAGTCAGCCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGCATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCT
TTATGGTCCTTTGAACAACAACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTGGCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGG
AGCACCAATCTCACTTGATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTGCCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGA
TGCTATCCTTCAGGTTCAGAGAGAAAGTGCATGTGGACCTCGATCTCCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTCGCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCC
GATGGTGACAGCCACAACGACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCAAACGAAAGTGATCCTCAGGA
AGGCATGGAAGATGAGTATGAACAAGAACAGAAAGACGACGCAGATCCGAAGGAAGAATCTATCTTGAAAGACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGT
GAAAGGAACTTCAGAGATTGTCGATCTTAAAACGAGTTTGGAGTGTGCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTAGAAACTGCTGGTGATGGCCTCATCACCTG
ACGCCTGCTAAGTTAATTCACAGAGTCTCATGGGAGACTGTAATGTGCCATATTATGTAAGGTTTTGTTAAGTTACATTCTCAGATGAGTGTAGGGAGCGGCGGGTAAAGTATTACAGATAATGCAATG
AGAGCCATATGAACTGCATGAGATGTGTACAGTGTCTTGCGATCAATACACTGGCTCCTTACTGAGAATATAATTATGTTACAAATGCCATTATTCCAGTAAACTGCTCAGTAACTGTAAGGACTGAAT
TGAGTAGTTAAGTCTAACTGTTGTTTATTGTCTAAATTTCTATTGCTTAGCAGCTATTTCTTATAAATTCATTTCC

>Ph_kni2_cJZ1209 [bp]  'w/o 1860, 1817
TTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGAAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAGACAGACGAAAGCTTTGTATTCGGCCATCATTTTTCAAAATCAGGTT
TTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGAAACAAGATAGCAAGCGAGATGAACCAACTGTGCAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTC
ACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAAACAGGACTTCGTGCAAGGCCTGTCGTCTC
AGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGTACAGCGGCACTGTCGGGCAGCTACCAACG
TTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGCTCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGA
ATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGCATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGTCGCCAAATTTCCCCATGGGACAAATAGGA
GCGGTTGAGCGTCGAACTCAGGAAGAAACTCTCAATGCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATG
CAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAAACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAGTCCGAGGAGTATTTGCTGAAGCTGAAATT
AGAATGAATTCAAACCCTCCATACGACTGGGATCCCAGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTTCAAAAGTTGCTCTCAGTGTAAGTGTTAGC
ACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAAATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAATCAGAAAGAAGGTCAGTAGGAAATCGTTCT
CCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCTAGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTCCTGAAGTCAGCCAGTAAAATATCTGCAAACAATACAGCT
ACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGCATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACAACAACATTTTCCCATTTGCATCGTATCCT
CCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTGGCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCAATCTCACTTGATGGTTTCAAGAGTCCCCTGTATTTGCCA
ACTGAGCTCAGAAACACAGTGTTGCCAAACACATTGCCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTCAGAGAGAAAGTGCATGTGGACCTCGATCT
CCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTCGCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAACGACGACATCCACAGCACCGGGTGAGCAG
CCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCAAACGAAAGTGATCCTCAGGAAGGCATGGAAGATGAGTATGAACAAGAACAGAAAGACGACGCAGAT
CCGAAGGAAGAATCTATCTTGAAAGACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGTGAAAGGAACTTCAGAGATTGTCGATCTTAAAACGAGTTTGGAGTGT
GCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTAGAAACTGCTGGTGATGGCCTCATCACCTGACGCCTGCTAAGTTAATTCACAGAGTCTCATGGGAGACTGTAATGT
GCCATATTATGTAAGGTTTTGTTAAGTTACATTCTCAGATGAGTGTAGGGAGCGGCGGGTAAAGTATTACAGATAATGCAATGAGAGCCATATGAACTGCATGAGATGTGTACAGTGTCTTGCGATCAA
TACACTGGCTCCTTACTGAGAATATAATTATGTTACAAATGCCATTATTCCAGTAAACTGCTCAGTAACTGTAAGGACTGAATTGAGTAGTTAAGTCTAACTGTTGTTTATTGTCTAAATTTCTATT

>Ph_kni2_cJZ1210 [bp] 'w/o 1860, 1817
TTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGAAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAGACAGACGAAAGCTTTGTATTCGGCCATCATTTTTCAAAATCAGGTT
TTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGAAACAAGATAGCAAGCGAGATGAACCAACTGTGCAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTC
ACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAAACAGGACTTCGTGCAAGGCCTGTCGTCTC
AGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGTACAGCGGCACTGTCGGGCAGCTACCAACG
TTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGCTCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGA
ATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGCATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGTCGCCAAATTTCCCCATGGGACAAATAGGA
GCGGTTGAGCGTCGAACTCAGGAAGAAACTCTCAATGCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATG
CAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAAACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAGTCCGAGGAGTATTTGCTGAAGCTGAAATT
AGAATGAATTCAAACCCTCCATACGACTGGGATCCCAGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTTCAAAAGTTGCTCTCAGTGTAAGTGTTAGC
ACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAAATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAATCAGAAAGAAGGTCAGTAGGAAATCGTTCT
CCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCTAGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTCCTGAAGTCAGCCAGTAAAATATCTGCAAACAATACAGCT
ACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGCATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACAACAACATTTTCCCATTTGCATCGTATCCT
CCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTGGCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCAATCTCACTTGATGGTTTCAAGAGTCCCCTGTATTTGCCA
ACTGAGCTCAGAAACACAGTGTTGCCAAACACATTGCCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTCAGAGAGAAAGTGCATGTGGACCTCGATCT
CCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTCGCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAACGACGACATCCACAGCACCGGGTGAGCAG
CCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCAAACGAAAGTGATCCTCAGGAAGGCATGGAAGATGAGTATGAACAAGAACAGAAAGACGACGCAGAT
CCGAAGGAAGAATCTATCTTGAAAGACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGTGAAAGGAACTTCAGAGATTGTCGATCTTAAAACGAGTTTGGAGTGT
GCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTAGAAACTGCTGGTGATGGCCTCATCACCTGACGCCTGCTAAGTTAATTCACAGAGTCTCATGGGAGACTGTAATGT
GCCATATTATGTAAGGTTTTGTTAAGTTACATTCTCAGATGAGTGTAGGGAGCGGCGGGTAAAGTATTACAGATAATGCAATGAGAGCCATATGAACTGCATGAGATGTGTACAGTGTCTTGCGATCAA
TACACTGGCTCCTTACTGAGAATATAATTATGTTACAAATGCCATTATTCCAGTAAACTGCTCAGTAACTGTAAGGACTGAATTGAGTAGTTAAGTCTAACTGTTGTTTATTGTCTAAATTTCTATT

>Ph_kni2_cJZ22 [bp] 'w/o 1813, 1816
GTTTTTTTGCTTTCAGAATAAATTCGGTTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGTAGTTCGGTGCTCCAGAATTAAGTCACTTTATTTAGACAGACGAAAGCTTTGTAT
TCGGCCATCATTTTTCAAAATCAGGTTTTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGATACAAGATAGCAAGCGAGATGAACCAACTGTGCAAGGTCTGTGCTGAACCT
GCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAAAC
AGGACTTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGTAC
AGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCGCACTGCGAGACGGCTCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAGTC
GAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGCATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGTCG
CCAAATTTCCCCATGGGACAAATAGGGGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAATGCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTTTG
```

```
GAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAAACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAGTC
CGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCCAGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTTCA
AAAGTTGCTCTCAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAAATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAATCA
GAAAGAAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCTAGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTCCTGAAGTCAGCC
AGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGCATCCCTGGCTTATTTCCTCATCGTTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACAAC
AACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTGGCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCAATCTCACTTGAT
GGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTGCCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTCAG
AGAGAAAGTGCCTGTGGACCTCGATCTCCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTCGCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAACG
ACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCAAACGAAAGTGATCCTCAGGAAGGCATGGAAGATAAGTAT
GAGCAAGAACAGAAAGACGACGCAGATCCGAAGGAAGAATCTATCTTGAAAGACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGTGAAAGGAACTTCAGAGATT
GTCGATCTTAAAACGAGTTTGGAGTGTGCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTAGAAACTGCTGGTGATGGCCTCATCACCTGACGCCTGCTAAGTTAATTC
ACAGAGTCTCATGGGAGACTGTAATGTGCCACATTATGTAAGGTTTTGTTAAGTTACATTCTCAGATGAGTGTAGGGAGCGGCGGGTAAAGTATTACAGATAATGCAATGAGAGCCATATGAACTGCAT
GAGATGTGTACAGTGTCTTGCGATCAATACACTGGCTCCTTACTGAGAATATAATTATGTTACAAATGCCATTATTCCAGTAAACTGCTCAGTAACT

>Ph_kni2_ORF21 'w/o 1814, 1815
CTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAA
ACGGGACTTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCACAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGT
ACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGCTCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAG
TCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGCATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGT
CGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAATGCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTT
TGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAAACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAG
TCCGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCCAGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTT
CAAAAGTTGCTCTTAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAAATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAAT
CAGAAAGAAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCTAGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTCCTGAAGTCAG
CCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGCATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACA
ACAACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTGGCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCAATCTCACTTG
ATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTGCCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTC
AGAGAGAAAGTGCATGTGGACCTCGATCTCCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTCGCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAA
CGACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCAAACGAAAGTGATCCTCAGGAAGGCATGGAAGATGAGT
ATGAGCAAGAACAGAAAGACGACGCAGATCCGAAGGAAGAATCTATCTTGAAAGACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGTGAAAGGAACTTCAGAGA
TTGTCGATCTTAAAACGAGTTTGGAGTGTGCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTA

>Ph_kni2_ORF23 'w/o 1814, 1815
CTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAA
ACAGGACTTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGT
ACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGTTCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAG
TCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGCATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGT
CGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAATGCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTT
TGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACACCTCTGCGGTCCAAACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTACGAAGAATAGGGGCTGAATACGAAAGAG
TCCGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCCAGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTT
CAAAAGTTGCTCTCAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAAATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAAGTTAAAT
CAGAAAGAAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTACTAGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTCCTGAAGTCAG
CCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGCATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACA
ACAACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTGGCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCAATCTCACTTG
ATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTGCCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAGAGAATTTTAGATGCTATCCTTCAGGTTC
AGAGAGAAAGTGCATGTGGACCTCGATCTCCTTTGGGAGGATTATCTCCTCCTGGAGTGACGCTCGCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAA
CGACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCAAACGAAAGTGATCCTCAGGATGGCATGGAAGATGAGT
ATGAACAAGAACAGAAAGACGATGCAGATCCGAAGGAAGAATCTATCATGAAAGACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGTGAAAGGAACTTCAGAGA
TTGTCGATCTTAAAACGAGTTTGGAGTGTGCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTA

>Ph_kni2_ORF31 'w/o 1814, 1815
CTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAA
ACAGGACTTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCATGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGT
ACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGCTCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAG
TCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGCATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGT
CGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAATGCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTT
TGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAAACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAG
TCCGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCCAGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTT
CAAAAGTTGCTCTCAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAAATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAAT
CAGAAAGGAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCTAGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTCCTGAAGTCAG
CCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGCATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACA
ACAACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTGGCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCAATCTCACTTG
ATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTGCCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTC
AGAGTGAAAGTGCATGTGGACCTCGATCTCCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTCGCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAA
CGACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCAAACGAAAGTGATCCTCAGGAAGGCATGGAAGATGAGT
ATGAACAAGAACAGAAAGACGACGCAGATCCGAAGGAAGAATCTATCTTGAAAGACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGTGAAAGGAACTTCAGAGA
TTGTCGATCTTAAAACGAGTTTGGAGTGTGCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTA

>Ph_kni2_5R01 'w/o 1806
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAACAAATTCGGTTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGAAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAG
ACAGACGAAAGCTTTGTATTCGGCCATCATTTTTCAAAATCAGGTTTTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGAAACAAGATAGCAAGCGAGATGAACCAACTGTG
CAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTG
TGTGATCAACAAACAAAA

>Ph_kni2_5R02 'SMARTIIA has bp exchange: AAGCAGTGCTATCAACGCAGACTACGCGGG, w/o 1806
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAATAAATTCGGTTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGTAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAG
ACAGACGAAAGCTTTGTATTCGGCCATCATTTTTCAAAATCAGGTTTTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGATACAAGATAGCAAGCGAGATGAACCAACTGTG
CAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTG
TGTGATCAACAAACAAAA

>Ph_kni2_5R03 'w/o 1806
GGAATTTTCACTGTTCCCTTTACCACGCTTTACTGAATAAATTCGGTTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGTAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAG
ACAGACGAAAGCTTTGTATTCGGCCATCATTTTTCAAAATCAGGTTTTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGGGATACAAGATAGCAAGCGAGATGAACCAACTGTG
CAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGGAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTG
TGTGATCAACAAACAAAA

>Ph_kni2_5R04 'w/o 1806
GGAATTTTCACCGTTCCCTTTACCACGTTTTACTGAACAAATTCGGTTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGAAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAG
ACAGACGAAAGCTTTGTATTCGGCCATCATTTTTCAAAATCAGGTTTTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGAAACAAGACAGCAAGCGAGATGAACCAACTGTG
```

```
CAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTG
TGTGATCAACAAACAAAA

>Ph_kni2_5R06 'w/o 1806
GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAACAAATTCGGTTAAAAATTTACCCGTACAAATAAGGGATGAAAATTGACAGATAATGAAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAG
ACAGACGAAAGCTTTGTATTCGGCCATCATTTTTCAAAATCAGGTTTTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGAAACAAGATAGCAAGCGAGATGAACCAACTGTG
CAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTG
TGTGATCAACAAACAAAA

>Ph_kni2_vsdP01 'w/o fw_CKV, re_LLQ, ref seq for vsdPCR seqs kni2
TGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAAACAGGAC
TTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAAT

>Ph_kni2_vsdP02 'w/o fw_CKV, re_LLQ
TGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAAACAGGAC
TTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAAT

>Ph_kni2_vsdP03 'w/o fw_CKV, re_LLQ
TGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCCTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAAACAGGAC
TTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAAT

>Ph_kni2_vsdP21 'w/o fw_CKV, re_LLQ
TGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGGAGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAAACAGGAC
TTCGTGCAAGGCCTGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAAT

>Ph_kni2_3R13 [bp] 'w/o 1801, seq incomplete 3'
TCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGTACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAGGACGGC
TCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGC
ATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGTCGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAAT
GCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAA
ACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAGTCCGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCC
AGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTTCAAAAGTTGCTCTCAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAA
ATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAATCAGAAAGAAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCT
AGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTCCTGAAGTCAGCCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGC
ATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACAACAACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTG
GCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCAATCTCACTCGATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTG
CCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTCAGAGAGAAAGTGCATGTGGACCTCGATCTCCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTC
GCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAACGACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGG

>Ph_kni2_3R16 [bp] 'w/o 1801 'seq incomplete 3'
TCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGTACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGC
TCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGC
ATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGTCGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAAT
GCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATGCAGTCACCTCGGCTCCTGGACCCCTCTGCGGTCCAA
ACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAGTCCGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCC
AGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTTCAAAAGTTGCTCTCAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAA
ATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAATCAGAAAGAAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCT
AGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTCCTGAAGTCAGCCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGC
ATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACAACAACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTG
GCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCAATCTCACTTGATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTG
CCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTCAGAGAGAAAGTGCATGTGGACCTCGATCTCCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTC
GCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAACGACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGG

>Ph_kni2_3R17 [bp] 'w/o 1801 'seq incomplete central
TCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGTACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGGCGGC
TCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGC
ATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGTCGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGGATTCAGGAAGAAACTCTCAAT
GCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACCCCTCTGC
CTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCAAACGAAGGTGATCCTCAGGAAGGCATGGAAGATGAGTATGAACAAGAACAGAAAGACGACGCAGATCCGAAGGAA
GAATCTATCTTGAAAGACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGTGAAAGGAACTTCAGAGATTGGCGATCTTAAAACGAGTTTGGAGTGTGCCCAGTCC
CTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTAGAAACTGCTGGTGATGGCCTCATCACCTGACGCCTGCTAAGTTAATTCACAGAGTCTCATGGGAGACTGTAATGTGCCATATTA
TGTAAGGTTTTGCTAAGTTACATTCTCAGATGAGTGTAGGGAGCGGCGGGTAAAGTATTACAGATAATGCAATGAGAGCCATATGAACTGCATGAGATGTGTACAGTGTCTTGCGATCAATACACTGGC
TCCCTACTGAGAATATAATTATGTTACAAATGCCATTATTCCAGTAAACTGCTCAGTAACTGTAAGGACTGAATTGAGTAGTTAAGTCTAACTGTTGTTTATTGTCTAAATTTCTATTGCTTAGCAGCT
ATTTCTTATAAATTCATTTCCTAAAAAAAAAAAAAAAAAAAAAAAAAA

>Ph_kni2_3R20 [bp] 'w/o 1801
TCACGGTACGGCCGCAGGTCTAATTGGTTCAAGGTCCATTGTTTACTACAAGAACAATGCGGGTACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGC
TCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTGTGTGC
ATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGTCGCCAAATTTCCCCATGGGGCAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAAT
GCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAA
ACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAGTCCGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCC
AGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTTCAAAAGTTGCTCTCAGTGTAAGTGTTTGCACTTCTCAAGAAGTAGACAGTAGCAGCTCAAATCAA
ATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAATCAGAAAGAAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCT
AGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTTCTGAAGTCAGCCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGC
ATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACAACAACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTG
GCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACTAATCTCACTTGATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTG
CCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTCAGAGAGAAAGTGCATGTGGACCTCGATCTCCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTC
GCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAACGACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGG

>Ph_kni2_3R21 [bp] 'w/o 1801
TCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTACTACAAGAACAATGCGGGTACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGC
TCCTTACCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGAATACCTATCACGTCTGCTGCCCCTGGGGGTATGTGC
ATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCGCACTTGTCGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAAT
GCACTACGGAGAGGTGTGGGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTCCTGGACCCCTCTGCGGTCCAA
ACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAGTCCGAGGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCC
AGATTGCAAGAAGCTCCGGTGTTCAGTGAAGCGTCGGTAGTGAAATCTGAATCTCCTCATAGTTCAAAAGTTGCTCTCAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAACTCAAATCAA
ATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAATCAGAAAGAAGGTCAGTAGGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCT
AGATTACGAGATGAGACAGGTGCTCCCATGACCGAAGAAGAACGATTATCTTTCCTGAAGTCAGCCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGC
ATCCCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACAACAACATTTTCCCATTTGCATCGTATCCTCCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTG
GCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCAATCTCACTTGATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTG
```

```
CCGGATGTTACAAAAGAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTCAGAGAGAAAGTGCATGTGGACCTCGATCTCCTCTGGGAGGATTATCTCCTCCTGGAGTGACGCTC
GCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAACGACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATG
GCCAAGCGTGACATTATGTATCAGCCAAACGAAAGTGATCTTCAGGAAGGCATGGAAGATGAGTATGACCAAGAACAGAAAGACGACGCAGATCCGAAGGAAGAATCTATCTTGAAAGACGAAGAGGTT
TTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAGAGAAATCCTAGTGAAAGGAACTTCAGAGATTGTCGATCTTAAAACGAGTTTGGAGTGTGCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCA
ATGAAGATAATGAAATTAGAAACTGCTGGTGATGGCCTCATCACCTGACGCCTGCTAAGTTAATTCACAGAGTCTCATGGGAGACTGTAATGTGCCATATTATGTAAGGTTTTGTTAAGTTACATTCTC
AGATGAGTGTAGGGAGCGGCGGGTAAAGTATTACAGATAATGCAATGAGAGCCATATGAACTGCATGAGATGTGTACAGTGTCTTGCGATCAATACACTGGCTCCTTACTGAGAATATAATTATGTTAC
AAATGCCATTATTCCAGTAAACTGCTCAGTAACTGTAAGGACTGAATTGAGTAGTTAAGTCTAACTGTTGTTTATTGTCTAAATTTCTATTGCTTAGCAGCTATTTCTTATAAATTCATTTCCCAAAAA
AAAAAAAAAAAAAAAAAAAAAAAA


>Ph_kni2_ref [aa] 'Ph_kni2_cJZ1209, exchanged 101V>A, 198T>I, 273F>S
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRD
ALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEP
VRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMT
EEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNK
KRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQPNESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVS
KEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT&

>Ph_kni2_cJZ1209 [aa]  ' ref seq
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTVGQLPTFRSPVTPTLQDGSLPFVNTQLRD
ALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRTQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEP
VRRIGAEYERVRGVFAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMT
EEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNK
KRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQPNESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVS
KEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT&

>Ph_kni2_cJZ1210 [aa]
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTVGQLPTFRSPVTPTLQDGSLPFVNTQLRD
ALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRTQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEP
VRRIGAEYERVRGVFAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMT
EEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNK
KRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQPNESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVS
KEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT&

>Ph_kni2_cJZ22 [aa]
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPALRDGSLPFVNTQLRD
ALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEP
VRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMT
EEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNK
KRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQPNESDPQEGMEDKYEQEQKDDADPKEESILKDEEVLEEEPVDERVS
KEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT&

>Ph_kni2_ORF21 [aa]

VCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNGTSCKACRLRKCLVVGMSKTGSRYGHRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDV
PLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGA
EYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLS
FLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNKKRILDA
ILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQPNESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVSKEILVK
GTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT&

>Ph_kni2_ORF23 [aa]
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRD
ALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDTSAVQTLRRSPPIAEP
VRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASTRLQDETGAPMT
EEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNK
KRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQPNESDPQDGMEDEYEQEQKDDADPKEESIMKDEEVLEEEPVDERVS
KEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT&

>Ph_kni2_ORF31
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSWYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRD
ALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEP
VRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMT
EEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNK
KRILDAILQVQSESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQPNESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVS
KEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT&

>Ph_kni2_5R01 [aa]
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQ

>Ph_kni2_5R02 [aa]
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQ

>Ph_kni2_5R03 [aa]
MNQLCKVCAEPAAGFHFGAFTCGGCKSFFGRTYNNLSQIHECKNGGQCVINKQ

>Ph_kni2_5R04 [aa]
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQ

>Ph_kni2_5R06 [aa]
MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQ

>Ph_Kni2_3R13 [aa]
CLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAV
ERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTS
QEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPV
WPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPI
DLT

>Ph_Kni2_3R16 [aa]
```

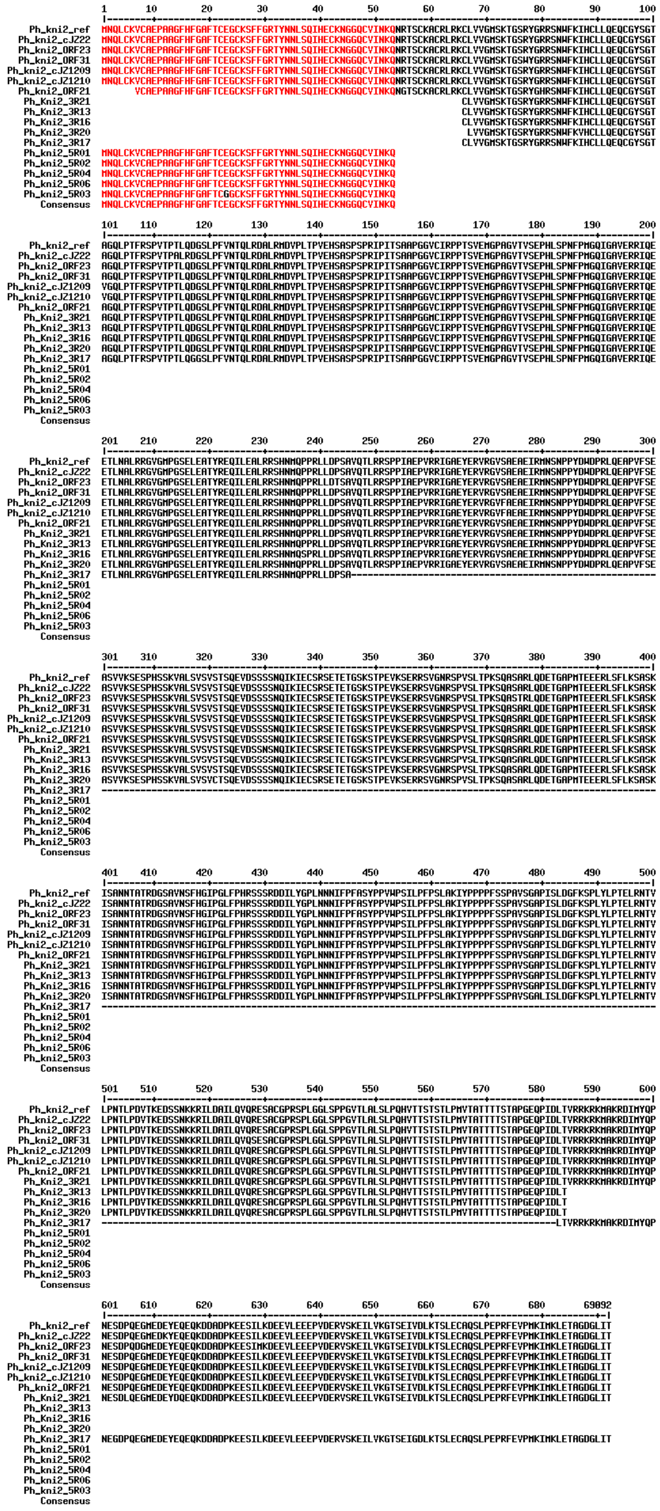
```
CLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAV
ERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQSPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTS
QEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPV
WPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPI
DLT

>Ph_Kni2_3R17 [aa]
CLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQGGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAV
ERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSA
LTVRRKRKMAKRDIMYQPNEGDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVSKEILVKGTSEIGDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT

>Ph_Kni2_3R20 [aa]
LVVGMSKTGSRYGRRSNWFKVHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVE
RRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVCTSQ
EVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVW
PSILPFPSLAKIYPPPPFSSPAVSGALISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPID
LT

>Ph_Kni2_3R21 [aa]
CLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGMCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAV
ERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSEASVVKSESPHSSKVALSVSVSTS
QEVDSSNSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLRDETGAPMTEEERLSFLKSASKISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPV
WPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPI
DLTVRRKRKMAKRDIMYQPNESDLQEGMEDEYDQEQKDDADPKEESILKDEEVLEEEPVDERVSREILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
```

# [PhD_A2_Parhyale_goi_sequences]

## [PhD_A2.12_Ph_kni2]

### [PhD_A2.12.2_Ph_kni2_alignment]

**“alignment_Ph_kni2_all_aa_sequences.gif”**



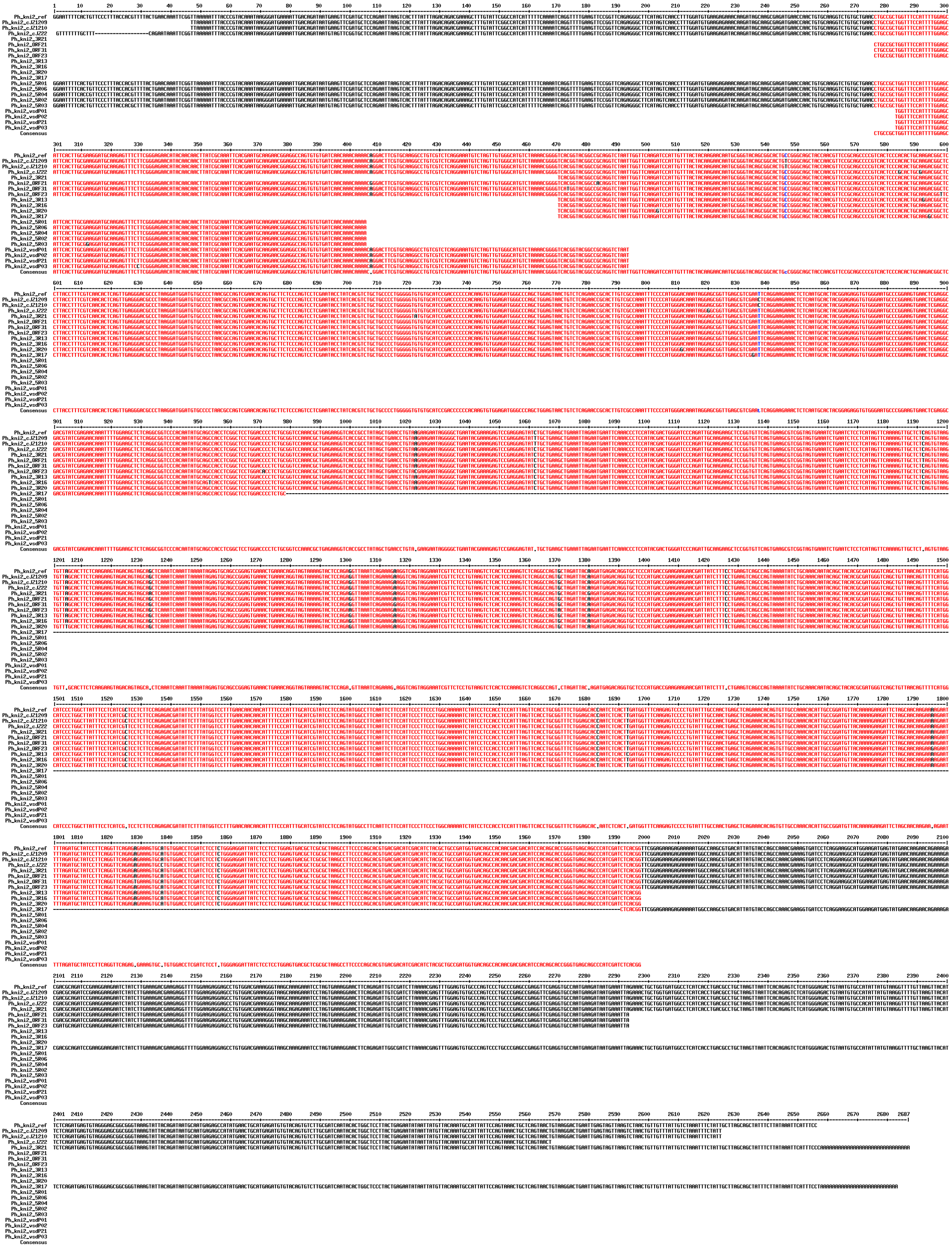
```
                1        10        20        30        40        50        60        70        80        90       100
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
    Ph_kni2_ref MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGT
  Ph_kni2_cJZ22 MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGT
  Ph_kni2_ORF23 MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGT
  Ph_kni2_ORF31 MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSWYGRRSNWFKIHCLLQEQCGYSGT
Ph_kni2_cJZ1209 MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGT
Ph_kni2_cJZ1210 MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGT
  Ph_kni2_ORF21       VCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQNGTSCKACRLRKCLVVGMSKTGSRYGHRSNWFKIHCLLQEQCGYSGT
   Ph_Kni2_3R21                                                                  CLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGT
   Ph_Kni2_3R13                                                                  CLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGT
   Ph_Kni2_3R16                                                                  CLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGT
   Ph_Kni2_3R20                                                                   LVVGMSKTGSRYGRRSNWFKVHCLLQEQCGYSGT
   Ph_Kni2_3R17                                                                  CLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGT
   Ph_kni2_5R01 MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQ
   Ph_kni2_5R02 MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQ
   Ph_kni2_5R04 MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQ
   Ph_kni2_5R06 MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQ
   Ph_kni2_5R03 MNQLCKVCAEPAAGFHFGAFTCGGCKSFFGRTYNNLSQIHECKNGGQCVINKQ
      Consensus MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYNNLSQIHECKNGGQCVINKQ

                101      110       120       130       140       150       160       170       180       190     200
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
    Ph_kni2_ref AGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
  Ph_kni2_cJZ22 AGQLPTFRSPVTPALRDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
  Ph_kni2_ORF23 AGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
  Ph_kni2_ORF31 AGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
Ph_kni2_cJZ1209 VGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRTQE
Ph_kni2_cJZ1210 VGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRTQE
  Ph_kni2_ORF21 AGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
   Ph_Kni2_3R21 AGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGMCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
   Ph_Kni2_3R13 AGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
   Ph_Kni2_3R16 AGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
   Ph_Kni2_3R20 AGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
   Ph_Kni2_3R17 AGQLPTFRSPVTPTLQGGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIRPPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQE
   Ph_kni2_5R01
   Ph_kni2_5R02
   Ph_kni2_5R04
   Ph_kni2_5R06
   Ph_kni2_5R03
      Consensus

                201      210       220       230       240       250       260       270       280       290     300
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
    Ph_kni2_ref ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSE
  Ph_kni2_cJZ22 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSE
  Ph_kni2_ORF23 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDTSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSE
  Ph_kni2_ORF31 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSE
Ph_kni2_cJZ1209 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVFAEAEIRMNSNPPYDWDPRLQEAPVFSE
Ph_kni2_cJZ1210 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVFAEAEIRMNSNPPYDWDPRLQEAPVFSE
  Ph_kni2_ORF21 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSE
   Ph_Kni2_3R21 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSE
   Ph_Kni2_3R13 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSE
   Ph_Kni2_3R16 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQSPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSE
   Ph_Kni2_3R20 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRLQEAPVFSE
   Ph_Kni2_3R17 ETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSA-------------------------------------------------------
   Ph_kni2_5R01
   Ph_kni2_5R02
   Ph_kni2_5R04
   Ph_kni2_5R06
   Ph_kni2_5R03
      Consensus

                301      310       320       330       340       350       360       370       380       390     400
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
    Ph_kni2_ref ASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASK
  Ph_kni2_cJZ22 ASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASK
  Ph_kni2_ORF23 ASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASTRLQDETGAPMTEEERLSFLKSASK
  Ph_kni2_ORF31 ASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASK
Ph_kni2_cJZ1209 ASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASK
Ph_kni2_cJZ1210 ASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASK
  Ph_kni2_ORF21 ASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASK
   Ph_Kni2_3R21 ASVVKSESPHSSKVALSVSVSTSQEVDSSNSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLRDETGAPMTEEERLSFLKSASK
   Ph_Kni2_3R13 ASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASK
   Ph_Kni2_3R16 ASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASK
   Ph_Kni2_3R20 ASVVKSESPHSSKVALSVSVCTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASK
   Ph_Kni2_3R17 ----------------------------------------------------------------------------------------------------
   Ph_kni2_5R01
   Ph_kni2_5R02
   Ph_kni2_5R04
   Ph_kni2_5R06
   Ph_kni2_5R03
      Consensus

                401      410       420       430       440       450       460       470       480       490     500
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
    Ph_kni2_ref ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
  Ph_kni2_cJZ22 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
  Ph_kni2_ORF23 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
  Ph_kni2_ORF31 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
Ph_kni2_cJZ1209 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
Ph_kni2_cJZ1210 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
  Ph_kni2_ORF21 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
   Ph_Kni2_3R21 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
   Ph_Kni2_3R13 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
   Ph_Kni2_3R16 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISLDGFKSPLYLPTELRNTV
   Ph_Kni2_3R20 ISANNTATRDGSAVNSFHGIPGLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGALISLDGFKSPLYLPTELRNTV
   Ph_Kni2_3R17 ----------------------------------------------------------------------------------------------------
   Ph_kni2_5R01
   Ph_kni2_5R02
   Ph_kni2_5R04
   Ph_kni2_5R06
   Ph_kni2_5R03
      Consensus

                501      510       520       530       540       550       560       570       580       590     600
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
    Ph_kni2_ref LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQP
  Ph_kni2_cJZ22 LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQP
  Ph_kni2_ORF23 LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQP
  Ph_kni2_ORF31 LPNTLPDVTKEDSSNKKRILDAILQVQSESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQP
Ph_kni2_cJZ1209 LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQP
Ph_kni2_cJZ1210 LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQP
  Ph_kni2_ORF21 LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQP
   Ph_Kni2_3R21 LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLTVRRKRKMAKRDIMYQP
   Ph_Kni2_3R13 LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLT
   Ph_Kni2_3R16 LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLT
   Ph_Kni2_3R20 LPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTTTSTAPGEQPIDLT
   Ph_Kni2_3R17 ----------------------------------------------------------------------------------LTVRRKRKMAKRDIMYQP
   Ph_kni2_5R01
   Ph_kni2_5R02
   Ph_kni2_5R04
   Ph_kni2_5R06
   Ph_kni2_5R03
      Consensus

                601      610       620       630       640       650       660       670       680       69092
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+-|
    Ph_kni2_ref NESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVSKEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
  Ph_kni2_cJZ22 NESDPQEGMEDKYEQEQKDDADPKEESILKDEEVLEEEPVDERVSKEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
  Ph_kni2_ORF23 NESDPQDGMEDEYEQEQKDDADPKEESIMKDEEVLEEEPVDERVSKEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
  Ph_kni2_ORF31 NESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVSKEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
Ph_kni2_cJZ1209 NESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVSKEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
Ph_kni2_cJZ1210 NESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVSKEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
  Ph_kni2_ORF21 NESDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVSKEILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
   Ph_Kni2_3R21 NESDLQEGMEDEYDQEQKDDADPKEESILKDEEVLEEEPVDERVSREILVKGTSEIVDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
   Ph_Kni2_3R13
   Ph_Kni2_3R16
   Ph_Kni2_3R20
   Ph_Kni2_3R17 NEGDPQEGMEDEYEQEQKDDADPKEESILKDEEVLEEEPVDERVSKEILVKGTSEIGDLKTSLECAQSLPEPRFEVPMKIMKLETAGDGLIT
   Ph_kni2_5R01
   Ph_kni2_5R02
   Ph_kni2_5R04
   Ph_kni2_5R06
   Ph_kni2_5R03
      Consensus
```

[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.12_Ph_kni2]

[PhD_A2.12.2_Ph_kni2_alignment]

## **“alignment_Ph_kni2_all_bp_sequences.gif”**

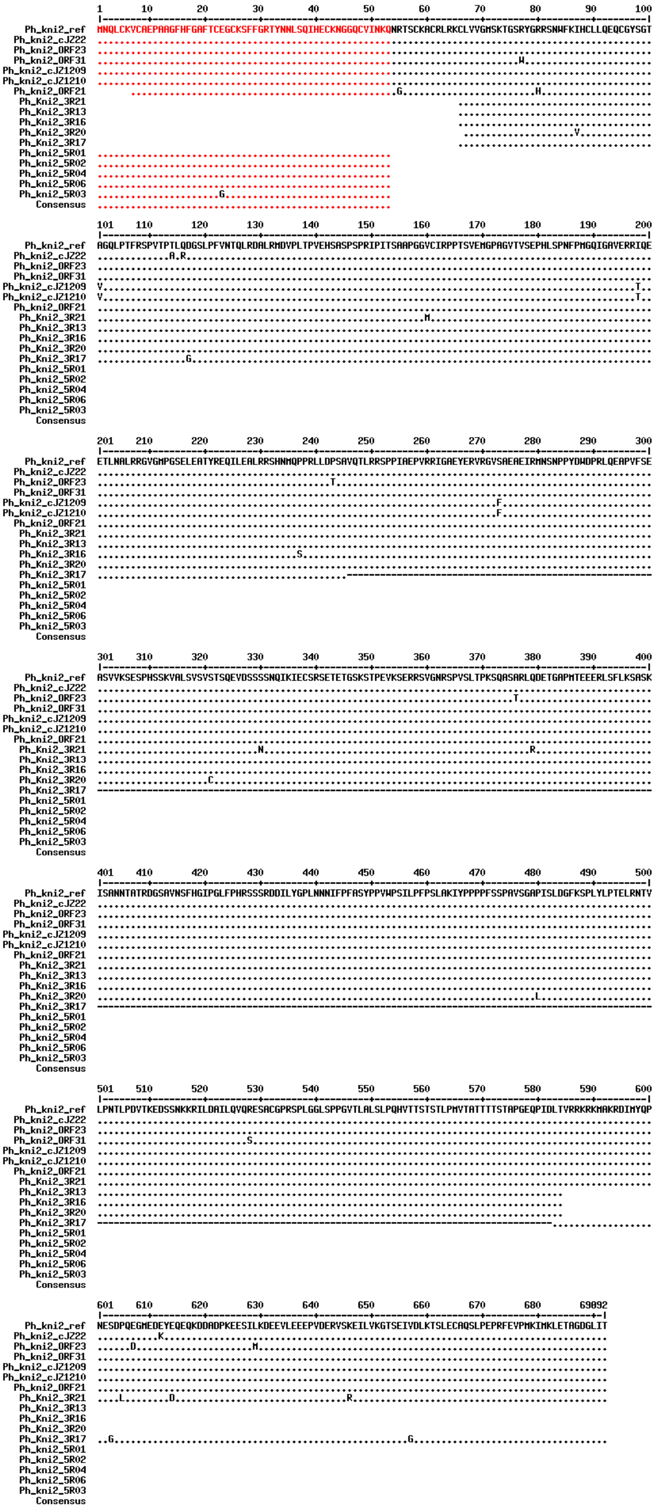


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.12_Ph_kni2]

[PhD_A2.12.2_Ph_kni2_alignment]

**“difference_alignment_Ph_kni2_all_aa_sequences.gif”**

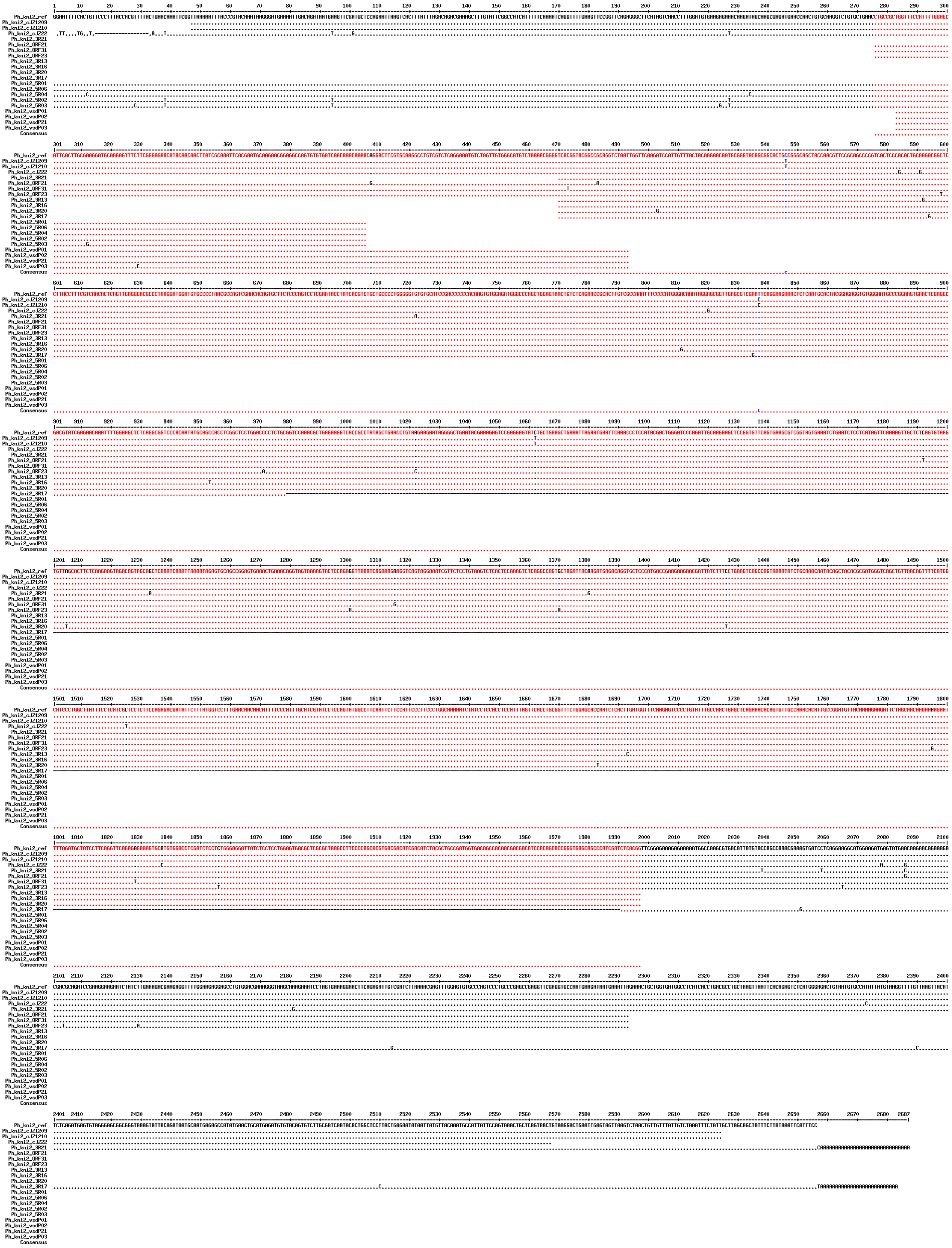


[PhD_A2_Parhyale_goi_sequences]

[PhD_A2.12_Ph_kni2]

[PhD_A2.12.2_Ph_kni2_alignment]

# “difference_alignment_Ph_kni2_all_bp_sequences.gif”



[PhD_A3_list_of_sequence_clones]

**“PhD_A3_List_of_Parhyale_goi_sequence_clones.pdf”**

## A3 List of *Parhyale hawaiensis* gene of interest sequence clones

**Table A1: List of *Parhyale hawaiensis* gene of interest sequence clones.** Shown are, from left to right: *Parhyale* gene of interest, name of clone, PCR reaction from which the clone derived, used oligonucleotides, length of the clone sequence (without oligonucleotide sequences), number of nucleotide exchanges compared to corresponding reference sequence (5.3.3.2), relative difference in nucleotide sequence, number of polymorphic nucleotide exchanges among all nucleotide exchanges, remaining non-polymorphic nucleotide exchanges and relative amount of non-polymorphic nucleotide exchanges. The last number was used to assess if a specific sequence in fact derived from the corresponding gene locus (or from the locus of a paralogous gene, 5.3.3.3).

| *P. h. gene o. i.* | **clone name** | **PCR reaction** | ***oligo-nucleotides*** | length | ***n. e.*** | ***rel. n. e.*** | **p.** | **n.-p. n. e.** | rel. n.-p. n. e. |
|---|---|---|---|---|---|---|---|---|---|
| *otd1* | **Ph_otd1_vsdP_QW12** | vsdPCR | *fw_QRR, re_QVW* | **109** | *1* | *0,9* | 0 | 1 | **0,9** |
| *otd1* | **Ph_otd1_vsdP_Phad04** | vsdPCR | *fw_QRR, re_QVW* | **109** | *1* | *0,9* | 0 | 1 | **0,9** |
| *otd1* | **Ph_otd1_vsdP_Phad05** | vsdPCR | *fw_QRR, re_QVW* | **109** | *3* | *2,8* | 3 | 0 | **0** |
| *otd1* | **Ph_otd1_5R01** | 1-step 5' RACE | *1005* | **605** | *11* | *1,8* | 4 | 7 | **1,2** |
| *otd1* | **Ph_otd1_5R03** | 1-step 5' RACE | *1005* | **605** | *8* | *1,3* | 6 | 2 | **0,3** |
| *otd1* | **Ph_otd1_5R04** | 1-step 5' RACE | *1005* | **605** | *2* | *0,3* | 1 | 1 | **0,2** |
| *otd1* | **Ph_otd1_3R09** | 1-step 3' RACE | *1000* | **1423** | *4* | *0,3* | 1 | 3 | **0,2** |
| *otd1* | **Ph_otd1_3R11** | 1-step 3' RACE | *1000* | **1440** | *3* | *0,2* | 0 | 3 | **0,2** |
| *otd1* | **Ph_otd1_3R12** | 1-step 3' RACE | *1000* | **1432** | *2* | *0,1* | 0 | 2 | **0,1** |
| *otd1* | **Ph_otd1_ORF17** | ORF isolation via ldPCR | *1014, 1015* | **1132** | *18* | *1,6* | 17 | 1 | **0,1** |
| *otd1* | **Ph_otd1_ORF21** | ORF isolation via ldPCR | *1014, 1015* | **1132** | *4* | *0,4* | 1 | 3 | **0,3** |
| *otd1* | **Ph_otd1_ORF201** | ORF isolation via ldPCR | *1014, 1015* | **1132** | *15* | *1,3* | 11 | 4 | **0,4** |
| *otd1* | **Ph_otd1_ORFs01** | ORF isolation via ldPCR | *1014, 1015* | **1132** | *3* | *0,3* | 1 | 2 | **0,2** |
| *otd1* | **Ph_otd1_ORFs02** | ORF isolation via ldPCR | *1014, 1015* | **1132** | *2* | *0,2* | 0 | 2 | **0,2** |
| *otd1* | **Ph_otd1_ORFs06** | ORF isolation via ldPCR | *1014, 1015* | **1132** | *21* | *1,9* | 16 | 5 | **0,4** |
| *otd1* | **Ph_otd1_cDNA25** | cDNA isolation via ldPCR | *1013, 1016* | **1950** | *5* | *0,3* | 0 | 5 | **0,3** |
| *otd1* | **Ph_otd1_cDNA26** | cDNA isolation via ldPCR | *1013, 1016* | **1941** | *35* | *1,8* | 31 | 4 | **0,2** |
| *otd1* | **Ph_otd1_cDNA27** | cDNA isolation via ldPCR | *1013, 1016* | **1955** | *16* | *0,8* | 13 | 3 | **0,2** |
| *otd1* | **Ph_otd1_cDNA29** | cDNA isolation via ldPCR | *1013, 1016* | **2170** | *46* | *2,1* | 24 | 22 | **1** |
| *otd1* | **Ph_otd1_cDNA34** | cDNA isolation via ldPCR | *1013, 1016* | **2165** | *6* | *0,3* | 2 | 4 | **0,2** |
| *otd2* | **Ph_otd2_vsdP_Pha104** | vsdPCR | *fw_QRR, re_QVW* | **109** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_vsdP_Pha105** | vsdPCR | *fw_QRR, re_QVW* | **109** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_5Rn143** | 2-step 5' RACE | *1105, 1107* | **566** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_5Rn164** | 1-step 5' RACE | *1105, 1107* | **542** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *otd2* | **Ph_otd2_5Rn167** | 1-step 5' RACE | *1105, 1107* | **570** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_5Rn169** | 1-step 5' RACE | *1105, 1107* | **221** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_5Rn171** | 1-step 5' RACE | *1105, 1107* | **339** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_5Rn200** | 1-step 5' RACE | *1105, 1107* | **957** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_5Rn203** | 1-step 5' RACE | *1105, 1107* | **957** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_5Rn204** | 1-step 5' RACE | *1105, 1107* | **957** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_3Rn81** | 2-step 3' RACE | *1100, 1101* | **3255** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *otd2* | **Ph_otd2_3Rn83** | 2-step 3' RACE | *1100, 1101* | **3260** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *hbn* | **Ph_hbn_vsdPms078** | vsdPCR | *fw_QRR, re_QVW* | **109** | *3* | *2,8* | 3 | 0 | **0** |
| *hbn* | **Ph_hbn_vsdPms093** | vsdPCR | *fw_QRR, re_QVW* | **109** | *3* | *2,8* | 3 | 0 | **0** |
| *hbn* | **Ph_hbn_vsdPms109** | vsdPCR | *fw_QRR, re_QVW* | **109** | *3* | *2,8* | 3 | 0 | **0** |
| *hbn* | **Ph_hbn_5RHH51** | 1-step 5' RACE | *1505* | **796** | *5* | *0,6* | 4 | 1 | **0,1** |
| *hbn* | **Ph_hbn_5RHH53** | 1-step 5' RACE | *1505* | **796** | *5* | *0,6* | 4 | 1 | **0,1** |
| *hbn* | **Ph_hbn_5RHH54** | 1-step 5' RACE | *1505* | **798** | *15* | *1,9* | 6 | 9 | **1,1** |
| *hbn* | **Ph_hbn_5Rn25** | 2-step 5' RACE | *1506, 1507* | **731** | *0* | *0* | 0 | 0 | **0** |
| *hbn* | **Ph_hbn_5Rn26** | 2-step 5' RACE | *1506, 1507* | **731** | *0* | *0* | 0 | 0 | **0** |
| *hbn* | **Ph_hbn_5Rn28** | 2-step 5' RACE | *1506, 1507* | **731** | *1* | *0,1* | 0 | 1 | **0,1** |
| *hbn* | **Ph_hbn_3RHH313** | 2-step 3' RACE | *1500, 1501* | **628** | *2* | *0,3* | 1 | 1 | **0,2** |
| *hbn* | **Ph_hbn_3RHH314** | 2-step 3' RACE | *1500, 1501* | **628** | *3* | *0,5* | 1 | 2 | **0,3** |
| *hbn* | **Ph_hbn_3RHH315** | 2-step 3' RACE | *1500, 1501* | **628** | *4* | *0,6* | 1 | 3 | **0,5** |
| *hbn* | **Ph_hbn_3RHH336** | 2-step 3' extension RACE | *1503, 1504* | **609** | *6* | *1* | 2 | 4 | **0,7** |
| *hbn* | **Ph_hbn_3RHH356** | 2-step 3' extension RACE | *1503, 1504* | **721** | *0* | *0* | 0 | 0 | **0** |
| *hbn* | **Ph_hbn_3RHH357** | 2-step 3' extension RACE | *1503, 1504* | **1311** | *0* | *0* | 0 | 0 | **0** |
| *hbn* | **Ph_hbn_3RHH358** | 2-step 3' extension RACE | *1503, 1504* | **710** | *1* | *0,1* | 0 | 1 | **0,1** |
| *hbn* | **Ph_hbn_3RHH361** | 2-step 3' extension RACE | *1503, 1504* | **1471** | *5* | *0,3* | 3 | 2 | **0,1** |
| *hbn* | **Ph_hbn_3RHH363** | 2-step 3' extension RACE | *1503, 1504* | **709** | *5* | *0,7* | 3 | 2 | **0,3** |
| *hbn* | **Ph_hbn_3RHH366** | 2-step 3' extension RACE | *1503, 1504* | **713** | *3* | *0,4* | 3 | 0 | **0** |
| *hbn* | **Ph_hbn_ORF01** | ORF isolation via ldPCR | *1514, 1515* | **1519** | *5* | *0,3* | 4 | 1 | **0,1** |
| *hbn* | **Ph_hbn_ORF02** | ORF isolation via ldPCR | *1514, 1515* | **1519** | *6* | *0,4* | 4 | 2 | **0,1** |
| *hbn* | **Ph_hbn_ORF05** | ORF isolation via ldPCR | *1514, 1515* | **1519** | *5* | *0,3* | 4 | 1 | **0,1** |
| *hbn* | **Ph_hbn_cDNA01** | cDNA isolation via ldPCR | *1512, 1516* | **2559** | *1* | *0* | 0 | 1 | **0** |
| *hbn* | **Ph_hbn_cDNA05** | cDNA isolation via ldPCR | *1512, 1516* | **2559** | *0* | *0* | 0 | 0 | **0** |
| *hbn* | **Ph_hbn_cDNA11** | cDNA isolation via ldPCR | *1512, 1516* | **2558** | *5* | *0,2* | 0 | 5 | **0,2** |
| *al1* | **Ph_al1_vsdP_Phad16** | vsdPCR | *fw_QRR, re_QVW* | **109** | *0* | *0* | 0 | 0 | **0** |
| *al1* | **Ph_al1_vsdP_Pha092** | vsdPCR | *fw_QRR, re_QVW* | **109** | *0* | *0* | 0 | 0 | **0** |
| *al1* | **Ph_al1_vsdP_Pha097** | vsdPCR | *fw_QRR, re_QVW* | **109** | *0* | *0* | 0 | 0 | **0** |
| *al1* | **Ph_al1_vsdP_Pha102** | vsdPCR | *fw_QRR, re_QVW* | **109** | *0* | *0* | 0 | 0 | **0** |
| *al1* | **Ph_al1_5Rn2** | 2-step 5' RACE | *1207, 1208* | **845** | *16* | *1,9* | 15 | 1 | **0,1** |
| *al1* | **Ph_al1_5Rn3** | 2-step 5' RACE | *1207, 1208* | **843** | *16* | *1,9* | 14 | 2 | **0,2** |
| *al1* | **Ph_al1_5Rn4** | 2-step 5' RACE | *1207, 1208* | **845** | *16* | *1,9* | 15 | 1 | **0,1** |
| *al1* | **Ph_al1_5Rn118** | 2-step 5' RACE | *1207, 1208* | **843** | *0* | *0* | 0 | 0 | **0** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *al1* | **Ph_al1_5Rn119** | 2-step 5' RACE | *1207, 1208* | **844** | *0* | *0* | 0 | 0 | **0** |
| *al1* | **Ph_al1_3Rn5** | 2-step 3' RACE | *1202, 1203* | **932** | *1* | *0,1* | 0 | 1 | **0,1** |
| *al1* | **Ph_al1_3Rn7** | 2-step 3' RACE | *1202, 1203* | **930** | *19* | *2* | 4 | 15 | **1,6** |
| *al1* | **Ph_al1_3Rn8** | 2-step 3' RACE | *1202, 1203* | **930** | *1* | *0,1* | 0 | 1 | **0,1** |
| *al1* | **Ph_al1_ORF13** | ORF isolation via ldPCR | *1214, 1215* | **1466** | *33* | *2,3* | 33 | 0 | **0** |
| *al1* | **Ph_al1_ORF14** | ORF isolation via ldPCR | *1214, 1215* | **1466** | *35* | *2,4* | 33 | 2 | **0,1** |
| *al1* | **Ph_al1_ORF15** | ORF isolation via ldPCR | *1214, 1215* | **1466** | *33* | *2,3* | 33 | 0 | **0** |
| *al1* | **Ph_al1_ORF16** | ORF isolation via ldPCR | *1214, 1215* | **1466** | *4* | *0,3* | 0 | 4 | **0,3** |
| *al1* | **Ph_al1_ORF17** | ORF isolation via ldPCR | *1214, 1215* | **1466** | *3* | *0,2* | 0 | 3 | **0,2** |
| *al1* | **Ph_al1_ORF18** | ORF isolation via ldPCR | *1214, 1215* | **1466** | *34* | *2,3* | 33 | 1 | **0,1** |
| *al1* | **Ph_al1_cDNA01** | cDNA isolation via ldPCR | *1213, 1216* | **1752** | *1* | *0,1* | 0 | 1 | **0,1** |
| *al1* | **Ph_al1_cDNA04** | cDNA isolation via ldPCR | *1213, 1216* | **1752** | *4* | *0,2* | 0 | 4 | **0,2** |
| *al1* | **Ph_al1_cDNA05** | cDNA isolation via ldPCR | *1213, 1216* | **1752** | *3* | *0,2* | 0 | 3 | **0,2** |
| *al2* | **Ph_al2_vsdP_Pha089** | vsdPCR | *fw_QRR, re_QVW* | **109** | *0* | *0* | 0 | 0 | **0** |
| *al2* | **Ph_al2_vsdP_Pha111** | vsdPCR | *fw_QRR, re_QVW* | **109** | *2* | *1,8* | 2 | 0 | **0** |
| *al2* | **Ph_al2_5Rg1** | 1-step 5' RACE | *1305* | **403** | *5* | *1,2* | 5 | 0 | **0** |
| *al2* | **Ph_al2_5Rp09** | 1-step 5' RACE | *1305* | **1559** | *28* | *1,8* | 26 | 2 | **0,1** |
| *al2* | **Ph_al2_5Rp11** | 1-step 5' RACE | *1305* | **1560** | *31* | *2* | 27 | 4 | **0,3** |
| *al2* | **Ph_al2_5Rp13** | 1-step 5' RACE | *1305* | **1559** | *27* | *1,7* | 27 | 0 | **0** |
| *al2* | **Ph_al2_5Rn09** | 2-step 5' RACE | *1305, 1306* | **1520** | *18* | *1,2* | 17 | 1 | **0,1** |
| *al2* | **Ph_al2_5Rn11** | 2-step 5' RACE | *1305, 1306* | **1520** | *17* | *1,1* | 17 | 0 | **0** |
| *al2* | **Ph_al2_5Rn12** | 2-step 5' RACE | *1305, 1306* | **1519** | *19* | *1,3* | 17 | 2 | **0,1** |
| *al2* | **Ph_al2_3Rn227** | 2-step 3' RACE | *1300, 1301* | **590** | *0* | *0* | 0 | 0 | **0** |
| *al2* | **Ph_al2_3Rfwx01** | 2-step 3' extension RACE | *1302, 1304* | **463** | *1* | *0,2* | 0 | 1 | **0,2** |
| *al2* | **Ph_al2_3Rfwx05** | 2-step 3' extension RACE | *1302, 1304* | **559** | *1* | *0,2* | 0 | 1 | **0,2** |
| *al2* | **Ph_al2_ORF03** | ORF isolation via ldPCR | *1314, 1315* | **1707** | *14* | *0,8* | 11 | 3 | **0,2** |
| *al2* | **Ph_al2_ORF04** | ORF isolation via ldPCR | *1314, 1315* | **1707** | *16* | *0,9* | 11 | 5 | **0,3** |
| *al2* | **Ph_al2_ORF05** | ORF isolation via ldPCR | *1314, 1315* | **1707** | *16* | *0,9* | 0 | 16 | **0,9** |
| *pby1* | **Ph_pby1_vsdP_PhQW11** | vsdPCR | *fw_QRR, re_QVW* | **109** | *1* | *0,9* | 1 | 0 | **0** |
| *pby1* | **Ph_pby1_vsdP_Phad02** | vsdPCR | *fw_QRR, re_QVW* | **109** | *1* | *0,9* | 1 | 0 | **0** |
| *pby1* | **Ph_pby1_vsdP_Phad06** | vsdPCR | *fw_QRR, re_QVW* | **109** | *2* | *1,8* | 1 | 1 | **0,9** |
| *pby1* | **Ph_pby1_vsdP_Phad08** | vsdPCR | *fw_QRR, re_QVW* | **109** | *2* | *1,8* | 1 | 1 | **0,9** |
| *pby1* | **Ph_pby1_vsdP_Phad14** | vsdPCR | *fw_QRR, re_QVW* | **109** | *3* | *2,8* | 1 | 2 | **1,8** |
| *pby1* | **Ph_pby1_5R0101** | 2-step 5' RACE | *1605, 1607* | **764** | *8* | *1* | 5 | 3 | **0,4** |
| *pby1* | **Ph_pby1_5n54** | 2-step 5' RACE | *1606, 1607* | **764** | *7* | *0,9* | 0 | 7 | **0,9** |
| *pby1* | **Ph_pby1_5n55** | 2-step 5' RACE | *1606, 1607* | **764** | *1* | *0,1* | 0 | 1 | **0,1** |
| *pby1A* | **Ph_pby1_3R2219** | 2-step 3' RACE | *1601, 1602* | **1308** | *2* | *0,2* | 0 | 2 | **0,2** |
| *pby1A* | **Ph_pby1_3Rf27** | 2-step 3' RACE | *1601, 1602* | **1304** | *10* | *0,8* | 7 | 3 | **0,2** |
| *pby1A* | **Ph_pby1_3Rf29** | 2-step 3' RACE | *1601, 1602* | **1304** | *31* | *2,4* | 10 | 21 | **1,6** |
| *pby1A* | **Ph_pby1A_ORF09** | ORF isolation via ldPCR | *1614, 1616* | **1599** | *18* | *1,1* | 14 | 4 | **0,3** |
| *pby1A* | **Ph_pby1A_ORF13** | ORF isolation via ldPCR | *1614, 1616* | **1599** | *24* | *1,5* | 16 | 8 | **0,5** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *pby1A* | **Ph_pby1A_ORF16** | ORF isolation via ldPCR | *1614, 1616* | **1599** | *1* | *0,1* | 0 | 1 | **0,1** |
| *pby1B* | **Ph_pby1_3R1413** | 2-step 3' RACE | *1601, 1602* | **951** | *5* | *0,5* | 3 | 2 | **0,2** |
| *pby1B* | **Ph_pby1_3Rf21** | 2-step 3' RACE | *1601, 1602* | **950** | *10* | *1,1* | 3 | 7 | **0,7** |
| *pby1B* | **Ph_pby1_3Rf316** | 2-step 3' RACE | *1601, 1602* | **928** | *5* | *0,5* | 3 | 2 | **0,2** |
| *pby1B* | **Ph_pby1B_ORF01** | ORF isolation via ldPCR | *1614, 1615* | **1132** | *1* | *0,1* | 0 | 1 | **0,1** |
| *pby1B* | **Ph_pby1B_ORF03** | ORF isolation via ldPCR | *1614, 1615* | **1132** | *13* | *1,1* | 7 | 6 | **0,5** |
| *pby1B* | **Ph_pby1B_ORF05** | ORF isolation via ldPCR | *1614, 1615* | **1132** | *5* | *0,4* | 1 | 4 | **0,4** |
| *pby2* | **Ph_pby2_vsdP_Phad081** | vsdPCR | *fw_QRR, re_QVW* | **109** | *n. a.* | *n. a.* | n. a. | n. a. | **n. a.** |
| *awh* | **Ph_awh_vsdP10** | vsdPCR | *fw_QRR, re_QVW* | **109** | *0* | *0* | 0 | 0 | **0** |
| *awh* | **Ph_awh_vsdP15** | vsdPCR | *fw_QRR, re_QVW* | **109** | *1* | *0,9* | 0 | 1 | **0,9** |
| *awh* | **Ph_awh5RCR1** | 2-step 5' RACE | *1405, 1406* | **2089** | *24* | *1,1* | 15 | 9 | **0,4** |
| *awh* | **Ph_awh5RCR2** | 2-step 5' RACE | *1405, 1406* | **2020** | *14* | *0,7* | 14 | 0 | **0** |
| *awh* | **Ph_awh5RCR5** | 2-step 5' RACE | *1405, 1406* | **2021** | *16* | *0,8* | 14 | 2 | **0,1** |
| *awh* | **Ph_awh5RCR6** | 2-step 5' RACE | *1405, 1406* | **2087** | *23* | *1,1* | 20 | 3 | **0,1** |
| *awh* | **Ph_awh_5Rn121** | 2-step 5' RACE | *1405, 1406* | **1861** | *60* | *3,2* | 60 | 0 | **0** |
| *awh* | **Ph_awh_5Rn122** | 2-step 5' RACE | *1405, 1406* | **1957** | *60* | *3,1* | 49 | 11 | **0,6** |
| *awh* | **Ph_awh_5Rn150** | 2-step 5' RACE | *1405, 1406* | **1813** | *54* | *3* | 51 | 3 | **0,2** |
| *awh* | **Ph_awh_3Rn21** | 2-step 3' RACE | *1400, 1401* | **1522** | *0* | *0* | 0 | 0 | **0** |
| *awh* | **Ph_awh_3Rn22** | 2-step 3' RACE | *1400, 1401* | **1377** | *8* | *0,6* | 8 | 0 | **0** |
| *awh* | **Ph_awh_3Rn78** | 2-step 3' RACE | *1400, 1401* | **1521** | *3* | *0,2* | 1 | 2 | **0,1** |
| *awh* | **Ph_awh_3RCR07** | 2-step 3' RACE | *1400, 1401* | **1571** | *32* | *2* | 24 | 8 | **0,5** |
| *awh* | **Ph_awh_3RCR09** | 2-step 3' RACE | *1400, 1401* | **1532** | *35* | *2,3* | 32 | 3 | **0,2** |
| *awh* | **Ph_awh_3RCR10** | 2-step 3' RACE | *1400, 1401* | **1535** | *26* | *1,7* | 19 | 7 | **0,5** |
| *awh* | **Ph_awh_ORF20** | ORF isolation via ldPCR | *1414, 1415* | **1058** | *5* | *0,5* | 0 | 5 | **0,5** |
| *awh* | **Ph_awh_ORF21** | ORF isolation via ldPCR | *1414, 1415* | **1058** | *0* | *0* | 0 | 0 | **0** |
| *awh* | **Ph_awh_ORF22** | ORF isolation via ldPCR | *1414, 1415* | **1058** | *1* | *0,1* | 0 | 1 | **0,1** |
| *six3* | **Ph_six3_vsdP13** | vsdPCR | *fw_QAM, re_PTQ* | **217** | *5* | *2,3* | 5 | 0 | **0** |
| *six3* | **Ph_six3_vsdP17** | vsdPCR | *fw_QAM, re_PTQ* | **217** | *6* | *2,8* | 5 | 1 | **0,5** |
| *six3* | **Ph_six3_vsdP19** | vsdPCR | *fw_QAM, re_PTQ* | **217** | *1* | *0,5* | 1 | 0 | **0** |
| *six3* | **Ph_six3_5R01** | 2-step 5' RACE | *2005, 2006* | **496** | *4* | *0,8* | 1 | 3 | **0,6** |
| *six3* | **Ph_six3_5R02** | 2-step 5' RACE | *2005, 2006* | **562** | *6* | *1,1* | 4 | 2 | **0,4** |
| *six3* | **Ph_six3_5R04** | 2-step 5' RACE | *2005, 2006* | **495** | *0* | *0* | 0 | 0 | **0** |
| *six3* | **Ph_six3_5R05** | 2-step 5' RACE | *2005, 2006* | **540** | *6* | *1,1* | 4 | 2 | **0,4** |
| *six3* | **Ph_six3_5R06** | 2-step 5' RACE | *2005, 2006* | **562** | *0* | *0* | 0 | 0 | **0** |
| *six3* | **Ph_six3_5R16** | 2-step 5' extension RACE | *2006, 2007* | **490** | *2* | *0,4* | 0 | 2 | **0,4** |
| *six3* | **Ph_six3_5R17** | 2-step 5' extension RACE | *2006, 2007* | **490** | *4* | *0,8* | 3 | 1 | **0,2** |
| *six3* | **Ph_six3_5R18** | 2-step 5' extension RACE | *2006, 2007* | **490** | *4* | *0,8* | 3 | 1 | **0,2** |
| *six3* | **Ph_six3_5R19** | 2-step 5' extension RACE | *2006, 2008* | **76** | *2* | *2,6* | 0 | 2 | **2,6** |
| *six3* | **Ph_six3_5R23** | 2-step 5' extension RACE | *2006, 2008* | **76** | *0* | *0* | 0 | 0 | **0** |
| *six3* | **Ph_six3_5R24** | 2-step 5' extension RACE | *2006, 2008* | **76** | *0* | *0* | 0 | 0 | **0** |
| *six3* | **Ph_six3_3R08** | 2-step 3' RACE | *2000, 2001* | **3067** | *19* | *0,6* | 16 | 3 | **0,1** |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *six3* | **Ph_six3_3R09** | 2-step 3' RACE | *2000, 2001* | **3070** | *19* | *0,6* | 16 | 3 | **0,1** |
| *six3* | **Ph_six3_3R10** | 2-step 3' RACE | *2000, 2001* | **3070** | *19* | *0,6* | 16 | 3 | **0,1** |
| *six3* | **Ph_six3_ORF10** | ORF isolation via ldPCR | *2014, 2015* | **1638** | *3* | *0,2* | 0 | 3 | **0,2** |
| *six3* | **Ph_six3_ORF11** | ORF isolation via ldPCR | *2014, 2015* | **1638** | *4* | *0,2* | 0 | 4 | **0,2** |
| *six3* | **Ph_six3_ORF12** | ORF isolation via ldPCR | *2014, 2015* | **1638** | *1* | *0,1* | 0 | 1 | **0,1** |
| *six3* | **Ph_six3_cDNA09** | cDNA isolation via ldPCR | *2013, 2016* | **3572** | *3* | *0,1* | 0 | 3 | **0,1** |
| *six3* | **Ph_six3_cDNA13** | cDNA isolation via ldPCR | *2013, 2016* | **3572** | *10* | *0,3* | 0 | 10 | **0,3** |
| *six4* | **Ph_six4_vsdP14** | vsdPCR | *fw_QAM, re_PTQ* | **217** | *0* | *0* | 0 | 0 | **0** |
| *six4* | **Ph_six4_vsdP16** | vsdPCR | *fw_QAM, re_PTQ* | **217** | *0* | *0* | 0 | 0 | **0** |
| *six4* | **Ph_six4_vsdP21** | vsdPCR | *fw_QAM, re_PTQ* | **217** | *0* | *0* | 0 | 0 | **0** |
| *six4* | **Ph_six4_5R13** | 2-step 5' RACE | *2105, 2106* | **1642** | *1* | *0,1* | 0 | 1 | **0,1** |
| *six4* | **Ph_six4_5R14** | 2-step 5' RACE | *2105, 2106* | **1564** | *6* | *0,4* | 0 | 6 | **0,4** |
| *six4* | **Ph_six4_5R15** | 2-step 5' RACE | *2105, 2106* | **1647** | *50* | *3* | 0 | 50 | **3** |
| *six4* | **Ph_six4_3R19** | 2-step 3' RACE | *2100* | **1179** | *2* | *0,2* | 0 | 2 | **0,2** |
| *six4* | **Ph_six4_3R20** | 2-step 3' RACE | *2100* | **1164** | *2* | *0,2* | 0 | 2 | **0,2** |
| *six4* | **Ph_six4_3R24** | 2-step 3' RACE | *2100, 2101* | **1092** | *2* | *0,2* | 0 | 2 | **0,2** |
| *six4* | **Ph_six4_ORF14** | ORF isolation via ldPCR | *2114, 2115* | **1556** | *2* | *0,1* | 0 | 2 | **0,1** |
| *six4* | **Ph_six4_ORF15** | ORF isolation via ldPCR | *2114, 2115* | **1556** | *2* | *0,1* | 0 | 2 | **0,1** |
| *six4* | **Ph_six4_ORF16** | ORF isolation via ldPCR | *2114, 2115* | **1556** | *1* | *0,1* | 0 | 1 | **0,1** |
| *gbx* | **Ph_gbxvsdP105** | vsdPCR | *fw_AFT, re_KIW* | **100** | *0* | *0* | 0 | 0 | **0** |
| *gbx* | **Ph_gbxvsdP106** | vsdPCR | *fw_AFT, re_KIW* | **100** | *0* | *0* | 0 | 0 | **0** |
| *gbx* | **Ph_gbxvsdP108** | vsdPCR | *fw_AFT, re_KIW* | **100** | *0* | *0* | 0 | 0 | **0** |
| *gbx* | **Ph_gbxvsdP109** | vsdPCR | *fw_AFT, re_KIW* | **100** | *1* | *1* | 0 | 1 | **1** |
| *gbx* | **Ph_gbxvsdP110** | vsdPCR | *fw_AFT, re_KIW* | **100** | *0* | *0* | 0 | 0 | **0** |
| *gbx* | **Ph_gbxvsdP111** | vsdPCR | *fw_AFT, re_KIW* | **100** | *0* | *0* | 0 | 0 | **0** |
| *gbx* | **Ph_gbxvsdP112** | vsdPCR | *fw_AFT, re_KIW* | **100** | *0* | *0* | 0 | 0 | **0** |
| *gbx* | **Ph_gbx_5R02** | 2-step 5' RACE | *2305, 2306* | **1408** | *2* | *0,1* | 0 | 2 | **0,1** |
| *gbx* | **Ph_gbx_5R03** | 2-step 5' RACE | *2305, 2306* | **1409** | *3* | *0,2* | 0 | 3 | **0,2** |
| *gbx* | **Ph_gbx_5R05** | 2-step 5' RACE | *2305, 2306* | **1409** | *3* | *0,2* | 1 | 2 | **0,1** |
| *gbx* | **Ph_gbx_3R03** | 2-step 3' RACE | *2301, 2303* | **1674** | *4* | *0,2* | 1 | 3 | **0,2** |
| *gbx* | **Ph_gbx_3R11** | 2-step 3' RACE | *2301, 2303* | **2026** | *8* | *0,4* | 1 | 7 | **0,3** |
| *gbx* | **Ph_gbx_3R12** | 2-step 3' RACE | *2301, 2303* | **2021** | *3* | *0,1* | 1 | 2 | **0,1** |
| *gbx* | **Ph_gbx_ORF01** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *4* | *0,2* | 0 | 4 | **0,2** |
| *gbx* | **Ph_gbx_ORF02** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *5* | *0,3* | 0 | 5 | **0,3** |
| *gbx* | **Ph_gbx_ORF03** | ORF isolation via ldPCR | *2314, 2315* | **1725** | *41* | *2,4* | 1 | 40 | **2,3** |
| *gbx* | **Ph_gbx_ORF04** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *2* | *0,1* | 0 | 2 | **0,1** |
| *gbx* | **Ph_gbx_ORF05** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *3* | *0,2* | 0 | 3 | **0,2** |
| *gbx* | **Ph_gbx_ORF07** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *3* | *0,2* | 0 | 3 | **0,2** |
| *gbx* | **Ph_gbx_ORF52** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *3* | *0,2* | 0 | 3 | **0,2** |
| *gbx* | **Ph_gbx_ORF53** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *2* | *0,1* | 0 | 2 | **0,1** |
| *gbx* | **Ph_gbx_ORF55** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *4* | *0,2* | 0 | 4 | **0,2** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *gbx* | **Ph_gbx_ORF56** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *2* | *0,1* | 0 | 2 | **0,1** |
| *gbx* | **Ph_gbx_ORF57** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *2* | *0,1* | 0 | 2 | **0,1** |
| *gbx* | **Ph_gbx_ORF58** | ORF isolation via ldPCR | *2314, 2315* | **1722** | *0* | *0* | 0 | 0 | **0** |
| *kni1* | **Ph_kni1_vsdP07** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *5* | *2,4* | 0 | 5 | **2,4** |
| *kni1* | **Ph_kni1_vsdP09** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *0* | *0* | 0 | 0 | **0** |
| *kni1* | **Ph_kni1_vsdP14** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *1* | *0,5* | 1 | 0 | **0** |
| *kni1* | **Ph_kni1_vsdP16** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *1* | *0,5* | 0 | 1 | **0,5** |
| *kni1* | **Ph_kni1_vsdP18** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *0* | *0* | 0 | 0 | **0** |
| *kni1* | **Ph_kni1_vsdP23** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *1* | *0,5* | 0 | 1 | **0,5** |
| *kni1* | **Ph_kni1_vsdP24** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *0* | *0* | 0 | 0 | **0** |
| *kni1* | **Ph_kni1_5R25** | 2-step 5' RACE | *1905, 1906* | **784** | *13* | *1,7* | 16 | 4 | **0,5** |
| *kni1* | **Ph_kni1_5R26** | 2-step 5' RACE | *1905, 1906* | **785** | *3* | *0,4* | 16 | 1 | **0,1** |
| *kni1* | **Ph_kni1_5R27** | 2-step 5' RACE | *1905, 1906* | **785** | *9* | *1,1* | 16 | 1 | **0,1** |
| *kni1* | **Ph_kni1_5R30** | 2-step 5' RACE | *1905, 1906* | **753** | *11* | *1,5* | 16 | 0 | **0** |
| *kni1* | **Ph_kni1_5R31** | 2-step 5' RACE | *1905, 1906* | **784** | *44* | *5,6* | 16 | 32 | **4,1** |
| *kni1* | **Ph_kni1_5R33** | 2-step 5' RACE | *1905, 1906* | **783** | *13* | *1,7* | 16 | 2 | **0,3** |
| *kni1* | **Ph_kni1_5R36** | 2-step 5' RACE | *1905, 1906* | **783** | *12* | *1,5* | 16 | 1 | **0,1** |
| *kni1* | **Ph_kni1_3R41** | 2-step 3' RACE | *1900, 1901* | **548** | *2* | *0,4* | 2 | 1 | **0,2** |
| *kni1* | **Ph_kni1_3R43** | 2-step 3' RACE | *1900, 1901* | **559** | *6* | *1,1* | 2 | 4 | **0,7** |
| *kni1* | **Ph_kni1_3R44** | 2-step 3' RACE | *1900, 1901* | **559** | *3* | *0,5* | 2 | 1 | **0,2** |
| *kni1* | **Ph_kni1_3R45** | 2-step 3' RACE | *1900, 1901* | **548** | *6* | *1,1* | 2 | 5 | **0,9** |
| *kni1* | **Ph_kni1_3Rfwxn08** | 2-step 3' extension RACE | *1903, 1904* | **621** | *8* | *1,3* | 5 | 3 | **0,5** |
| *kni1* | **Ph_kni1_3Rfwxn10** | 2-step 3' extension RACE | *1903, 1904* | **763** | *3* | *0,4* | 5 | 2 | **0,3** |
| *kni1* | **Ph_kni1_ORF02** | ORF isolation via ldPCR | *1914, 1915* | **1150** | *4* | *0,3* | 6 | 4 | **0,3** |
| *kni1* | **Ph_kni1_ORF03** | ORF isolation via ldPCR | *1914, 1915* | **1150** | *10* | *0,9* | 6 | 5 | **0,4** |
| *kni1* | **Ph_kni1_ORF04** | ORF isolation via ldPCR | *1914, 1915* | **1150** | *7* | *0,6* | 6 | 2 | **0,2** |
| *kni1* | **Ph_kni1_ORF07** | ORF isolation via ldPCR | *1914, 1915* | **1150** | *11* | *1* | 6 | 6 | **0,5** |
| *kni1* | **Ph_kni1_ORF08** | ORF isolation via ldPCR | *1914, 1915* | **1150** | *5* | *0,4* | 6 | 1 | **0,1** |
| *kni1* | **Ph_kni1_ORF11** | ORF isolation via ldPCR | *1914, 1915* | **1150** | *2* | *0,2* | 6 | 2 | **0,2** |
| *kni1* | **Ph_kni1_ORF12** | ORF isolation via ldPCR | *1914, 1915* | **1150** | *7* | *0,6* | 6 | 3 | **0,3** |
| *kni1* | **Ph_kni1_cDNAJZB1** | cDNA isolation via ldPCR | *1913, 1919* | **2179** | *24* | *1,1* | 17 | 7 | **0,3** |
| *kni1* | **Ph_kni1_cDNAJZB2** | cDNA isolation via ldPCR | *1913, 1919* | **2178** | *50* | *2,3* | 46 | 4 | **0,2** |
| *kni1* | **Ph_kni1_cDNAJZB5** | cDNA isolation via ldPCR | *1913, 1919* | **2179** | *0* | *0* | 0 | 0 | **0** |
| *kni2* | **Ph_kni2_vsdP01** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *0* | *0* | 0 | 0 | **0** |
| *kni2* | **Ph_kni2_vsdP02** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *0* | *0* | 0 | 0 | **0** |
| *kni2* | **Ph_kni2_vsdP03** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *1* | *0,5* | 0 | 1 | **0,5** |
| *kni2* | **Ph_kni2_vsdP21** | vsdPCR | *fw_CKV, re_LLQ* | **211** | *0* | *0* | 0 | 0 | **0** |
| *kni2* | **Ph_kni2_5R01** | 2-step 5' RACE | *1805, 1806* | **405** | *0* | *0* | 3 | 0 | **0** |
| *kni2* | **Ph_kni2_5R02** | 2-step 5' RACE | *1805, 1806* | **405** | *3* | *0,7* | 3 | 0 | **0** |
| *kni2* | **Ph_kni2_5R03** | 2-step 5' RACE | *1805, 1806* | **405** | *6* | *1,5* | 3 | 3 | **0,7** |
| *kni2* | **Ph_kni2_5R04** | 2-step 5' RACE | *1805, 1806* | **405** | *2* | *0,5* | 3 | 2 | **0,5** |

|  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|
| *kni2* | **Ph_kni2_5R06** | 2-step 5' RACE | *1805, 1806* | **405** | *0* | *0* | 3 | 0 | **0** |
| *kni2* | **Ph_Kni2_3R13** | 2-step 3' RACE | *1800, 1801* | **1528** | *2* | *0,1* | 0 | 2 | **0,1** |
| *kni2* | **Ph_Kni2_3R16** | 2-step 3' RACE | *1800, 1801* | **1528** | *1* | *0,1* | 0 | 1 | **0,1** |
| *kni2* | **Ph_Kni2_3R17** | 2-step 3' RACE | *1800, 1801* | **1202** | *6* | *0,5* | 0 | 6 | **0,5** |
| *kni2* | **Ph_Kni2_3R20** | 2-step 3' RACE | *1800, 1801* | **1528** | *5* | *0,3* | 0 | 5 | **0,3** |
| *kni2* | **Ph_Kni2_3R21** | 2-step 3' RACE | *1800, 1801* | **2218** | *7* | *0,3* | 0 | 7 | **0,3** |
| *kni2* | **Ph_kni2_ORF21** | ORF isolation via ldPCR | *1814, 1815* | **2018** | *4* | *0,2* | 1 | 3 | **0,1** |
| *kni2* | **Ph_kni2_ORF23** | ORF isolation via ldPCR | *1814, 1815* | **2018** | *10* | *0,5* | 0 | 10 | **0,5** |
| *kni2* | **Ph_kni2_ORF31** | ORF isolation via ldPCR | *1814, 1815* | **2018** | *3* | *0,1* | 0 | 3 | **0,1** |
| *kni2* | **Ph_kni2_cJZ1209** | cDNA isolation via ldPCR | *1860, 1817* | **2578** | *3* | *0,1* | 3 | 0 | **0** |
| *kni2* | **Ph_kni2_cJZ1210** | cDNA isolation via ldPCR | *1860, 1817* | **2578** | *3* | *0,1* | 3 | 0 | **0** |
| *kni2* | **Ph_kni2_cJZ22** | cDNA isolation via ldPCR | *1813, 1816* | **2548** | *18* | *0,7* | 4 | 14 | **0,5** |

[PhD_A4_oligonucleotides]

**“PhD_A4.1_degenerate_oligonucleotides.pdf”**

# A4 Oligonucleotides

## A4.1 Degenerate Oligonucleotides

**Table A2:** **List of degenerate oligonucleotides.** Shown are: the name of the degenerate oligonucleotide (left), its nucleic acid sequence from 5’ to 3’ in degenerate nucleic acid code (centre) and the protein sequence it encodes (right).

| oligonucleotide | 5‘-3‘ sequence | corresponding protein sequence |
|---|---|---|
| fw_AFT | MGIMGIMGIACIGCITTYAC | RRRTAFT |
| fw_CKV | TGYAARGTNTGYGGNGARCCNGCNGC | CKVCGEPAA |
| fw_QAM | CARGCIATGTGGYTIGARGCICAYTA | QAMWLEAHY |
| fw_QRR | CAGMGGMGGGARMGIACIACITTYAC | QRRERTTFT |
| re_KIW | AACCADATYTTIACYTGIAC | EVQVKIW |
| re_LLQ | TGNARNARRCARTGDATYTTRAACCA | WFKIHCLLQ |
| re_PTQ | TTYTTRAACCARTTICCIACYTGIGTIGG | PTQVGNWFKN |
| re_QVW | GCCCKCCKRTTYTTRAACCAIACYTG | QVWFKNRRA |

[PhD_A4_oligonucleotides]

**“PhD_A4.2_specific_oligonucleotides.pdf”**

## A4.2 Specific oligonucleotides

**Table A3: List of specific oligonucleotides.** Shown are the identification number of the specific oligonucleotide (left) and its nucleic acid sequence from 5’ to 3’ (right).

| oligonucleotide | 5‘-3‘ sequence |
|---|---|
| 1000 | CGGCAAGACGCGGTATCCCGACATCTTC |
| 1005 | CTGGTAGGTTGATTTTGATCGCAACTTCCTCTCTC |
| 1013 | AGTTGTGCGAACGTAAGCTTGTCGAATACG |
| 1014 | GATGCACATGAACGGAGATCCCATGGCCTACC |
| 1015 | TTGAAGGGCGTATCTATAAGAATGCCGTTC |
| 1016 | TCGCAAGTAGAAAAGGTCAAACATATCGTAACAGTG |
| 1030 | ACTGGTGCGGCCGCCATACCCTTTCGACGAACGTGG |
| 1031 | ACTGGTCCATGGCATACCCTTTCGACGAACGTGG |
| 1032 | ACTGGTGGTACCCACGTGAAGGTCATTACCAGTTCAGCGATTG |
| 1033 | ACTACCGAGGATCCGTATACGAGGTCATTACCAGTTCAGCGATTG |
| 1100 | GCGCAACTGGACATCCTGGAGACCC |
| 1101 | GGAAGTTGCGATCAAGATCAATCTCCCTGAG |
| 1105 | GATGTCGGGATACCGCGTCTTGCCG |
| 1107 | CAGGGTCTCCAGGATGTCCAGTTGCGC |
| 1202 | GCGTTCTCCAGAACACATTACCCAGATGTC |
| 1203 | GGCAATGAAAATAGGTCTCACGGAAGCC |
| 1207 | GGCTTCCGTGAGACCTATTTTCATTGCC |
| 1208 | GACATCTGGGTAATGTGTTCTGGAGAACGC |
| 1213 | GGAATTTTCACTGTTCCCTTTACCACGTTTTACTG |
| 1214 | TATGGTGAATTTCAGCGAACTCGCTGTGCAGGG |
| 1215 | TTCATGTTTTGATCTCTCTTTTGGGTGAACATCGCG |
| 1216 | GGGGAAGTGTTTTCTTTGATGTTTAGTCTAACCAGC |
| 1300 | CCGGATGTGTTTACCAGGGAGGAGTTGGC |
| 1301 | GAAGATAGGCCTGACGGAGGCACGC |
| 1302 | CTTCCCTCAGCTTATCCTCGACCAGCGCTACC |
| 1304 | AGCGCAACTGTCACACCGATCCAAAATAGACCC |
| 1305 | GCGTGCCTCCGTCAGGCCTATCTTC |
| 1306 | CACATCCGGATAATGAGTCCTGGAAAACAC |
| 1314 | AATGGGAATTCTGGAAGAGTGCAGCGTC |

| oligonucleotide | 5‘-3‘ sequence |
|---|---|
| 1315 | CTCACGCTGTAATTTCGCCAGCTTTTCG |
| 1400 | ACGAGCAGCTGCAGGTGCTCCAGGC |
| 1401 | CCCTGACAGCCAGGACCTGGAGCGC |
| 1405 | CTTGCTCAGCCCCGTGAGCTGTGCG |
| 1406 | GCGCTCCAGGTCCTGGCTGTCAGGG |
| 1414 | GATGATGGCCAGCGCAACAGTTCTGTTGCC |
| 1415 | GTCATGCTTCTCCAGGCATGAGCGTCATAAGCG |
| 1500 | CGTTCCAACTCCACCAGCTGGAGCG |
| 1501 | CCCAATATCCTGACGTCTTCACTCGAGAGG |
| 1503 | ATTCCGGTTTAGCCACAACGGCACCTC |
| 1504 | CGTTTCCGCACACAATGTTGCCGTTATC |
| 1505 | CGCGTGCCTCGGATAGGTCAAGGTTGAGTG |
| 1506 | GCCAACTCCTCTCGAGTGAAGACGTCAG |
| 1507 | CTCGAACGCCCGCTCCAGCTGGTGG |
| 1512 | TTTACTGAGTCGAGGAGATACTTACCAAAGTG |
| 1514 | TTGACTGGTGACATAAACTCTAGACCG |
| 1515 | CGCCTGACACGTCACAACAATCG |
| 1516 | GTATGTCATGGGGTTTATCCCTTGATCACGAT |
| 1601 | GGAAATTCTGGAGCAGGAGTTCTGCAGC |
| 1602 | CGCGAGCACCTGGCAGAGAGGACTG |
| 1605 | CGCGAGTCGTAATGTCCGGGTATTGGTTGC |
| 1606 | GCCTCAGTGAGCCCAGTCCTCTCTGC |
| 1607 | GCTGCAGAACTCCTGCTCCAGAATTTCC |
| 1614 | AATGGAGTGCATGTATAGCTTTAGGCCTCAG |
| 1615 | ACTAATAAGAAGCTTCCGGGGGATTCCCGC |
| 1616 | CTCAAGCGAGGCCCACGGACATCGG |
| 1800 | GGAGCATTCACTTGCGAAGGATGCAAGAGT |
| 1801 | TGTCTAGTTGTGGGCATGTCTAAAACGGGG |
| 1805 | TTAGACCTGCGGCCGTACCGTGACCCCG |
| 1806 | TGAGACGACAGGCCTTGCACGAAGTCCTG |
| 1813 | TCACTGTTCCCTTTACCACGTTTTACTG |
| 1814 | GATGAACCAACTGTGCAAGGTCTGTGCTGAAC |
| 1815 | GTCAGGTGATGAGGCCATCACCAGCAGTTTC |
| 1816 | CAACAGTTAGACTTAACTACTCAATTCAGTCCTTAC |
| 1817 | GGAAATGAATTTATAAGAAATAGCTGCTAAGC |
| 1860 | CCACGTTTTACTGGACAAATTCGG |
| 1864 | CGTCAACACTCAGTTGAGGGACGC |
| 1865 | CTCCGTAGTGCATTGAGAGTTTCTTCC |
| 1900 | GGAGCTTTTACGTGTGAAGGCTGCAAGTCC |
| 1901 | CGGAAATGCCTTTTCGTTGGAATGTCCAAAAGT |

| oligonucleotide | 5‘-3‘ sequence |
|---|---|
| 1903 | CACTCCGGCTTTTCGTTACTACAAGCATCAC |
| 1904 | CAATTGCACCACCTCAACAATCGCCATCACC |
| 1905 | CTTGATCCACTTTTGGACATTCCAACGAAAAGG |
| 1906 | CTCCCGAAGAAGGACTTGCAGCCTTCACAC |
| 1913 | AGTTGTGCTCGTGCTGGCCG |
| 1914 | CATGACAGTCGCAAATTATCTCAGGATGAATCAG |
| 1915 | TGCATCACCGATTCCGAAACTATCAAGCATTGG |
| 1918 | GGCTAATAGTTTGCTATCGGGTTGTG |
| 2000 | CTTTGGGACCCGTCGACAAGTACCGCGTCAG |
| 2001 | GAGCAGAAGTCCCACTGTTTCAAGGAACGCACG |
| 2005 | GCCAGTTCCCGCTTCCTGCTGGGGTTTGGGTAC |
| 2006 | AGGGACCGCGTGCGTTCCTTGAAACAGTGGGAC |
| 2007 | TTCCTGACGCGGTACTTGTCGACGGGTCCC |
| 2008 | GACAGCCCCGCCAGGAGCGCCATG |
| 2013 | AGTGCTGCAAGCGCCACCGACAC |
| 2014 | GACCTTCTGCGCGCCCTGCAGCCCAAC |
| 2015 | CGAGAGCTAATAGGAGACAATAAAAATGTTGAG |
| 2016 | GGATCCGAATGGCAACGTTCGGC |
| 2100 | TGCCGTCGACAAGTACCGCATACGCAAG |
| 2101 | ACCATCTATTGCTTCAAAGAGAAGTCCAGGATG |
| 2105 | GTGTGAGGCCGGTCTGGCGGGAC |
| 2106 | CTTCTCATCAGGGGTCGGGTAGCGGTTC |
| 2114 | AGTTCATGATTGAACTTTGTTTTTAAATTGTATCAAG |
| 2115 | CTCCATGAAACTGTCCCAATCGATGGTACGTG |
| 2301 | CGCAGAGCACCGCGGCACCGGCAG |
| 2303 | CCAAACGAAGAAATATCTGACGCTGAGCGAG |
| 2305 | TCACTGAGGTGCAGTGTCTGAGCAATGTGC |
| 2306 | ACCGCTCGCTCAGCGTCAGATATTTCTTCG |
| 2314 | AGTGCGCATTCGATCATTGTTCATCCCAGG |
| 2315 | AGTGCACGATGATATTCTTTATCACCAATCTGTTC |
| 6114 | ACTGGTCATATGCGGCCGTGACTCTAGATCATAATCAG |
| 6115 | TACCGACGTACGGCGGCCGCTTAAGATACATTGATGAGTTTGGAC |
| 6950 | CTCAAGGCGCGCCTTCCGCGGTTACTGTAACCGCAGGGGCAAAAG |
| 6951 | GAATACCATGGTTGTGGATTGTTCTGTGAAC |
| 8100 | ACTGGTGCGGCCGCATGGCCTCCTCCGAGGACGTC |
| 8101 | ACTGGTCCATGGATGGCCTCCTCCGAGGACGTC |
| 8104 | TACCGAGGTACCCTACAGGAACAGGTGGTGGCG |
| 8105 | ACTACCGAGGATCCCTACAGGAACAGGTGGTGGCG |
| 8250 | GTTAATTAAGCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGTGAGTACCAGTATTAAA |
| 8251 | ATCTGAGCTCGACCAGGATGGGCACCACACCTGAGAGGACGAGAACAAAGGAAAC |

| oligonucleotide | 5‘-3‘ sequence |
|---|---|
| 8258 | AGATGAGCTCGACGGCGACGTAAACGGCC |
| 8259 | CATCCTGCAGGTTACTTGTACAGCTCG |
| 9600 | ACTACCGAGGATCCAAGCACAGATGCTTCGTTCAG |
| 9601 | ACTGGTGGTACCCGAACCCCAGAGTCCCGC |
| 9700 | CGCGCCGAATTCGTTTAAACACGTGCGTACGGGCCCACGCGTCCATGGCCGG |
| 9701 | CCATGGACGCGTGGGCCCGTACGCACGTGTTTAAACGAATTCGG |

[PhD_A4_oligonucleotides]

**“PhD_A4.3_siRNA_molecules.pdf”**

## A4.3 siRNA

**Table A4: List of siRNA molecules.** Shown are, from left to right: the *Parhyale* gene sequence from which a siRNA molecule was derived, its name, the 5’ to 3’ sequence, the corresponding position within the respective cDNA and the mRNA region in which the position lay. siRNA molecules were synthesized as STEALTH siRNA duplexes. Here, only the nucleic acid sequence in orientation of the mRNA/cDNA is shown.

| | | | | |
|---|---|---|---|---|
| ***Ph kni1*** | **si1** | CAGTCGCAAATTATCTCAGGATGAA | *703-727* | 5’ ORF |
| ***Ph kni1*** | **si2** | GGCCAGTGTGTGATCAACAAGCAAA | *858-882* | 5’ ORF encoding Zinc Finger domain |
| ***Ph kni1*** | **si8** | ACCTCAATCTTTACCTGCCCGTCAT | *1607-1631* | 3’ ORF |
| ***Ph six3*** | **si4** | CGTGCCCTTGTTGCTTTCCACTTAG | *339-363* | 5’ ORF encoding *Six* domain |
| ***Ph six3*** | **si6** | GACGCAGATACTGCCTTGAAGTGCA | *1014-1038* | 3’ ORF |
| ***Ph six3*** | **si9** | GGCACTTCTTGGAGCACCAAATCTT | *1137-1161* | 3’ ORF |

# [PhD_A5_plasmid_sequences]

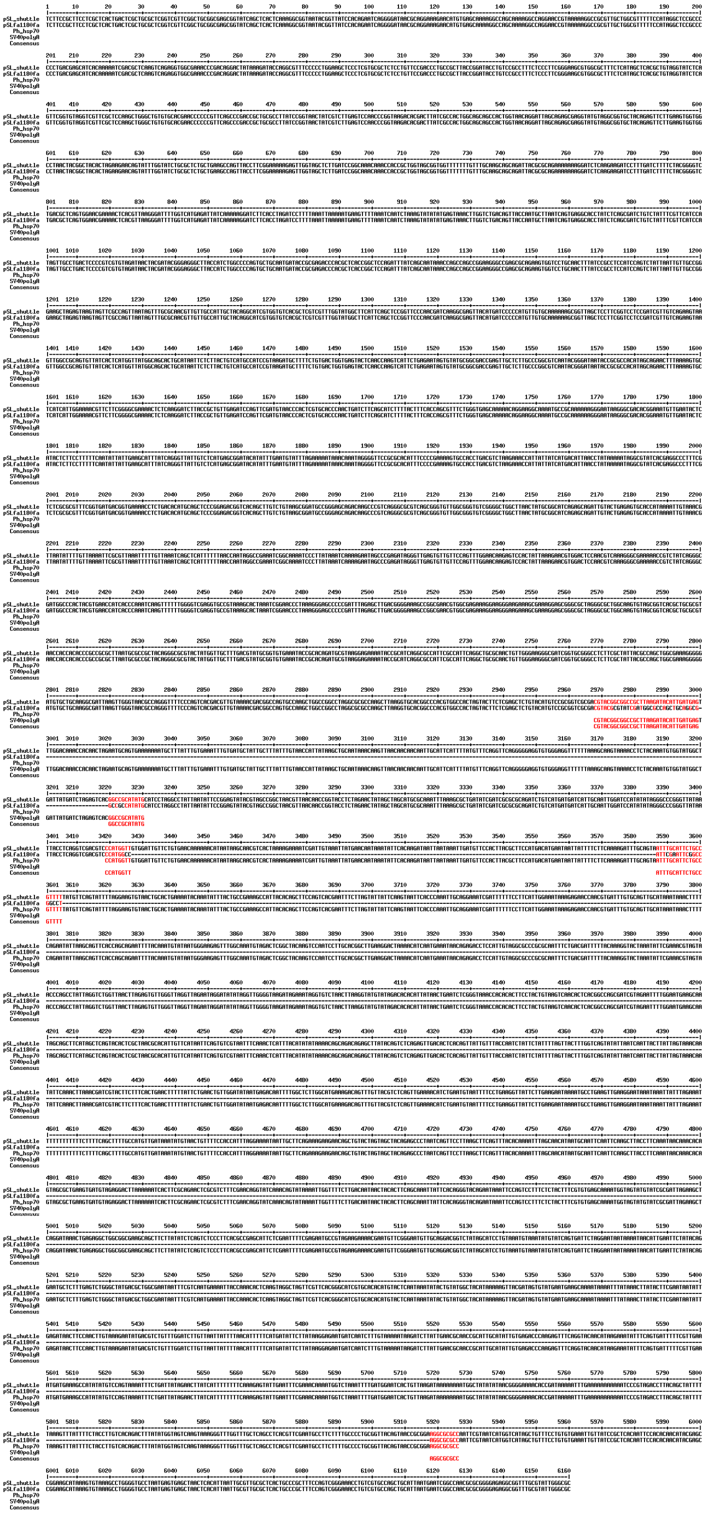
## “pSL_shuttle.txt”

```
>pSL_shuttle 'pSL_Phhsp70_SV40 clone# fc100+16
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGA
GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGC
TGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAA
GATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAA
TCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACT
ACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCA
ACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGT
ATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA
CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGC
CCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCC
ACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTC
TTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAC
GTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACA
GCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAAAATT
GTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTT
GTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGG
TGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGT
GTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTACTATGGTTGCTTTGACGTATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCA
TCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGT
TTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTGGCCGGCCTAGGCGCGCCAAGCTTAAGGTGCACGGCCCACGTGGCCACTAGTACTTCTCGAGCTCTGTACATGTCCGCGGTCGCGA
CGTACGGCGGCCGCTTAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATA
AACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATCTAGAGTCACGGCCGC
ATATGCATCCTAGGCCTATTAATATTCCGGAGTATACGTAGCCGGCTAACGTTAACAACCGGTACCTCTAGAACTATAGCTAGCATGCGCAAATTTAAAGCGCTGATATCGATCGCGCGCAGATCTGTC
ATGATGATCATTGCAATTGGATCCATATATAGGGCCCGGGTTATAATTACCTCAGGTCGACGTCCCATGGTTGTGGATTGTTCTGTGAACAAAAAACATAATAAGCAACGTCACTAAAAGAAAATCGAT
TGTAAATTATGAACAATAAATATTCACAAGATAATTAATAAATTGATGTTCCACTTACGCTTCCATGACATGAATAATTATTTTCTTCAAAAGATTTGCAGTAATTTGCATTCTGCCGTTTTTATGTTC
AGTATTTTAGGAAGTGTAACTGCACTGAAAATACAAATATTTACTGCCGAAAGCCATTACACAGCTTCCAGTCACGAATTTCTTAGTATTATTCAAGTAATTCACCCAAATTGCAGGAAATCGATTTTT
TCCTTCATTGGAAATAAAGAGAACCAACGTGATTTGTGCAGTTGCATAAATAAACTTTTCAGAATATTAAGCAGTTCACCAGCAGAATTTTACAAATGTATAATGGGAAGAGTTTGGCAAATGTAGACT
CGGCTACAAGTCCAATCCTTGCACGGCTTGAAGGACTAAAACATCAATGAAATAACAGAGACCTCCATTGTAGGCGCCCGCGCAATTTCTGACGATTTTTACAAAGGTACTAAATATTCGAAACGTAGT
AACCCAGCCTATTAGGTCTGGTTAACTTAGAGTGTTGGGTTAGGTTAGAATAGGATATATAGGTTGGGGTAAGATAGAAATAGGTGTCTAACTTAAGGTATGTATAGACACACATTATAACTGAATCTC
GGGTAAACCACACACTTCCTACTGTAAGTCAACACTCACGGCCAGCGATCGTAGAATTTTGGAATGAAGCAATAGCAGCTTCATAGCTCAGTACACTCGCTAACGCACATTGTTCATAATTCAGTGTCG
TAATTTCAAACTCATTTACATATATAAAACAGCAGACAGAGCTTATACAGTCTCAGAGTTGACACTCACAGTTATTGTTTACCAATCTATTCTATTTTAGTTACTTTGGTCAGTATATTAATCAATTAC
TTATTAGTAAACAATATTCAAACTTAAACGATCGTACTTCTTTCACTGAACTTTTTATTCTGAACTGTTGGATATAATGAGACAATTTTGGCTCTTGGCATGAAAGACAGTTTGTTACGTCTCAGTTGA
AAACATCTGAATGTAATTTTCCTGAAGGTTATTCTTGAAGAATAAAATGCCTGAAGTTGAAGGAATAAATAAATTATTTAGAAATTTTTTTTTTTTCTTTTCAGCTTTTGCCATGTTGATAAATATGTA
ACTGTTTTCCACCATTTAGGAAAATAATTGCTTCAGAAAGAAGAACAGCTGTACTAGTAGCTACAGAGCCCTAATCAGTTCCTTAAGCTTCAGTTTACACAAAATTTAGCAACATAATGCAATTCAATT
CAAGCTTACCTTCAAATAACAAACACAGTAGCGCTGAAGTGATGTAGAGGACTTAAAAAATCACTTCGCAGAACTCGCGTCTTTCGAACAGGTATCAAACAGTATAAAATTGGTTTTCTTGACAATAAC
TACACTTCAGCAAATTATTCACAGGGTACAGAATAAATTCCAGTCCTTTCTCTACTTTCGTGTGAGCAAAATGGTAGTATGTATCGCGATTAGAAGCTCAGGATAAACTGAGAGGCTGGCGGCGAAGCA
GCTTCTTATATCTCAGTCTCCCTTCACGCCGAGCATTCTCGAATTTTCGAGAATGCCGTAGAAGAAAACGAATGTTCGGGAATGTTGCAGGACGGTCTATAGCATCCTGTAAATGTAAATATGTATCAG
TGATTCTAGGAATAATAAAATAACATTGAATTCTATACAGGAATGCTCTTTGAGTCTGGGCTATGACGCTGGCGAATAATTTCGTCAATGAAAATTACCAAACACTCAAGTAGGCTAGTTCGTTCACGG
GCATCGTGCACACATGTACTCAATAAATATACTGTATGGCTACATAAAAAGTTACGATAGTGTATGAATGAAGCAAAATAAAATTTATAAACTTATACTTCGAATAATATTGAGATAACTTCCAACTTG
TAAAGAATATGACGTCTGTTTGGATCTTGTTAATTATTTTAACATTTTTCATGATATTCTTATAAGGAGAATGATCAATCTTTGTAAAAATAAGATCTTATTGAACGCAACCGCATTGCATATTGTGAG
ACCCAAGAGTTTCAGGTACAACATAAGAAATATTTCAGTGATTTTTCGTTGAAATGATGAAAGCCATATATGTCCAGTAAAATTTCTGATTATAGAACTTATCATTTTTTTTCAAAGAGTATTGAATTT
CGAAACAAAATGGTCTAAATTTTGATGGAATCACTGTTAAGATAAAAAAAATGGCTATATATAACGGGGAAAACACCGATAAAAATTTGAAAAAAAAAAATCCCGTAGACCTTACAGCTATTTTTAAAG
TTTATTTTCTACCTTGTCACAGACTTTATATGGTAGTCAAGTAAAGGGTTTGGTTTGCTCAGCCTCACGTTCGAATGCCTTCTTTTGCCCCTGCGGTTACAGTAACCGCGGAAGGCGCGCCAATTCGTA
ATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTT
GCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC

>Ph_hsp70 'amplified via PCR (6950/6951), using pSL_Hsp70_En2 as template (courtesy of Anastassios Pavlopoulos)
CCATGGTTGTGGATTGTTCTGTGAACAAAAAACATAATAAGCAACGTCACTAAAAGAAAATCGATTGTAAATTATGAACAATAAATATTCACAAGATAATTAATAAATTGATGTTCCACTTACGCTTCC
ATGACATGAATAATTATTTTCTTCAAAAGATTTGCAGTAATTTGCATTCTGCCGTTTTTATGTTCAGTATTTTAGGAAGTGTAACTGCACTGAAAATACAAATATTTACTGCCGAAAGCCATTACACAG
CTTCCAGTCACGAATTTCTTAGTATTATTCAAGTAATTCACCCAAATTGCAGGAAATCGATTTTTTCCTTCATTGGAAATAAAGAGAACCAACGTGATTTGTGCAGTTGCATAAATAAACTTTTCAGAA
TATTAAGCAGTTCACCAGCAGAATTTTACAAATGTATAATGGGAAGAGTTTGGCAAATGTAGACTCGGCTACAAGTCCAATCCTTGCACGGCTTGAAGGACTAAAACATCAATGAAATAACAGAGACCT
CCATTGTAGGCGCCCGCGCAATTTCTGACGATTTTTACAAAGGTACTAAATATTCGAAACGTAGTAACCCAGCCTATTAGGTCTGGTTAACTTAGAGTGTTGGGTTAGGTTAGAATAGGATATATAGGT
TGGGGTAAGATAGAAATAGGTGTCTAACTTAAGGTATGTATAGACACACATTATAACTGAATCTCGGGTAAACCACACACTTCCTACTGTAAGTCAACACTCACGGCCAGCGATCGTAGAATTTTGGAA
TGAAGCAATAGCAGCTTCATAGCTCAGTACACTCGCTAACGCACATTGTTCATAATTCAGTGTCGTAATTTCAAACTCATTTACATATATAAAACAGCAGACAGAGCTTATACAGTCTCAGAGTTGACA
CTCACAGTTATTGTTTACCAATCTATTCTATTTTAGTTACTTTGGTCAGTATATTAATCAATTACTTATTAGTAAACAATATTCAAACTTAAACGATCGTACTTCTTTCACTGAACTTTTTATTCTGAA
CTGTTGGATATAATGAGACAATTTTGGCTCTTGGCATGAAAGACAGTTTGTTACGTCTCAGTTGAAAACATCTGAATGTAATTTTCCTGAAGGTTATTCTTGAAGAATAAAATGCCTGAAGTTGAAGGA
ATAAATAAATTATTTAGAAATTTTTTTTTTTTCTTTTCAGCTTTTGCCATGTTGATAAATATGTAACTGTTTTCCACCATTTAGGAAAATAATTGCTTCAGAAAGAAGAACAGCTGTACTAGTAGCTAC
AGAGCCCTAATCAGTTCCTTAAGCTTCAGTTTACACAAAATTTAGCAACATAATGCAATTCAATTCAAGCTTACCTTCAAATAACAAACACAGTAGCGCTGAAGTGATGTAGAGGACTTAAAAAATCAC
TTCGCAGAACTCGCGTCTTTCGAACAGGTATCAAACAGTATAAAATTGGTTTTCTTGACAATAACTACACTTCAGCAAATTATTCACAGGGTACAGAATAAATTCCAGTCCTTTCTCTACTTTCGTGTG
AGCAAAATGGTAGTATGTATCGCGATTAGAAGCTCAGGATAAACTGAGAGGCTGGCGGCGAAGCAGCTTCTTATATCTCAGTCTCCCTTCACGCCGAGCATTCTCGAATTTTCGAGAATGCCGTAGAAG
AAAACGAATGTTCGGGAATGTTGCAGGACGGTCTATAGCATCCTGTAAATGTAAATATGTATCAGTGATTCTAGGAATAATAAAATAACATTGAATTCTATACAGGAATGCTCTTTGAGTCTGGGCTAT
GACGCTGGCGAATAATTTCGTCAATGAAAATTACCAAACACTCAAGTAGGCTAGTTCGTTCACGGGCATCGTGCACACATGTACTCAATAAATATACTGTATGGCTACATAAAAAGTTACGATAGTGTA
TGAATGAAGCAAAATAAAATTTATAAACTTATACTTCGAATAATATTGAGATAACTTCCAACTTGTAAAGAATATGACGTCTGTTTGGATCTTGTTAATTATTTTAACATTTTTCATGATATTCTTATA
AGGAGAATGATCAATCTTTGTAAAAATAAGATCTTATTGAACGCAACCGCATTGCATATTGTGAGACCCAAGAGTTTCAGGTACAACATAAGAAATATTTCAGTGATTTTTCGTTGAAATGATGAAAGC
CATATATGTCCAGTAAAATTTCTGATTATAGAACTTATCATTTTTTTTCAAAGAGTATTGAATTTCGAAACAAAATGGTCTAAATTTTGATGGAATCACTGTTAAGATAAAAAAAATGGCTATATATAA
CGGGGAAAACACCGATAAAAATTTGAAAAAAAAAAATCCCGTAGACCTTACAGCTATTTTTAAAGTTTATTTTCTACCTTGTCACAGACTTTATATGGTAGTCAAGTAAAGGGTTTGGTTTGCTCAGCC
TCACGTTCGAATGCCTTCTTTTGCCCCTGCGGTTACAGTAACCGCGGAAGGCGCGCC

>SV40polyA 'amplified via PCR (6114/6115), using pMi[3xP3_EGFP] as template (Horn C and Wimmer EA)
CGTACGGCGGCCGCTTAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATA
AACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATCTAGAGTCACGGCCGC
ATATG
```

```
>pSLfa1180fa 'Horn C and Wimmer EA, 1999
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGA
GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGC
TGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAA
GATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAA
TCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACT
ACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCA
ACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGT
ATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA
CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGC
CCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCC
ACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTC
TTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAC
GTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACA
GCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAAAATT
GTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTT
GTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGG
TGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGT
GTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTACTATGGTTGCTTTGACGTATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCA
TCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGT
TTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTGGCCGGCCTAGGCGCGCCAAGCTTAAGGTGCACGGCCCACGTGGCCACTAGTACTTCTCGAGCTCTGTACATGTCCGCGGTCGCGA
CGTACGCGTATCGATGGCGCCAGCTGCAGGCGGCCGCCATATGCATCCTAGGCCTATTAATATTCCGGAGTATACGTAGCCGGCTAACGTTAACAACCGGTACCTCTAGAACTATAGCTAGCATGCGCA
AATTTAAAGCGCTGATATCGATCGCGCGCAGATCTGTCATGATGATCATTGCAATTGGATCCATATATAGGGCCCGGGTTATAATTACCTCAGGTCGACGTCCCATGGCCATTCGAATTCGGCCGGCCT
AGGCGCGCCAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAA
CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC
```

[PhD_A5_plasmid_sequences]

## “alignment_pSLshuttle_pSLfa1180fa_Ph-hsp70_SV40.gif”



# [PhD_A5_plasmid_sequences]

## “pMi[3xP3_DsRed; hsp70_hp-otd1].txt”

```
>Ph_otd1_hairpin '700bp fragment used for hairpin generation (from Ph_otd1 cDNA positions 1072-1771; 3' fraction of ORF and
3'UTR; excluding homeobox and low complexity regions)
CATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTATGCAGACTACGGGACACTACACGCCGATGCCT
CCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTCAGATGGCCGTC
GGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCGAAAGCTTTCACAATCTTTGAACGGCATTCTTA
TAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGAGTTAGCGGATTGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTGTTGTGGAGACGCTCCGTTGTTTGTTACGATAA
GGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCCAAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTAAGGCGAGGATAGAGGAGCATACCTGCGGAAAGATGAGTA
CTTGTGTGAACTGCTGTGGCAATTCTGATAACAATCGCTGAACTGGTAATGACCT

>pMi[3xP3_DsRed; hsp70_hp-otd1] 'also termed pMi_3P3DsRed_04VS_dtoph_shhP for illustration of its structure
GGTACCAGTGCTTGAATGCTAAATGTTTTCAATTTTTCGCCATTAAGACAAGCCTACACAAATGCTTCTATAAATTATGCCAAGCACGTTAGCAGCTTCTACGAGCCCCAACCACTATTAATTCGAACA
GCATGTTTTTTTGCAGTGCGCAATGTTTAACACACTATATTATCAATACTACTAAAGATAACACATACCAATGCATTTCGTCTCAAAGAGAATTTTATTCTCTTCACGACGAAAAAAAAAGTTTTGCTC
TATTTCCAACAACAACAAAAATATGAGTAATTTATTCAAACGGTTTGCTTAAGAGATAAGAAAAAAGTGACCACTATTAATTCGAACGCGGCGTAAGCTTACCTTAATCTCAAGAAGAGCAAAACAAAA
GCAACTAATGTAACGGAATCATTATCTAGTTATGATCTGCAAATAATGCTGCAGAGCTGGGGGATCCCCGGCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCGAGCTCGCC
CGGGGATCTAATTCAATTAGAGACTAATTCAATTAGAGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCGACCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAA
TTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGGGTACCGCTAGAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCC
ACCATGGTGCGCTCCTCCAAGAACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCCAC
AACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAG
CTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGTTCATCGGCGTG
AACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCCCTGAAGCTGAAGGACGGC
GGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAG
TACGAGCGCACCGAGGGCCGCCACCACCTGTTCCTGTAGCGGCCGCTTAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTAT
TGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGG
TATGGCTGATTATGATCTAGAGTCACGGCCGCCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGTACACCGAACTGTA
TGCAGACTACGGGACACTACACGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCCTCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGA
GCCACATGACCAACTTTTCTCAGATGGCCGTCGGTCATCATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCCG
AAAGCTTTCACAATCTTTGAACGGCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGAGTTAGCGGATTGCAATAGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTG
TTGTGGAGACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGTTATAGCATCGCCATGACTTTGTGCTGCTCCAAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTAAGGCGAG
GATAGAGGAGCATACCTGCGGAAAGATGAGTACTTGTGTGAACTGCTGTGGCAATTCTGATAACAATCGCTGAACTGGTAATGACCTTCACTACGAGGTCATTACCAGTTCAGCGATTGTTATCAGAAT
TGCCACAGCAGTTCACACAAGTACTCATCTTTCCGCAGGTATGCTCCTCTATCCTCGCCTTAACCTTTTATTGTTTTACAGCGCAATACTGCTGCTGAACTTGGAGCAGCACAAAGTCATGGCGATGCT
ATAACTGCTTCTTGTTCAGCCTTATCGTAACAAACAACGGAGCGTCTCCACAACAGAAACTCTAAACCTTCAACACACCAAGCAATCCCTATTGCAATCCGCTAACTCTAAACAGCAAGTAATAAACAA
AATATTTGAAGGGCGTATCTATAAGAATGCCGTTCAAAGATTGTGAAAGCTTTCGGCTGCGGTACCTCCGTAATCCATACATTCCTGTTTGGTCGGAGCCGCTCTGTACTGCTGGGCTGCTGTCATTTG
GGAATTCATATTATGATGACCGACGGCCATCTGAGAAAAGTTGGTCATGTGGCTCGTGTTGAAATGATTTGGCTGGTGGAGGTACTCAATGTTACCGTAGTATGAGGGGGTGTAGCTCTGGTTCCATGA
AGGGGGGGTCTTGGCATGAGGAGGCATCGGCGTGTAGTGTCCCGTAGTCTGCATACAGTTCGGTGTACCCAGGTCCGACCGTAAAGCTGTCGATTGCCATAGGCTGGAAGAATAGGCGTTGGGGTCGAC
CACGTTCGTCGAAAGGGTATGCCATGGTTGTGGATTGTTCTGTGAACAAAAAACATAATAAGCAACGTCACTAAAAGAAAATCGATTGTAAATTATGAACAATAAATATTCACAAGATAATTAATAAAT
TGATGTTCCACTTACGCTTCCATGACATGAATAATTATTTTCTTCAAAAGATTTGCAGTAATTTGCATTCTGCCGTTTTTATGTTCAGTATTTTAGGAAGTGTAACTGCACTGAAAATACAAATATTTA
CTGCCGAAAGCCATTACACAGCTTCCAGTCACGAATTTCTTAGTATTATTCAAGTAATTCACCCAAATTGCAGGAAATCGATTTTTTCCTTCATTGGAAATAAAGAGAACCAACGTGATTTGTGCAGTT
GCATAAATAAACTTTTCAGAATATTAAGCAGTTCACCAGCAGAATTTTACAAATGTATAATGGGAAGAGTTTGGCAAATGTAGACTCGGCTACAAGTCCAATCCTTGCACGGCTTGAAGGACTAAAACA
TCAATGAAATAACAGAGACCTCCATTGTAGGCGCCCGCGCAATTTCTGACGATTTTTACAAAGGTACTAAATATTCGAAACGTAGTAACCCAGCCTATTAGGTCTGGTTAACTTAGAGTGTTGGGTTAG
GTTAGAATAGGATATATAGGTTGGGGTAAGATAGAAATAGGTGTCTAACTTAAGGTATGTATAGACACACATTATAACTGAATCTCGGGTAAACCACACACTTCCTACTGTAAGTCAACACTCACGGCC
AGCGATCGTAGAATTTTGGAATGAAGCAATAGCAGCTTCATAGCTCAGTACACTCGCTAACGCACATTGTTCATAATTCAGTGTCGTAATTTCAAACTCATTTACATATATAAAACAGCAGACAGAGCT
TATACAGTCTCAGAGTTGACACTCACAGTTATTGTTTACCAATCTATTCTATTTTAGTTACTTTGGTCAGTATATTAATCAATTACTTATTAGTAAACAATATTCAAACTTAAACGATCGTACTTCTTT
CACTGAACTTTTTATTCTGAACTGTTGGATATAATGAGACAATTTTGGCTCTTGGCATGAAAGACAGTTTGTTACGTCTCAGTTGAAAACATCTGAATGTAATTTTCCTGAAGGTTATTCTTGAAGAAT
AAAATGCCTGAAGTTGAAGGAATAAATAAATTATTTAGAAATTTTTTTTTTTCTTTTCAGCTTTTGCCATGTTGATAAATATGTAACTGTTTTCCACCATTTAGGAAAATAATTGCTTCAGAAAGAAG
AACAGCTGTACTAGTAGCTACAGAGCCCTAATCAGTTCCTTAAGCTTCAGTTTACACAAAATTTAGCAACATAATGCAATTCAATTCAAGCTTACCTTCAAATAACAAACACAGTAGCGCTGAAGTGAT
GTAGAGGACTTAAAAAATCACTTCGCAGAACTCGCGTCTTTCGAACAGGTATCAAACAGTATAAAATTGGTTTTCTTGACAATAACTACACTTCAGCAAATTATTCACAGGGTACAGAATAAATTCCAG
TCCTTTCTCTACTTTCGTGTGAGCAAAATGGTAGTATGTATCGCGATTAGAAGCTCAGGATAAACTGAGAGGCTGGCGGCGAAGCAGCTTCTTATATCTCAGTCTCCCTTCACGCCGAGCATTCTCGAA
TTTTCGAGAATGCCGTAGAAGAAAACGAATGTTCGGGAATGTTGCAGGACGGTCTATAGCATCCTGTAAATGTAAATATGTATCAGTGATTCTAGGAATAATAAAATAACATTGAATTCTATACAGGAA
TGCTCTTTGAGTCTGGGCTATGACGCTGGCGAATAATTTCGTCAATGAAAATTACCAAACACTCAAGTAGGCTAGTTCGTTCACGGGCATCGTGCACACATGTACTCAATAAATATACTGTATGGCTAC
ATAAAAAGTTACGATAGTGTATGAATGAAGCAAAATAAAATTTATAAACTTATACTTCGAATAATATTGAGATAACTTCCAACTTGTAAAGAATATGACGTCTGTTTGGATCTTGTTAATTATTTTAAC
ATTTTTCATGATATTCTTATAAGGAGAATGATCAATCTTTGTAAAAATAAGATCTTATTGAACGCAACCGCATTGCATATTGTGAGACCCAAGAGTTTCAGGTACAACATAAGAAATATTTCAGTGATT
TTTCGTTGAAATGATGAAAGCCATATATGTCCAGTAAAATTTCTGATTATAGAACTTATCATTTTTTTTCAAAGAGTATTGAATTTCGAAACAAAATGGTCTAAATTTTGATGGAATCACTGTTAAGAT
AAAAAAAATGGCTATATATAACGGGAAAACACCGATAAAAATTTGAAAAAAAAAAATCCCGTAGACCTTACAGCTATTTTTAAAGTTTATTTTCTACCTTGTCACAGACTTTATATGGTAGTCAAGTA
AAGGGTTTGGTTTGCTCAGCCTCACGTTCGAATGCCTTCTTTTGCCCCTGCGGTTACAGTAACCGCGGAAGGCGCGCCTAGGGGGGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGGACGTTACAC
AATTCTAATATTAATTAAATTATTGTTTTAAGTATGATAGTAAATCACATTACGCCGCGTTCGAATTAATAGTGGTCACTTTTTTCTTATCTCTTAAGCAAACCGTTTGAATAAATTACTCATATTTTT
GTTGTTGTTGGAAATAGAGCAAAACTTTTTTTTTCGTCGTGAAGAGAATAAAATTCTCTTTGAGACGAAATGCATTGGTATGTGTTATCTTTAGTAGTATTGATAATATAGTGTGTTAAACATTGCGCA
CTGCAAAAAAAAACATGCTGTTCGAATTAATAGTGGTTGGGGCTCGTAGAAAACGAAAAATATCTTAAGCTAGCATAGAGAATGGAGCAAAACTCAATTTGATGCGAGCTCCAATTCGCCCTATAGTGAG
TCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGC
ACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCG
CCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTT
GATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCG
GTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTC
GGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAAC
ATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGG
ATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTC
GCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCA
ACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTG
ACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGC
GCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGG
GGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTT
AATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTT
TTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATA
CTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGT
TGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAA
GCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCC
ACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTG
CGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAC
CGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTAC
ACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTG
```

[PhD_A5_plasmid_sequences]

## “pMi[3xP3_EGFP; hsp70_hp-DsRed].txt”

```
>pMi[3xP3_EGFP; hsp70_hp-DsRed]
GGTACCAGTGCTTGAATGCTAAATGTTTTCAATTTTTCGCCATTAAGACAAGCCTACACAAATGCTTCTATAAATTATGCCAAGCACGTTAGCAGCTTCTACGAGCCCCAACCACTATTAATTCGAACA
GCATGTTTTTTTGCAGTGCGCAATGTTTAACACACTATATTATCAATACTACTAAAGATAACACATACCAATGCATTTCGTCTCAAAGAGAATTTTATTCTCTTCACGACGAAAAAAAAAGTTTTGCTC
TATTTCCAACAACAACAAAAATATGAGTAATTTATTCAAACGGTTTGCTTAAGAGATAAGAAAAAAGTGACCACTATTAATTCGAACGCGGCGTAAGCTTACCTTAATCTCAAGAAGAGCAAAACAAAA
GCAACTAATGTAACGGAATCATTATCTAGTTATGATCTGCAAATAATGCTGCAGAGCTGGGGGATCCCCGGCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCGAGCTCGCC
CGGGGATCTAATTCAATTAGAGACTAATTCAATTAGAGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCGACCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAA
TTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGGGTACCGCTAGAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCC
ACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTC
AAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGC
ATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAAC
ATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTT
TACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCA
CAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAAGCTTATCGATACGCGTACGGCGCGCCTAGGCCGGCCGAATTGGCCGCGGAATTCGATTTTAC
TGTAACCGCAGGGGCAAAAGAAGGCATTCGAACGTGAGGCTGAGCAAACCAAACCCTTTACTTGACTACCATATAAAGTCTGTGACAAGGTAGAAAATAAACTTTAAAAATAGCTGTAAGGTCTACGGG
ATTTTTTTTTTTCAAATTTTTATCGGTGTTTTCCCCGTTATATATAGCCATTTTTTTTATCTTAACAGTGATTCCATCAAAATTTAGACCATTTTGTTTCGAAATTCAATACTCTTTGAAAAAAAATGA
TAAGTTCTATAATCAGAAATTTTACTGGACATATATGGCTTTCATCATTTCAACGAAAAATCACTGAAATATTTCTTATGTTGTACCTGAAACTCTTGGGTCTCACAATATGCAATGCGGTTGCGTTCA
ATAAGATCTTATTTTTACAAAGATTGATCATTCTCCTTATAAGAATATCATGAAAAATGTTAAAATAATTAACAAGATCCAAACAGACGTCATATTCTTTACAAGTTGGAAGTTATCTCAATATTATTC
GAAGTATAAGTTTATAAATTTTATTTTGCTTCATTCATACACTATCGTAACTTTTTATGTAGCCATACAGTATATTTATTGAGTACATGTGTGCACGATGCCCGTGAACGAACTAGCCTACTTGAGTGT
TTGGTAATTTTCATTGACGAAATTATTCGCCAGCGTCATAGCCCAGACTCAAAGAGCATTCCTGTATAGAATTCAATGTTATTTTATTATTCCTAGAATCACTGATACATATTTACATTTACAGGATGC
TATAGACCGTCCTGCAACATTCCCGAACATTCGTTTTCTTCTACGGCATTCTCGAAAATTCGAGAATGCTCGGCGTGAAGGGAGACTGAGATATAAGAAGCTGCTTCGCCGCCAGCCTCTCAGTTTATC
CTGAGCTTCTAATCGCGATACATACTACCATTTTGCTCACACGAAAGTAGAGAAAGGACTGGAATTTATTCTGTACCCTGTGAATAATTTGCTGAAGTGTAGTTATTGTCAAGAAAACCAATTTTATAC
TGTTTGATACCTGTTCGAAAGACGCGAGTTCTGCGAAGTGATTTTTTAAGTCCTCTACATCACTTCAGCGCTACTGTGTTTGTTATTTGAAGGTAAGCTTGAATTGAATTGCATTATGTTGCTAAATTT
TGTGTAAACTGAAGCTTAAGGAACTGATTAGGGCTCTGTAGCTACTAGTACAGCTGTTCTTCTTTCTGAAGCAATTATTTTCCTAAATGGTGGAAAACAGTTACATATTTATCAACATGGCAAAAGCTG
AAAAGAAAAAAAAAAAATTTCTAAATAATTTATTTATTCCTTCAACTTCAGGCATTTTATTCTTCAAGAATAACCTTCAGGAAAATTACATTCAGATGTTTTCAACTGAGACGTAACAAACTGTCTTTC
ATGCCAAGAGCCAAAATTGTCTCATTATATCCAACAGTTCAGAATAAAAAGTTCAGTGAAAGAAGTACGATCGTTTAAGTTTGAATATTGTTTACTAATAAGTAATTGATTAATATACTGACCAAAGTA
ACTAAAATAGAATAGATTGGTAAACAATAACTGTGAGTGTCAACTCTGAGACTGTATAAGCTCTGTCTGCTGTTTTATATATGTAAATGAGTTTGAAATTACGACACTGAATTATGAACAATGTGCGTT
AGCGAGTGTACTGAGCTATGAAGCTGCTATTGCTTCATTCCAAAATTCTACGATCGCTGGCCGTGAGTGTTGACTTACAGTAGGAAGTGTGTGGTTTACCCGAGATTCAGTTATAATGTGTGTCTATAC
ATACCTTAAGTTAGACACCTATTTCTATCTTACCCCAACCTATATATCCTATTCTAACCTAACCCAACACTCTAAGTTAACCAGACCTAATAGGCTGGGTTACTACGTTTCGAATATTTAGTACCTTTG
TAAAAATCGTCAGAAATTGCGCGGGCGCCTACAATGGAGGTCTCTGTTATTTCATTGATGTTTTAGTCCTTCAAGCCGTGCAAGGATTGGACTTGTAGCCGAGTCTACATTTGCCAAACTCTTCCCATT
ATACATTTGTAAAATTCTGCTGGTGAACTGCTTAATATTCTGAAAAGTTTATTTATGCAACTGCACAAATCACGTTGGTTCTCTTTATTTCCAATGAAGGAAAAAATCGATTTCCTGCAATTTGGGTGA
ATTACTTGAATAATACTAAGAAATTCGTGACTGGAAGCTGTGTAATGGCTTTCGGCAGTAAATATTTGTATTTTCAGTGCAGTTACACTTCCTAAAATACTGAACATAAAAACGGCAGAATGCAAATTA
CTGCAAATCTTTTGAAGAAAATAATTATTCATGTCATGGAAGCGTAAGTGGAACATCAATTTATTAATTATCTTGTGAATATTTATTGTTCATAATTTACAATCGATTTTCTTTTAGTGACGTTGCTTA
TTATGTTTTTTGTTCACAGAACAATCCACAACCATGGATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGG
CGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAA
GCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTC
CTTCATCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACTATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGA
GATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTC
CCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCGCCGAGGGCCGCCACCACCTGTTCCTGTAGCGTAGTGACTACAGGAACAGGTGGTGGCGGCCCTCGGCGCGCTCGTACTGCTCCACGAT
GGTGTAGTCCTCGTTGTGGGAGGTGATGTCCAGCTTGGAGTCCACGTAGTAGTAGCCGGGCAGCTGCACGGGCTTCTTGGCCATGTAGATGGACTTGAACTCCACCAGGTAGTGGCCGCCGTCCTTCAG
CTTCAGGGCCTTGTGGATCTCGCCCTTCAGCACGCCGTCGCGGGGGTACAGGCGCTCGGTGGAGGCCTCCCAGCCCATAGTCTTCTTCTGCATTACGGGGCCGTCGGAGGGGAAGTTCACGCCGATGAA
CTTCACCTTGTAGATGAAGGAGCCGTCCTGCAGGGAGGAGTCCTGGGTCACGGTCACCACGCCGCCGTCCTCGAAGTTCATCACGCGCTCCCACTTGAAGCCCTCGGGGAAGGACAGCTTCTTGTAGTC
GGGGATGTCGGCGGGGTGCTTCACGTACACCTTGGAGCCGTACTGGAACTGGGGGGACAGGATGTCCCAGGCGAAGGGCAGGGGGCCGCCCTTGGTCACCTTCAGCTTGGCGGTCTGGGTGCCCTCGTA
GGGGCGGCCCTCGCCCTCGCCCTCGATCTCGAACTCGTGGCCGTTCACGGAGCCCTCCATGCGCACCTTGAAGCGCATGAACTCCTTGATGACGTCCTCGGAGGAGGCCATGCGGCCGCGACTCTAGAT
CATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGG
TTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAAGCTTGGCGCGCCTAGGGGGGGATCCACTAGT
TCTAGAGCGGCCGCCACCGCGGGACGTTACACAATTCTAATATTAATTAAATTATTGTTTTAAGTATGATAGTAAATCACATTACGCCGCGTTCGAATTAATAGTGGTCACTTTTTTCTTATCTCTTAA
GCAAACCGTTTGAATAAATTACTCATATTTTTGTTGTTGTTGGAAATAGAGCAAAACTTTTTTTTTCGTCGTGAAGAGAATAAAATTCTCTTTGAGACGAAATGCATTGGTATGTGTTATCTTTAGTAG
TATTGATAATATAGTGTGTTAAACATTGCGCACTGCAAAAAAAAACATGCTGTTCGAATTAATAGTGGTTGGGGCTCGTAGAAAACGAAAAATATCTTAAGCTAGCATAGAGAATGGAGCAAAACTCAAT
TTGATGCGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTT
TCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTA
GTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT
TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAA
TATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATA
ATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATC
CCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGC
TGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGA
GCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT
GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAA
GCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATA
TATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGA
AAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGT
AACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGC
TGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGA
ACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGC
CTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTT
TTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAG
CGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAG
TTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAAT
TAACCCTCACTAAAGGGAACAAAAGCTG
```

## [PhD_A5_plasmid_sequences]

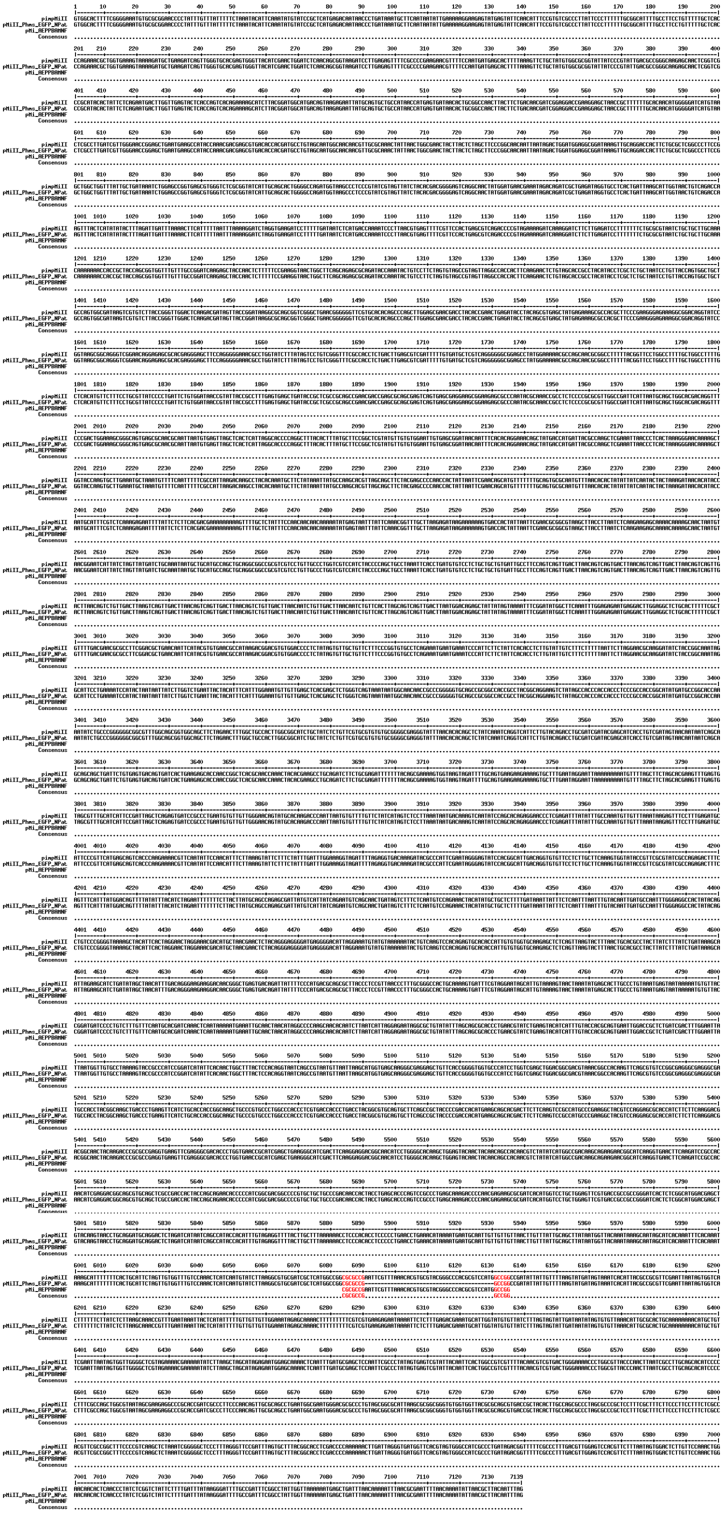
### **“pimpMiII.txt”**

```
>pimpMiII
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTA
TGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTT
ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAG
AGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATA
ACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAA
ACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAG
GACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTA
TCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAA
AACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTT
GAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGC
AGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGT
GTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTG
TCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACA
TGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCC
CAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCAC
CCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAAC
AAAAGCTGGTACCAAGTGCTTGAAATGCTAAATGTTTTCAATTTTTCGCCATTAAGACAAGCCTACACAAATGCTTCTATAAATTATGCCAAGCACGTTAGCAGCTTCTACGAGCCCCAACCACTATTA
ATTCGAACAGCATGTTTTTTTGCAGTGCGCAATGTTTAACACACTATATTATCAATACTACTAAAGATAACACATACCAATGCATTTCGTCTCAAAGAGAATTTTATTCTCTTCACGACGAAAAAAAAA
GTTTTGCTCTATTTCCAACAACAACAAAAATATGAGTAATTTATTCAAACGGTTTGCTTAAGAGATAAGAAAAAAGTGACCACTATTAATTCGAACGCGGCGTAAGCTTACCTTAATCTCAAGAAGAGC
AAAACAAAAGCAACTAATGTAACGGAATCATTATCTAGTTATGATCTGCAAATAATGCTGCATGCCAGCTGCAGGCGGCCGCGTCGTCCTGTTGCCCTGGTCGTCCATCTACCCCAGCTGCCTAAATTC
ACCTGATGTGTCCTCTGCTGCTGTGATTGCCTTCCAGTCAGTTGACTTAACAGTCAGTGACTTAACAGTCAGTTGACTTAACAGTCAGTTGACTTAACAGTCTGTTGACTTAAGTCAGTTGACTTAACA
GTCAGTTGACTTAACAGTCTGTTGACTTAACAATCTGTTGACTTAACAATCTGTTCACTTAGCAGTCAGTTGACTTAATGGACAGAGCTATTATAGTAAAATTTCGGATATGGCTTCAAATTTGGAGAG
AATGAGGACTTGGAGGCTCTGCACTTTTTCGCTGTTTTGACGAACGCGCCTTCGGACGCTGAACAATTCATACGTGTGAACGCCATAAGACGGACGTGTGGACCCCTCTATAGTGTTGCTGTTCTTTCC
CGGTGTGCCTCAGAAATGAATGAAATCCCATTCTTCTATTCACACCTCTTGTATTGTCTTTCTTTTTAATTCTTAGGAACGCAAGGATATCTACCGGCAAATAGGCATTCCTGAAAATCCATACTAATA
ATTATCTTGGTCTGAATTACTACATTTCATTTGGAAATGTTGTTGAGCTCACGAGCTCTGGGTCAGTAAATAATGGCAACAACCGCCCGGGGGTGCAGCCGCGGCCACCGCCTACGGCAGGAAGTCTAT
AGCCACCCACCACCCTCCCGCCACCGGCATATGATGCCGGCACCAAAATATCTGCCCGGGGGGCGGCGTTTGGCAGCGGTGGCAGCTTCTAGAACTTTGGCTGCCACTTGGCGGCATCTGCTATCTCTG
TTCGTGCGTGTGTGCGGGGCGAGGGTATTTAACACACAGCTCTATCAAATCAGGTCATTCTTGTACAGACCTGCGATCGATACGAGCATCACCTGTCGATAGTAACAATAATCAGCAGCAGCAGCTGAT
TCTGTGAGTGACAGTGATCACTGAAGAGCACCAACCGGCTCACGCAACCAAACTACACGAAGCCTGCAGATCTTCTGCGAGATTTTTTTACAGCGAAAAGTGGTAAGTAGATTTTGCAGTGAAGAAGAA
AAGTGCTTTGAATAGGAATTAAAAAAAAATGTTTTAGCTTCTAGCACGAAGTTTGAGTGTAGCGTTTGCATCATTCCGATTAGCTCAGAGTGATCCGCCCTGAATGTGTTGTTGGGAACAGTATGCACA
AGACCCAATTAATGTGTTTTGTTCTATCATAGTCTCCTTAAATAATGACAAAGTCAATATCCAGCACAGAGGAACCCTCGAGATTTATATTTGCCAAATGTTGTTTAAATAAGAGTTTCCTTTGAGATG
CATTCCCGTTCATGAGCAGTCACCCAAGAAAACGTTCAATATTCCAACATTTCTTAAAGTATTCTTTCTATTTGATTTGGAAAGGTAGATTTTAGAGGTGACAAAGATACGCCCATTCGAATAGGGAGT
ATCCACGGCATTGACAGGTGTGTTCCTCTTGCTTCAAAGTGGTATACCGTTCGCGTATCGCCAGAGACTTTCAGTTTCATTTATGGACAGTTTTATATTTACATCTAGAATTTTTTTCTTACTTATGCA
GCCAGAGCGATTATGTCATTATCAGAATGTCAGCAACTGATAGTCTTTCTCAATGTCCAGAAACTACATATGCTGCTCTTTTGATAAATTATTTCTCAATTTAATTTGTACAATTGATGCCAATTTGGG
AGGCCACTATACAGCTGTCCCGGGGTAAAAGCTACATTCACTAGGAACTAGGAAACGACATGCTAACGAACTCTACAGGGAGGGGATGAGGGGACATTAGGAAATGTATGTAAAAAATACTGTCAAGTC
CACAGAGTGCACACCATTGTGTGGTGCAAGAGCTCTCAGTTAAGTACTTTAACTGCACGCCTACTTATCTTTATCTGATAAAGCAATTAGAAGCATCTGATATAGCTAACATTTGACAGGGAAGAAGGA
CAACGGGCTGAGTGACAGATTATTTTCCCATGACGCAGCGCTTACCCTCCGTTAACCCTTTGCGGGCCACTGCAAAAGTGATTTCGTAGGAATAGCATTGTAAAAGTAACTAAATATGAGCACTTGCCC
TGTAAATGAGTAATAAAAATGTGTTACCGGATGATCCCCTGTCTTTGTTTCAATGCACGATCAAACTCAATAAAAATGAAATTGCAACTAACATAGGCCCCAAGCAACACAATCTTAATCATTAGGAGA
ATAGGCGCTGTATATTTAGCAGCGCACCCTGAACGTATCTGAAGTACATCATTTGTACCACGCAGTGAATTGGACCGCTCTGATCGACTTTGGAATTATTAATGGTTGTGCCTAAAAGTACCGCCCATC
CGGATCATATTCACAACTGGCTTTACTCCACAGGTAATCAGCCGTAATGTTAATTAAGCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAA
ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCT
ACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATA
TCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTAACCTGCAGGATGCAGGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTT
GTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGC
GTGCGATCGCTCATGGCCGGCGCGCCGAATTCGTTTAAACACGTGCGTACGGGCCCACGCGTCCATGGCCGGCCGATATTATTGTTTTAAGTATGATAGTAAATCACATTACGCCGCGTTCGAATTAAT
AGTGGTCACTTTTTTCTTATCTCTTAAGCAAACCGTTTGAATAAATTACTCATATTTTTGTTGTTGTTGGAAATAGAGCAAAACTTTTTTTTTCGTCGTGAAGAGAATAAAATTCTCTTTGAGACGAAA
TGCATTGGTATGTGTTATCTTTAGTAGTATTGATAATATAGTGTGTTAAACATTGCGCACTGCAAAAAAAACATGCTGTTCGAATTAATAGTGGTTGGGGCTCGTAGAAAACGAAAAATATCTTAAGCT
AGCATAGAGAATGGAGCAAAACTCAATTTGATGCGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTA
ATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAGCG
CGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGG
GGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCA
CGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAA
AATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAG

>pMi_AEPPBAMNF '9700
CGCGCCGAATTCGTTTAAACACGTGCGTACGGGCCCACGCGTCCATGGCCGG

>pMiII_Phms_EGFP_NPatel 'courtesy of Nipam Patel
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTA
TGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTT
ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAG
AGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATA
ACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAA
ACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAG
GACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTA
TCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAA
AACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTT
GAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGC
AGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGT
GTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTG
TCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACA
TGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCC
CAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCAC
CCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAAC
```

AAAAGCTGGTACCAAGTGCTTGAAATGCTAAATGTTTTCAATTTTTCGCCATTAAGACAAGCCTACACAAATGCTTCTATAAATTATGCCAAGCACGTTAGCAGCTTCTACGAGCCCCAACCACTATTA
ATTCGAACAGCATGTTTTTTTGCAGTGCGCAATGTTTAACACACTATATTATCAATACTACTAAAGATAACACATACCAATGCATTTCGTCTCAAAGAGAATTTTATTCTCTTCACGACGAAAAAAAAA
GTTTTGCTCTATTTCCAACAACAACAAAAATATGAGTAATTTATTCAAACGGTTTGCTTAAGAGATAAGAAAAAAGTGACCACTATTAATTCGAACGCGGCGTAAGCTTACCTTAATCTCAAGAAGAGC
AAAACAAAAGCAACTAATGTAACGGAATCATTATCTAGTTATGATCTGCAAATAATGCTGCATGCCAGCTGCAGGCGGCCGCGTCGTCCTGTTGCCCTGGTCGTCCATCTACCCCAGCTGCCTAAATTC
ACCTGATGTGTCCTCTGCTGCTGTGATTGCCTTCCAGTCAGTTGACTTAACAGTCAGTGACTTAACAGTCAGTTGACTTAACAGTCAGTTGACTTAACAGTCTGTTGACTTAAGTCAGTTGACTTAACA
GTCAGTTGACTTAACAGTCTGTTGACTTAACAATCTGTTGACTTAACAATCTGTTCACTTAGCAGTCAGTTGACTTAATGGACAGAGCTATTATAGTAAAATTTCGGATATGGCTTCAAATTTGGAGAG
AATGAGGACTTGGAGGCTCTGCACTTTTTCGCTGTTTTGACGAACGCGCCTTCGGACGCTGAACAATTCATACGTGTGAACGCCATAAGACGGACGTGTGGACCCCTCTATAGTGTTGCTGTTCTTTCC
CGGTGTGCCTCAGAAATGAATGAAATCCCATTCTTCTATTCACACCTCTTGTATTGTCTTTCTTTTTAATTCTTAGGAACGCAAGGATATCTACCGGCAAATAGGCATTCCTGAAAATCCATACTAATA
ATTATCTTGGTCTGAATTACTACATTTCATTTGGAAATGTTGTTGAGCTCACGAGCTCTGGGTCAGTAAATAATGGCAACAACCGCCCGGGGGTGCAGCCGCGGCCACCGCCTACGGCAGGAAGTCTAT
AGCCACCCACCACCCTCCCGCCACCGGCATATGATGCCGGCACCAAAATATCTGCCCGGGGGGCGGCGTTTGGCAGCGGTGGCAGCTTCTAGAACTTTGGCTGCCACTTGGCGGCATCTGCTATCTCTG
TTCGTGCGTGTGTGCGGGGCGAGGGTATTTAACACACAGCTCTATCAAATCAGGTCATTCTTGTACAGACCTGCGATCGATACGAGCATCACCTGTCGATAGTAACAATAATCAGCAGCAGCAGCTGAT
TCTGTGAGTGACAGTGATCACTGAAGAGCACCAACCGGCTCACGCAACCAAACTACACGAAGCCTGCAGATCTTCTGCGAGATTTTTTTACAGCGAAAAGTGGTAAGTAGATTTTGCAGTGAAGAAGAA
AAGTGCTTTGAATAGGAATTAAAAAAAAATGTTTTAGCTTCTAGCACGAAGTTTGAGTGTAGCGTTTGCATCATTCCGATTAGCTCAGAGTGATCCGCCCTGAATGTGTTGTTGGGAACAGTATGCACA
AGACCCAATTAATGTGTTTTGTTCTATCATAGTCTCCTTAAATAATGACAAAGTCAATATCCAGCACAGAGGAACCCTCGAGATTTATATTTGCCAAATGTTGTTTAAATAAGAGTTTCCTTTGAGATG
CATTCCCGTTCATGAGCAGTCACCCAAGAAAACGTTCAATATTCCAACATTTCTTAAAGTATTCTTTCTATTTGATTTGGAAAGGTAGATTTTAGAGGTGACAAAGATACGCCCATTCGAATAGGGAGT
ATCCACGGCATTGACAGGTGTGTTCCTCTTGCTTCAAAGTGGTATACCGTTCGCGTATCGCCAGAGACTTTCAGTTTCATTTATGGACAGTTTTATATTTACATCTAGAATTTTTTTCTTACTTATGCA
GCCAGAGCGATTATGTCATTATCAGAATGTCAGCAACTGATAGTCTTTCTCAATGTCCAGAAACTACATATGCTGCTCTTTTGATAAATTATTTCTCAATTTAATTTGTACAATTGATGCCAATTTGGG
AGGCCACTATACAGCTGTCCCGGGGTAAAGCTACATTCACTAGGAACTAGGAAACGACATGCTAACGAACTCTACAGGGAGGGGATGAGGGGACATTAGGAAATGTATGTAAAAAATACTGTCAAGTC
CACAGAGTGCACACCATTGTGTGGTGCAAGAGCTCTCAGTTAAGTACTTTAACTGCACGCCTACTTATCTTTATCTGATAAAGCAATTAGAAGCATCTGATATAGCTAACATTTGACAGGGAAGAAGGA
CAACGGGCTGAGTGACAGATTATTTTCCCATGACGCAGCGCTTACCCTCCGTTAACCCTTTGCGGGCCACTGCAAAAGTGATTTCGTAGGAATAGCATTGTAAAAGTAACTAAATATGAGCACTTGCCC
TGTAAATGAGTAATAAAAATGTGTTACCGGATGATCCCCTGTCTTTGTTTCAATGCACGATCAAACTCAATAAAAATGAAATTGCAACTAACATAGGCCCCAAGCAACACAATCTTAATCATTAGGAGA
ATAGGCGCTGTATATTTAGCAGCGCACCCTGAACGTATCTGAAGTACATCATTTGTACCACGCAGTGAATTGGACCGCTCTGATCGACTTTGGAATTATTAATGGTTGTGCCTAAAAGTACCGCCCATC
CGGATCATATTCACAACTGGCTTTACTCCACAGGTAATCAGCCGTAATGTTAATTAAGCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAA
ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCT
ACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATA
TCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTAACCTGCAGGATGCAGGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTT
GTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGC
GTGCGATCGCTCATGGCCGGCGCGCCGGCCGGCCGATATTATTGTTTTAAGTATGATAGTAAATCACATTACGCCGCGTTCGAATTAATAGTGGTCACTTTTTTCTTATCTCTTAAGCAAACCGTTTGA
ATAAATTACTCATATTTTTGTTGTTGTTGGAAATAGAGCAAAACTTTTTTTTTCGTCGTGAAGAGAATAAAATTCTCTTTGAGACGAAATGCATTGGTATGTGTTATCTTTAGTAGTATTGATAATATA
GTGTGTTAAACATTGCGCACTGCAAAAAAAACATGCTGTTCGAATTAATAGTGGTTGGGGCTCGTAGAAAACGAAAAATATCTTAAGCTAGCATAGAGAATGGAGCAAAACTCAATTTGATGCGAGCTC
CAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAA
TAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACT
TGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCT
CGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAAC
ACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAAT
TTAG

[PhD_A5_plasmid_sequences]

## “alignment_pimpMiII_pMiII_9700.gif”



[PhD_A5_plasmid_sequences]

## “pimpMiII[mse_EG-kni2intron-FP].txt”

```
>pimpMiII[mse_EG-kni2intron-FP]
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTA
TGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTT
ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAG
AGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATA
ACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAA
ACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAG
GACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTA
TCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAA
AACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTT
GAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGC
AGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGT
GTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTG
TCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACA
TGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCC
CAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCAC
CCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAAC
AAAAGCTGGTACCAAGTGCTTGAAATGCTAAATGTTTTCAATTTTTCGCCATTAAGACAAGCCTACACAAATGCTTCTATAAATTATGCCAAGCACGTTAGCAGCTTCTACGAGCCCCAACCACTATTA
ATTCGAACAGCATGTTTTTTTGCAGTGCGCAATGTTTAACACACTATATTATCAATACTACTAAAGATAACACATACCAATGCATTTCGTCTCAAAGAGAATTTTATTCTCTTCACGACGAAAAAAAAA
GTTTTGCTCTATTTCCAACAACAACAAAAATATGAGTAATTTATTCAAACGGTTTGCTTAAGAGATAAGAAAAAAGTGACCACTATTAATTCGAACGCGGCGTAAGCTTACCTTAATCTCAAGAAGAGC
AAAACAAAAGCAACTAATGTAACGGAATCATTATCTAGTTATGATCTGCAAATAATGCTGCATGCCAGCTGCAGGCGGCCGCGTCGTCCTGTTGCCCTGGTCGTCCATCTACCCCAGCTGCCTAAATTC
ACCTGATGTGTCCTCTGCTGCTGTGATTGCCTTCCAGTCAGTTGACTTAACAGTCAGTGACTTAACAGTCAGTTGACTTAACAGTCAGTTGACTTAACAGTCTGTTGACTTAAGTCAGTTGACTTAACA
GTCAGTTGACTTAACAGTCTGTTGACTTAACAATCTGTTGACTTAACAATCTGTTCACTTAGCAGTCAGTTGACTTAATGGACAGAGCTATTATAGTAAAATTTCGGATATGGCTTCAAATTTGGAGAG
AATGAGGACTTGGAGGCTCTGCACTTTTTCGCTGTTTTGACGAACGCGCCTTCGGACGCTGAACAATTCATACGTGTGAACGCCATAAGACGGACGTGTGGACCCCTCTATAGTGTTGCTGTTCTTTCC
CGGTGTGCCTCAGAAATGAATGAAATCCCATTCTTCTATTCACACCTCTTGTATTGTCTTTCTTTTTAATTCTTAGGAACGCAAGGATATCTACCGGCAAATAGGCATTCCTGAAAATCCATACTAATA
ATTATCTTGGTCTGAATTACTACATTTCATTTGGAAATGTTGTTGAGCTCACGAGCTCTGGGTCAGTAAATAATGGCAACAACCGCCCGGGGGTGCAGCCGCGGCCACCGCCTACGGCAGGAAGTCTAT
AGCCACCCACCACCCTCCCGCCACCGGCATATGATGCCGGCACCAAAATATCTGCCCGGGGGGCGGCGTTTGGCAGCGGTGGCAGCTTCTAGAACTTTGGCTGCCACTTGGCGGCATCTGCTATCTCTG
TTCGTGCGTGTGTGCGGGGCGAGGGTATTTAACACACAGCTCTATCAAATCAGGTCATTCTTGTACAGACCTGCGATCGATACGAGCATCACCTGTCGATAGTAACAATAATCAGCAGCAGCAGCTGAT
TCTGTGAGTGACAGTGATCACTGAAGAGCACCAACCGGCTCACGCAACCAAACTACACGAAGCCTGCAGATCTTCTGCGAGATTTTTTTACAGCGAAAAGTGGTAAGTAGATTTTGCAGTGAAGAAGAA
AAGTGCTTTGAATAGGAATTAAAAAAAAATGTTTTAGCTTCTAGCACGAAGTTTGAGTGTAGCGTTTGCATCATTCCGATTAGCTCAGAGTGATCCGCCCTGAATGTGTTGTTGGGAACAGTATGCACA
AGACCCAATTAATGTGTTTTGTTCTATCATAGTCTCCTTAAATAATGACAAAGTCAATATCCAGCACAGAGGAACCCTCGAGATTTATATTTGCCAAATGTTGTTTAAATAAGAGTTTCCTTTGAGATG
CATTCCCGTTCATGAGCAGTCACCCAAGAAAACGTTCAATATTCCAACATTTCTTAAAGTATTCTTTCTATTTGATTTGGAAAGGTAGATTTTAGAGGTGACAAAGATACGCCCATTCGAATAGGGAGT
ATCCACGGCATTGACAGGTGTGTTCCTCTTGCTTCAAAGTGGTATACCGTTCGCGTATCGCCAGAGACTTTCAGTTTCATTTATGGACAGTTTTATATTTACATCTAGAATTTTTTTCTTACTTATGCA
GCCAGAGCGATTATGTCATTATCAGAATGTCAGCAACTGATAGTCTTTCTCAATGTCCAGAAACTACATATGCTGCTCTTTTGATAAATTATTTCTCAATTTAATTTGTACAATTGATGCCAATTTGGG
AGGCCACTATACAGCTGTCCCGGGGTAAAAGCTACATTCACTAGGAACTAGGAAACGACATGCTAACGAACTCTACAGGGAGGGGATGAGGGGACATTAGGAAATGTATGTAAAAAATACTGTCAAGTC
CACAGAGTGCACACCATTGTGTGGTGCAAGAGCTCTCAGTTAAGTACTTTAACTGCACGCCTACTTATCTTTATCTGATAAAGCAATTAGAAGCATCTGATATAGCTAACATTTGACAGGGAAGAAGGA
CAACGGGCTGAGTGACAGATTATTTTCCCATGACGCAGCGCTTACCCTCCGTTAACCCTTTGCGGGCCACTGCAAAAGTGATTTCGTAGGAATAGCATTGTAAAAGTAACTAAATATGAGCACTTGCCC
TGTAAATGAGTAATAAAAATGTGTTACCGGATGATCCCCTGTCTTTGTTTCAATGCACGATCAAACTCAATAAAAATGAAATTGCAACTAACATAGGCCCCAAGCAACACAATCTTAATCATTAGGAGA
ATAGGCGCTGTATATTTAGCAGCGCACCCTGAACGTATCTGAAGTACATCATTTGTACCACGCAGTGAATTGGACCGCTCTGATCGACTTTGGAATTATTAATGGTTGTGCCTAAAAGTACCGCCCATC
CGGATCATATTCACAACTGGCTTTACTCCACAGGTAATCAGCCGTAATGTTAATTAAGCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGTGAGTACCAGTATTAAAATTTTTCACTTAAAGGGGAAA
CAATTCTTCCAAAGCACACAAAGTGGGAGCAACCACCATAGCTGTGGAACAAATACTAATGAAGTAGCAACATGAAACATAGCAGTTATGGAATTTATTGTCTCATCAAAGACTAAATATTAAATAAAC
TTAGTACTTAGTACTGAAATATGCCTGGCATTCACACACAATCATTAACGGATTCAACTTAGCACATTGTGGGATGTTAAATGAGATTGATCTGCTGTATACAGTACACGTACATATACAATGTAGAAT
GGCAGTTTTGCATAAAAAGCATAGCAAATAAGCATAAATTATTTTTAAACTCTCCTACTCTGTTATCGAAAATTAAAAAAAAAAAAATTTACTTCAGTTTTTGGTTCTTTTACATGAAGTCAATTATGT
ACCACACGATACCTTGACTGTCCACTGAAATTTACAACTTGGGCAGATGGCAAATAGACTCAACTTCGATAAAATCTTTTCCAGTCAACGGAACGACCTCAAACAAACACTAAACAGTTCTTGGTGATA
CAAGAAATGTGTTTTCCAGCAGGAGACAGAAACGCGTAGCTATATACCAACTGTGGACCGACGATGATTGAAGCTTGGGACAAGAAAATGAACTAAAATTGGCTAAAATACATTTTATAATTTACCGTT
GGCCAGTGTTCACGCACGCATGTTTCTTTCCAATAAATAAATTCCTTGAGTCTCCAAGAGCAATTCATTTTGCTTCAAGTCGGTACGTTGAAGGGAACAAATTTTATTGAGGATTTATCAATTGGCCAG
CAGCCTCATTTGCGAGTAGTGGACAGTTAATGTGTCGTGTGGTCCATGAATCAAATTAAATTGATTTAGTTGTCAGATAACACCAGCTTGTAAATTCCCATCTTTGATTATTACTCTTCACCTAAACGA
TGTTTCCTTTGTTCTCGTCCTCTCAGGTGTGGTGCCCATCCTGGTCGAGCTCGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT
GAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC
CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGA
TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAACCTGCAGGATGCAGGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTAC
TTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA
ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGCGTGCGATCGCTCATGGCCGGCGCGCCGAATTCGTTTAAACACGTGCGTACGGGCCCACGCGTCC
ATGGCCGGCCGATATTATTGTTTTAAGTATGATAGTAAATCACATTACGCCGCGTTCGAATTAATAGTGGTCACTTTTTTCTTATCTCTTAAGCAAACCGTTTGAATAAATTACTCATATTTTTGTTGT
TGTTGGAAATAGAGCAAAACTTTTTTTTTCGTCGTGAAGAGAATAAAATTCTCTTTGAGACGAAATGCATTGGTATGTGTTATCTTTAGTAGTATTGATAATATAGTGTGTTAAACATTGCGCACTGCA
AAAAAAACATGCTGTTCGAATTAATAGTGGTTGGGGCTCGTAGAAAACGAAAAATATCTTAAGCTAGCATAGAGAATGGAGCAAAACTCAATTTGATGCGAGCTCCAATTCGCCCTATAGTGAGTCGTA
TTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCG
CCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCC
TTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGG
TGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTC
TTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAG
```

[PhD_A6_phylogeny_database]

**“PhD_A6.1_List_of_species.pdf”**

# A6 Phylogeny database

## A6.1 List of species included in phylogenetic analyses

Table A5: **List of species included in phylogenetic analyses.** The taxonomic classification of the species was adopted from the NCBI taxonomy database (Sayers et al., 2009; Benson et al., 2010; see below). Shown are (from left to right): common name, scientific name, phylum, class and order (or, if no order is classified, family) to which the species belongs and taxonomy identification number.

| **common name** | *scientific name* | ***phylum*** | ***class*** | ***order (family)*** | **taxonomy ID** |
|---|---|---|---|---|---|
| **pea aphid** | ***Acyrthosiphon pisum*** | *Arthropoda* | *Insecta* | *Hemiptera* | 7029 |
| **mosquito** | ***Anopheles gambiae*** | *Arthropoda* | *Insecta* | *Diptera* | 7165 |
| **western honey bee** | ***Apis mellifera*** | *Arthropoda* | *Insecta* | *Hymenoptera* | 7460 |
| **nematode** | ***Caenorhabditis elegans*** | *Nematoda* | *Chromadorea* | *Rhabditida* | 6239 |
| **sea louse** | ***Caligus rogercresseyi*** | *Arthropoda* | *Maxillopoda* | *Siphonostomatoida* | 217165 |
| **white-tufted-ear marmoset** | ***Callithrix jacchus*** | *Chordata* | *Mammalia* | *Primates* | 9483 |
| **Mediterranean fruit fly** | ***Ceratitis capitata*** | *Arthropoda* | *Insecta* | *Diptera* | 7213 |
| **zebra fish** | ***Danio rerio*** | *Chordata* | *Actinopterygii* | *Cypriniformes* | 7955 |
| **fruit fly** | ***Drosophila grimshawi*** | *Arthropoda* | *Insecta* | *Diptera* | 7222 |
| **vinegar fly** | ***Drosophila melanogaster*** | *Arthropoda* | *Insecta* | *Diptera* | 7227 |
| **fruit fly** | ***Drosophila mojavensis*** | *Arthropoda* | *Insecta* | *Diptera* | 7230 |
| **fruit fly** | ***Drosophila pseudoobscura*** | *Arthropoda* | *Insecta* | *Diptera* | 7237 |
| **fruit fly** | ***Drosophila virilis*** | *Arthropoda* | *Insecta* | *Diptera* | 7244 |
| **velvet worm** | ***Euperipatoides kanangrensis*** | *Onychophora* | *n. c.* | *Peripatopsidae (family)* | 488523 |
| **Asian lady beetle** | ***Harmonia axyridis*** | *Arthropoda* | *Insecta* | *Coleoptera* | 115357 |
| **deer tick** | ***Ixodes scapularis*** | *Arthropoda* | *Arachnida* | *Ixodida* | 6945 |
| **house mouse** | ***Mus musculus*** | *Chordata* | *Mammalia* | *Rodentia* | 10090 |
| **parasitoid wasp** | ***Nasonia vitripennis*** | *Arthropoda* | *Insecta* | *Hymenoptera* | 7425 |
| **starlet sea anemone** | ***Nematostella vectensis*** | *Cnidaria* | *Anthozoa* | *Actiniaria* | 45351 |

| | | | | | |
|---|---|---|---|---|---|
| **milkweed bug** | ***Oncopeltus fasciatus*** | *Arthropoda* | *Insecta* | *Hemiptera* | 7536 |
| **beach hopper** | ***Parhyale hawaiensis*** | *Arthropoda* | *Malacostraca* | *Amphipoda* | 317513 |
| **human body louse** | ***Pediculus humanus corporis*** | *Arthropoda* | *Insecta* | *Phthiraptera* | 121224 |
| **Dumeril's clam worm** | ***Platynereis dumerilii*** | *Annelida* | *Polychaeta* | *Phyllodocida* | 6359 |
| **Acorn worm** | ***Saccoglossus kowalevskii*** | *Hemichordata* | *Enteropneusta* | *Harrimaniidae (family)* | 10224 |
| **centipede** | ***Strigamia maritima*** | *Arthropoda* | *Chilopoda* | *Geophilomorpha* | 126957 |
| **red flour beetle** | ***Tribolium castaneum*** | *Arthropoda* | *Insecta* | *Coleoptera* | 7070 |
| **parasitic nematode** | ***Trichinella spiralis*** | *Nematoda* | *Enoplea* | *Trichocephalida* | 6334 |
| **African clawed frog** | ***Xenopus laevis*** | *Chordata* | *Amphibia* | *Anura* | 8355 |

## In acknowledgment of:

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W. (2010) 'GenBank', *Nucleic acids research* 38(Database issue): D46-51.

Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S. et al. (2009) 'Database resources of the National Center for Biotechnology Information', *Nucleic acids research* 37(Database issue): D5-15.

[PhD_A6_phylogeny_database]

**"PhD_A6.2_List_of_non-Parhyale_sequences.pdf"**

## A6.2 List of non-*Parhyale* protein sequences included in phylogenetic analyses

**<u>Table A6:</u> List of non-*Parhyale* protein sequences included in phylogenetic analyses.** Shown are (from left to right): name tag of protein sequences used in the phylogenetic calculations in this work, (homologous) gene the respective protein sequence derives from, species the protein sequence was isolated from and NCBI reference sequence number, reflecting accession number and version of the sequence.

| name used in phylogeny | *gene/homolog* | *species* | reference |
|---|---|---|---|
| Agam_AWH | *arrowhead* | *Anopheles gambiae* | XM_316578.4 |
| Agam_OPTIX | *optix* | *Anopheles gambiae* | XM_559661.2 |
| Agam_SIX4 | *six4* | *Anopheles gambiae* | XM_309580.4 |
| Agam_SO | *sine oculis* | *Anopheles gambiae* | XM_320814.4 |
| Agam_UNPG | *unplugged* | *Anopheles gambiae* | XM_308835.4 |
| Amel_AWH | *arrowhead* | *Apis mellifera* | XM_001121365.1 |
| Amel_EG | *eagle* | *Apis mellifera* | XM_001120662.1 |
| Amel_KNI | *knirps* | *Apis mellifera* | XM_001120531.1 |
| Amel_KNRL | *knirps related* | *Apis mellifera* | XM_395932.2 |
| Amel_SIX2 | *six2* | *Apis mellifera* | XM_396811.3 |
| Amel_SIX3 | *six3* | *Apis mellifera* | XM_623761.2 |
| Amel_SIX4 | *six4* | *Apis mellifera* | XM_001120698.1 |
| Cele_H-32 | *ceh-32* | *Caenorhabditis elegans* | NM_073557.3 |
| Cele_H-33 | *ceh-33* | *Caenorhabditis elegans* | NM_072019.1 |
| Cele_H-34 | *ceh-34* | *Caenorhabditis elegans* | NM_072018.5 |
| Cele_LIM-4 | *lim-4* | *Caenorhabditis elegans* | NM_076268.1 |
| Crog_KNI | *knirps* | *Caligus rogercresseyi* | ACO11525.1 |
| Dmel_AL | *aristaless* | *Drosophila melanogaster* | NM_164382.1 |
| Dmel_AP | *apterous* | *Drosophila melanogaster* | NM_165445.3 |
| Dmel_AWH | *arrowhead* | *Drosophila melanogaster* | NM_168042.1 |
| Dmel_B-H1 | *BarH1* | *Drosophila melanogaster* | NM_078663.2 |
| Dmel_CG11085 | *CG11085* | *Drosophila melanogaster* | NM_132587.3 |
| Dmel_CG32105 | *CG32105* | *Drosophila melanogaster* | NM_168497.1 |
| Dmel_CG34340 | *CG34340* | *Drosophila melanogaster* | NM_001103605.1 |
| Dmel_CG34367 | *CG34367* | *Drosophila melanogaster* | NM_001103670.2 |
| Dmel_CG4328 | *CG4328* | *Drosophila melanogaster* | NM_140310.2 |
| Dmel_CG9876 | *CG9876* | *Drosophila melanogaster* | NM_137912.1 |
| Dmel_DR | *drop* | *Drosophila melanogaster* | NM_057976.2 |
| Dmel_EG | *eagle/egon* | *Drosophila melanogaster* | NM_079482.3 |
| Dmel_ERR | *estrogen-related receptor* | *Drosophila melanogaster* | NM_139926.2 |
| Dmel_EXEX | *extra-extra* | *Drosophila melanogaster* | NM_139907.2 |

| | | | |
|---|---|---|---|
| Dmel_EY | *eyeless* | *Drosophila melanogaster* | NM_166789.1 |
| Dmel_GSB | *gooseberry* | *Drosophila melanogaster* | NM_079139.2 |
| Dmel_GSB-N | *gooseberry-neuro* | *Drosophila melanogaster* | NM_079138.1 |
| Dmel_GSC | *goosecoid* | *Drosophila melanogaster* | NM_001144290.1 |
| Dmel_HBN | *homeobrain* | *Drosophila melanogaster* | NM_176240.1 |
| Dmel_KNI | *knirps* | *Drosophila melanogaster* | NM_079463.2 |
| Dmel_KNRL | *knirps related* | *Drosophila melanogaster* | NM_176374.1 |
| Dmel_LIM1 | *Lim1* | *Drosophila melanogaster* | NM_132277.2 |
| Dmel_LIM3 | *Lim3* | *Drosophila melanogaster* | NM_165277.1 |
| Dmel_MU | *munster* | *Drosophila melanogaster* | NM_057982.3 |
| Dmel_OPTIX | *optix* | *Drosophila melanogaster* | NM_079956.2 |
| Dmel_OTP | *orthopedia* | *Drosophila melanogaster* | NM_001103918.2 |
| Dmel_PHDP | *Putative homeodomain protein* | *Drosophila melanogaster* | NM_079110.2 |
| Dmel_PRDA | *paired, isoform A* | *Drosophila melanogaster* | NM_078832.2 |
| Dmel_PRDB | *paired, isoform B* | *Drosophila melanogaster* | NM_164990.2 |
| Dmel_REPO | *reverse polarity* | *Drosophila melanogaster* | NM_057678.3 |
| Dmel_RX | *retinal homeobox* | *Drosophila melanogaster* | NM_166413.2 |
| Dmel_SIX4 | *six4* | *Drosophila melanogaster* | NM_140999.2 |
| Dmel_SLOU | *slouch* | *Drosophila melanogaster* | NM_057309.2 |
| Dmel_SO | *sine oculis* | *Drosophila melanogaster* | NM_057385.4 |
| Dmel_SV | *shaven* | *Drosophila melanogaster* | NM_079894.3 |
| Dmel_TOY | *twin of eyeless* | *Drosophila melanogaster* | NM_079899.3 |
| Dmel_TUP | *tailup/islet* | *Drosophila melanogaster* | NM_057427.2 |
| Dmel_UNPG | *unplugged* | *Drosophila melanogaster* | NM_057798.2 |
| Dpse_ SIX4 | *six4* | *Drosophila pseudoobscura* | XM_001352978.2 |
| Dpse_AWH | *arrowhead* | *Drosophila pseudoobscura* | XM_001353513.2 |
| Dpse_EG | *eagle* | *Drosophila pseudoobscura* | XM_001353471.2 |
| Dpse_KNI | *knirps* | *Drosophila pseudoobscura* | XM_001352966.2 |
| Dpse_KNRL | *knirps related* | *Drosophila pseudoobscura* | XM_001352963.2 |
| Dpse_OPTIX | *optix* | *Drosophila pseudoobscura* | XM_001360625.2 |
| Dpse_SO | *sine oculis* | *Drosophila pseudoobscura* | XM_001360175.2 |
| Dpse_UNPG | *unplugged* | *Drosophila pseudoobscura* | XM_001361091.1 |
| Drer_Gbx1 | *gastrulation brain homeobox 1* | *Danio rerio* | NM_174861.2 |
| Drer_Gbx2 | *gastrulation brain homeobox 2* | *Danio rerio* | NM_152964.1 |
| Drer_Lhx6 | *lhx6* | *Danio rerio* | NM_001004015.1 |
| Drer_Lhx8a | *lhx8a* | *Danio rerio* | NM_001003980.1 |
| Drer_Six1a | *sine oculis homeobox homolog 1a* | *Danio rerio* | NM_001009904.1 |
| Drer_Six1b | *sine oculis homeobox homolog 1b* | *Danio rerio* | NM_207095.1 |
| Drer_Six2.1 | *sine oculis homeobox homolog 2.1* | *Danio rerio* | NM_131783.2 |
| Drer_Six3a | *sine oculis homeobox homolog 3a* | *Danio rerio* | NM_131362.1 |
| Drer_Six3b | *sine oculis homeobox homolog 3b* | *Danio rerio* | NM_131363.1 |
| Drer_Six4.1 | *sine oculis homeobox homolog 4.1* | *Danio rerio* | NM_131717.1 |
| Drer_Six4.2 | *sine oculis homeobox homolog 4.2* | *Danio rerio* | NM_131718.1 |
| Drer_Six4.3 | *sine oculis homeobox homolog 4.3* | *Danio rerio* | NM_131720.1 |
| Drer_Six6a | *sine oculis-related homeobox 6a* | *Danio rerio* | NM_201105.1 |
| Drer_Six6b | *sine oculis-related homeobox 6b* | *Danio rerio* | NM_001020585.1 |
| Drer_Six7 | *sine oculis homeobox homolog 7* | *Danio rerio* | NM_131354.2 |
| Drer_Six9 | *sine oculis homeobox homolog 9* | *Danio rerio* | NM_001130608.1 |
| Drer_Zgc_195096 | *zgc:195096* | *Danio rerio* | NM_001128734.1 |

| | | | |
|---|---|---|---|
| Dvir_AWH | *arrowhead* | *Drosophila virilis* | XM_002046523.1 |
| Dvir_EG | *eagle* | *Drosophila virilis* | XM_002048604.1 |
| Dvir_KNI | *knirps* | *Drosophila virilis* | XM_002048059.1 |
| Dvir_KNRL | *knirps related* | *Drosophila virilis* | XM_002048061.1 |
| Dvir_OPTIX | *optix* | *Drosophila virilis* | XM_002050350.1 |
| Dvir_SIX4 | *six4* | *Drosophila virilis* | XM_002048047.1 |
| Dvir_SO | *sine oculis* | *Drosophila virilis* | XM_002049752.1 |
| Dvir_UNPG | *unplugged* | *Drosophila virilis* | XM_002059265.1 |
| Isca_Xvent-2B | *xvent-2B* | *Ixodes scapularis* | XM_002404613.1 |
| Mmus_Alx1 | *ALX homeobox 1* | *Mus musculus* | NM_172553.4 |
| Mmus_Alx3 | *aristaless-like homeobox 3* | *Mus musculus* | NM_007441.2 |
| Mmus_Alx4 | *aristaless-like homeobox 4* | *Mus musculus* | NM_007442.2 |
| Mmus_Arx | *aristaless related homeobox* | *Mus musculus* | NM_007492.3 |
| Mmus_Esr1 | *estrogen receptor 1 (alpha)* | *Mus musculus* | NM_007956.4 |
| Mmus_Esr2 | *estrogen receptor 2 (beta)* | *Mus musculus* | NM_207707.1 |
| Mmus_Gbx1 | *gastrulation brain homeobox 1* | *Mus musculus* | NM_015739.2 |
| Mmus_Gbx2 | *gastrulation brain homeobox 2* | *Mus musculus* | NM_010262.3 |
| Mmus_Gsc | *goosecoid homeobox* | *Mus musculus* | NM_010351.1 |
| Mmus_Gsc2 | *goosecoid homeobox 2* | *Mus musculus* | NM_029469.1 |
| Mmus_Isl1 | *ISL1 transcription factor, LIM/homeodomain* | *Mus musculus* | NM_021459.4 |
| Mmus_Isl2 | *insulin related protein 2* | *Mus musculus* | NM_027397.3 |
| Mmus_Lhx1 | *LIM homeobox protein 1* | *Mus musculus* | NM_008498.2 |
| Mmus_Lhx2 | *LIM homeobox protein 2* | *Mus musculus* | NM_010710.3 |
| Mmus_Lhx3 | *LIM homeobox protein 3* | *Mus musculus* | NM_001039653.1 |
| Mmus_Lhx4 | *LIM homeobox protein 4* | *Mus musculus* | NM_010712.2 |
| Mmus_Lhx5 | *LIM homeobox protein 5* | *Mus musculus* | NM_008499.5 |
| Mmus_Lhx6 | *LIM homeobox protein 6* | *Mus musculus* | NM_001083125.1 |
| Mmus_Lhx8 | *LIM homeobox protein 8* | *Mus musculus* | NM_010713.2 |
| Mmus_Lhx9 | *LIM homeobox protein 9* | *Mus musculus* | NM_001042577.1 |
| Mmus_Lmx1a | *LIM homeobox transcription factor 1 alpha* | *Mus musculus* | NM_033652.5 |
| Mmus_Lmx1b | *LIM homeobox transcription factor 1 beta* | *Mus musculus* | NM_010725.2 |
| Mmus_Meox1 | *mesenchyme homeobox 1* | *Mus musculus* | NM_010791.3 |
| Mmus_Meox2 | *mesenchyme homeobox 2* | *Mus musculus* | NM_008584.3 |
| Mmus_Mnx1 | *motor neuron and pancreas homeobox 1* | *Mus musculus* | NM_019944.2 |
| Mmus_Nkx1-2 | *NK1 transcription factor related, locus 2* | *Mus musculus* | NM_009123.2 |
| Mmus_Otp | *orthopedia* | *Mus musculus* | NM_011021.3 |
| Mmus_Pax2 | *paired box gene 2* | *Mus musculus* | NM_011037.3 |
| Mmus_Pax3 | *paired box gene 3* | *Mus musculus* | NM_008781.4 |
| Mmus_Pax6 | *paired box gene 6* | *Mus musculus* | NM_013627.4 |
| Mmus_Pax7 | *paired box gene 7* | *Mus musculus* | NM_011039.2 |
| Mmus_Phox2a | *paired-like homeobox 2a* | *Mus musculus* | NM_008887.2 |
| Mmus_Phox2b | *paired-like homeobox 2b* | *Mus musculus* | NM_008888.3 |
| Mmus_Prrxl1 | *paired related homeobox protein-like 1* | *Mus musculus* | NM_001001796.4 |
| Mmus_Rax | *retina and anterior neural fold homeobox* | *Mus musculus* | NM_013833.2 |
| Mmus_Six1 | *sine oculis-related homeobox 1 homolog* | *Mus musculus* | NM_009189.3 |
| Mmus_Six2 | *sine oculis-related homeobox 2 homolog* | *Mus musculus* | NM_011380.2 |
| Mmus_Six3 | *sine oculis-related homeobox 3 homolog* | *Mus musculus* | NM_011381.4 |
| Mmus_Six4 | *sine oculis-related homeobox 4 homolog* | *Mus musculus* | NM_011382.2 |
| Mmus_Six5 | *sine oculis-related homeobox 5 homolog* | *Mus musculus* | NM_011383.1 |

| | | | |
|---|---|---|---|
| Mmus_Six6 | *sine oculis-related homeobox 6 homolog* | *Mus musculus* | NM_011384.4 |
| Mmus_Uncx | *UNC homeobox* | *Mus musculus* | NM_013702.3 |
| Nvec_GBX | *gbx* | *Nematostella vectensis* | ABF61779.1 |
| Nvec_SIX3 | *six3* | *Nematostella vectensis* | XM_001625109.1 |
| Nvec_SIX4.1 | *six4.1* | *Nematostella vectensis* | XM_001623084.1 |
| Nvec_SIX4.2 | *six4.2* | *Nematostella vectensis* | XM_001634946.1 |
| Nvec_SO | *sine oculis* | *Nematostella vectensis* | XM_001626384.1 |
| Nvit_KNI | *knirps* | *Nasonia vitripennis* | XM_001604919.1 |
| Nvit_UNPG | *unplugged* | *Nasonia vitripennis* | XP_001602252.1 |
| Ofas_KNI1 | *knirps 1* | *Oncopeltus fasciatus* | Jonas Schwirz |
| Ofas_KNI2 | *knirps 2* | *Oncopeltus fasciatus* | Jonas Schwirz |
| Pdum_GBX | *gbx* | *Platynereis dumerilii* | AJ505024.1 |
| Phum_GBX1 | *Homeobox protein GBX-1, putative, mRNA* | *Pediculus humanus* | XM_002423021.1 |
| Tcas_ KNI | *knirps* | *Tribolium castaneum* | NM_001128495.1 |
| Tcas_AWH | *arrowhead* | *Tribolium castaneum* | XM_966109.1 |
| Tcas_EG | *eagle* | *Tribolium castaneum* | NM_001114367.1 |
| Tcas_OPTIX | *optix/six3* | *Tribolium castaneum* | NM_001113467.1 |
| Tcas_SIX4 | *six4* | *Tribolium castaneum* | EFA01346.1 |
| Tcas_SO | *sine oculis* | *Tribolium castaneum* | XM_967074.2 |
| Tcas_UNPG | *unplugged* | *Tribolium castaneum* | XM_969836.1 |

## [PhD_A6_phylogeny_database]

### [PhD_A6.3_phylogeny_ClustalW2_alignments]

### “alignment_aristaless-group.txt”

```
Alignment of the HD protein sequences of various aristaless-group genes from different species.
This was done using the ClustalW2 algorithm (Larkin et al., 2007).
For details and references regarding the species' protein sequences, see Appendix (A7).


Phaw_HBN      -GTDLEKPRKQVRRSRTTFTTFQLHQLERAFEKTQYPDVFTREELALNLDLSEARVQVWFQNRRAKWRKREKAMGRESPS-
Dmel_HBN      DLSDMERPRK-VRRSRTTFTTFQLHQLERAFEKTQYPDVFTREDLAMRLDLSEARVQVWFQNRRAKWRKREKFMNQDKAG-
Dmel_RX       -GQDDNCAKKKHRRNRTTFTTYQLHELERAFEKSHYPDVYSREELAMKVNLPEVRVQVWFQNRRAKWRRQEKSESLRLGL-
Mmus_Rax      -LSEEEPPKKKHRRNRTTFTTYQLHELERAFEKSHYPDVYSREELAGKVNLPEVRVQVWFQNRRAKWRRQEKLEVSSMKL-
Dmel_REPO     -KKKGDPNGIKKKKTRTTFTAYQLEELERAFERAPYPDVFAREELAIKLNLSESRVQVWFQNRRAKWRKHEPPRK-TGYIK
Mmus_Uncx     -DPDKESPGCKRRRTRTNFTGWQLEELEKAFNESHYPDVFMREALALRLDLVESRVQVWFQNRRAKWRKKENTKKGPGRP-
Dmel_OTP      -NNNNSMQQQDQKRHRTRFTPAQLNELERCFSKTHYPDIFMREEIAMRIGLTESRVQVWFQNRRAKWKKRK-KTTNVFRTP
Mmus_Otp      -QAGQQQGQQKQKRHRTRFTPAQLNELERSFAKTHYPDIFMREELALRIGLTESRVQVWFQNRRAKWKKRK-KTTNVFRAP
Dmel_CG34367  -PLDTSLVNTKQRRSRTNFTLDQLNELERLFEETHYPDAFMREELSQRLGLSEARVQVWFQNRRAKCRKHENQMHKGFLV-
Mmus_Alx4     -KTDSESNKGKKRRNRTTFTSYQLEELEKVFQKTHYPDVYAREQLAMRTDLTEARVQVWFQNRRAKWRKRERFGQMQQVR-
Mmus_Alx1     -KCDSNVSSSKKRRHRTTFTSLQLEELEKVFQKTHYPDVYVREQLALRTELTEARVQVWFQNRRAKWRKRERYGQIQQAK-
Mmus_Alx3     -SMELAKTKSKKRRNRTTFSTFQLEELEKVFQKTHYPDVYAREQLALRTDLTEARVQVWFQNRRAKWRKRERYGKMQEGR-
Phaw_AL1      -GDEVTPNKKKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSHHLA-
Phaw_AL2      -GDTAGPTKRKQRRYRTTFTSYQLEELEKVFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPSSHPY-
Dmel_AL       EADEYAP-KRKQRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQEKVGPQSHPY-
Mmus_Arx      -DSEEGLLKRKQRRYRTTFTSYQLEELERAFQKTHYPDVFTREELAMRLDLTEARVQVWFQNRRAKWRKREKAGAQTHPP-
Dmel_MU       --MDKGTMKRKQRRYRTTFNTLQLQELERAFQRTHYPDVFFREELAVRIDLTEARVQVWFQNRRAKWRKQEKIGGLGGDY-
Dmel_CG34340  -IHDETFVRRKQRRNRTTFTLQQLEELETAFAQTHYPDVFTREDLAMKINLTEARVQVWFQNRRAKWRKAERLKDEQRKR-
Mmus_Prrxl1   -DFDDGFLRRKQRRNRTTFTLQQLEALEAVFAQTHYPDVFTREELAMKINLTEARVQVWFQNRRAKWRKTERGASDQEPG-
Mmus_Phox2b   -DHGGLNEKRKQRRIRTTFTSAQLKELERVFAETHYPDIYTREELALKIDLTEARVQVWFQNRRAKFRKQERAAAAAAAA-
Mmus_Phox2a   -EPSGLHEKRKQRRIRTTFTSAQLKELERVFAETHYPDIYTREELALKIDLTEARVQVWFQNRRAKFRKQERAASAKGAA-
Dmel_PHDP     -DGLSLTDKSKQRRIRTTFTSNQLNELEKIFLETHYPDIYTREEIASKLHLTEARVQVWFQNRRAKFRKQERHAIYIMKD-
Dmel_CG9876   -PAGGGHHSRKPRRNRTTFSSAQLTALEKVFERTHYPDAFVREELATKVHLSEARVQVWFQNRRAKFRRNERSVGSRTLL-
Mmus_Gsc      -QLLNQLHCRRKRRHRTIFTDEQLEALENLFQETKYPDVGTREQLARKVHLREEKVEVWFKNRRAKWRRQKRSSSEESEN-
Mmus_Gsc2     -GTSGPGPQRRTRRHRTIFSEEQLQALEALFVQNQYPDVGTRERLAVRIRLREERVEVWFKNRRAKWRHQKRASSSRLLP-
Dmel_GSC      -HLGHGPPPKRKRRHRTIFTEEQLEQLEATFDKTHYPDVVLREQLALKVDLKEERVEVWFKNRRAKWRKQKREEQERLRK-
```

# [PhD_A6_phylogeny_database]

## [PhD_A6.3_phylogeny_ClustalW2_alignments]

### “alignment_Arrowhead-Lhx68_proteins.txt”

```
Alignment of the protein sequence fractions of various arrowhead and Lhx6/8 genes from different species, encompassing LIM1, LIM2
and HD domains.
This was done using the ClustalW2 algorithm (Larkin et al., 2007).
For details and references regarding the species' protein sequences, see Appendix (A7).

Dmel_AWH        TELRSCAACGEPISDRFFLEVGGCSWHAHCLRCCMCMCPLDRQQSCFIRERQVYCK-----------
ADYSKNFGAKCSKCCRGISASDWVRRARELVFHLACFACDQCGRQLSTGEQFALMDDRVLCKAHYLETVEGGTTSSDEGCDGDGYHKS----------
KTKRVRTTFTEEQLQVLQANFQIDSNPDGQDLERIASVTGLSKRVTQVWFQNSRARQKKHIH-AGKNKIREPEGSS---------FARHINLQLTYSFQ
Dpse_AWH        TELRSCAACGEPISDRFFLEVGGCSWHAHCLRCCMCMCPLDRQQSCFIRERQVYCK-----------
ADYSKNFGAKCSKCCRGISASDWVRRARDLVFHLACFACDQCGRQLSTGEQFALMDDRVLCKAHYLETVEGGTTSSDEGCDGDGYHKS----------
KTKRVRTTFTEEQLQVLQANFQIDSNPDGQDLERIASVTGLSKRVTQVWFQNSRARQKKHIH-AGKNKIREPEGSS---------FARHINLQLTYSFQ
Dvir_AWH        ----------------------------------MCMCPLDRQQSCFIRERQVYCK-----------
ADYSKNFGAKCSKCCRGISASDWVRRARDLVFHLACFACDQCGRQLSTGEQFALMDDRVLCKAHYLETVEGGTTSSDDGCDGDGYHKS----------
KTKRVRTTFTEEQLQVLQANFQIDSNPDGQDLERIASVTGLSKRVTQVWFQNSRARQKKHIH-AGKNKIREPEGSS---------FARHINLQLTYSFQ
Agam_AWH        KELRSCTACGEPISDQFLLDVGGCSWHSACLRCCICHTPLDHQPSCFLRERQIYCK-----------
TDYTKRFGTKCARCSRTISATDWVRRARDLIFHLACFACDSCGRQLSTGEQFALVDDKVLCKTHYSEMFDCG-TSSDDGCEADGYQKNN---------
KTKRVRTTFTEEQLQILQANFNIDSNPDGQDLERIASVTGLSKRVTQVWFQNSRARQKKHVQGTGVYPYLEGEMNP---------FARHINLQLSYTFQ
Tcas_AWH        TEYRLCSACGEPISDKFLLEVSGRSWHARCLRCCVCQLQLDRQPSCFIRDRAIYCK-----------
ADYAKSFGAKCSVCSRGISSSDWVRKARDHVYHLACFACAACHRQLSTGEEFALHEDRVLCKAHYLETLDGGTTSSDDGCDAEGYHKN----------
KAKRVRTTFTEEQLQVLQANFQLDSNPDGQDLERIAQITGLSKRVTQVWFQNSRARQKKHMH-TGKIKTQSREDNG---------FGRHINLHLTYSFQ
Amel_AWH        ANVMECGGCGERVRERTVLCVGGRTWHSRCLRCFACARPLHDQHSCFLKGMRLYCR-----------
HDYALTFGAKCAKCGRSVGAGDWVRRARERVYHLACFACDACSRQLSTGEQFALLDARLLCKAHYLDVVEGNNTSSDEGGDSESGHKSGN--------
KAKRVRTTFTEEQLSVLQANFQLDSNPDGQDLERIAHVTGLSKRVTQVWFQNSRARQKKHLH-TGKMKGQHVHQSPPNSNASTGDFGRHINLHLTYSFQ
Phaw_AWH        PTTESCAGCARIITDRFLLRVNCLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCK-----------
SDYSRNFGARCAKCCRAIGSTDWVRRARDRVYHLACFACDACKRQLSTGEEFALHESRVLCKQHYLEAVEGGATSNDENTDGEGGQRK-----------
SKRVRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVTQVWFQNSRARQKKHMTSKKHHQGLERGLHDITG-----SLARTIDLHFMYSFR
Cele_LIM-4      AVIVICTQCQHQIQDKFFLSIDGRNYHENCLQCSTCENPLSN--
KCFYKDKTFYCKGCYFRTHVTSTASSCRELGPKCASCDRTIQATDWVRRARNYVYHLACFSCNQCKRQLSTGEEYALQEGNLLCKQHFLELVEGDSGVSS--------QKA----------
KTKRVRTTFAEDQLSVLQTYFNRDSNPDGADLEKIASMTGLSKRVTQVWFQNSRARQKKWHQ-KSEGDNGDSQRSS---------VGPSSPSQKSDSSS
Mmus_Lhx6       AGKNICSSCGLEILDRYLLKVNNLIWHVRCLECSVCRTSLRQQNSCYIKNKEIYCK-----------
MDYFSRFGTKCARCGRQIYASDWVRRARGNAYHLACFACFSCKRQLSTGEEFGLVEEKVLCRIHYDTMIENLKRAAENGNGLTLEGAVPSEQDS-
QPKPAKRARTSFTAEQLQVMQAQFAQDNNPDAQTLQKLADMTGLSRRVIQVWFQNCRARHKKHT-PQHP--VPPSGAPPTR-------LPSALSDDIHYSPF
Drer_Lhx6       TGKNVCASCGLEILDRYLLKVNNLIWHVRCLECSVCRTSLRQHSSCYIKNKEIFCK-----------
MDYFSRFGTKCARCGRQIYASDWVRRARGNAYHLACFACFSCKRQLSTGEEFGLVEEKVLCRIHYDTMVENLKRAAESGSGLTLEGAVPSEQDS-
QPKPAKRARTSFTAEQLQVMQAQFAQDNNPDAQTLQKLADMTGLSRRVIQVWFQNCRARHKKHT-PQHN--GPPQAHPQAR-------MPPSLPDELHYSPF
Mmus_Lhx8       PGKCVCSSCGLEIVDKYLLKVNDLCWHVRCLSCSVCRTSLGRHTSCYIKDKDIFCK-----------
LDYFRRYGTRCSRCGRHIHSTDWVRRAKGNVYHLACFACFSCKRQLSTGEEFALVEEKVLCRVHFDCMLDNLKREVENGNGISVEGALLTEQDVNHPKPAKRARTSFTADQLQVMQAQFAQDNNPDAQT
LQKLAERTGLSRRVIQVWFQNCRARHKKHVSPNHSSSAPVTAVPSSR-------LSPPMLEEMAYSAY
Drer_Lhx8a      -EKAICTSCGTEIVDKYLLKVNDMCWHVRCLSCSVCQTSLGRHISCYIKEKEIFCK-----------
LDYFRKYGTRCAHCGRNIHSNDWVRRAKGNTYHLACFACFSCKRQLSTGEEFALVDERVLCRVHYDCMLDNLKRAMENGKGVNVEGAVPSEQEVNQPKPAKRARTSFTADQLQVMQAQFAQDNNPDAQT
LQKLAERTGLSRRVIQVWFQNCRARHKKHVSPNHSSAAPVSSLQSAH-------LSPPLIDELQYTAF
```

# [PhD_A6_phylogeny_database]

## [PhD_A6.3_phylogeny_ClustalW2_alignments]

### “alignment_Arrowhead-LIM-HD-group.txt”

Alignment of the protein sequence fractions of various LIM-HD-group genes from Parhyale hawaiensis, Drosophila melanogaster and mouse, encompassing LIM1, LIM2 and HD domains.
This was done using the ClustalW2 algorithm (Larkin et al., 2007).
For details and references regarding the species' protein sequences, see Appendix (A7).

```
Phaw_AWH     CAGCARIITDRFLLRVN-CLSWHQSCLRCCVCQLALDRQPSCFIRDQNIYCKSDYSRNFGAR-CAKCCRAIGSTDWVRR-ARDRVYHLACFACDACKRQLSTGEEFALHES-
RVLCKQHYLEAVEG--------------------------------------------------------------------------GATSNDEN----------TDGEGGQR-
KSKRVRTTFTDEQLQVLQANFQIDSNPDSQDLERIAQLTGLSKRVTQVWFQNSRARQKKHMTSKK
Dmel_AWH     CAACGEPISDRFFLEVG-GCSWHAHCLRCCMCMCPLDRQQSCFIRERQVYCKADYSKNFGAK-CSKCCRGISASDWVRR-ARELVFHLACFACDQCGRQLSTGEQFALMDD-
RVLCKAHYLETVEG--------------------------------------------------------------------------GTTSSDEG----------
CDGDGYHKSKTKRVRTTFTEEQLQVLQANFQIDSNPDGQDLERIASVTGLSKRVTQVWFQNSRARQKKHIHAGK
Mmus_Lhx6    CSSCGLEILDRYLLKVN-NLIWHVRCLECSVCRTSLRQQNSCYIKNKEIYCKMDYFSRFGTK-CARCGRQIYASDWVRR-ARGNAYHLACFACFSCKRQLSTGEEFGLVEE-
KVLCRIHYDTMIEN--------------------------------------------------------------------------LKRAAENGNGLTLEGAVPSEQDS-
QPKPAKRARTSFTAEQLQVMQAQFAQDNNPDAQTLQKLADMTGLSRRVIQVWFQNCRARHKKHT-PQH
Mmus_Lhx8    CSSCGLEIVDKYLLKVN-DLCWHVRCLSCSVCRTSLGRHTSCYIKDKDIFCKLDYFRRYGTR-CSRCGRHIHSTDWVRR-AKGNVYHLACFACFSCKRQLSTGEEFALVEE-
KVLCRVHFDCMLDN--------------------------------------------------------------------------
LKREVENGNGISVEGALLTEQDVNHPKPAKRARTSFTADQLQVMQAQFAQDNNPDAQTLQKLAERTGLSRRVIQVWFQNCRARHKKHVSPNH
Mmus_Lhx9    CAGCGGKISDRYYLLAV-DKQWHLRCLKCCECKLALESELTCFAKDGSIYCKEDYYRRFSVQRCARCHLGISASEMVMR-ARDSVYHLSCFTCSTCNKTLTTGDHFGMKDS-
LVYCRAHFETLLQG---------EYPPQLSYTELAAKSGG----------------LALPYFNGTGT--VQKGRPRKRKSPALGVDIVNYNSG--
CNENEADHLDRDQQPYPPSQKTKRMRTSFKHHQLRTMKSYFAINHNPDAKDLKQLAQKTGLTKRVLQVWFQNARAKFRRNLLRQE
Mmus_Lhx2    CAGCGGKISDRYYLLAV-DKQWHMRCLKCCECKLNLESELTCFSKDGSIYCKEDYYRRFSVQRCARCHLGISASEMVMR-ARDLVYHLNCFTCTTCNKMLTTGDHFGMKDS-
LVYCRLHFEALLQG---------EYPAHFNHADVAAAAAAAAAAKSAGLGSAGANPLGLPYYNGVGT--VQKGRPRKRKSPGPGADLAAYNAALSCNENDAEHLDRDQ-
PYPSSQKTKRMRTSFKHHQLRTMKSYFAINHNPDAKDLKQLAQKTGLTKRVLQVWFQNARAKFRRNLLRQE
Dmel_AP      CSGCGRQIQDRFYLSAV-EKRWHASCLQCYACRQPLERESSCYSRDGNIYCKNDYYSFFGTRRCSRCLASISSNELVMR-ARNLVFHVNCFCCTVCHTPLTKGDQYGIIDA-
LIYCRTHYSIAREGDTASSSMSATYPYSAQFGSPHNDSSSPHSDPSRSIVPTGIFVPASHVINGLPQPARQKGRPRKRKPK----DIEAFTANIDLNTEYVDFGRGSH--
LSSSSRTKRMRTSFKHHQLRTMKSYFAINHNPDAKDLKQLSQKTGLPKRVLQVWFQNARAKWRRMMMKQD
Mmus_Lhx4    CAGCNQHILDKFILKVL-DRHWHSSCLKCADCQMQLAD--RCFSRAGSVYCKEDFFKRFGTK-CTACQQGIPPTQVVRK-
AQDFVYHLHCFACIICNRQLATGDEFYLMEDGRLVCKEDYETAK-------------------------------------------------------------------------------------
----QNDD-------SEAGAKRPRTTITAKQLETLKNAYKNSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKEKR--LKKD
Mmus_Lhx3    CAGCDQHILDRFILKAL-DRHWHSKCLKCSDCHVPLAE--RCFSRGESVYCKDDFFKRFGTK-CAACQLGIPPTQVVRR-
AQDFVYHLHCFACVVCKRQLATGDEFYLMEDSRLVCKADYETAK-------------------------------------------------------------------------------------
----QRE--------AEATAKRPRTTITAKQLETLKSAYNTSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKEKR--LKKD
Dmel_LIM3    CGGCHELILDRFILKVL-ERTWHAKCLQCSECHGQLND--KCFARNGQLFCKEDFFKRYGTK-CSACDMGIPPTQVVRR-
AQDNVYHLQCFLCAMCSRTLNTGDEFYLMEDRKLICKRDYEEAK-------------------------------------------------------------------------------------
----AKGLYLDGSLDGDQPNKRPRTTITAKQLETLKTAYNNSPKPARHVREQLSQDTGLDMRVVQVWFQNRRAKEKR--LKKD
Mmus_Lhx5    CAGCERPILDRFLLNVL-DRAWHIKCVQCCECKTNLSE--KCFSREGKLYCKNDFFRRFGTK-CAGCAQGISPSDLVRK-
ARSKVFHLNCFTCMVCNKQLSTGEELYVIDENKFVCKDDYLS-------SSSLKEGSLNSVSSCTD---------------------------------------
RSLSPDLQDPLQDDPKETDNSTSSDKETANNENE--EQNSGTKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERR--MKQL
Mmus_Lhx1    CAGCKRPILDRFLLNVL-DRAWHVKCVQCCECKCNLTE--KCFSREGKLYCKNDFFRCFGTK-CAGCAQGISPSDLVRR-
ARSKVFHLNCFTCMMCNKQLSTGEELYIIDENKFVCKEDYLSN------SSVAKENSLHSATTGSD---------------------------------------
PSLSPDSQDPSQDDAKDSESANVSDKEGGSNEND--DQNLGAKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERR--MKQL
Dmel_LIM1    CAGCNKPILDKFLLNVL-ERAWHASCVRCCECLQPLTD--KCFSRESKLYCRNDFFRRYGTK-CSGCGQGIAPSDLVRK-
PRDKVFHLNCFTCCICRKQLSTGEQLYVLDDNKFICKDDYLLGKAPSCGHNSLSDSLMGSASEDDDDDDPPHLRATALGLGVLGPNGPDSAGGPLGTSD---
ISVQSMSTDSKNTHDDSDQGSLDGDPDGRGDSQAENKSPDDANGSKRRGPRTTIKAKQLEVLKTAFNQTPKPTRHIREQLAKETGLPMRVIQVWFQNKRSKERR--MKQI
Mmus_Isl2    CVGCGSQIHDQFILRVSPDLEWHAACLKCAECSQYLDETCTCFVRDGKTYCKRDYVRLFGIK-CAQCQVGFSSSDLVMR-ARDSVYHIECFRCSVCSRQLLPGDEFSLREH-
ELLCRADHGLLLERAAAGSPRSPGP----------------------------------------------------------LPGARGLHLP-DAGSGRQPSLRTHVHKQAE--
KTTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKSILMK
Mmus_Isl1    CVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFVRDGKTYCKRDYIRLYGIK-CAKCSIGFSKNDFVMR-ARSKVYHIECFRCVACSRQLIPGDEFALRED-
GLFCRADH-DVVERASLGAGDPLSP----------------------------------------------------------LHPARPLQMAAEPISARQPALRPHVHKQPE--
KTTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMMK
Dmel_TUP/ISL CVGCGGQIHDQYILRVAPDLEWHAACLKCQECRQFLDESCTCFVRDGKTYCKRDYVRLFGTK-CDKCGNSFSKNDFVMR-
AKTKIFHIECFRCSACARQLLPGDEFALRDAGALYCKEDH-DVLEKSSQSSLTSSSVESNNNISSSN--------------------------------------
NNNTNLSNNNHSSELGSMSDSGSESGSHKSIRDKRPSGPSDGKPTRVRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTSLSPRVIRVWFQNKRCKDKKKTIQMK
Mmus_Lmx1a   CEGCQRVISDRFLLRLN-DSFWHEQCVQCASCKEPLET--TCFYRDKKLYCKYHYEKLFA-VKCGGCFEAIAPNEFVMR-AQKSVYHLSCFCCCVCERQLQKGDEFVLKEG-
QLLCKGDYEKERELLSLVSPAASDS-----------------------------------------------------------GKSDDEESLCKSAHGAG---
KGASEDGKDHKRPKRPRTILTTQQRRAFKASFEVSSKPCRKVRETLAAETGLSVRVVQVWFQNQRAKMKKLARRQQ
Mmus_Lmx1b   CEGCQRPISDRFLMRVN-ESSWHEECLQCAACQQALTT--SCYFRDRKLYCKQDYQQLFA-AKCSGCMEKIAPTEFVMR-ALECVYHLGCFCCCVCERQLRKGDEFVLKEG-
QLLCKGDYEKEKDLLSSVSPDESDS-----------------------------------------------------------
VKSEDEDGDMKPAKGQGSQSKGSGDDGKDPRRPKRPRTILTTQQRRAFKASFEVSSKPCRKVRETLAAETGLSVRVVQVWFQNQRAKMKKLARRHQ
Dmel_CG32105 CEGCGQKIHDRFLMNVG-DANWHEQCLACCYCGMQLHH--TCYVRNSKLYCKMDYDRLFG-VKCSSCCHAILPQELVMRPIPNFVFHLPCFVCYACRLPLQKGEQFMLRDG-
QLFCY-RHDLEKEMFLAAAAAQHCG-----------------------------------------------------------FVGLDEEDLLRPR------------DGR--
RGPKRPRTILTSQQRKQFKASFDQSPKPCRKVREALAKDTGLSVRVVQVWFQNQRAKMKKIQRKAK
Dmel_CG4328  CAHCCQPICDRYIMRVV-ENSFHEGCLKCTACSLHLVH--SCYAREGKLYCRVDYERLYIRNHCLGCGLKIAADELVMR-CHENVFHLKCFACVVCGALLKKGEQYVVKQG-
QLFCRFDYEKEVEMLQGYDFYGDEL-----------------------------------------------------------FPPKLDGR----------------------
RGPKRPRTILNTQQRRAFKASFEVSPKPCRKVRENLAKDTGLSLRIVQVWFQNQRAKVKKIQKKAK
```

# [PhD_A6_phylogeny_database]

## [PhD_A6.3_phylogeny_ClustalW2_alignments]

### “alignment_knirps-family.txt”

Alignment of coherent fragments (translation start until end of the PIDL motifs) of various arthropod knirps-family encoded proteins together with those murine estrogen receptors that have the highest sequence similarity to knirps encoded proteins, as well as their putative Drosophila ortholog.

This was done using the ClustalW2 algorithm (Larkin et al., 2007).

For details and references regarding the species' protein sequences, see Appendix (A7).

```
Amel_KNI    MNQLCRVCGEPAAGFHFGAFTCEGCKSFFGRTYN--NLGSISECKNGGVCVINKKNRTACKACRLRKCLMVGMSKSGSRYGRRSNWFKIHCLLQEQSHQAQTRLIKDP---------
---------------------------------------------------------------------------------------------------------------------------------
--------------------------KSAYEKAFGSLDVANNNNNNNENPGTTNATGVVGIVTTTTTTANSYHHHHHQHHQDRFPKRDELRPQDIPAQLRAPDIPTRAAQDPLLRPMDLVRAPHELLR-
-------------------------------------------------PEVSRFPMWRGPPFFHPALTHMQLLNTPFFPFQQRFIVPYVNQVGAPQMVASLTSSSSDS--------------------
-------------------------------VSPRSTPSPKRDDDNRSTRDECREVGSGVCQRSQMEAYDKSIAFLRS---------------------------------------------------
-----------------------------------------------LGPEQDEPIDLSMKA
Crog_KNI    RNQICRVCDEPAAGYHFGAFTCEGCKSFFGRTCGKSNLNSISECKNNYSCVIDKKNRTSCKACRLRKCLRVGMSKAGSRYGRRSNWFKMHCVMS-----------------------
---------------------------------------------------------------------------------------------------------------------------------
-------------------------------------------GSNKGNSAASVVAGSKGISRVEILKK------------------------------------------------------------
-----------------------------------------------------RSETPRGDNISHSSATFSVQFPWWYFVLQRQFLASYEN--------------------------------------
-----------------------------------SLKKALRDEDECGT--------------------------------------------------------------------------------
---------------------------------------------------NDGPMDLSQRS
Dmel_KNI    MNQTCKVCGEPAAGFHFGAFTCEGCKSFFGRSYN--
NISTISECKNEGKCIIDKKNRTTCKACRLRKCYNVGMSKGGSRYGRRSNWFKIHCLLQEHEQAAAAAGKAPPLAGGVSVGG------------------------------------------------
-----------------------------------------------------------------------------------------------
APSASSPVGSPHTPGFGDMAAHLHHHHQQQQQQQ---------VPRHPHMPLLGYPSYLSDPS--------------AALPFFSMMGGVPHQSPFQLPPHLLFPGYHASAAAAAASA------------
---------------------------ADAAYRQEMYKHRQSVDSVESQNRFSPASQPPVVQPTSSAR----QSPIDVCLEEDVHS---------------------------
VHSHQSSASLLHPIAIRATP--------------------TTPTSSS-PLSFAAKMQSLSPVSVCSIGGETTSVVPVHP--------------------------------------------------
------------------------------------------------PTVSAQEGPMDLSMKT
Dpse_KNI    MNQTCKVCGEPAAGFHFGAFTCEGCKSFFGRSYN--NISTISECKNDGKCVIDKKNRTTCKACRLRKCYNVGMSKGGSRYGRRSNWFKIHCLLQEHEQAAAAAGKVQSAGG------
---------------------------------------------------------------------------------------------------------------------------------
----------------VTSSPVGSPHTPGFGDMAAAAAHLHHQHQQHQQQ-QQAQQQQQRHPHMPLLGYPSYLGEPFGA---RHPADAAAAAALPFFSMMATP--
QSAFQLPPHLLFPGYHASAAAAAA-----------------------------------------
ADAAYRQEMYKHRQSVDSAGSVESHSRYTPPSVAAVPSPPAPVRPQTPIDVCLQEEDVQSVHSHQSHHSQHSHHSQQSHQSQQSQQSHQSHHSQTMHHPIAIRATP--------------------
TAQLSSSSPLSFAAKMQSLSPVSVCSIGGETAVHHQQHQQ-----------------------------------------------------------------------------------------
---EQPAVPAAAAQDGPMDLSMKT
Dvir_KNI    MNQTCKVCGEPAAGFHFGAFTCEGCKSFFGRSYN--NISTISECKNDGKCIIDKKNRTTCKACRLRKCYNVGMSKGGSRYGRRSNWFKIHCLLQEHEQAAAAAAGKGPGHATG----
---------------------------------------------------------------------------------------------------------------------------------
----------------------SPMSSPGFGDLAAHLQQQQQQHQQQQQQQHQHQQQQQRHPHLPPLLGYTGYHLPEHFGARHPADAAAAAAALPFFSMMATP--
QSAFQLPPHLLFPGYHASAAAAAA-----------------------------------------
ADAAYRQEMYKHRQSVDSAASAESHSRYTPPTVATVPQQSQPQPAASPIDVCLGADDDVQSQHSHSHS------------------
HSQSQSPHTIHTPVAIRATPPQQLLPGLTTASHSSSASPTPSKSQSSSPLSFTAKMQSLSPVSVCSIGGETAAANAAAS--------------------------------------------------
------------------------------------------------AAAAAQDGPMDLSMKT
Phaw_KNI1   MNQLCKVCSEPAAGFHFGAFTCEGCKSFFGRTYN--NLSQIHECKNGGQCVINKQNRTSCKACRLRKCLFVGMSKSGSRYGRRSNWFKIHCALQEQAAN------------------
---------------------------------------------------------------------------------------LNRSTEGAGNINLQLSAAR-----------------------
------------------------------ENLIRLSPSRCVDDEGKLGSP-------ALSSPDSNTSDTSVDVEPRRNISGTSIQTSHSQLTLAQHESLVARELSKRTGNALATVGGRSLEPPVPG--
-----------------------------------TDFVAGLSLYGIHSGFSLLQAS-------QLYQRLYPPFLYPGIRPVMDPIAPPQQSPSPTQIKSLPIK-------------------
SPSNPPSPAHSLPPP----------------------STKREQVSNPSSPQ---SLPARHFSDRTAHSPSSSSPSPPLISRKRA---------------------------------------------
---------------------------------------------------ADSPPEQDAPIDLSVKR
Phaw_KNI2   MNQLCKVCAEPAAGFHFGAFTCEGCKSFFGRTYN--
NLSQIHECKNGGQCVINKQNRTSCKACRLRKCLVVGMSKTGSRYGRRSNWFKIHCLLQEQCGYSGTAGQLPTFRSPVTPTLQDGSLPFVNTQLRDALRMDVPLTPVEHSASPSPRIPITSAAPGGVCIR
PPTSVEMGPAGVTVSEPHLSPNFPMGQIGAVERRIQEETLNALRRGVGMPGSELEATYREQILEALRRSHNMQPPRLLDPSAVQTLRRSPPIAEPVRRIGAEYERVRGVSAEAEIRMNSNPPYDWDPRL
QEAPVFSEASVVKSESPHSSKVALSVSVSTSQEVDSSSSNQIKIECSRSETETGSKSTPEVKSERRSVGNRSPVSLTPKSQASARLQDETGAPMTEEERLSFLKSASKISANNTATRDGSAVNSFHGIP
GLFPHRSSSRDDILYGPLNNNIFPFASYPPVWPSILPFPSLAKIYPPPPFSSPAVSGAPISL---------
DGFKSPLYLPTELRNTVLPNTLPDVTKEDSSNKKRILDAILQVQRESACGPRSPLGGLSPPGVTLALSLPQHVTTSTSTLPMVTATTT-----------------------------------------
-------------------------------------------------------TSTAPGE-QPIDLTVRR
Nvit_KNI    MNQQCKVCGEPAAGFHFGAFTCEGCKSFFGRTYN--NLGSISECKNGGVCVINKKNRTACKACRLRKCLMVGMSKSGSRYGRRSNWFKIHCLLQEHKR-------------------
---------------------------------------------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------AAAVLAATKEEDRVNSLQQQSPANTTPPRNSTPPSFP--
--------------------------------------------------YKDLVSSSAMFGLVAAGQKDPSAMASASSLEVFRPFLPIYKRTTDTPSDSGASSVES------------
DEHDSRSNSAFSYFKQEGNSP-------------PRRKINATVTLSTDYHHHQHHHHGLVYPPSELVTPATSHHSLRGGDLLLASPSP-----------------------------------------
-------------------------------------------------------GGLAVEQMEPIDLSLKS
Amel_EG     MNQQCKVCGEPAAGFHFGAFTCEGCKSFFGRTYN--NLGSISECKNGGVCVINKKNRTACKACRLRKCLMVGMSKSGSRYGRRSNWFKIHCLLQEQTNGNQSVTSGGAS--------
---------------------------------------------------------------------------------------------------------------------------------
--------------------PFGPGFLPGFLPQAPGQQSSGIYGKDGKDRASPSQEDLALQSHLLHVAMLKQERDRGNKSPHPDDVALQNQLRLLAWQKERERVGGNQQPP---GTPPRDTLTPPFY--
--------------------------------------------------KQQQQPQHPVSPILPMNFTQKDGVPTSP-SDVFRPFLPTYKRTADTPSDSGASSVEGGDG---------
QDTESRSNSALSYFKASTSSPTL------SERDFPAKKINATVTLTTGFHPHSS--LSLIYPPPGLITPAASQLSPRGGDLLLVSPSP-----------------------------------------
-------------------------------------------------------GGLAVEQDEPIDLSVRS
Ofas_KNI2   MNQQCKVCNEPAAGFHFGAFTCEGCKSFFGRTYN--NLGSISECKNNGECVINKKNRTSCKACRLRKCLLVGMSKSGSRYGRRSNWFKIHCLLQEQGAS------------------
---------------------------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------------------------------FSPFPKEEPIK--
--------------------------------------------------SEPEDQQEGVSFRSSPEISRKCVTPTIGFQPYPRPVAPLFLEPWPREQG------------------------------
-------------------------------------------------------------------------ASPSP---------------------------------------------------
---------------------------------------------GRVALEQEAPIDLSVKK
Dmel_EG     MNQLCKVCGEPAAGFHFGAFTCEGCKSFFGRTYN--NIAAIAGCKHNGDCVINKKNRTACKACRLRKCLLVGMSKSGSRYGRRSNWFKIHCLLQEQQTTSGLGGGSS----------
---------------------------------------------------------------------------------------------------------------------------------
---------------------VGSGSGGGVSSASLEQLARLQQASNQARQTYQDKTNPCIKSATATTSPRIEGAAVGTG-----------IGGGASPSFLQAAKLHHQRQLKLDSRLSNTPSDSGAS--
--------------------------------------------------SAGDPNEDGVTSVLGGQIATPSSTNATSLPKLDLR-HPNFPATSEPDADMQRQ--------------------------
-------------------------RHQELLEIFRSHSEPLYSSFAPFSHLPPVLLAAGVPQLPIFKDQFKAELLFPT---------------------------------------------------
---------------------------------------------TSS-PELEEPIDLSFRS
```

```
Dpse_EG     MNQLCKVCGEPAAGFHFGAFTCEGCKSFFGRTYN--NIAAIAGCKHNGDCVINKKNRTACKACRLRKCLLVGMSKSGSRYGRRSNWFKIHCLLQEQQSTAGTASGTAG---------
---------------------------------------------------------------------------------------------------------------------------------
---------------------SGPGGVGGVSSANLEQLARLQQASHQARQTYQDKTNPCIKSATSSSPTMGGVGAAGATSSAASPSLPGFLQAAKYLNDRQQHQKHQQRQLKLESRLSNTPSDSGAS--
--------------------------------------------------SAGDPNEDGATSVLSGNTTGVISTAASMPKLGQRQGCPPFPATSEPDADLAAQQQRQ----------------------
-------------------------RHQELLEIFRSHSEPLYSSFAPFSHLPPVLLAAGVPPLPIFKDQFKAELLFPA---------------------------------------------------
---------------------------------------------ASS-PELDEPIDLSLRS
Dvir_EG     MNQLCKVCGEPAAGFHFGAFTCEGCKSFFGRTYN--NIPAIAGCKHNGDCVINKKNRTACKACRLRKCLLVGMSKSGSRYGRRSNWFKIHCLLQEQQGGGGGAGTGS----------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------SSVNNANLDQLARLQQASNQARQTYQDKTNPCIKSATAATSPSIALQTCSAVAPAAASTPSPSLPGFLQAAKYLNEQQHR--QLKLESRLSNTPSDSGAS--
--------------------------------------------------SAGDPTDDGTS---------VLSATTSTIRKAMRQSSSAYAGSSEADIAEQQQR-------------------------
-------------------------RHKELLDIFRSHSEPLYSSFTPFT-LPPALMASSMPPLPIFKEQFKSELLFPF---------------------------------------------------
---------------------------------------------PATSPAEEEPIDLSLRS
Tcas_EG     MNQLCKVCGEPAAGFHFGAFTCEGCKSFFGRTYN--NISSISECKNNGECVINKKNRTACKACRLRKCLLVGMSKSGSRYGRRSNWFKIHCLLQEQQQAG-----------------
---------------------------------------------------------------------------------------------------------------------------------
--------------------------------ANMAAAANLLRPHPYLSPLFHSAAPRTVSRETKSS-----------------------------------------------------ESDSGAS--
--------------------------------------------------SADPDDE------------LRSSSAFSYIKPSSERDYFSVKKLPSPASDSECFDRRK----------------------
-------------------------FISALVNAPSVSPASLPSISPSGGFLPKWLPPLQGFSDPSAWRDLWARGPVTT---------------------------------------------------
---------------------------------------------AAAAPDQVEPIDLSLKS
Dmel_KNRL   MNQTCKVCGEPAAGFHFGAFTCEGCKSFFGRSYN--
NLSSISDCKNNGECIINKKNRTACKACRLKKCLMVGMSKSGSRYGRRSNWFKIHCLLQEQQQQAVAAMAAHHNSQQAGGGS------------------------------------------------
-----------------------------------------------------------------------------------------SGGSGGGQGMPNGVKGMSGVPPPAAAAAALGMLGHPGG-
YPGLYAVANAGGSSRSKEELMMLGLDGSVEYGSHKHP-VVASPSVSSPDSHNSDSSVEVSSVRGNPLLHLGGKSNSGGSSSGADGSHSGGGGGGGGGVTPGRPPQMRKDLSP---
FLPLPFPGLASMPVMPPPAFLPPSHLLFPGYHPALYSHHQGLLKPTPEQQQ--AAVAAAAVQHLFNSSGAGQRFAPG---------TSPFANHQQHHKEEDQPAPARSPSTHANNN-------------
-HLLTNGG-----AADELTKRFYLDAVLKSQQQ-SPP-PTTKLP--PHSKQDYSISALVTPNSESG---RERVKSRQNEEDDEARADGIIDG-----------
AEHDDEEEDLVVSMTPPHSPAQQEERTPAGEDP------RPSPGQDNPIDLSMKT
Dvir_KNRL   MNQTCKVCGEPAAGFHFGAFTCEGCKSFFGRSYN--
NLSSISDCKNNGECIINKKNRTACKACRLKKCLMVGMSKSGSRYGRRSNWFKIHCLLQEQQQQAVAAMAAHHNSQQAGAGG------------------------------------------------
-----------------------------------------------------------------------------------------NATVGG--
ASNGCVKTMAGVSAPSAAAAALGMLGHATGGYPGLYATG----PPRSKEELMMLGLDGSGDYGGIKHPSLVASPSVSSPDSHNSDSSVEVSSVRGNSLLQLGGKPNAATDGSQAG-----------
SGTAPGRPPQLRKDLAP---FLPLPFPGLTSMPVMPPPAFLPPSHLLFPGYHPALYSHHQGLLKPTPEQQQQ-AAVAAAAVQHLFNSSNASGQRFVP---------QLHATAPFTGHKEEAALSPNHS-
--NSNNN--------------NLLPNGSRTATSSADELTKRFYLDAVLKSQQH-SPPPLATKLP--PHSKQDYSISALVTPNSESG---
RERLKSRQSNGEADDEERETATGSDKGVKEELEPGEEDQEEEDLVVSMTPPRSPAQQQ---PAQLVD------VHTAAQDNPIDLSMKT
Dpse_kNRL   MNQTCKVCGEPAAGFHFGAFTCEGCKSFFGRSYN--
NLSSITDCKNNGECIINKKNRTACKACRLKKCLMVGMSKSGSRYGRRSNWFKIHCLLQEQQQQAVAAMAAHHNSQQAGGGGGGGGSGGP----------------------------------------
-----------------------------------------------------------------------------------------SGGPGGMSGMPNGVKGMSGQGS-
AAAAAALGMLGHPGGYPGLYAAAAAGGGGPRSKEELMMLGLEVSGDYGMLKHPSVVASPSVSSPDSHNSDSSVEVS-VRGNPLLHLGAKPNAGGGSASAPGGDAN----
QSTSTAPGRPPQLQQRKDLPTFHLPLPFPGLGSMSVMPPPPFLPPSHLLFPGYHPALYQHHQGLLKPTPEQQQQQAAVAAAAVQHLFNSSGAAGQRGFA---------
PTAASSPFGNHHPLEEAPPGHPHNHSSSNTNNSQSQHQNQNQHQHHILANGGKGVSAADELTKRFYLDAVLKSQQHGSPPPMATKLAPLQHTKQDYSISALVTPNSESGGGGRVKSRQSNGGQGEDAAR
GDTSDKEENVDETEDEHDEEEDEEEDLVVSMTPPHSPAQQLQDQDHDHDQDQDLALSPSVAQDNPIDLSMKT
Tcas_KNI    MNQTCKVCGEPAAGFHFGAFTCEGCKSFFGRSYN--NLSSISECKNNGECVINKKNRTACKACRLRKCLMVGMSKSGSRYGRRSNWFKIHCLLQEQQQ-----------
QQQAGGGG-------------------------------------------------------------------------------------------------------------------------
--------------------------IQTGLKPPQKTPSPHQHLGLGLLNQPFGHPLMHHITK-------TKEELMLLGLDEYKNS---------ASPSVSSPESHNSDSSVEISDARRIPLFPG----
-------------------------------------------------------LLPPTFLPPHGLLFPPGYPPLYP---GLLQPTNNNNR------------LMRN---------------------
---------------------------------------------------HNPGVEAFNKRVFLDAVLQSQRSPTPEVEAPPP---------------------------------------------
------------------------------------------------------VASPLQEDPIDLSMKT
Amel_KNRL   MNQQCKVCGEPAAGFHFGAFTCEGCKSFFGRSYN--NLSSISECKNGGECVINKKNRTACKACRLRKCLLVGMSKSGSRYGRRSNWFKIHCLLQEQQQ-------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------QQQQQQQHYQQTLSNQPLTPQRKPMSPPIGLHPGQRKDDVLMLGLDDYKNST---------
SPSISSPESHNSDSSVEISERRAAFSGHPGFRPLHSQLQPSVADLSALS-----------------------
KEMMSLPLGFHLGGMSLIPPAFLPPPSLTMFSPYHLYATSHGASHPLVPNHSSNLLRHSPVIEDTSSTISTATTTTTTT---------ISIDSKDSCGNNNNNNNNNNNNNDRYAHNN-----------
---NVQTRQSSNESEIGETYTKRFYLDAVLKSQRTQEGSTAASTKE-----------------------------------------------------------------------------------
HSENESPICSPGLETDCPPPQDNPIDLSIKS
Ofas_KNI1   MNQQCKVCGEPAAGFHFGAFTCEGCKSFFGRSYN--NLGSISECKNNGECVINKKNRTSCKACRLRKCLLVGMSKSGSRYGRRSNWFKIHCLLQEQKQ-------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------KDGSEQKLQMQIASIN------------------RSKDELLLLGLDEYKQAS---------SPSVSSPDSISDSSDSPYRSHLKELSGG-------------
------------------------------------------YLPYMFPHPAYLPPLYPSVAGSYFLPGG--------------------------------------------------------
RCYDDNFNHRIRS--------------------------------------------------------------------------------------------------------------------
----------------------------------------------------PSPDSPIDLSLKG
Mmus_Esr1   ETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQG---
HNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDLEGRNEMGASGDMRAANLWPS--------------------------------------------------
--------------------------------------------------------------------------------------------------
PLVIKHTKKNSPALSLTADQMVSALLDAEPPMIYSEYDPSRPFSEASMMGLLTNLADRELVHMINWAKRVPGFGDLNLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDR-----------
--------------------------------NQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSL----------------------------------------
---------------EEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHRRLAQLLLILSH----------------------------------------------------------------------
--------------------------IRHMSNKGMEHLYNMKC
Mmus_Esr2   DAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQG---HNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRS-----------
ASEQVHCLNKAKR--------------------------------------------------------------------------------------------------------------------
--------------------------------TSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVLVSR-
PSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRS-------------------------
SEDPHWHVAQTKSAVPRDEGKCVEGILEIFDMLLATTARFRELKLQHKEYLCVKAMILLNSSMYPLATASQEAE-------------------------------------------------------
SSR-KLTHLLNAVTDALVWVISKSGISSQQQSVRLANLLMLLSH-------------------------------------------------------------------------------------
-----------VRHISNKGMEHLLSMKC
Dmel_ERR    LRRLCLVCGDVASGFHYGVASCEACKAFFKRTIQG---NIEYTCPANNECEINKRRRKACQACRFQKCLLMGMLKEGVRLDRVRGGRQKYRRNPVSNSYQT----------------
---------------------------------------------------------------------------------------------------------------------------------
------------------------
MQLLYQSNTTSLCDVKILEVLNSYEPDALSVQTPPPQVHTTSITNDEASSSSGSIKLESSVVTPNGTCIFQNNNNNDPNEILSVLSDIYDKELVSVIGWAKQIPG------------------------
-------------------------FIDLPLNDQMKLLQVSWAEILTLQLTFRSLPFNGKLCFATDVWMDEHLAKECGYTEFYYHC-------------------------------------------
-----------VQIAQRMERISPRREEYYLLKALLLANCDILLDDQSSLRAFR----------------------------------------------------------------------------
----------------------DTILNSLNDVVYLLRH
```

# [PhD_A6_phylogeny_database]

## [PhD_A6.3_phylogeny_ClustalW2_alignments]

### “alignment_paired-Pax-group.txt”

```
Alignment of the complete protein sequences of various paired/Pax-group genes from Parhyale hawaiensis, Drosophila melanogaster
and mouse.
This was done using the ClustalW2 algorithm (Larkin et al., 2007).
For details and references regarding the species' protein sequences, see Appendix (A7).

Dmel_TOY    -----------------------------------------------------------------------------------------------------------
MMLTTEHIMHGHPHSS---------------------------------------VGQSTLFGCSTAGHSGINQLGGVYVNGRPLPDSTRQKIVELAHSGARPCDISRILQ--------------
VSNGCVSKILGRYYETGSIKPRAIGGSKPR---VATTPVVQKIADYKRECPSIFAWEIRDRLLSEQVCNSDNIPSVSSINRVLRNLASQKEQQAQ----------------------------------
-----------------------------QQNESVYEKLRMFNGQTGG----------------------------------------------------
WAWYPSNTTTAHLTLPPAASVVTSPANLS------------------GQADRDDVQKRELQFSVEVSHTNSHDSTS-DGNSEHNSSG----DEDSQMRLRLKRKLQRNRTSFSNEQIDSLEK-------
----------------------EFERTHYPDVFARERLADKIGLPEARIQVWFSNRRAKWRREEKMRTQRRSADTVDGSGRTSTA-------NNPSGTTASSSVATSNNSTPGI---------------
------------VNSAINVAERTSSALVSNSLPEASN---------------------GPTVLGGEANTTHTSSESPPL-----QPAAPRLPLNSG---------------------------------
--FNTMYSSIPQPIATMAENYNSSL----------GSMTPSCLQQRD---AYPYMFHDPLSLGSPYVSA-----------HHRNTACN-----------------------------------------
------------PSAAHQQPPQHGVYTNSSPMPSS-------------NTGVISAGVSVPVQISTQNV----------SDLTGSNYWPRLQ------------
Mmus_Pax6   ---------------------------------------------------------------------------------------------------------------------
------------------------------------------------------
MQNSHSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQTHADAKVQVLDNENVSNGCVSKILGRYYETGSIRPRAIGGSKPR---
VATPEVVSKIAQYKRECPSIFAWEIRDRLLSEGVCTNDNIPSVSSINRVLRNLASEKQQMG------------------------------------------------------------------
ADGMYDKLRMLNGQTGS----------------------------------------------------WGTRPG--------WYPGTSVP-------------------------GQPTQDGCQQQ--
----EGGGENTNSISS-NGED----------SDEAQMRLQLKRKLQRNRTSFTQEQIEALEK-----------------------------
EFERTHYPDVFARERLAAKIDLPEARIQVWFSNRRAKWRREEKLRNQRRQAS------------------NTPSHIPISSSFSTS----------------------------------
VYQPIPQPTTPVSSFTSG------------------------------SMLG-RTDTALTNTYS-------------ALP-----------------------------------------------
PMPSFTMANNLP---------------MQPPVPSQTS---SYSCMLPT-----SPSVNG-----------RSYDTYT----------------------------------------------------
---------PPHMQTHMNSQPMGTSGT----------TSTGLISPGVSVPVQVPG-------------SEPDMSQYWPRLQ------------
Dmel_EY     -----------------------------------------------------------------------------------
MRNLPCLGTAGGSGLGGIAGKPSPTMEAVEASTASHPHSTS------------------------------------
SYFATTYYHLTDDECHSGVNQLGGVFVGGRPLPDSTRQKIVELAHSGARPCDISRILQ--------------VSNGCVSKILGRYYETGSIRPRAIGGSKPR---
VATAEVVSKISQYKRECPSIFAWEIRDRLLQENVCTNDNIPSVSSINRVLRNLAAQKEQQSTGSGSSSTSAGNSISAKVSVSIGGNVSNVASGSRGTLSSSTDLMQTATPLNSSESGGASNSGEGSEQE
AIYEKLRLLNTQHAAGPGPLEPARAAPLVGQSPNHLGTRSSHPQLVHGNHQALQQHQQQSWPPRHYSGSWYPTSLSEIPISSAPNIASVTAYASGPSLAHSLSPPNDIESLASIGHQRNCPVATEDIHL
KKELDGHQSDETGSGEGENSNGGASNIGNTEDDQARLILKRKLQRNRTSFTNDQIDSLEK-----------------------------
EFERTHYPDVFARERLAGKIGLPEARIQVWFSNRRAKWRREEKLRNQRRTPNSTGASATSSSTSATASLTDSPNSLSACSSLLSGSAGGPSVSTINGLSSPSTLSTNVNAPTLGAGIDSSESPTPIPHI
RPSCTSDNDNGRQSEDCRRVCSP-------------CPLGVGGHQNTHHIQSNGHAQGHALVPAISPRLNFNSGS----------------------------------
FGAMYSNMHHTALSMSDSYGAVTPIPSFNHSAVGPLAPPSPIPQQGDLTPSSLYPCHMTLRPPPMAP-----------
AHHHIVPGDGGRPAGVGLGSGQSANLGASCSGSGYEVLSAYALPPPPMASSSAADSSFSAASSASANVTPHHTIAQESCPSPCSSASHFGVAHSSGFSSDPISPAVSSYAHMSYNYASSANTMTPSSAS
GTSAHVAPGKQQFFASCFYSPWV
Phaw_PBY1A  ----------------------------------------------------------------------------------------
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPST------------------------------------
ANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALK--------------
VSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTN--------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------TDSDSEE----------DTEPGLTIRRKQRRSRTSFTAIQLEILEQ-----------------------------
EFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKAS-----NSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFN------------------------
RYGLQAPASDAVVPPVKIEHSP-------------------------------SSSCEDSSSSPNQPSYHP------------------PAYQGTSLL--------------------------
YPRQQVDGFGFNYNTPYNPAVHPGYPFGDPGYPGQMYLGHVAAQQAFQAPGRSGQSEVQTPM------------SGDRPSAGQELNQRHRSRDSLQFPGPQST------------------
YGYPQPLSNNTSDVTATRLSPPCTSYAGASEQQQQGPVTTSAANFWQPSNLKPEPGTSSLGSSSFATAANGGSFCENQTGTSSSADQQRPMSVGLA-----------
Phaw_PBY1B  ----------------------------------------------------------------------------------------
MECMYSFRPQPFGPMGVSVPFPSASAFPPTAPLPST------------------------------------
ANLDHENNDFDNAHANSRVNQLGGKFVNGRPLPSNVRREIIALAAQGIRPCNISKALK--------------
VSHGCVSKILNRYQETGSIKPGVLAGVARSLPPRLDERIVQRIWSFNNCHGVMTLEQLRERLVAEFTGTPVNVPPLPVLGKGLRPYRDQLPFLECTN--------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------TDSDSEE----------DTEPGLTIRRKQRRSRTSFTAIQLEILEQ-----------------------------
EFCSNQYPDITTREHLAERTGLTEARVQVWFSNRRARQRKAS-----NSPEEAAQRPANNRNGGSPVAAISPFDPAVPGPSHGLDGGYGMFN------------------------
RYGLQAPASDAVVPPVKIEHSP-------------------------------SSSCEDSSSSPNQPSYHP------------------PAYQ------------------------------------
-DGPGFLFPMQGQAAVLN-MSFSD----------------------------------------------GDEPEN---------------YSG-----------------------------------
------NPPEASY---------------------------------------------------------------------------------
Dmel_PRDA   ---------------------------------------------------------------------------------------------------------------------
------------------------------------------------------MNSGQGRVNQLGGVFINGRPLPNNIRLKIVEMAADGIRPCVISRQLR--------------
VSHGCVSKILNRYQETGSIRPGVIGGSKPR---IATPEIENRIEEYKRSSPGMFSWEIREKLIREGVCDRSTAPSVSAISRLVRGRDAPLDNDMSSA--------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------SGSPAGDGTKASSSCGSDVSGGHHNNGKPSDEDISDC----------ESEPGIALKRKQRRCRTTFSASQLDELER-----------------------------
AFERTQYPDIYTREELAQRTNLTEARIQVWFSNRRARLRKQHTSVSGGAPGGAAASVSHVAASSSLPSVVSSVPSMAP-LAMMPGSLDPATVY---------------
QQQYDFYGSHANISVSAAAPMASSNLSPGITTTPPHHHQFYNPSANTASYIMPGENGNTTPTGNIIVSSYETQLGSVYG-----------------
TETETHQTMPRNESPNESVSSAFGQLPPTPNSLSAVVSGAGVTSSSGANSGADPSQSLANASAGSEELSAALKVESVDLIAASQSQLYGGWSSMQALRPNAPLSPEDSLNSTSSTSQALDVTAHQMFHP
YQHTPQYASY------------------PAPGHAHSHHGHPHAPHPHAHPHPQYAGAHPHYPPPSSSAHFMPQNFNAAAFPSPSKVNYTTMPPQPFYPSWY----------------------------
------
Dmel_PRDB   ----------------------------------------------------------------------------------------------------------
MTVTAFAAAMHRP------------------------------------------FFNGYSTMQDMNSGQGRVNQLGGVFINGRPLPNNIRLKIVEMAADGIRPCVISRQLR--------------
VSHGCVSKILNRYQETGSIRPGVIGGSKPR---IATPEIENRIEEYKRSSPGMFSWEIREKLIREGVCDRSTAPSVSAISRLVRGRDAPLDNDMSSA--------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------SGSPAGDGTKASSSCGSDVSGGHHNNGKPSDEDISDC----------ESEPGIALKRKQRRCRTTFSASQLDELER-----------------------------
AFERTQYPDIYTREELAQRTNLTEARIQVWFSNRRARLRKQHTSVSGGAPGGAAASVSHVAASSSLPSVVSSVPSMAP-LAMMPGSLDPATVY---------------
QQQYDFYGSHANISVSAAAPMASSNLSPGITTTPPHHHQFYNPSANTASYIMPGENGNTTPTGNIIVSSYETQLGSVYG-----------------
TETETHQTMPRNESPNESVSSAFGQLPPTPNSLSAVVSGAGVTSSSGANSGADPSQSLANASAGSEELSAALKVESVDLIAASQSQLYGGWSSMQALRPNAPLSPEDSLNSTSSTSQALDVTAHQMFHP
YQHTPQYASY------------------PAPGHAHSHHGHPHAPHPHAHPHPQYAGAHPHYPPPSSSAHFMPQNFNAAAFPSPSKVNYTTMPPQPFYPSWY----------------------------
------
Dmel_GSB-N  ------------------------------------------------------------------------------------------------------------
MDMSSANSLRP------------------------------------------LFAGYP-----FQGQGRVNQLGGVFINGRPLPNHIRLKIVEMAASGVRPCVISRQLR--------------
```

```
VSHGCVSKILNRYQETGSIRPGVIGGSKPK---VTSPEIETRIDELRKENPSIFSWEIREKLIKEGFAD---PPSTSSISRLLRGSDR------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------
------------------GS--EDGRKDYTING--ILGGR-------DSDISDT----------ESEPGIPLKRKQRRSRTTFTAEQLEALER------------------------------
AFSRTQYPDVYTREELAQTTALTEARIQVWFSNRRARLRKH----SGGSNSGLSPMNS---GSSNVGVGVGLSGATAP-LGYGPLG-----------------------------
VGSMAGYSPAPGTTATGAGMNDGV------HHAAHAPSS------------HHSAATAAAAAHHHTQMGGYDL------------------VQSAAQHGFP----------GGFAQP-------------
-GHFGSQNYYHQDYSKLTIDDFSK--------LTADSVSKISPS-LHLSDNYSKLEAP---------------SNWSQAA----------YHAAANYN--------------------------------
----AHVAQHQLNDYAAAAAHGNPASAYSHPLPTQGQAKYWS-------------------------------------------------------
Phaw_PBY2   -------------------------------------------------------------------------------------------------------------------
MTIRP------------------------------------------IFTGYP-----FQGQGRVNQLGGVFINGRPLPNHIRLKIVEMAAAGVRPCVISRQLR---------------
VSHGCVSKILNRYQETGSIRPGVIGGSKPK---ISAPDIEKKIDDYKKENPGAFSWEIRDKLVKDGIVDKTDAPSISSISRIIRGGGGGGGGGGGGVG--------------------------------
----------------------------------------------------------------------------------------------------------------------------------
-----------------TGRREDDPRKDHSIDG--ILAEDG-------SDVSDT----------ESEPGLTLKRKQRRSRTTFTAEQLEILER------------------------------
SFEKTQYPDVYTREELAQKARLTEARVQVWFSNRRARLRKH------LNSQQLATLTSPMAPSAAAAAAAASAAAAAPYMAHHQYP---------------------------------
AQMCDPSSAMTASAAYHAGATSWQAGAATCYDYSASTLS----GVDHSSLMSHTMASMPYAAMGGAPDGG------------WG--AAAATKS-----------------------------------
AMGSWCATGMAAPTGFEQLGGQLTGQLGSLLPAQSYNVSAAHATFN-ASVSSPSVP--------------TSSPDKPQ----------TQAATASSV-------------------------------
NGASSSGQASSVTPHTQAAAVHAAHAVAASSYPSYFSQFADPSSIMCGRVH----------------------------------------------
Dmel_GSB    --------------------------------------------------------------------------------------------------------------
MAVSALNMTP------------------------------------------YFGGYP-----FQGQGRVNQLGGVFINGRPLPNHIRRQIVEMAAAGVRPCVISRQLR---------------
VSHGCVSKILNRFQETGSIRPGVIGGSKPR---VATPDIESRIEELKQSQPGIFSWEIRAKLIEAGVCDKQNAPSVSSISRLLRGSSGSG----------------------------------------
----------------------------------------------------------------------------------------------------------------------------------
--------------------------TSHSIDG--ILGGGAGS--VGSEDESED----------DAEPSVQLKRKQRRSRTTFSNDQIDALER------------------------------
IFARTQYPDVYTREELAQSTGLTEARVQVWFSNRRARLRKQ------LNTQQVPSFAP---TSTSFGATPTTSAAPAPNMGMSLYS---------------------------------
SQSWPSSGAYENHAAYGG-----------------------------------SVASMSPASS------------------------TSGTSS-----------------------------------
AAHSPVQTQAQQPG------------TGSEFMTSTYGVGSSNATYPSAAYSMPQTP--------------ATSAEQLR----------SQFASAA-----------------------------------
ASGSHHPSTWDSYNFAGSFFPPASAAGNHISGYHHQVDQKSSMMTTAPTYPYFGF----------------------------------------
Mmus_Pax7   ----------------------------------------------------------------------------------------------------------
MAALPGAVPRMMRPGPGQN------------------------------------YPRTGFPLEVSTPLGQGRVNQLGGVFINGRPLPNHIRHKIVEMAHHGIRPCVISRQLR---------------
VSHGCVSKILCRYQETGSIRPGAIGGSKPRQ--VATPDVEKKIEEYKRENPGMFSWEIRDRLLKDGHCDRSTVPSVSSISRVLRIKFGKKEDDEEGD---------------------------------
----------------------------------------------------------------------------------------------------------------------------------
-----------------KKEEDGEKKAKHSIDG--ILGDKGNR----LDEGSDV----------ESEPDLPLKRKQRRSRTTFTAEQLEELEK------------------------------
AFERTHYPDIYTREELAQRTKLTEARVQVWFSNRRARWRKQ------AGANQLAAFNHLLPGGFPPTGMPTLPPYQLPDSTYPTTTISQ-----------------------------
DGGSTVHRPQPLPPSTMHQGG------------------------------LAAAAAAADTSSAYGARHS-----------------FSSYSDS-----------------------------------
FMNPGAPSNHMNP-VSNGLSPQVMSILSNPSAVPPQPQADFSISPLHGGLDSASSIS-------------ASCSQRA-----------DSIKPGDS----------------------------------
-LPTSQSYCPPTYSTTGYSVDPVAGYQYSQYGQTAVDYLAKNVSLSTQRRMKLGEHSAVLGLLPVETGQAY------------------------
Mmus_Pax3   ----------------------------------------------------------------------------------------------------------
MTTLAGAVPRMMRPGPGQN------------------------------------YPRSGFPLEVSTPLGQGRVNQLGGVFINGRPLPNHIRHKIVEMAHHGIRPCVISRQLR---------------
VSHGCVSKILCRYQETGSIRPGAIGGSKPKQ--VTTPDVEKKIEEYKRENPGMFSWEIRDKLLKDAVCDRNTVPSVSSISRILRSKFGKGEEEEADLER-------------------------------
----------------------------------------------------------------------------------------------------------------------------------
-----------------KEAEESEKKAKHSIDG--ILSERASAP--QSDEGSDI----------DSEPDLPLKRKQRRSRTTFTAEQLEELER------------------------------
AFERTHYPDIYTREELAQRAKLTEARVQVWFSNRRARWRKQ------AGANQLMAFNHLIPGGFPPTAMPTLPTYQLSETSYQPTSIPQAV--------------------------
SDPSSTVHRPQPLPPSTVHQST------------------------------IPSNADSSSAYCLPSTRHG-----------------FSSYTDS-----------------------------------
FVPPSGPSNPMNPTIGNGLSPQVMGLLTNHGGVPHQPQTDYALSPLTGGLEPTTTVS-------------ASCSQRL-----------EHMKNVDS----------------------------------
-LPTSQPYCPPTYSTAGYSMDPVTGYQYGQYGQ------SKPWTF--------------------------------------------------
Dmel_SV
MLIMDIQTSTHHIHGLGTHELQHRILHPHILHSTQEETLNTSTGQLEHDSQHLQQHHLTHHQQQDVSPTLHNLQNTRTGDSLSTTINTNQNQHGHQHLSGSNMYTSSQMEDKSKANKYDEYSSRTLSNI
SDANTTPSANNFITQSQGIEWITAMNDIQNGAEDSHSSQGSISGDGHGGVNQLGGVFVNGRPLPDVVRQRIVELAHNGVRPCDISRQLR---------------
VSHGCVSKILSRYYETGSFKAGVIGGSKPK---VATPPVVDAIANYKRENPTMFAWEIRDRLLAEAICSQDNVPSVSSINRIVRNKAAEKAKHVHHHQQ-------------------------------
----------------------------------------------------------------------------------------------------------------------------
HHVSQSLGGGHIATESVDSSTGTIGEPQPPTSNSSANSVNTNVSASASVHASIPTSGTDSVQVSVGHINANSNETTHINSTAEQRTTGYSINGILGIQHGHHSHNNNNNSSVNNNNNTESSCKRKRIEA
HDENHDTNIHSDNDDGKRQRMSTYSGDQLYTNIWSGKWCIKDDHKLLAELGNLTASTGNCPATYYEASNGFSTTPISGSGATASGNDTSMLYD-
SITTISQTQSSIYTPAIGPSIGTGSLTPLVPISMHEMKLSANSIQEQTVPPFYTALAF------------DGNYTSMTSLENCSSLVGQEHIVMPESSDSNTLCPISTRIVPDITETNSTR--------
VKEPLTNSDGCSEDNNKEPEKSNSSQSSDHIASPHLHHIGEEQLRGNRRSNLNASTHPSSLIPLQPSGGSSLLNINPSENRSDVELNLSNNVGLYSTPTVLPSFNHYSAGCSSVVPGSD----------
-----------------------YAYNP-AYTQYGGAYGSYGYGTGSGLINSSYYYESGQTQSPLTHDLRSPLVATRANSLASAASPGSGSACTKSESSDIFLA------------
Mmus_Pax2   -------------------------------------------------------------------------------------------------------------
MDMHCKADPFSAMHR----------------------------------------------------HGGVNQLGGVFVNGRPLPDVVRQRIVELAHQGVRPCDISRQLR---------------
VSHGCVSKILGRYYETGSIKPGVIGGSKPK---VATPKVVDKIAEYKRQNPTMFAWEIRDRLLAEGICDNDTVPSVSSINRIIRTK--------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------
-----------------VQQPFHPTPDGAG----TGVTAPG--HTIVPST-----------------ASPPVSSASNDPVGSYSINGILGIPR-------------------------------------
----SNGEKRKREEVEVYTDP---AHIRGG-----GGLHLVWTLRDVSE------------------------GSVPNGDSQS--------------------------GVDSLR----------
KHLRADTFTQQQLEALDRVFER------------PSYPDVFQASEHIKSEQGNEYSLP--------------ALTPGLDEVKSS--------------------LSASANPELGSNVSGTQTYPVVT---
------------GRDMTSTTLPGYPPHVPPTGQ-----------------------GSYPTSTLAG-----------MVPGSE----------------------------------
FSGNPYSHPQYTAYNEAWRFSN-PALLSSPYYYS----------------AAPRGSAPAAAAAAYDRH-------------------------
```

[PhD_A6_phylogeny_database]

[PhD_A6.3_phylogeny_ClustalW2_alignments]

## **“alignment_sine-oculis-Six-family.txt”**

```
Alignment of the entire protein sequences of various sine oculis/Six-group genes from different species.
This was done using the ClustalW2 algorithm (Larkin et al., 2007).
For details and references regarding the species' protein sequences, see Appendix (A7).

Phaw_SIX3        ----------------------MALLAG-----------LSAYSAALLGPAPHGGAPGPVLPS-------------------------------------------------
---------------------------------------------------------------------------------------------SNVSLPSLSFTATQVAAVCETLE----
ESGDIERLGRFLWSLPVAHP----
HLKDLNKHEAVIRARALVAFHLGNFRELYSLIECNRFTRASYPRLQALWLESHYQEAERLRGRPLGPVDKYRVRKKFPFPRTIWDGEQKSHCFKERTRSLLREAYLQDPYPNPSRKRELAQATTLTPTQ
VGNWFKNRRQRDRAAAIKN---------------RLHQQ-------
SSSISSSAKLETGPRSPGADDDEDDLINPGSPTPFEDDVSEDEVSLGQQSPLSKLGQPPLAAFDGVITGDPPQHGGSKGPLDDLSKDAD-----
TALKCSSAGSSFSLSEGASSDTSNSAFSSPYPPTSLAAGTSWSTKSSSQE-------IFLPASPVPLPPDRLFRI-----------
NPDGDHSALRLPLPGALFSPFASALPPRLPLAFPPHLSTWSQYAPFPTISTPGDSTYHRSFPAPLSIDSLISPARNKREAPPRSPSPHTVTSKTEPEDPGIGADL--
TTAQGDGPSLRSPTQSPVAKRLHTSPGDTSISDQQPRPHIFQPIKLEFAPSTETS--------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
--------------
Nvec_SIX3        ----------------------------------------------MFAP--------------------------------------------------------------
-------------------------------------------------------------------------------------------------LPALSFSAHQIAQVCETLE----
ESGDVERLARFLWSLPVAPG----
TLEALGKHESVLRARAIVAFHMGNFRDLYHILETHRFTRESHAKLQAMWLEAHYQEAERLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPTKKRELAQATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------R----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------
Drer_Six6b       ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------MFQLPILNFSPQQVAGVCETLE----
ESGDIERLGRFLWSLPVAPA----
ACEVLNRNESVLRARAIVAFHTGNFRELYHILENHKFTKESHSKLQALWLESHYQEAEKLRGRPLGPVDKYRVRKKFPLPKTIWDGEQKTHCFKERTRHLLREWYLQDPYPNPSKKRELAQATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------RLQQQ-------VLSNGS--------------------------------------------------------------------------
VRSLSGE--------------DGAVDRLGNASSPEASLSSK---------AAASAISITSSDSE-------CDI-------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
Drer_Six6a       ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------MFQLPILNFSPQQVAGVCETLE----
ESGDVERLGRFLWSLPVAPS----
ACDVLGKNESVLRARAVVAFHAGNFRELYHILENHKFTKDSHAKLQALWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKTHCFKERTRHLLREWYLQDPYPNPSKKRELAQATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------RLQQQ-------VLSGGS--------------------------------------------------------------------------
VRSLGDD--------------DTTVDRLGPASSPEVSLSSK---------AATSAISITSSDSE-------CDI-------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
Mmus_Six6        ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------MFQLPILNFSPQQVAGVCETLE----
ESGDVERLGRFLWSLPVAPA----
ACEALNKNESVLRARAIVAFHGGNYRELYHILENHKFTKESHAKLQALWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKTHCFKERTRHLLREWYLQDPYPNPSKKRELAQATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------RLQQQ-------VLSQGP--------------------------------------------------------------------------
GRVLRSEG-------------EGTPEVLGVASSPAASLSSK---------AATSAISITSSDSE-------CDI-------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
Drer_Six3a       ----------------------MVFRSP-----------LELYPSHFFLPNFADR---PLLLASSAP---------------------------------------------
-------------------------------------------------------------------------------------STRSPEDLSMFQLPTLNFSPEQVASVCETLE----
ETGDIERLGRFLWSLPVAPG----
ACEAINKHESILRARAVVAFHTGNFRDLYHILENHKFTKDSHGKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPSKKRELAQATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------RLQHQ-------AIGQNG--------------------------------------------------------------------------
MRSLSESG----------CAPRSSAESPSTAASPTTSVSSMTE----RVDTGTSILSVTSSDSE-------CDV-------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
Mmus_Six3        ----------------------MVFRSP-----------LDLYSSHFLLPNFADSHHCSLLLASSGGGSGASGGGGGAGGGGGGNRAGGGGAGGAGG---------------
-------------------------------------------------------------------------------GSGGGGSRAPPEELSMFQLPTLNFSPEQVASVCETLE----
ETGDIERLGRFLWSLPVAPG----
ACEAINKHESILRARAVVAFHTGNFRDLYHILENHKFTKESHGKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPSKKRELAQATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------RLQHQ-------AIGPSG--------------------------------------------------------------------------
MRSLAEPG----------CPTHGSAESPSTAASPTTSVSSLTE----RADTGTSILSVTSSDSE-------CDV-------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
Drer_Six3b       ----------------------MVFRSP-----------LELYPSHLFLPNFADR---PLLLAGSIP---------------------------------------------
-------------------------------------------------------------------------------------RARSPEDLPMFQLPTLNFSAEQVASVCETLE----
ETGDIERLGRFLWSLPVAPG----
ACDAINKHESIQRARAVVAYHTGSFRELYHILETHKFTKDSHGKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKTHCFKERTRGLLREWYLQDPYPNPSKKRELAQATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------RLQHH-------GLGQSS--------------------------------------------------------------------------
LRSMSESG----------CTPHSSAESPCAAASPTTSVSSMNE----RGDGGT-ILSVTDSDSD-------FDV-------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
```

```
---------------------------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
Agam_OPTIX       -----------------------------------------PVQPHP-----------ILAPS-------------------------------------------------
---------------------------------------------------------------------------------------------PLFALPTLNFTASQVATVCETLE----
ESGDIERLARFLWSLPVAHP----
NVTELDRSEAVLRARAIVAYHTGHFRELYSILERHKFTKTSHGKLQAMWLEAHYHEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPTKKRELAQATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------R----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------
Amel_SIX3        ---------------------------------------MSAVAPAAGTGATAPSANGPIVPA-------------------------------------------------
---------------------------------------------------------------------------------------------PIFALPTLSFTVNQVATVCETLE----
ESGDIERLARFLWSLPVAHP----
NIQELNQSEAVLRARAIVAFHSGHYRELYAILERHKFTKDSHGKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPGKKRELAAATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------RMQQQ------SGTGNGNT-------------------------------------------------------------------------
RRALSPGPGGSDSEDSDISLGGTSPASPSSSPPPTHQPSPVPLGPLGPLPPPSLNFRFDPSQSQ-------
FRFNHATAFKFPPTGFRFGPHSPQHNHPNISGGGIFRLTETSPFQAVGPRGDLGSRLPDPLGLPPPRLHPHEGLLHHPHVQHQLPERISDGSRLSDGGISRLSDSLSEAHSPPL-
MATNLSVTGPLRVAVPSPH--TPSSRGSPPPPPASQP-TPQGQSQGQGLIRVATPPPPNPKPPQIHRPFSPA---------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------
Tcas_OPTIX       ----------------------MALGLGSGLSSGPGSGNHSETSSSPGISVPIPVAPAPLVPN-------------------------------------------------
---------------------------------------------------------------------------------------------PMFALPTLNFTVAQVAAVCETLE----
ESGDIERLARFLWSLPVAHP----
NIGELNKNEAVLRARAIVCFHSGNFRELYSILESNKFSKSSHHKLQEIWQLAHYQEAEKLRGRALGPVDKYRVRKKFPWPRTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPTKKRELAQATGLTPTQ
VGNWFKNRRQRDRAAAAKN---------------RMQQQQLAAQLAHHGGGGR-------------------------------------------------------------------------
QPPLSPTP--SDS-DSDISLGAHSP--PISSPGPMRFSNS---------PTPITSFRFGP---------------HSPG-PMPP-QFRFGNHSP----PGFRQDRKSR--NSSPIN-
VDTTSNYNNSAVN------LRLTSNESPIDVDSTQEKSHSPIRVDSSPPPMNLRIHSDNTKIETPLALRLTSSHQLTTPLRIQVASPH--RMAAPPSPPHTSLHLG-TPHG-----
GIVRIAPPSPSNP--PVLHRPFSPPRLT-----------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------
Dmel_OPTIX       ----------------------MAVG--------------PTEGKQPPSESFSPTHHQIIAPS-------------------------------------------------
---------------------------------------------------------------------------------------------PILAVPTLAFSAAQVEIVCKTLE----
DSGDIERLARFLWSLPVALP----
NMHEILNCEAVLRARAVVAYHVGNFRELYAIIENHKFTKASYGKLQAMWLEAHYIEAEKLRGRSLGPVDKYRVRKKFPLPPTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPTKKRELAKATGLNPTQ
VGNWFKNRRQRDRAAAAKN---------------RIQHS---QNSSGMGCRSRR-----------------------------------------------------------------------
ADGAASPTPSDSSDSDISLGTHSPVPSSLQLQHSPGSTSNGANDREESLSVDDDKPRDLSGSLPLPLSLPLPLASPTHTPPQLPPG---
YGGGAGAGPGGPLTGPGCLPPFKLDAATSLFSAGCYLQSFSNLKEMSQQFPIQPIVLRPH---PQLPQSLALNGASGGPP-LHHPAYAAAYSVECVPGGHGPPHPPPKLRINSPEKLNSTAVAAAAS--
-VGGGGGN--------QHHEPTTTGYHHSGQLMLHRPFSTSPELKHSAPEIT-----------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
----
Dpse_OPTIX       ----------------------MAVG--------------PTEGKQPPSESFSPTHHQIIAPS-------------------------------------------------
---------------------------------------------------------------------------------------------PILAVPTLAFSAAQVEIVCKTLE----
DSGDIERLARFLWSLPVALP----
NMHEILNCEAVLRARAVVAYHVGNFRELYAIIENHKFTKASYGKLQAMWLEAHYIEAEKLRGRSLGPVDKYRVRKKFPLPPTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPTKKRELAKATGLNPTQ
VGNWFKNRRQRDRAAAAKN---------------RLQHN---QNS-GMGCRSRR-----------------------------------------------------------------------
ADGAASPTPSDSSDSDISLGTHSPVPSSLQLQHSPGSTSNGANDRDESLSVDDDKPRDLSSSLPL--------
ASPGLTPPQLPHGGVVFGGAGGGGGGPPLPGPGCLPPFKLDAATSLFSAGCYLQSFSNLKEISQQFPIQPIVLRPHPQLPQLPQPLALNGGSGGPPHLHHPAYAAY--
VECVPSHAQQHPPPPKLRINSPEKLNSTAVAAAAAGAGSGGGGANPGAV--NHHHHEATTTGYHHSGQLLLHRPFSTSPELKHSAPEIT----------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------
Dvir_OPTIX       ----------------------MAVG--------------PTEGKQPPSETFSPTHHQIIAPS-------------------------------------------------
---------------------------------------------------------------------------------------------PILAVPTLAFSAAQVEIVCKTLE----
DSGDIERLARFLWSLPVALP----
NMHEILNCEAVLRARAVVAYHVGNFRELYAIIENHKFTKASYGKLQAMWLEAHYIEAEKLRGRSLGPVDKYRVRKKFPLPPTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPTKKRELAKATGLNPTQ
VGNWFKNRRQRDRAAAAKN---------------RIQHN---QNNSGMGCRSRR-----------------------------------------------------------------------
ADGAASPTPSDSSDSDISLGTHSPVPSSLQLQHSPGSMSNGAN--EESLSVDDDKPRDLSSALPS------------PTPPQLPSA---YGAAAGAGAGG-LP-
AGCLPPFKLDAASSLFNAGCYLQSFSNLKEMSQQFPIQPIVIRPH---PQLPQPLALNGSAGGAPHLHHPAYAAAYSVECG--
VVAPPPPPPKLRINSPEKLNSTAVAAAAAAVAACAGAAHEISSSGNNTTTTTTTTGYHHSAQLLLHRPFSTSPELKHSAPEIT----------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------
Drer_Six7        ----------------------------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------------MFPLPMFTPEQVARVCENLE----
ETGDIERLGRFLWSLPAAVPGS--
AGELLNRHESVMRARALVAFHGGNFEALYQILQSHRFTRESHAKLQDLWLDAHYREAERLRGRPLGPVEKYRIRKKFPLPRTIWDGEQKTHCFKERTRSLLREWYLQDPYPNPSRKRHLAQATGLTPTQ
VGNWFKNRRQRDRAASAKN---------------RLQQD----------------------------------------------------------------------------------------
SSLLPSG--------------SSPECSSSEHNPLLRGSSPR-------HPDSPGNSDCSSGPR------------------------------------------------------------------
-------------------------------GTGASTPDISVSSDSEFES-------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
Cele_H-32        -----------MFTPEQFTKVMSQLGNFSQLGQMFQPGNVAMLQALQANGASSTPSLFPAMPSVIPS---------------------------------------------
-------------------------------------------------------------------------------------------LAAPSSPTTSNLTADQIVKTCEQLE----
TDGDVDGLFRFMCTIPPQKT------
QEVAGNEAFLRARALVCFHASHFRELYAILENNKFSPKYHPKLQEMWHEAHYREQEKNRGKSLCAVDKYRVRKKYPMPRTIWDGEQKTHCFKERTRSLLREWYLKDPYPNPPKKKELANATGLTQMQVG
NWFKNRRQRDRAAAAKN---------------KQNIIG-----VELKKTS---------------------------------------------------------------------------
SDMSDSDDDFEDSMTDSPSPIDEPKDLSKSHIPKLSPTLLP--------KMATPFDMFAAAAN---------------------------------------------------------------
PLMMLNLNPALYMQFHNFFNTMRNPQIDEEENSETTVEVEADIEPPKKRSKLSIDEILNIKSEVSPSQ-----------
CSPCSNESLSPKRAVKTEEVKKEDDEAAEEDSRSVKSETSEDPKHSSPKSTTSQSE-------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------
Cele_H-34        -----------------------------------------MQQSYNPQNS-------------------------------------------------------------
------------------------------------------------------------------------------------------------
LTATTSYSEQEIVCICESLFNEGLQTGRTEQLANFIYNLP----------
QCYQVMESVLKAQALVYFTTQNWKMLYKLLECSKFSPHNHTVLQNLWLDAHYKEAAKTKDRELGAVCKYRIRKKNPFPNTIWDGEETNYCFKSKSRNVLRDAYKKCQYPSVEDKRRLAQQTELSIIQVS
NWFKNKRQRERAAGQLD---------------RSSAR----------------------------------------------------------------------------------------
SNDSDDG-----------------SSGCESKPPMNIDSPAP-------PPLPTSFDLQPYYPS------------------------------------------------------------------
------------------------------------PYTFAPHCDFSYIQNL-----------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
Dmel_SO          ----------------------MLQHPATDFYDLAAANAAAVLTARHTPPYSPTGLSGSVALHNNNNNNSSTS------NNNNSTLDIMAHNGG----GAGGGLHLN-
SSSNGG-----------------------------------------------------------------------GGG------GVVSGG--GSGGRENLPSFGFTQEQVACVCEVLQ----
```

```
QAGNIERLGRFLWSLPQCDKLQ--------
LNESVLKAKAVVAFHRGQYKELYRLLEHHHFSAQNHAKLQALWLKAHYVEAEKLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRSVLRDWYSHNPYPSPREKRDLAEATGLTTTQVSNWFK
NRRQRDRAAEHKD--------GSTDKQHLDSSSDSEMEGSMLPSQSAQHQ-----------QQQQQQQHSPGNSS-------------------------------------------------
GNNNG-----------LHQQQLQHVAAEQGLQHHPHQP-------HPASNIANVAATKS-------------------------------------------------------------
SGGGGGGGVS-----------------------------AAAAAQMQMPPLTAAVAY-----------------------------SHLHSVMGAMPMTA-------MYDMGEYQHL------------
---------------------------------------------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------------
Dpse_SO          ----------------------MLQHPATDFYDLAAANAAAVLTARHTPPYSPTGLSGSVALLNNNNNNNNSSSTGNSNNNNNNNLDIMTHNGV----
VVGGAVGAGGLHLTGG-----------------------------------------------------------------------SNG------GVTGGG--
ASGGRENLPSFGFTQEQVACVCEVLQ----QAGNIERLGRFLWSLPQCDKLQ--------
LNESVLKAKAVVAFHRGQYKELYRLLEHHHFSTQNHAKLQALWLKAHYVEAEKLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRSVLRDWYSHNPYPSPREKRDLAEATGLTTTQVSNWFK
NRRQRDRAAEHKD--------GSTDKQHLDSSSDSEMEGSMLPSQSAQQQ-----------QQQQQQ--SP------------------------------------------------------
NNNG-----------LHQQQLQHAAAEQSLQHHPHQQH------HPAATGPTSVVSKTG----------------------------------------------------
HSGAGGGGGGGGGGGGGG-----------------------------AATATQMQMPPLSAAVAY-----------------------------SHLHSVMGAMPMT--------MYDMGEYQHL----
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------
Dvir_SO          ----------------------MLQHPATDFYDLAAANAAAVLTARHTPPYSPTGLSGSVVALHNNNNNSTSN-------
NNNNNLDMLTHNGGGVGVGLGGSVTATGLHLSGG-----------------------------------------------------------------------
GGGSRNGGPGAGGGGSAATAGRETLPSFGFTQEQVACVCEVLQ----QAGNIERLGRFLWSLPQCDKLQ--------
LNESVLKAKAVVAFHRGQYKELYRLLEHHHFSAQNHAKLQALWLKAHYVEAEKLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRSVLRDWYAHNPYPSPREKRDLAEATGLTTTQVSNWFK
NRRQRDRAAEHKD--------GSTDKQHLDSSSDSEMEGNMLSSQSAQQQHAAATHHSHHSQQQQQQSPSSGNNNN---------------------------------------------
INNNNSNGNNS----NSLHQHQQQLQHVAAEQSLQHHQHQQ-------HAAS---AAAASKS------------------------------------------------------------------
SAAAA-----------------------------AVAVAASQMPPLSAAVAY-----------------------------SHLHSVMGAMPMY----------DMGEYQHL-----------------
---------------------------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
Agam_SO          ----------------------MLQHPSTDYYELSS-----VAIKQTPPTFSPQSGGGGGGVGDGGTGG-------------DADLELIAQNGS------------------
-------------------------------------------------------------------------------------------AADRESLPSFGFTQEQVACVCEVLQ----
QAGNVERLGRFLWSLPQCDKLQ--------
LHESVLKAKAVVAFHRGNFKELYRLLEHHQYSPHNHAKLQALWLKAHYVEAEKLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRSVLRDWYTHNPYPSPREKRELAEATGLTTTQVSNWFK
NRRQRDRAAEHKDT-------GESDKQHLDSSSDSDMEGPSLASMASGNG----------VGGPGGVLHAT------------------------------------------------------
TGGG----------------MHLGTSGGTLTSAASGGV------TAGVLGGGTTVGPG----------------------------------------------------------------
AGGGVGL-----------------------------GSTGKSTSIGSLSAAAAY-----------------------------SHFHNVMGSMPLY----------DIGDYQHL---------------
---------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
Nvec_SO          ----------------------------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------------MLPSFSFTPEQVACVCEVLQ----
QSGDIERLGRFLWSLPECETIQ--------
KNESVLKAKAIVSFHQQNFQELYRILENNNFSPNAHPKLQSLWLKAHYMEAEKLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRNILREWYSHNPYPSPREKRELAEGTGLTTTQVSNWFK
NRRQRDRAAEAKIR-------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Tcas_SO          -----------------------MDEAINRECLKFALRTRESAMLSAGLGYGVSG---PVTGTTP---------------PYSPPP-PARHSP-------------------
-------------------------------------------------------------------------------------------------LPSFGFTQEQVACVCEVLQ----
QAGNVERLGRFLWSLPACDKLH--------
NNESVLKAKAIVAFHRGNFKELYKILESHQFSPHNHAKLQALWLKAHYIEAERLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRSVLRDWYSHNPYPSPREKRELADATGLTTTQVSNWFK
NRRQRDRAAEHK-----------------DSG-----------QSNSDKP-------------------------------------------------------------------------------
---------------HLDSSGSDSES-------E------HKLSTYTPSGAVG------------------------------------------------------------------PPQLAP----
------YG--------------------------GLGAPSLG------------------------------QPLYQAPGGL--------------LHDYQAL--------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Amel_SIX2        ----------------------MAYGGTTGNGGAMGPSSGGTDVLGSTGDYSPQGTPTPLTGHSAGPVEAT---------TYGPGTGPASYSPQD-----------------
----------------------------------------------------------------------------------SPASSAGGGSGSNGQLSSFGFTQEQVACVCEVLQ----
QAGSVERLGRFLWSLPACTRLH--------
RHESVLKAKAIVAFHRGHFKELYRILESHTFSPHNHQKLQALWLKAHYIEAERLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRSVLRDWYATNPYPSPREKRELAESTGLTTTQVSNWFK
NRRQRDRAAEHSGQREDKTHGGGLGDSSSESGDDTKRQAVGQQQQQQQQPSSCAVQQQQQQQQMVDHQQTSGMYQLPPPV------------------------------------
SVSSPHSYHQGHGSTL----SHPHTPTHHMQHHDTGSESGDDKRNLH------HQLQHHLQVGAHGTH----------------------------------------------------
GLVHPTATLPPPPPSCAQQKSQLDYMQYYGPQDYLITPTSADLQKSLPHTATSMQMGAIAPSMGGLSPMGPSTMLPNTMQSVNTMNTMSMNGIGMGAYQGMGSMGMSTSHGSFHNGLVLGDYHSL----
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------
Drer_Six1b       ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------MSMLPSFGFTQEQVACVCEVLQ----
QGGNLERLGRFLWSLPACDHLH--------
KNESVLKAKAVVAFHRGNFRELYKILESHQFSPHNHPKLQQLWLKAHYIEAEKLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRGVLREWYTHNPYPSPREKRELAEATGLTTTQVSNWFK
NRRQRDRAAEAKE---------------RENSENNNTGANKQNQLSPLDGG------------------------------------------------------------------------------
---------------KSLMSSSEDEFSPPQSPDQ------NS---VLLLQGNM-------------------------------------------------------------------SHPGA----
--------------------------SAYSMTGLGAAQSVHG----------------------------MQGHPHQLQDSLLGPLT-------SSLVDLGS---------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Mmus_Six1        ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------MSMLPSFGFTQEQVACVCEVLQ----
QGGNLERLGRFLWSLPACDHLH--------
KNESVLKAKAVVAFHRGNFRELYKILESHQFSPHNHPKLQQLWLKAHYVEAEKLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRGVLREWYAHNPYPSPREKRELAEATGLTTTQVSNWFK
NRRQRDRAAEAKE---------------RENTENNNSSSNKQNQLSPLEGG------------------------------------------------------------------------------
---------------KPLMSSSEEEFSPPQSPDQ------NS---VLLLQSNM-------------------------------------------------------------------GHARS----
--------------------------SNYSLPGLTASQPSHG----------------------------LQAHQHQLQDSLLGPLT-------SSLVDLGS---------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Drer_Six1a       ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------MSILPSFGFTQEQVACVCEVLQ----
QGGNLERLGRFLWSLPACDHLH--------
KNESVLKAKAVVAFHRGNFRELYKILESYQFSTHNHPKMQQLWLKAHYVEAEKLRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRSVLREWYTHNPYPSPREKRELAEATGLTTTQVSNWFK
NRRQRDRAAEAKE---------------RENGENNGVGG-KQSQRSPLDGV------------------------------------------------------------------------------
---------------KSLMSSSEDEFSPPQSPEH------SS---VLLLQGTM-------------------------------------------------------------------NNPGA----
--------------------------PAYPMPGLGAPHPLHG----------------------------MQGHPHQIQDSLLGSLT-------SSLVDLGS---------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Drer_Six2.1      ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------MSMLPTFGFTQEQVACVCEVLQ----
```

```
QGGNIERLGRFLWSLPACEHLH--------
KNESVLKAKAVVAFHRGNFRELYKILESHQFSPHNHPKLQQLWLKAHYIEAEKLRGRPLGAVGKYRVRRKFPLPRSIWDGEETSYCFKEKSRSVLREWYTHNPYPSPREKRELAEATGLTTTQVSNWFK
NRRQRDRAAEAKE---------------RENNENSNTNSHNPLTSSMNGN-------------------------------------------------------------------------------
---------------KTILGSSDDDKTPSGTPDH------TSSSPALLLTSNS-------------------------------------------------------------------GLQSL----
--------------------------HGLAPPPGPSAIPVPS--------------------------VDSVHHHHSLHDTILNPMS-------SNLVDLGS---------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Mmus_Six2        ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------MSMLPTFGFTQEQVACVCEVLQ----
QGGNIERLGRFLWSLPACEHLH--------
KNESVLKAKAVVAFHRGNFRELYKILESHQFSPHNHAKLQQLWLKAHYIEAEKLRGRPLGAVGKYRVRRKFPLPRSIWDGEETSYCFKEKSRSVLREWYAHNPYPSPREKRELAEATGLTTTQVSNWFK
NRRQRDRAAEAKE---------------RENSENSNSSSHNPLASSLNGSG------------------------------------------------------------------------------
---------------KSVLGSSEDEKTPSGTPDH------SSSSPALLLSPPPP-----------------------------------------------------------------PGLPSL----
--------------------------HSLGHPPGPSAVPVPVPGG---------------------GGADPLQHHHSLQDSILNPMS-------ANLVDLGS---------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Drer_Zgc_195096  ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------MSMPPTFGFTQEQVACVCEVLQ----
QGGSIERLGRFLWSLPACEHLH--------
KNESVLKAKAVVAFHRGNFRELYKVLESHQFSPHNHPKLQQLWLKAHYIEAEKLRGRPLGAVGKYRVRRKFPLPRSIWDGEETSYCFKEKSRCVLKEWYTHNPYPSPREKRELAEATGLTTTQVSNWFK
NRRQRDRAAEAKE---------------REN-EGANPNGHNPLTSHVNEN-------------------------------------------------------------------------------
---------------KSLCESSDDDKSPAGTPDH------TNMSPAILMPSSS-------------------------------------------------------------------GLPPL----
--------------------------HSFAPPPGPSASISDT----------------------------QSHHHFSMHDGLLNPMT-------ASLVELGS---------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Drer_Six9        ----------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------------MAMGFSPEQVACVCEVLL----
QSGSMDRLSSFLCSLPSISTSSNMYMGFGQSQESVLKARAAVAFHHCRFTELYALLEGNVFSPRSHPLLQQLWLRAHYMEAELQRGRPLGAVGKYRIRRKFPLPRTIWDGEETSYCFKEKSRSVLREWY
CRKPYPSPREKRDLAAATGLTATQVSNWFKNRRQRDRAATSRQ------------------------------GTSAG---------------------------------------------------
-----------------------------------------------AFLSSDEEISPPGSPR-------TLFSCSQQLSAHP----------------------------------------------
-------------------------------------------------------------PPLRHLGPAHY---------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------------------------------
Mmus_Six5        ----------------------MATSPAEPSAGPAARGEAAAATEEQEEEARQLLQTLQAAEGEAAAAG-------------AGDAAAAADSGS------------------
-----------------------------------------------------------------------------------PSGPGSPRETVTEVPTGLRFSPEQVACVCEALL----
QAGHAGRLSRFLGALPPAERLR--------
GSDPVLRARALVAFQRGEYAELYQLLESRPFPAAHHAFLQDLYLRARYHEAERARGRALGAVDKYRLRKKFPLPKTIWDGEETVYCFKERSRAALKACYRGNRYPTPDEKRRLATLTGLSLTQVSNWFK
NRRQRDRTGTGGG-APCKSESDGNPTTEDESSRSPEDLERGVASMAAEAPAQSSIFLAGATSPATCPASSSILVNGSFLA------------------------------------
ASSPPAVLLNGSPVIINSLALGENSSLGPLLLTGGSAPQPQPSLQGV------SEAKNSLVLDPQTGE----------------------------------------------------
VRLDEAQSEAPETKGVHGTTGEEIPGALPQVVPGPPPASTFPLTPGAVPAVAAPQVVPLSPSSGYPTGLSPTSPRLNLPQVVPTSQVVTLPQAVGPLQLLAAGPGSPVKVAAAAGPTNVHLINSSVGVT
ALQLPSSTAPGNFLLANPVSGSPIVTGVAVQQGKIILTATFPTSMLVSQVLPPAPSLALPLKQEPAITVPEGALPVGPSPTLPEGHTLGPISTQPLPPASVVTSGTSLPFSPDSSGLLSSFSAPLPEGL
MLSPAAVPVWPAGLELSTGVEGLGTQATHTVLRLPDPDPQGLLLGATTGTEVDEGLEAEAKVLTQLQSVPVEEPLEL----------------------------
Mmus_Six4        ----------------------MSSSSPTGQIASAADIKQENGMESASEGQEAHREVAGGAAAGLSPPAPAPFPLEPGDAAAASRVSREEGAAAAGAADQVQLH--------
------------------------------------------------------------------------------------SELLGRHQHAAAAQPPLAFSPDHVACVCEALQ----
QGGNLDRLARFLWSLPQSDLLR--------
GNESLLKARALVAFHQGIYPELYSILESHSFESANHPLLQQLWYKARYTEAERARGRPLGAVDKYRLRRKFPLPRTIWDGEETVYCFKEKSRNALKELYKQNRYPSPAEKRHLAKITGLSLTQVSNWFK
NRRQRDRNPSETQSK------------SESDGNPSTEDESSKGHEDLSPHP---------------------------------------------------------------------LSGASDG-
VTNLSL---SSHVEP-VYMQQIGNAKI-SLSSSGVLLNGSLVPASTSPVFLNGN-
SFIQGHNGVILNGLNVGNTQTVSLNPPKMSSNIVGNGIAMTDILGSTSQDVKEFKVLQSSAVNSAATTSYSPSAPVSFPGLIPCTEVKREGIQTVASQDGGSV-
VTFTTPVQINQYGIVQIPNSGANGQFLNGSIGFSPLQLPPVSVAASQGNLSVTPSTSDGSTFTSEPATVQHGKLFLSPLTPSAVVYTVPNSGQTVGAVKQEGLERGLVFSQLMPVNHSAQVNASLSSEN
LSGSGLHPLTSSLVNVSAAHGFSLTPPTLLNPTELNPDLAESQPVSAPVASKCTVSSVSNTNYATLQNCSLIPGQDLLSGPMTQAALGEIVPTAEEQVSHASTAVHQDFVREQRLVLQSVPNIKENFLQ
NSENKATNNLMMLDSKSKYVLDGMVEAGCEDLGTDKKELAKLQTVQLDEDMQDL
Nvec_SIX4.1      ----------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------DQVACVCDALR----
QAGDIERLSRFLWSLPPDDLLN--------
GSESVLKARAIVSFHRGRYREVYNILETNEFDPSSHELLQCLWYKAHYSEAEKLRGRSLGAVDKYRIRRKFPLPRTIWDGEETVYCFKEKARAALKDCYEQNKYPTPQEKRLIAKQTNLTLKQVSNWFK
NRRQRDRIPSNKR--------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Drer_Six4.2      ----------------------MSVSSS--EVTVPGDIKKEN--VKYTELLESSTDSQE-HVKLLTLDSP---------VATGMRVGQQTSPPAR----KMPTM--------
------------------------------------------------------------------------------------DSLFAHSSPAVTATS-LAFSPEQVACVCEALQ----
QGGNVDRLARFLWSLPQSDLLR--------
GNESILRAQALVAFHQARYQELYSILESHSFSPSCHSALQDLWYKARYTEAEKARGRPLGAVDKYRLRRKFPLPRTIWDGEETVYCFKERSRNALKDLYKQNRYPSPAEKRNLAKITGLSLTQVSNWFK
NRRQRDRNPSETQSKK----------VSESDGNHSTEDESSKGQESLSPCP---------------------------------------------------------------------
MSTCADGPVSNTALPLCSGSLDAGVIVQQVADSRT-SISPPSVIFNGVSVHSPAS-VFHNGTPSFLSSTGHVLFSGMNLG-LQSLACRP----TDVEMDGAAQE--KGGTADPALAYSSFHCS-VN---
----GTDVEVKEEDLR-----KDVSVQDQAPSIGSAFSITPSNGLQIDSYSLVTEGNK---------------VVLPSLQLTVSSSPSSVTQGVVDCS------SPVCQERLQLTSLEPSSTLYSLGN-
IHTLSAVKKEPLEP-------------------------PGGFLYHLGYS-----------------SDPTHSSPLLAQTDLP---------------TEFRG--------------------------
-----HTDHMTSSVNGDFLCAVS----------------EGGKGE----------------MEEE-------EEKQLTKLKTVHLEEEMTDL
Drer_Six4.1      ----------------------MCCLKPR-GVTYQDAFKMSSSSSEVISADEIKRESRG--SVKLSVLDP----------------------------AELPME--------
------------------------------------------------------------------------------------NAELLDCSPAS-----LAFSPEQVACVCEALM----
QGGNVDRLARFLWSLPQSDLLR--------
GNESILKAQAIVAFHHARYQELYCILENHSFSPSNHSSLQDMWYKARYTEAEKARGRPLGAVDKYRLRRKYPLPRTIWDGEETVYCFKERSRNALKDMYKRNRYPSPAEKRNLAKMTGLSLTQVSNWFK
NRRQRDRNPSEAQSK------------SESDGNHSTEDESSKGQEDLSPRP---------------------------------------------------------------------
LSSGSSSSVPHGTAPPTVYSDGGTTVIQQIGDVKMPSHSGGGMSYNGDLASVNAH--YINSGAYGHSHSSNALLNGLGATGGHFLTFDP------QRVSHHSQERLLGGSSVSYAPFS-
MGVSETASGKIDTMQTLVSNGEDGVIS-------SVVSFAS--APSTPSTTSGGLHLSSYSVVN---------------------LPGNESTMGLPPLMLPPSS----------
TSSTHGSAPLGNVVSDSSQHSSQTLLYSQ-TVKQEPLDV-------------------------PGSYAYP--------------------SDMPLDQGSNLGYTAS----------------------
-------------------------------LFLSSSSSSGSLPSTIT------------------------------------------------------ATVTSTLSNSDVHHA
Cele_H-33        ----------------------------------------------------------------------------------------------------MQLN--------
------------------------------------------------------------------------------------SSSFHPHHFTCDTTR---YSEEQVACICEALS----N--
DARKLSQFVWTVLERDEMR--------
NNQYILKAQAFLAFHSNNFKELYRIIESHHFASEHHLPLQEWWLNAHYHEAEKIRGRQLGAVGKYRIRRKYPLPRTIWDGEETSYCFRDKSRVLLRDWYCRNSYPSPREKRELAEKTHLTVTQVSNWFK
NRRQRDRAG-------------------VPEPKDCLKDISEEEDLKLIRKT---------------------------------------------------------------------ASKLSNS--
-----------------FHNPSDLSSYSAAAAAATFPG-----------------FYMNYNDMMIG--------------------------------AGTSYQSL-----------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Drer_Six4.3      ----------------------MASLSLESTEQTENGPKESAQNEAKVKEETDPDEVSEQLLQS------------------------------------------------
-----------------------------------------------------------------------------------------------FQNSALSFSTDQVACLCEALL----
```

```
QAGNVDRLWRFLATIPPSADLLR-------
GNETLLKAQALVAFHRDEFKELYAILDSHDFHPSNHGFLQDLYLKARYKEAERSRGRSLGAVDKYRLRKKFPLPKTIWDGEETVYCFKEKSRNALKECYKINRYPTPAEKKNLAKVTGLSLTQVSNWFK
NRRQRDRTPSGTNSK------------
SESDGNHSTEDEASKGDLEDITDKPAAQETGSSNASLISVPGAPCGTGQLVLNSPGGFLTSPHPLLLNGSPILSGAGTGVIINGLTLSDGHTVTLSPVTANAPLLLNGAQVISKDERGISDLEAQGSLT
TVVLNPQASLTSTIPLSLSEDTKSTSSN-----
VSPLDFISLPEALKNTDNTHPVPANSISLPTISTSVISSTSLSSGILAPNSISASIPGSVSSVISSPVVPLQPTQPPEIVVLGKAEPLSSTCSSFSSPQVLSLPQVVPSIQGVPVSQLMQPPSGTTVSS
RPQLVPVSPVNPQLSQVSIPQFPTQTFHMGPRLAQAQPQNGSVTLNTSSALTLPQIADGQVTQVLTPQLGNESNLASLPQIQTSLGTQVIPISSPTQVVPISQTKDSSQMVPLSLPQLMPVSSITGTTG
TFSFPQVVPATPSLSIPSHGGAFQILTSGAGTGLSVAQGPIRLSPIGPPQSVPTGTIPGVQLLNSGVIKLPTASPGNILLAGGIGSSPILSVQNGKLILTIPAGIQFASMPVKSVPDNSVSSNSNLDSQ
ISTVHSLENQPAPPLIARNSQTL
Nvec_Six4.2      ----------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------MPFSMEQIISVCECLQ----
NSGNIERLARFLWSLPKDSEEIH-------
ACETILVAKAVVAFHQNNFKELYSILESRKFQRSEHEKLQCLWRTAHYIEAERVRGRPLGAVGKYRVRRKFPLPRTIWDGEETTYCFKEKSRSILNKAYVDSPYPTPREKHELAKMTDLTVTQVSNWFK
NKRQRVRAAEIRKR-------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Phaw_SIX4        --
MKTEHPRPSSAESYGSFSIPSPDGLSTPNPRLPHYSSMNSSSSSVANFSNLSNFQSTASNLCSRSFKNDTYTPPPHIGMGMSPMSSQSSVTTPPLSSSPLSPSEPLRNITNNPHLIHGSAASPFHYGAL
AANTSGSTAPCAVSQSTSSSCTTNISNLSKSPAIKPQPTSPHIATVTPEPHHRESPSLSTPKKTPSSSASPSSPSSSPVKTSQFNKEQIDCICEVLI----QARDMDKLSKFLNNLPASHFATD-----
-
AVSEVILRAKVEVAFAKGNYKEVYNILEATSFHANYHTHLQGMWYKAHYKEAERVRQRPLGAVDKYRIRKKYPLPRTIWDGEETIYCFKEKSRMALKDCYRQNRYPTPDEKKTLSRQTGLTLTQVSNWF
KNRRQRDRNPAPRPDLLLG--------------------------------------------------------------------------------------------------------------
-----------------AHMDHGGLAMDHHVHPDMRAMYHHAKMAAHAYDPCKMGT-------------------------------------------------------------------------
-------------------------GVHPAYDPNNVMCMSMRGHRADAILHHLPNVAMGSYPYSSYDLQSVAATHGNSHQDMREDKWTAAGVQ------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------
Amel_SIX4        --------MPD-----------SSDQLSSPN------SECNSQIGVMHRNTSAGQFDDSS---------------------------------
MPGLALTHQTHSQNLTSASN--------GGGSLQY------------PQELTTCSSANTSTNIGNIG-------------------------GLSLG-------------------
ITASNYTPDQISCMCKALS----QRQDIEKLTRLLWSLPPGELFR--------
RDENVLIARATVAFHRGAYHELYSILESHPFSSDRHPELQQMWYKSHYREAEKIRGRPLGPVDKYRLRKKYPLPKTIWDGEEVVYCFKERSRNALKQCYMRTKYPLSEEKKNLAKETGLTLTQVANWFK
NRRQRDRTPQTRT--------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Dmel_SIX4        MFDKNLDGNNLSVSIGGDLDSTSSGGTSSDHS---
AVHQDNLSSPMAYGSLFLPNAGYRGNLSCKTVLQLDKFAPYEGVEKDHLLERRFQDITNDYDKSPPPTASTTPTHYPSLNSIIFENGSSGN-LGDLNGNTKTD-LCAGLQRSG-GGLGGN-AGSG----
---------------------GHLISNL--TAAHNMSAVSSFPIDAKMLQFSTDQIQCMCEALQ----QKGDIEKLTTFLCSLPPSEFFK--------
TNESVLRARAMVAYNLGQFHELYNLLETHCFSIKYHVDLQNLWFKAHYKEAEKVRGRPLGAVDKYRLRKKYPLPKTIWDGEETVYCFKEKSRNALKDCYLTNRYPTPDEKKTLAKKTGLTLTQVSNWFK
NRRQRDRTPQ-QRPDIMS---------------------------------------------------------------------------------------------------------------
-----------------VLPVG--QLDGNGFPRMFN---------------------------------------------------------------------------------------------
--------------------------APSYYPETIFNGQ------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Dpse_SIX4        MFEKNMDTNNLSASISGDLDSTSSGGTSSDHS---
AAQQDNLSSPMAYGSLFLPNPGYRGNLSCKTVLQLDKFGPYEAAEKDHLLERRFQDITNDYDKSPPPTASTTPTHYPTLNSIMFENGVSGGPLGDLNGNTKTDSLCAGLQRGGAGGLGGN-PVTG----
---------------------GHLISNL--TAGHNMSAVGNFPIDAKMLQFSTDQIQCMCEALQ----QKGDIEKLTTFLCSLPPSEFFK--------
TNESVLRARAMVAYNLGQFHELYNLLETHCFSIKYHVDLQNLWFKAHYKEAEKVRGRPLGAVDKYRLRKKYPLPKTIWDGEETVYCFKEKSRNALKDCYLTNRYPTPDEKKTLAKKTGLTLTQVSNWFK
NRRQRDRTPQ-QRSDIMS---------------------------------------------------------------------------------------------------------------
-----------------VLPVG--QLDGNGFPRMFN---------------------------------------------------------------------------------------------
--------------------------APSYYPETIFNGQ------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Dvir_SIX4        MFEKNMDSTNLSVSINGDLDSTSSAGTASDHS---AVQHDNISSPMAYGSLFLPNAGYRGNISCKTVLQLDKFAPYETSEKD-
LIERRFQDITNDYDKSPPPTAATTPTHYPTLNSIIFENG--NSSLSDLNGNTKTDALCAGLQRAPTSTLSGTGASSG-------------------------GHLISNL--
TAAHNMSAVGSFPIDAKMLQFSTDQIQCMCEALQ----QKGDIEKLTNFLCSLPPSEFFK--------
TNESVLRARAMVAYNLGQFHELYSLLETHCFSMKYHVDLQNLWFKAHYKEAEKVRGRPLGAVDKYRLRKKYPLPKTIWDGEETVYCFKEKSRNALKDCYLTNRYPTPDEKKTLAKKTGLTLTQVSNWFK
NRRQRDRTPQ-QRTDLM----------------------------------------------------------------------------------------------------------------
------------------LPVG--QLDANRFPRMFN---------------------------------------------------------------------------------------------
--------------------------APSYYPESIFNGQ------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Agam_SIX4        -------------------------------------TDNNNSS-------------------------------------------YLQSLQ------PPPPHPQPPSS--
--------------------------------------------AASG-------------------------GHPLG-------SSNGPIG-PDLERKCFSFSPEQIQCMCEALQ----
QEGDLEKLATFLWSLSPSDLIS--------
GNESLLRARALVAYHRGLYHELYAVLESHFFSPKYHADLQALWFKAHYREAEKVRGRPLGAVDKYRLRKKYPLPKTIWDGEETVYCFKEKSRNALKDCYTRNRYPTPDEKKTLAKKTGLTLTQVSNWFK
NRRQRDRTPQSSRPDLMM---------------------------------------------------------------------------------------------------------------
----------------SVLPVTPDQMDGS-YQRLFN---------------------------------------------------------------------------------------------
--------------------------VANYGPAHGYHANDMYAVQ------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
Tcas_SIX4        -----MEVDATSSTSNDSNNGNFCQNLNYYNENLFFTTNGKLDSFFNSDKSARKNGENEKVARKFNCYKKDDFS-------------
GNNNVLKSKSDDVNKRVSLSVANYN------------------------------KKVDVNSENRKNLIDFFN-------------------------KKNMDNN------
SIHLNNVDSYGKKSIGFSPEQVQCMCEALH----QRGDIERLATFLWSLPPSELLR--------
GNESILRARAAVAFYRGSYHELYSILESHAFNQRWHAELQTLWFKAHYNEAEKVRGRPLGAVDKYRLRKKYPLPKTIWDGEETVYCFKERSRNALKECYARNRYPTPDEKRALAKRTGLTLTQVSNWFK
NRRQRDRTPQPRPELILGN--------------------------------------------------------------------------------------------------------------
----------------MSLNQSGMMSAHQGSLDMVAFQ-------------------------------------------------------------------------------------------
------------------------AASKLFDSNCGVASCYSDYQVP-----------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------
```

## [PhD_A6_phylogeny_database]

### [PhD_A6.3_phylogeny_ClustalW2_alignments]

# **“alignment_Unplugged-Gbx-proteins.txt”**

```
Alignment of protein sequence fractions of various unplugged/gbx genes from different species, encompassing the HD domain and
VPIPV motifs.
This was done using the ClustalW2 algorithm (Larkin et al., 2007).
For details and references regarding the species' protein sequences, see Appendix (A7).

Dvir_unpg          SNSSKSRRRRTAFTSEQLLELEREFHAKKYLSLTERSQIATSLKLSEVQVKIWFQNRRAKWKRVKAGLTSHGLG---RNGSASGT--------------
KIVVPIPVHVNRFAVRSQHQQLEKMCLSG
Dpse_unpg          SSSSKSRRRRTAFTSEQLLELEREFHAKKYLSLTERSQIATSLKLSEVQVKIWFQNRRAKWKRVKAGLTSHGLG---RNGSASGT--------------
KIVVPIPVHVNRFAVRSQHQQLEKMCLSG
Dmel_unpg          SSNSKSRRRRTAFTSEQLLELEREFHAKKYLSLTERSQIATSLKLSEVQVKIWFQNRRAKWKRVKAGLTSHGLG---RNGTTSGT--------------
KIVVPIPVHVNRFAVRSQHQQLEKMCLSG
Agam_unpg          AENSKSRRRRTAFTSEQLLELEREFHAKKYLSLTERSQIATSLKLSEVQVKIWFQNRRAKWKRVKAGLNSHGLGN--RNASGSGTG-------
TTGTANKIVVPIPVHVNRFAIRSQHQQMEKMNLVG
Mmus_gbx2          TSTGKNRRRRTAFTSEQLLELEKEFHCKKYLSLTERSQIAHALKLSEVQVKIWFQNRRAKWKRVKAGNANSKT------GEPSRN-------------
PKIVVPIPVHVSRFAIRSQHQQLEQ-ARP-
Drer_gbx2          TSTGKNRRRRTAFTSEQLLELEKEFHCKKYLSLTERSQIAHALKLSEVQVKIWFQNRRAKWKRVKAGNVNSKT------GEPSRN-------------
PKIVVPIPVHVSRFAIRSQHQQLEQ-ARP-
Mmus_gbx1          TPGGKSRRRRTAFTSEQLLELEKEFHCKKYLSLTERSQIAHALKLSEVQVKIWFQNRRAKWKRIKAGNVSSRS------GEPVRN-------------
PKIVVPIPVHVNRFAVRSQHQQMEQGARP-
Drer_gbx1          GSAGKSRRRRTAFTSEQLLELEKEFHCKKYLSLTERSQIAHALKLSEVQVKIWFQNRRAKWKRIKAGNVNNRS------GEPVRN-------------
PKIVVPIPVHVNRFAVRSQHQQIEPGSRP-
Pdum_gbx           SGGGKSRRRRTAFTSEQLLELEKEFHSKKYLSLTERSQIAHNLKLSEVQVKIWFQNRRAKWKRVKAGYINGRAPTTTSPGEQNNK-------------
PKIIVPIPVHVNRMAIRSQHQQLERSGRLI
Nvit_unpg          GAGSKARRRRTAFTSEQLLELEREFHAKKYLSLTERSHIAHALKLSEVQVKIWFQNRRAKWKRVKAGLTSGGAGN--
PAARHSGGGSSAGGHQSNGGGPRIVVPIPVHVSRLAVRSHHHHLEKCPQPA
Tcas_unpg          GTGTKARRRRTAFTSEQLLELEREFHAKKYLSLTERSQIASALRLSEVQVKIWFQNRRAKWKRVKAGLGAG------PHQPKSGG------QQKS----
KLVVPIPVHVNRFAVRSQHQQLERALGDL
Phaw_gbx           GTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVKAGLVGS------SGGGLVGG----
TNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSA
Isca_Xvent-2B      GVTSLQRRRRTAFSSEQLLELEKEFHSKKYLSLTERSQIAHALQLTEVQVKIWFQNRRAKWKRVKAGLTAGR-------GPHSHP-----------
DAPRIVVPIPVHVNRVALRNQQDAKGYVLSLA
Nvec_gbx           LSSASRRRKRTAFTSKQLLQLEREFHNKKYVSLEERSVIATNLNLTEVQVKIWFQNRRAKWKRVRS-LSHPAHG----NAKKINV-------------
PKLVVPIPIHGNRFVAMNQANAL-------
```

## [PhD_A6_phylogeny_database]

### [PhD_A6.3_phylogeny_ClustalW2_alignments]

# **“alignment_Unplugged-Gbx-similar_proteins.txt”**

```
Alignment of the complete protein sequences of various unplugged/gbx and similar genes from different species.
This was done using the ClustalW2 algorithm (Larkin et al., 2007).
For details and references regarding the species' protein sequences, see Appendix (A7).


Mmus_Gbx1         ----------------------------------------------------------
MSPEGDTPNPPARPLPPYERESAAPEPHCAEPAPRRHLGPRARSEPCVSARGARRGPGSAMQRAAGGGAPGGSGGSSGGPGAAFSIDSLIGPPP-PRSGHLLYTGYPMFMPYRPLVLP------
QALAPAPLP-------------------------AGLPPLAPLASFAGRLSNTFCAGLGQAVPSMVALTTALP-SFAEPPDAYYGPPELAAAAASTASRSNPEPAARR----------------------
-------------------------------TDGALDAEELLPAREKVTEPPPPPPPPHFSETFPSLPAEGKVYSSDEEKLEPPAGDPAGSEQEEEGS----------
GGDSEDSFLDSSAGGPGALLGPKPKLKGSPGTGAEEG--TPVATG---------------------------------------------------------------------------
VTTPGGKSRRRRTAFTSEQLLELEKEFHCKKYLSLTERSQIAHALKLSEVQVKIWFQNRRAKWKRIK-------AGNVSSRSGE---------PVRNPKIVVPIPVHVNRFAVRSQHQQMEQGARP----
----------------------------------------------------------------------------------------------------------------------------------
-----------------------------
Mmus_Gbx2         ----------------------------------------------------------MS----AAFPPS------------------------------------------
----LMMMQRPLG---------SS----TAFSIDSLIGSPPQPSPGHFVYTGYPMFMPYRPVVLPPPPPPPPALPQAALQ-------------------------PALPPAHPHHQIP-
SLPTGFCSSLAQGMALTSTLMATLPGGFSASP-----QHQEAAAARKFAPQPLPGGGNFD-----------------------------------------------------
KAEALQADAEDGKAFLAKEGSLLAFSAAEAVQASLVGAVRGQGKDESKVE---DDPKGKEE-----------------SFSLESDVDYSSDD--NLPGQTAHKEEDPGHALEETPQSGGAAG--------
------------------------------------------------------------------
STTSTGKNRRRRTAFTSEQLLELEKEFHCKKYLSLTERSQIAHALKLSEVQVKIWFQNRRAKWKRVK-------AGNANSKTGE---------PSRNPKIVVPIPVHVSRFAIRSQHQQLEQ-ARP----
----------------------------------------------------------------------------------------------------------------------------------
-----------------------------
Dmel_UNPG         ------------------------------------------------------------------
MERPALLQNGEIGTMESPTTRLASKPFPKPFSIESLIANQTPATATPPSPPEERDQEQEAEQEQELSARAMVASSALGLTQFPLYNPWLHGYFAQNHERLTHLIAGGCYLPSSPAG-HPAAQQPQA---
----------------------QAQPQPPPPHPPTHALEKQLPPTLPHPLDTRFLPFNPAAAGVAPTDLSYRRLAELMNQDYVHSLSVHARLQHMA----------------------------------
-------------------AAGRMHEDQANPGMAQLQEPTPPQAHSSPAKSGSHSPMEPALDVGMDEDFECSGDSCSDISLTMS------------
PRNYNGEMDKSRNGAYTNSDSEDCSDDEGAQSRHEGGGMGGKDSQG------------------------------------------------------------------------
NGSSSNSKSRRRRTAFTSEQLLELEREFHAKKYLSLTERSQIATSLKLSEVQVKIWFQNRRAKWKRVK-------AGLTSHGLGRNG-------
TTSGTKIVVPIPVHVNRFAVRSQHQQLEKMCLSGPKPDLRKKLSAEAIGGFEKFSGSTNASSPSGGPVGLGVGVGVGVGVGLGVSTPLSLARSIY-----------------------------------
------------------------------------------------------------------
Mmus_Meox1        ----------------------------------------------------------------------------------------------------------------
------------------------------------------------------MDPVANSCVRNP-----QPPAPVWGC--------------------------
LRNPHSEDSSASGLSHYPPTPFSFHQKSDFPATAAYPD---FSASCLAATPHSLPRTERIFNEQHPAFPQTPDW-----------------------------------------------------
HFP-----ISEAGQRLNLGP-AGSAREMGAGSPGLVDGTAGLG-------------------------------------EDCMVLGTIAN-----------ETEKKS-SRRKKERSDNQENGG------
-----------------------------------------------------------------
GKPEGSSKARKERTAFTKEQLRELEAEFAHHNYLTRLRRYEIAVNLDLSERQVKVWFQNRRMKWKRVK-------GGQ------------------------PVSPQEQD-------REDGDSAASP---
----SSE---------------------------------------------------------------------------------------------------------------------------
------------------------------
Mmus_Meox2        ----------------------------------------------------------------------------------------------------------------
------------------------------------------------------MEHPLFGCLRSP-----HATAQGLHP--------------------------FSQSSLALHGRSDHMSYP----
ELSTSSSSCIIAGYPNEEGMFASQHHRGHHHHHHHHHHHHQQQQHQALQSNW-----------------------------------------------------
HLPQMSSPPSAARHSLCLQPDSGGPPELGSSPPVLCSNSSSLG-------------------------------------SSTPTGAACAPGDYGRQALSPADVEKRSGSKRKSDSSDSQE-GN------
-----------------------------------------------------------------
YKSEVNSKPRKERTAFTKEQIRELEAEFAHHNYLTRLRRYEIAVNLDLTERQVKVWFQNRRMKWKRVK-------GGQQGAAAREK--------
ELVNVKKGTLLPSELSGIGAATLQQTGDSLANEDSRDSDHSSEHAHL-----------------------------------------------------------------------------------
------------------------------------------------------------------
Phaw_GBX          -------------------------------------------------------------------------
MLASSEMTTCDSPPKSPTTKMGPNSKSFAMEYLLTISGQDKNDCNRSKSMKTSAASTLLSRETDTHPNHLPASHMLDISRQLSQITQISQNLALPHLYNSCLLTGG----MASVGLPPF----------
---------------YGHPPLAPHYPYPAPSTVPFPNPITGSLPNGFSAGGLQP--ALNPPFLSLKDFGPGAGSRYFGGSPTAGGENPR-----------------------------------------
------------NAILRSTLSSETPADLPSPFSGATNGDESSHISDLDFKESTKS---------------------------------------
NELVDDLQIDPVDSDDESHVNGNSAGLGGGGAPLDSSTTTGA------------------------------------------------------------------------
EHRGTGSKNRRRRTAFTSEQLMELEREFQTKKYLTLSERSHIAQTLHLSEVQVKIWFQNRRAKWKRVK-------
AGLVGSSGGGLVGGTNKGGAGGGHKIIVPIPVHVNRLSNLSQAHQIEKCRGSAAGSNSDGSLSRPTESSSAADDDGKLANQRMATEACTSTKHRNKSANLSSSFLNSETSESSIVGVNQHSQDQQQQHR
ALQSRHSAFQGLSSSTSSCSSGHSELSNLGLSSGRNGSALEAHLLGHTKEVTEQLKSMRNGGRTSTTSSSTSLQCYFQNHERISP
Mmus_Mnx1         ----------------------------------------------------------------------------------------------------------------
---MEKSKNFRIDALLAVDPPRAASTQSAPLALVTSLATTVSGPGRGGSGGGGTSSGASRSCSPASS----EATAAPGDR--------------------------
LRAESPSPPRLLAAHCALLPKPGFLGAGGGGGAAGGPGTPHHHAHPGAAAAAAAAAAAAAAGGLALGLHPGGAQ-----------------------------------------------------
GGAGLPAQAALYGHPVYSYSAAAAAAALAGQHPALSYSYPQVQG-----------------------------------AHPAHPADPIKLGASTFQLDQWLRASTAGMILPKMPDFSSQAQ--------
-------------------------------------------------------------------
SNLLGKCRRPRTAFTSQQLLELEHQFKLNKYLSRPKRFEVATSLMLTETQVKIWFQNRRMKWKRSKKAKEQAAQEAEKQKGGGGGTGKGGSEEKTEEELMGPPVSGDKASGRRLRDLRDSDPDEDEDDE
EEDNFPYSNGAGAHAASSDCSSEDDSPPPRLGGPGHQPLPQ-----------------------------------------------------------------------------------------
----------------------------
Dmel_EXEX         ----------------------------------------------------
MSSAQVSQQPPPPPLTASSSTSMPSQLLLSHIQHHSSSLHHQRSHQRSSPPAQDYDQGSNQSTLSPRSQITPSPPAASPTHSAATTGSPVSTYTPALSPGGAQDSSQTESGQSSQMPLLVAPMPVGTGP
PHSHPPPPHPR-------------------LYVEGFPPPHHVPHPHPALGTYPLPRLPLMLPANPPGVPSGAAAPPPVSPQPAANHLSHSGATMATTTLLPPAQSQPKKSF-------------------
-------CIDALLAKSQHQSGEPQPIIVDDRLAALHYARDQAELNHAFVTASNAAAAAQAAASAAGGISEQEALQRIRDSREYDSPSPDGMSRSESPT----------
SSHRSSPPISPGCEDQQQPHGALHPQHLSMRMSDFHDEFKKPIPPHSPIRPQDFPLYAGGHPYQLLAQGGSAFHRPLDPSGKPIPIPMGHNFMPSQLQFEFLARAGMLHHRIPELAAYPHHAILGKTRR
PRTAFTSQQLLELEKQFKQNKYLSRPKRFEVASGLMLSETQVKIWFQNRRMKWKRSK----------------------------------KAQQEAKERAKANQQQQQQQQTPSAAS------------
----------------------------------------------------------------------------------------------------------------------------------
-------------------
Dmel_CG11085      ----------------------------------------------------------------------------------------------------------------
--------------------------------------------------MQSSKSFLIRDLLG------------------------------------------
DLINRRQTDSELELSNDDSDIDIEDRSTPDSTAAGCQQELLLSHHHRRFTHHDESSVESCLSATRGPGSGTGS---------------------------------------------------------
---------------GGGGGGGGGGGVASGLSAAAAAAG-------------------------------------VAAGLLAAAASGANGDRDANGGSGPGSGGGTSGGYAEHKLQLSK----------
-----------------------------------------------------------------
SGRKPRRRRTAFTHAQLAYLERKFRCQKYLSVADRSDVAETLNLSETQVKTWYQNRRTKWKRQN-----------------------------
QLRLEQLRHQATMEKDFVVQDGGGAGGLGCCPSGLSSSFSAAAAAAAAAASNPCNFLTSAAAAAIFRNVGYVHGCPM-----------------------------------------------------
--------------------------------------------------------------
```

```
Dmel_B-H1         ---------------------------------------------------------------------------------------
MKDSMSILTQTPSEPNAAHPQLHHHLSTLQQQHHQHHLHYGLQPPAVAHSIHSTTTMSSGGSTTTASGIGKPNRSRFMINDILAGSAAAAFYKQQQHHQQLHHHN-------------------
NNNNSGSSGGSSPAHSNNNNNINGDNCEASNVAGVGVLPSALHHPQPHPPTHPHTHPHALMHPHGKLGHFPPTAGGNGLN------------------------------------------------
VAQYAAAMQQHYAAAAAAAAARNNAAAAAAAAAAAAAAAGVAAPPVDGGVDGGVG------------------------LAPPAGGDLDDSSDYHEENEDCDSGNMDDHSVCSNGGKDDDGNSVKSG--
-----------------------------------------------------------------
STSDMSGLSKKQRKARTAFTDHQLQTLEKSFERQKYLSVQERQELAHKLDLSDCQVKTWYQNRRTKWKRQTAVGLELLAEAGNFAAFQRLYGGSPYLGAWPYAAAAGAAHGATPHTNIDIYYRQAAAAA
AMQKPLPYNLYAGVPSVGVGVGVGVGPAPFSHLSASSSLSSLSSYYQSAAAAASAANPGGPHPVAPPPSVGGGSPPSGLVKPIPAHSASASPPPRPPSTPSPTLNPGSPPGRSVDSCSQSDDEDQIQV-
---------------------------------
Mmus_Nkx1-2       ---------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------------------------
MLAWQDVGAKAAPSHHKISFSVLDILDPQKFTRAALPPVRLAALEAKKSLEEVEAGQDACSGNPIGSQETPD---------------------------------------------------------
-------AVGRGIDPGSPVEGSEAEEEEEEAEDAGRAHQ-------------------------------------PERWQGVHEGSPEARAVAVGTEESGAEGLPASPGSPGSPRPRR----------
-----------------------------------------------------------RRAESSCAKPRRARTAFTYEQLVALENKFRATRYLSVCERLNLALSLSLTETQVKIWFQNRRTKWKKQN-
-----PGADGAVQAG------------GGAPQPGTPGAVAGGGGSATGSSPGPPVPGALPYQTFPTYPATNVLFPAASFPLTTAANGSPFTPFLGPSYLTPFYAPHL----------------------
------------------------------------------------------------------------------------------
Dmel_SLOU
MVMLQSPAQKASDSASAQNTAVGGLMSPNSNPDSPKSNTSPDVASADSVVSGTGGGSTPPAAKIPKFIISANGAAVAGKQEQELRYSLERLKQMSSESGSLLSRLSPLQEDSQDKEKPNHNNNNSLTNH
NANSNTRRSQSPPASVGSVSFSSPAQQRKLLELNAVRHLARPEPLQHPHAALLQQHPHLLQNPQFLAAAQQHMHHHQHQHHQHPAHPHSHQHPHPHPHPHPHPHPSAVFHLRAPSSSSTAPPSPATSPL
SPPTSPAMHSDQQMSPPIAPPQNPPHSSQPPQQQQVAAPSDMDLERIKLVAAVAARTTQASSTSALASASNSVSNASISISNSSSGSPSGRDLSDYGFRIQLGGLAAAAAAAAATSRQIAAATYARSDT
SEELNVDGNDEDSNDGSHSTPSVCPVDLTRSVNSSAAANPSSASTSASSDRDAATKRLAFSVENILDPNKFTGNKLPSGPFGHPRQWSYERDEEMQERLDDDQSEDMSAQDLNDMDQDDMCDDGSDIDD
PSSETDSKKGGSRNGDGKSGGGGGGGSKPRRARTAFTYEQLVSLENKFKTTRYLSVCERLNLALSLSLTETQVKIWFQNRRTKWKKQN---------------------------
PGMDVNSPTIPPPGGGSFGPGAYASGLLYSHAVPYPPYGPYFHPLGAHHLSHSHS--------------------------------------------------------------------------
----------------------------------------------------------------
Dmel_DR           ---------------------------------------------------------
MLKLSPASMTVTGLRQTMTSPTVPPSSNTPAGNLIITSSSSNSGSNSGSNMSSGNMTSSNLTNLSPSHPAGLNALASPTSPSALLLAHQQHLLQQHQQHQQQQQQQQQAAALQLAAVHPPAHHLHKTTS
RLSNFS-------------------VASLLADTRPRTPPNQAADGPQNLTSSAATSPISQASSTPPPPPASAAAQVPANTFHPAAVAHHAHLLQAAHAAAAAHAQH-----------------------
-------------------------
QAMAAQLRQQQQQADARANSPPASTSSTPSSTPLGSALGSQGNVASTPAKNERHSPLGSHTDSELEYDEEMLQDHEADHDEEEDSIVDIEDMNADDSPRSTPDGLDGSGKSLESPHGPPPGSHMQSTIL
SPAALASGHVPIRPTPFSALAAAAVAWTGMGGGVPWPGTRQMPPFGPPGMFPGAGFGGDANEPPRIKCNLRKHKPNRKPRTPFTTQQLLSLEKKFREKQYLSIAERAEFSSSLRLTETQVKIWFQNRRA
KAKRLQ----------------------------EAEIEKIKMAALGRGAPGAQWAMAGYFHPSLMHLG------------------------------------------------------------
-------------------------------------------------------------------------------------------------
```

[PhD_A7_abbreviations]

**“PhD_A7_Abbreviations.pdf”**

# A7 Abbreviations

| | |
|---|---|
| cDNA | complementary DNA |
| DsRed | *Discosoma* species red fluorescent protein |
| dsRNA | double-strand RNA |
| EGFP | enhanced green fluorescent protein |
| hp | hairpin |
| mRNA | messenger RNA |
| siRNA | small interfering RNA |
| ORF | open reading frame |
| RNAi | RNA interference |
| UTR | untranslated region |
| WT | wild-type |