# Molecular Studies on Head Development of the Amphipod Crustacean *Parhyale hawaiensis*

Dissertation submitted in partial fulfillment of the requirements for the degree
"doctor rerum naturalium"
of the Georg-August-University Göttingen

from
Bernhard Schmid

München,
Germany

Göttingen, 2011

D7

| | |
|---|---|
| First referee/Referent: | Prof. Dr. Ernst A. Wimmer |
| Second referee/Korreferent: | Dr. Nikola-Michael Prpic-Schäper |
| Day of oral examination/<br>Tag der mündlichen Prüfung: | July 5, 2011/ 05.07.2011 |

Statement of Authorship

I hereby certify that the dissertation,

"**Molecular Studies on Head Development of the Amphipod Crustacean *Parhyale hawaiensis*",**

was composed by myself and is based on my work, with no other sources and help than quoted and acknowledged in the text.

Göttingen, 16. Juni 2011

(Bernhard Schmid)

Für

**Kristina,**

**Nadine**

und

**Stefanie**

# Acknowledgments

First of all, I want to deeply thank my family for their love and great support. You made it possible for me to follow this path.

I would like thank Prof. Dr. Ernst A. Wimmer, who allowed me to work on my project in the Department of Developmental Biology at the Georg-August-University. I want to express my deep gratitude for his excellent supervision, his great scientific and personal advice and his unwavering support.

I want to thank Gregor Bucher for his constant support, his great advice, his readiness to discuss and, above all, his friendship.

Thanks, big time, to Sebastian Heeger, for his support, his advice and help, for his critical review of parts of my thesis and for always giving me a safe haven when I needed one.

I want to express my thankfulness to my students, who I was privileged to supervise: Claudia Rödel, who was supporting me on *Ph awh*, Hendrikje Hein, who worked with me on *Ph hbn*, and Eva-Maria Rieß, who assisted me a lot with *Parhyale* work. My special thanks go to Jonas Schwirz, for his excellent work on *Parhyale knirps*, among many, many other shared projects. It was great working with you all!

Next, I want to thank William E. Browne for his collaboration and, in particular, for his excellent introduction to the technical side of *Parhyale* in the first place. Thanks for the time in Hawai'i!

Also, I want to thank all other members of the *Parhyale* community, for their support, advice and provision of plasmids, protocols and animals. In particular, I would like to thank Anastassios Pavlopoulos, Cassandra Extavour, Michalis Averof, Ron Parchem, Nipam Patel and Matthias Gerberding.

Last, but never least, I want to thank Maria Wiese, Evgenia Ntini, Beate Preitz, Marc Florian Schetelig, Kerstin Meier, Sebastian Kittelmann, Nico Posnien, Anna Gilles and all past and present members of the Department of Developmental Biology for such a great and inspiring time.

# Table of Contents


**1 Summary 1**

**2 Introduction 3**

**2.1 Comparative developmental studies in the field of Evo-Devo 4**
2.1.1 The hour glass model of embryonic development and the phylotypic stage 4
2.1.2 The Urbilaterian as an example of reconstructed ancestral populations 5
2.1.3 Plasticity vs. ancestrality: exemplifying challenges for the origin of insects 6

**2.2 Head development 7**
2.2.1 The developmental basis of the adult bilaterian head 7
2.2.2 Head segmentation 8
2.2.3 The vertebrate neural plate: the arthropod head equivalent? 10

**2.3 Phylogenetic relations of insects and crustaceans 11**

**2.4 Direct comparison of malacostracan and insect head segments and CNS elements 12**

**2.5 Characterisation of the amphipod *Parhyale hawaiensis* 14**
2.5.1 Ecology and habitat of *Parhyale hawaiensis* 14
2.5.2 General aspects of *Parhyale hawaiensis* development 15
2.5.3 The *Parhyale hawaiensis* body design 15
2.5.4 *Parhyale hawaiensis* mating behaviour 17
2.5.5 Early developmental processes of *Parhyale hawaiensis* 18
2.5.5.1 *Parhyale hawaiensis* development from S1 through S7 18
2.5.5.2 *Parhyale hawaiensis* germ disc condensation during S8 19
2.5.5.3 Emergence of head lobes in *Parhyale hawaiensis* at S9 19
2.5.6 Technical repertoire of the genetic system *Parhyale hawaiensis* 20

**2.6 Conserved genetic factors of head development 21**
2.6.1 *orthodenticle* as a conserved player in anterior head and brain development 22
2.6.2 *paired*-class homeodomain transcription factors and head development 23
2.6.3 *optix*/*six3* and *unplugged*/*gbx* genes: conserved interaction partners of *orthodenticle* 25
2.6.4 *knirps* genes as novel players in arthropod head development 27

**2.7 Aims of this work 29**

**3 Results 30**

**3.1 Hallmarks of head morphogenesis in *Parhyale hawaiensis*** 30
3.1.1 Establishment and regionalisation of the embryonic head lobes (S11 and S12) 30
3.1.2 Condensation and fusion of the head lobes (S13-15) 32
3.1.3 Establishment of lateral pre-antennal hemispheres, the stomodeum

and the labrum (S16-20) 35
3.1.4 Three-dimensional head character morphogenesis (S21-23) 39

**3.2 *Parhyale orthodenticle* and other *paired*-class homeobox genes involved in head development 43**
3.2.1 *Parhyale hawaiensis orthodenticle* genes 43
3.2.1.1 Isolation and characterisation of *Parhyale hawaiensis orthodenticle* genes 44
3.2.1.2 Isolation of *Ph otd1* 45
3.2.1.3 Characterisation of *Ph otd1* 46
3.2.1.4 Isolation and characterisation of *Ph otd2* 48
3.2.1.5 Phylogeny and Expression of *Ph otd1 and Ph otd2* 49
3.2.2 *Parhyale hawaiensis* genes of the *aristaless* group 51
3.2.2.1 Isolation and characterisation of *Parhyale hawaiensis aristaless* group genes 51
3.2.2.2 Isolation of *Ph hbn* 52
3.2.2.3 Characterisation of *Ph hbn* 53
3.2.2.4 Isolation of *Ph al1* 55
3.2.2.5 Characterisation of *Ph al1* 56
3.2.2.6 Isolation of *Ph al2* 58
3.2.2.7 Characterisation of *Ph al2* 59
3.2.2.8 Phylogeny of the *Parhyale hawaiensis aristaless* group genes *Ph hbn*, *Ph al1* and *Ph al2* 62
3.2.2.9 Expression of *Ph hbn* 65
3.2.2.10 Expression of *Ph al1* and *Ph al2* 73
3.2.3 *Parhyale hawaiensis pax*-box genes 79
3.2.3.1 Isolation and characterisation of *Parhyale hawaiensis pax-box* genes 79
3.2.3.2 Isolation of *Ph pby1* 80
3.2.3.3 Characterisation of *Ph pby1* 81
3.2.3.4 Phylogeny of *Ph pby1A* and *Ph pby1B* 86
3.2.3.5 Expression of *Ph pby1* 89
3.2.3.6 Expression of *Ph otd1* and *Ph pby1* at stages 8 and 9 91
3.2.4 *Parhyale hawaiensis arrowhead* 92
3.2.4.1 Isolation of *Ph awh* 93
3.2.4.2 Characterisation of *Ph awh* 94
3.2.4.3 Phylogeny of *Ph awh* 97
3.2.4.4 Expression of *Ph awh* 101

**3.3 *Parhyale hawaiensis sine oculis/Six* and *unplugged/gbx* genes 106**
3.3.1 *Parhyale hawaiensis sine oculis/Six* genes 106
3.3.1.1 Isolation and characterisation of *Parhyale hawaiensis so/Six* genes 106
3.3.1.2 Isolation of *Ph six3* 107
3.3.1.3 Characterisation of *Ph six3* 108
3.3.1.4 Isolation of *Ph six4* 111
3.3.1.5 Characterisation of *Ph six4* 111

3.3.1.6 Phylogeny of *Parhyale hawaiensis so/Six* genes 114
3.3.1.7 Expression of *Ph six3* 119
3.3.1.8 Expression of *Ph six4* 128
3.3.2 *Parhyale hawaiensis gbx* 135
3.3.2.1 Isolation of *Ph gbx* 135
3.3.2.2 Characterisation of *Ph gbx* 136
3.3.2.3 Phylogeny of *Ph gbx* 139
3.3.2.4 Expression of *Ph gbx* 143

**3.4 *Parhyale hawaiensis knirps* genes 146**
3.4.1 Isolation and characterisation of *Parhyale hawaiensis knirps* genes 146
3.4.2 Isolation of *Ph kni1* 147
3.4.3 Characterisation of *Ph kni1* 148
3.4.4 Isolation of *Ph kni2* 151
3.4.5 Characterisation of *Ph kni2* 151
3.4.6 Phylogeny of *Parhyale hawaiensis knirps* genes 155
3.4.7 Expression of *Ph kni1* 158
3.4.8 Expression of *Ph kni2* 173

**3.5 Techniques to address gene loss-of-function via RNA interference in *Parhyale hawaiensis* 178**
3.5.1 RNAi mediated loss-of-function by injection of dsRNA 178
3.5.1.1 RNAi mediated by injection of dsRNA in embryos 178
3.5.1.2 RNAi mediated by injection of dsRNA in adult females 179
3.5.2 RNAi mediated by heat shock inducible transgene-based expression of hairpin RNAs 180
3.5.3 Splicing of a *Ph kni2* intron in the context of an artificial EGFP expression construct 183
3.5.4 RNAi mediated loss-of-function by injection of siRNA 186
3.5.4.1 Assessment of unspecific injection phenotypes 187
3.5.4.2 Effects of siRNA injection targeting *Ph six3* 189
3.5.4.3 Effects of siRNA injection targeting *Ph kni1* 199
3.5.4.4 Addressing systemic siRNA-mediated RNAi 208
3.5.4.5 Conception of a feeding assay in order to examine behavioural phenotypes after siRNA-mediated RNAi 209

**4 Discussion 211**

**4.1 Molecular aspects of candidate gene isolation in *Parhyale hawaiensis* and its implications for comparative developmental analyses 211**
4.1.1 In need of an isogenic *Parhyale hawaiensis* wild-type population 211
4.1.2 Splicing diversity and transcript isoforms of *Parhyale* candidate genes for head development 212
4.1.3 Consequences of varying transcription factor family repertoires 212

**4.2 Evaluation of functional approaches for developmental studies in *Parhyale hawaiensis* 215**
4.2.1 Sensitivity of *Parhyale* to injection-based techniques 215

4.2.2 Feasibility of siRNA-mediated RNAi in *Parhyale* 216
4.2.3 Overcoming RNAi- and injection-related disadvantages of *Parhyale hawaiensis*: a technical perspective 220

**4.3 Anterior-posterior axial polarity: the amphipod way 220**

**4.4 Early anterior regionalization in *Parhyale hawaiensis* 223**
4.4.1 Involvement of *Ph otd1* in early anterior regionalization 223
4.4.2 *Ph hbn* represents a conserved factor involved in anterior head and brain patterning 225

**4.5 *Ph six3* and *Ph gbx* do not specify procephalic head regions 227**

**4.6 Development of the anterior median area and the stomodeal field in *Paryhale hawaiensis* 228**
4.6.1 Two populations of cells comprise the anterior median region in *Parhyale hawaiensis* 228
4.6.2 Reconciling stomodeal development based on *Ph six3*, *Ph six4* and *Ph kni1* expression 229
4.6.3 Preliminary *Ph six3* loss-of-function data suggests a crucial role in the development of median anterior protocerebral derivates and the foregut 231
4.6.4 *Ph six4* is expressed during mesoderm and muscle development and might contribute to an ancestral bilaterian anterior neuronal region 233

**4.7 *Ph awh* is a highly conserved representative of the family of LIM-HD encoding transcription factors 235**

**4.8 *Ph gbx* expression suggests a conserved role in deutocerebral patterning and VNS development 236**

**4.9 *Ph al1* is not involved in head development 238**

**4.10 Findings from *Parhyale knirps* genes suggest conserved roles in SNS development and appendage branching but not in head patterning and segmentation 240**

**4.11 Conclusions and Outlook 242**

**5 Materials and Methods 244**

**5.1 *Parhyale hawaiensis* culture 244**

**5.2 Molecular biology 245**
5.2.1 Isolation of *Parhyale hawaiensis* candidate gene homologs 245
5.2.2 Preparation of *Parhyale hawaiensis* cDNA collections 246
5.2.3 Polymerase chain reactions (PCR) 246
5.2.3.1 Varied-stringency degenerate PCR (vsdPCR) 247
5.2.3.2 Depletive vsdPCR 247
5.2.3.3 Rapid elongation of cDNA ends (RACE) 248
5.2.3.4 Long-distance PCR 248
5.2.4 Oligonucleotides 249
5.2.5 DNA purification 249

5.2.6 Cloning 249
5.2.6.1 Cloning of PCR amplificates 250
5.2.6.2 Cloning of *pSL_shuttle* 250
5.2.6.3 Cloning of *pMi[3xP3_DsRed; hsp70_hp-otd1]* 251
5.2.6.4 Cloning of *pMi[3xP3_EGFP; hsp70_hp-DsRed]* 252
5.2.6.5 Cloning of *pimpMiII[mse_EGFP]* 252
5.2.6.6 Cloning of *pimpMiII[mse_EG-kni2intron-FP]* 252

**5.3 Sequence analysis 253**
5.3.1 Sequence alignments 253
5.3.2 Phylogenetic calculations 253
5.3.3 Evaluation of gene sequences 254
5.3.3.1 Test of sequence novelty and assessment of identity and homology of gene sequences 254
5.3.3.2 Evaluation of sequence polymorphisms and transcript isoforms 255
5.3.3.3 Assessment of paralogy 256

**5.4 Embryology and Histology 256**
5.4.1 Harvesting of *Parhyale hawaiensis* embryos 256
5.4.2 Fixation and dissection of *Parhyale hawaiensis* embryos 256
5.4.3 Whole-mount *in situ* hybridization (WMISH) 257
5.4.4 WMISH targeting two different transcripts (double *in situ*) 258
5.4.5 Nuclear labelling and mounting of *Parhyale hawaiensis* embryos 259
5.4.6 Injection of *Parhyale hawaiensis* embryos 259
5.4.7 Injection of *Parhyale hawaiensis* adult females 260

**5.5 Genetics 261**
5.5.1 Transgenesis of *Parhyale hawaiensis* embryos 261
5.5.2 Conducting siRNA-mediated RNAi in *Parhyale hawaiensis* 261
5.5.3 Monitoring of injected embryos 262
5.5.4 Analysing pereopod phenotypes of hatched *Parhyale* after siRNA-mediated RNAi targeting *Ph kni1* 262

**5.6 Microscopy 263**

**6 References 264**

**A Appendix 283**
A1 Published data 283
A2-A6 Directory structure of the enclosed electronic appendix 314
A7 Abbreviations 316

**Curriculum Vitae 317**

# 1 Summary

The adult head of bilateral-symmetric animals is an anterior, usually tagmatised morphological unit. In segmented phyla, the development of the anterior head segments underlies different genetic and molecular mechanisms compared to the clade-specific segmentation networks of more posterior segments. Similar activity and functional conservation of several identified genetic regulators of head development suggest that elements of the underlying genetic network may have been already present in the bilaterian ancestor. However, modelling of an ancestral head development network has been impeded by fragmentary species sampling and distinct species-specific differences. To potentially identify conserved aspects of head development, I cloned and analysed several homologs of known head regulators from the amphipod crustacean *Parhyale hawaiensis*.

*orthodenticle* (*otd*) is discussed as an ancestral anterior determinant. In contrast to insects, *Ph otd1* is not active while anterior-posterior polarity is determined. Importantly, this process takes place during holoblastic cleavage in *Parhyale* which differs greatly from *the* superficial cleavage resulting in a syncytial blastoderm typical for insects. It is likely that *Ph otd1* patterns the prospective pre-antennal and ocular region similar to insects.

Anterior adjacent to *otd*, *optix/six3* genes are active in head and brain patterning in many species. In contrast, *Ph six3* expression is found in few cells that lie medially within the *Ph otd1* domain during regional head patterning. However, an anterior-median area like in insect blastoderm stages does not exist until *Parhyale* segmentation. It arises from two different populations of cells. During brain regionalisation, *Ph six3* is strongly expressed in median parts of the protocerebrum, the prospective hemi-ellipsoid bodies (HE). At this moment of development, *Ph otd1* and *Ph six3* are expressed in mutually exclusive regions, in fact reflecting the situation also found in other species. Accordingly, siRNA-mediated RNAi targeting *Ph six3* causes reduction of the HE and malformation of the foregut, but not deletion of anterior-most head cuticle and labrum, as observed for *Tribolium six3*.

Posterior to *otd*, *unplugged/gbx* genes are involved in early regionalisation (*Platynereis dumerilii*), in establishing the isthmic organiser via interaction with *otd*/*otx* (vertebrates) and in patterning the deutocerebral neuromere (insects). *Ph gbx* follows the insect situation.

Representatives of the *knirps* (*kni*) family of genes have so far only been found in arthropods. Although involved in head patterning and segmentation in insects, their exact functions diverge. *Ph kni1* and *Ph kni2* are not expressed in a comparable fashion. Their

expression indicates, though, that they share important roles in the development of the stomatogastric nervous system with insect *kni*. Intriguingly, loss of *Ph kni1* function during development affects the morphogenesis and arrangement of segmental appendages in hatched *Parhyale*. In *Drosophila*, *kni* is involved in tracheal development. *Tribolium kni* expression argues in favour of functional conservation of insect *kni* in this regard. Patterning of crustacean legs and insect tracheal placodes indicates a possible evolutionary link of these structures.

Several additional *paired*-class homeobox genes with putatively conserved roles in head development were examined in this work. With regard to their expression patterns, *Ph al1*, *Ph awh*, *Ph hbn* and *Ph six4* appear to act similar to their insect homologs.

The determination of expression patterns in *Parhyale hawaiensis* was practicable. Great efforts were made to address gene function by means of RNA interference, also in collaborating *Parhyale* laboratories. However, neither injection of double-strand RNA in embryos and adult females, nor transgene-based inducible expression of RNA hairpins or use of siRNA proved to be efficient or systematically applicable. Therefore, studies in *Parhyale* can provide valuable contributions to specific issues in comparative developmental and evolutionary biology, but its emergence as true model system is implausible.

# 2 Introduction

Darwin's „On the origin of species" (1859) marked a watershed moment in the history of natural science. Relying entirely on macroscopic observation of animals' morphology and the conclusions he drew from there, he formulated the scientific theory of evolution: that the great diversity of extant species evolved over the course of generations from common ancestral populations through mechanisms of natural selection. Although from today's view, his experimental and informational resources were extremely limited, a surprising lot of the hypotheses he introduced have been proven valid by recent studies (e.g., Cohen, 2010; Gayon, 2010; Presgraves, 2010).

Of course, the theory of evolution cannot be addressed by classical, manipulative experimentation. It does, however, allow to model the phylogeny of populations, species and clades. The predicted phylogenetic relationship of any two extant representatives, which reflects the path of natural selection, adaptation and radiation, can be tested by comparison of specific sets of characters and features. Traditionally, anatomic, morphological and embryologic characters were used to compare two species and determine the nature and the degree of their phylogenetic relation. This approach, however, is prone to biased interpretations (Scholtz and Edgecombe, 2006)

Today, a great repertoire of genetic, molecular and bioinformatic techniques are well-established. They are used to support or challenge the findings from traditional phylogenetic analyses (Giribet, 2002). Importantly, an increasing number of extant species are being made available for molecular approaches, filling the gaps in the still insufficiently sampled animal phylogeny (e.g., Mallatt and Giribet, 2006) and helping identify reliable and evident phylogenetic relations of species, clades and phyla (e.g., Dunn et al., 2008; Regier et al., 2010).

The field of comparative developmental biology seeks to address phylogenetic relation, thereby reconciling evolutionary processes, by comparing gene sets and their involvement in genetic networks in different species. Termed "Evo-Devo" by some (Raff, 2000) it profits greatly from emerging model systems, which become accessible for molecular, genetic and functional techniques.

## 2.1 Comparative developmental studies in the field of Evo-Devo

Inspired by von Baer's findings from comparative embryology (1828-1837), Garstang recognized the relevance of comparative developmental studies for animal phylogenies, when he postulated that it is the processes of embryonic development which provide the material for evolutionary adaptation, and hence, that ontogeny drives phylogeny (1922). Studies in the field of Evo-Devo pick up on this hypothesis as they follow molecular and genetic experimental approaches to demonstrate conserved and divergent aspects of embryonic development of different species and use these findings to evaluate the nature and degree of relation these species have on the base of these aspects (Gould, 1977). However, when addressing the phylogenetic relationship of two species by comparison of their development, several implications and caveats need to be considered, such as the comparability of different modes of embryogenesis depending on specific stages; the modeling of hypothetical ancestors and evolutionary plasticity, which may mask and impede the identification of ancestral traits.

### 2.1.1 The hour glass model of embryonic development and the phylotypic stage

Evolutionary adaptation to specific environmental parameters and ecological niches may result in ostensibly substantial differences in the modes of fertilization, oogenesis, egg lay and early development on the one hand, and highly divergent morphologies and structures in adult animals on the other, even among closely related species. Conversely, distantly related species may exhibit seemingly identical modes of early development and morphologically similar adult structures due to convergent developmental processes based on similar ecological requirements.

For example, polyembrony is found in several species of parasitic wasps, but not in closely related non-parasitic hymenopterans (Grbic, 2003). In contrast, bryozoans, which constitute a distantly related lophotrochozoan clade compared to the ecdysozoan wasps (insecta, arthropoda), exhibit polyembryony as well (Hughes et al., 2005). In the parasitic wasps, this specific mode of early zygotic development is used to create specified morphs or castes of siblings that fend off rival parasite species on a host, while the remaining fraction of embryonic clones develops into normal larvae (Donnell and Strand, 2006). This represents

just one example of an early developmental character that would infer misleading interpretations of phylogenetic relationships if used.

In contrast, von Baer's laws on embryonic development (1828), which he inferred from his observations in comparative embryology, suggested a phase of development during which the embryos of different species belonging to a given monophyletic clade are highly similar to each other with regard to morphology and inner organization. These findings are summarized and extended in the hour glass model of development which says that, depending on the taxonomic level regarded, embryos of different species undergo highly similar developmental processes during a specific stage of embryogenesis, while the developmental processes preceding and succeeding this stage are dominated by species-specific characteristics and, therefore, may diverge greatly (for an overview, see Domazet-Loso and Tautz, 2010; Kalinka et al., 2010). Sander (1983) suggested the term "phylotypic stage" for this moment high similarity during development (reviewed in Slack et al., 1993). For molecular comparative approaches, this implies that the genetic mechanisms underlying developmental processes that take place prior to the phylotypic stage and after it may show high divergence even within closely related species while the phylotypic stage itself represents a suitable interface for comparing these mechanisms.

### 2.1.2 The Urbilaterian as an example of reconstructed ancestral populations

The theory of evolution and, consequently, the models of animal phylogeny imply that any two species derive from a single ancestral population that already possessed the synapomorphic characters of the species considered. However, fossil record of these populations that would substantiate proposed phylogenies is not always available (Budd, 2008; Morris, 2000). In particular, the last common ancestor of all bilaterian species, termed the "Urbilaterian" (De Robertis, 2008) is a hypothetical model regarding conserved bilaterian aspects of development and morphology that this ancestor putatively exhibited. However, this model ancestor illustrates the various recent debates about common genetic and developmental repertoire of bilaterians. Undisputedly, the Urbilaterian possessed the genetic mechanisms to establish the anterior-posterior as well as the dorso-ventral body axis (De Robertis and Sasai, 1996; Finnerty, 2003), since these represent the conceptual prerequisites for bilaterian symmetry. Interestingly, the two major bilaterian phyla, protostomes (including

lophotrochozoans and ecdysozoans) and deuterostomes (including chordates) show an inversion of the dorso-ventral axis in relation to each other, which is corroborated by a highly conserved underlying genetic mechanism (Arendt and Nübler-Jung, 1994, 1999). Also, it is widely accepted that triploblasty (the existence of the ectodermal, mesodermal and endodermal germ layers), is a novelty shared by bilaterian species (Burton, 2008). In contrast, non-bilaterians, such as cnidarians, sponges and ctenophores, specify only two germ layers during development, ectoderm and endoderm (Martindale et al., 2004).

The questions whether the Urbilaterian had a condensed anterior brain, a basic mechanism of segmentation and proximal-distal appendages remain controversial. Numerous models including varying sets of ancestral characters and proposing convergent developments to different extents or secondary losses have been proposed (de Rosa et al., 2005; Kimmel, 1996; Miller and Ball, 2009; Valentine, 2006). In particular, the configuration of the anterior located head, comprising stomodeum, brain, sensory organs and specialized feeding appendages may be the consequence of highly conserved, ancestral genetic networks or the result of independent, convergent evolutionary developments that occurred several times during bilaterian radiation, possibly as a result of strong functional and ecologic constraints (Erwin and Davidson, 2002).

### 2.1.3 Plasticity vs. ancestrality: exemplifying challenges for the origin of insects

Defining the origin of insects within the phylum of arthropods exemplifies the problems of evolutionary and phylogenetic studies mentioned above. The hexapods which include the insect clade are a well defined monophyletic group (e.g., Regier et al., 2008; Timmermans et al., 2008). However, the positioning of them with regard to the other arthropod clades has been controversial, partly also because almost no fossil record exists (Glenner et al., 2006) and, therefore, the nature of a common ancestor of hexapods and another arthropod clade has long remained an issue of debate. A detailed description of the phylogeny of arthropods will be presented below (2.3). Here, I want to present the difficulty in identifying shared ancestral traits within groups of species as opposed to determining evolutionary plasticity that hampers the former.

With regard to arthropods, the genetic and molecular regulatory networks that underlie development are best understood in the higher dipteran *Drosophila melanogaster* (Wolpert et al., 2002). However, increasing understanding of these processes in non-dipteran insects, in particular the red flour beetle *Tribolium castaneum* (Posnien et al., 2010; Schröder et al., 2008), suggests that *Drosophila* exhibits a highly derived state of development that does, in many aspects, not represent an ancestral insect mode. In addition, even within higher dipterans, single elements of development may be subject to considerable plasticity, as was shown for the early expression of the head development gene *orthodenticle* (Schetelig et al., 2008), which is, as a molecular item, highly conserved within bilaterians (e.g., Ranade et al., 2008). In fact, this specific finding goes well with the implications and prognoses of the hour glass model of development and the concept of the phylotypic stage (2.1.1; Kalinka et al., 2010), since the plasticity of *otd* expression in higher dipterans is most evident for the earliest processes of embryogenesis (Schetelig et al., 2008). However, patterning processes that occur during the phylotypic stage of insects, i.e. the fully segmented germ rudiment (Sander, 1976, 1983), may also reveal substantial differences in the functions of individual genetic players. The evolution of the highly conserved *Hox*-gene *Sex combs reduced* (*Scr*) in hemimetabolous insects illustrates this (Passalacqua et al., 2010).

Two major procedures are currently discussed and used as solutions for addressing ambiguity of findings from comparative studies, as mentioned above. Increasing the sample number, i.e. the number of species of a clade that are examined, and analyzing the situation in an outgroup, i.e. a species that represents a sister group of the regarded clade, help overcome the problem of identifying ancestral traits of a group vs. aspects of evolutionary plasticity (e.g., see West-Eberhard, 2005).

## 2.2 Head development

### 2.2.1 The developmental basis of the adult bilaterian head

The “head” is a morphological and functional, usually tagmatised unit located at the anterior end of the main body axis of bilaterian species, in locomotive direction. Around the feeding opening, i.e. the stomodeum, and the oesophagus, specialised feeding organs and

appendages, supra- and suboesophageal elements of the condensed brain and sensory organs are arranged. Although the morphological appearance of adult bilaterian heads vary greatly, the underlying molecular and genetic mechanisms must provide comparable prerequisites for head establishment during development even in divergent species. This implies that comparative studies on head development need to focus on several coordinated and successive phases and processes during embryogenesis.

### 2.2.2 Head segmentation

Importantly, the genetic mechanisms underlying head development have to be closely linked to several other basic processes of bilaterian development. In this regard, head development requires relative anterior-posterior information and, from a molecular and genetic point of view, signalling input provided by the axial patterning systems of bilaterians. In segmented phyla, such as arthropods, annelids and chordates, the prospective head develops from a set of corresponding anterior segments (e.g., Minelli, 2001; Tautz, 2004). Interestingly, the modes of segmentation, in particular of the more posterior segments, vary considerably between these phyla: chordates show mesodermal segmentation resulting from a posterior oscillating system (reviewed in Pourquie, 2001; Rida et al., 2004), whereas annelids show ectodermal and mesodermal segmentation that results from subsequent budding processes (reviewed in Tautz, 2004). In contrast, arthropod species show ectodermal segmentation (e.g., Minelli, 2001; Tautz, 2004). It is an issue of current discussion whether the varying modes of ectodermal segmentation derive from a conserved arthropod-specific sub-regionalisation mechanism that works top-to-bottom, compared to the bottom-to-top mode of chordate segmentation, or if the differences observed in the segmentation of various arthropods represent independently evolved segmentation machineries (reviewed in Tautz, 2004).

Species belonging to the segmented phyla share an intriguing characteristic: although single elements of these phylum-specific segmentation mechanisms appear to be present during anterior segmentation in a given species (Jouve et al., 2002; Minelli, 2001; Tautz, 2004), these anterior segments, from which the head will arise, appear to depend in general on a less understood alternative genetic network compared to trunk segmentation (Tautz, 2004). In particular, in contrast to the divergent posterior segmentation mechanisms, this anterior,

based on its putative evolutionary origin also termed primary segmentation mechanism (Tautz, 2004), includes several highly conserved factors, even among distantly related bilaterians (2.6). Therefore, it has been argued that it supposedly represents an ancestral segmentation mechanism that has already been present in the bilaterian ancestor (Reichert and Simeone, 2001; Tautz, 2004). In detail, the distinct and different nature of ancestral anterior primary segmentation and divergent posterior secondary segmentation mechanisms is quite obvious when regarding insect head segments, as exemplified by the situation in *Drosophila melanogaster*.

The posterior head, the gnathocephalon, is constituted by the mandibular, the maxillary and the labial segments. While the mandibular segment reflects a developmental "hinge" between anterior and posterior patterning systems and integrates input from both (Cohen and Jürgens, 1990; Grossniklaus et al., 1994; Vincent et al., 1997), the maxillary and the labial segments are governed by the identical genetic network that is responsible for the trunk segments. In *Drosophila*, this network consists of a hierarchic regulatory cascade of maternal genes, gap genes, pair-rule genes and segment polarity genes (St Johnston and Nüsslein-Volhard, 1992). This patterning mechanism results in reiterated segments along the anterior-posterior axis of the embryo that acquire specific identities by *Hox*-gene activity (e.g., Lawrence and Morata, 1994; McGinnis and Krumlauf, 1992). In stark contrast, the anterior, procephalic segments of the head, namely the pre-antennal/ocular region, the antennal and the intercalary segments (Rogers and Kaufman, 1996, 1997), lack integral elements of this genetic cascade. Activity of pair-rule genes is absent in all procephalic segments. Instead, activity of head gap genes is required for procephalic segment polarity gene expression (Cohen and Jürgens, 1990; Crozatier et al., 1999; Grossniklaus et al., 1994; Wimmer et al., 1997; Wimmer et al., 1993). The intercalary is the anterior-most segment that is specified by a *Hox* gene, in particular by *labial* (Diederich et al., 1989; Merrill et al., 1989). Although the procephalic segments show genetic mechanisms involving the segment polarity genes, the regulatory details are unique for each procephalic segment boundary; in general, they differ greatly from the uniform genetic mechanisms of segment polarisation in the gnathocephalic and subsequent posterior segments (Gallitano-Mendel and Finkelstein, 1997).

However, even though the principal mechanisms of anterior segmentation in insects may resemble an ancestral bilaterian program, the factual gene regulation networks of anterior patterning and segmentation vary greatly among different insect species. To date, comprehensive data is available chiefly for *Drosophila melanogaster* and *Tribolium*

*castaneum* (e.g., Peel, 2008). In order to elucidate the conserved aspects of head development in insects, the situation in an outgroup needs to be examined. Hexapods, including insects, form a clade within the pancrustacea (2.3), i.e. they are closest related to traditional crustacean clades. For this reason, the amphipod *Parhyale hawaiensis* (malacostraca) represents a promising comparative system to identify conserved elements of head development in insects.

### 2.2.3 The vertebrate neural plate: the arthropod head equivalent?

Findings from studying head development in a non-insect arthropod species, such as *Parhyale hawaiensis*, may well provide important conclusions not only for the situation in the insect, pancrustacean (2.3) or arthropod ancestor, but also for representatives of distantly related bilaterian phyla. In this regard, vertebrate models, such as zebrafish or mouse, are intensively studied and of great relevance for comparative developmental studies (e.g., Bally-Cuif and Boncinelli, 1997; Schilling and Knight, 2001). However, the vertebrate head cannot be easily homologized with the insect head segments.

Specifically, only the developing vertebrate brain shows elements of segment-like regional patterning and subdivision (Rubenstein et al., 1994; Shimamura et al., 1995). It develops from the neural plate, which resembles the vertebrate neurectodermal platform (reviewed in Colas and Schoenwolf, 2001; Smith and Schoenwolf, 1997; Vieira et al., 2010) and is specified dorsally through dorso-ventral patterning, similar to the neurectoderm of arthropods (reviewed in Lichtneckert and Reichert, 2005). While the vertebrate neural plate invaginates to form the neural tube (for a detailed overview, see Gilbert, 2006), the nervous system of arthropods is formed by entirely different developmental processes. In insects and malacostracans, neuroblast populations arise from the neurectoderm via segregation. During neurogenesis, brain and other central nervous system structures are formed from these (e.g., Ungerer and Scholtz, 2008).

In contrast, many ectodermal structures and cells of the vertebrate head, including melanocytes, craniofacial cartilage, bone and muscle tissue are generated by migratory cells of the neural crest (Graham et al., 2004; Hanken and Gross, 2005; Schilling, 1997). The neural crest is specified through signalling from the epidermis and the neural tube (reviewed in Knecht and Bronner-Fraser, 2002). Conclusively, I will treat the vertebrate neural plate and

its derivatives as the correspondent developmental field for comparing bilaterian head development.

## 2.3 Phylogenetic relations of insects and crustaceans

Traditionally, crustaceans were believed to constitute one major arthropod sub-phylum along the other sub-phyla hexapods (including insects), myriapods and chelicerates (for an overview, see Edgecombe, 2010). For comparative analyses of head development, using the "crustacean" *Parhyale hawaiensis* as an outgroup of hexapods, recent findings regarding the phylogenetic relation of these two clades has to be considered carefully. Hexapods and, in particular, insects represent a well-supported group of terrestrial arthropods (e.g., Regier et al., 2008; Timmermans et al., 2008). However, the evolutionary origin of insects is highly debated, partially also because fossil data that would allow to link them to the other major arthropod sub-phyla is almost completely absent (Glenner et al., 2006).

According to traditional models, hexapods have been considered close relatives of myriapods, forming the taxon "atelocerata" (e.g., Telford and Thomas, 1995). This phylogenetic position was based on highly similar morphological characters that both exclusively terrestrial groups share, in particular the head appendages and nature of their tracheal respiration systems (reviewed in Edgecombe, 2010). However, recent morphological and, importantly, novel molecular data suggest a close, monophyletic relationship of hexapods and crustaceans (e.g., Dunn et al., 2008; Regier et al., 2010). The clade that combines these two groups was originally termed "pancrustacea" (Shultz and Regier, 2000). The fact that crustaceans and hexapods have a unique structure of ommatidia in the compound eye in common, has coined the alternative name "tetraconata" for this group (Dohle, 2001). In this work, I will refer to the term "pancrustacea".

Apart from that, the concept of crustaceans as a monophyletic group has been challenged: phylogenetic analyses based on broad molecular sampling suggest several crustacean paraphyla, positioning hexapods as one of these within the pancrustacea (e.g., Koenemann et al., 2010; Regier et al., 2005; Regier et al., 2010). The question which crustacean class embodies the sister group to the hexapods is an issue of ongoing debate. Comparative analyses of neuronal characters place the hexapods close to the malacostracans

which include the amphipod *Parhyale hawaiensis* (Strausfeld, 2005). Alternatively, phylogenetic calculations have placed the branchipods (including the water flea *Daphnia pulex* and the fairy shrimp *Artemia salina*) closer to the hexapods than the malacostracans (Regier et al., 2005). A recently performed comprehensive phylogenomic analysis of nuclear protein-coding sequences supports a group termed "Xenocarida" (comprising the more enigmatic crustacean taxa Remipedia and Cephalocarida) as the sister group of hexapods. According to this model, malacostracans and branchipods (constituting the group "Vericrustacea") are phylogenetically equidistant to insects and represent the sister group of the clade combining hexapods and Xenocarida, which has been named "Miracrustacea" (Regier et al., 2010). Although, accordingly, the exact phylogenetic relations of insects and amphipods remain heavily debated, it is apparent that within arthropods, *Parhyale hawaiensis* represents a well-suited and closely related outgroup to the better-studied insects *Drosophila* and *Tribolium*.

## 2.4 Direct comparison of malacostracan and insect head segments and CNS elements

Although adult insect head morphologies may vary greatly as a consequence of specific functional requirements, they derive from the same set of segments, each of which acquires a typical, insect-specific identity. Even within arthropods, the segments of the anterior part of the embryo can be homologized, although they acquire entirely different functional and appendage properties (reviewed in Scholtz and Edgecombe, 2006).

As a basic consideration for comparing head development in insects with the outgroup *Parhyale hawaiensis*, I want to directly compare and homologize anterior head segments from insect and malacostracan body plans. The anterior most region of both insects and malacostracans is the pre-antennal/ocular segmental region (reviewed in Scholtz and Edgecombe, 2006). There is open debate of as how many segments, probably cryptic ones, are part of the pre-antennal area. Insects have been suggested to have at least an ocular and possibly a labral segmental unit (Scholtz and Edgecombe, 2006). In *Parhyale*, engrailed activity has been found in a spot putatively indicating the eye field. This suggests a similar situation as in insects, possibly pointing to an ocular segmental remnant (Browne et al., 2005).

In this work, I will refer to all parts anterior of the *Parhyale* first antennal segment as pre-antennal.

The first antennal segment of a malacostracan corresponds to the antennal segment of insects. Accordingly, the second antennal segment corresponds to the intercalary segment in insects which does not carry appendages. The proposed tripartition of the supraoesophageal brain in pancrustaceans (e.g., Hirth, 2010; Hirth et al., 2003; Lichtneckert and Reichert, 2005; Scholtz and Edgecombe, 2006) goes well with this: the intercalary/ second antennal segment contributes the tritocerebrum, the antennal/1$^{st}$ antennal to the deutocerebrum and the pre-antennal head region of both insects and crustaceans to the protocerebrum. The protocerebrum has been suggested to further subdivide in the prosocerebrum and the originally anterior-most archicerebrum (Damen et al., 1998; Remane et al., 1975).

Posterior of the second antennal/intercalary segment, the mandibular segment can be identified in all pancrustaceans. It is the anterior-most gnathocephalic segment. Following posterior, the insect head includes the maxillary segment, corresponding to the first maxillary segment in malacostracans, and the labial segment, corresponding to the second maxillary malacostracan segment. The adult *Parhyale* head tagma also includes the first thoracic segment which bears maxillipeds. In insects, the first thoracic segment carries a pair of walking legs. Interestingly, the transition of procephalon and gnathocephalon spans the same segments in insects and malacostracans. The distinction between procephalic and gnathocephalic segments is paralleled in the anterior brain: derivates of the procephalic segments form the supraoesophageal tripartite brain, with the tritocerebrum retaining one suboesophageal neuropil, while the ganglia of the gnathocephalic segments are entirely suboesophageal. This resembles the situation in all pancrustaceans (Harzsch, 2004; Scholtz and Edgecombe, 2006). In contrast to the pancrustacean brain design, the position of the stomodeum in relation the segments homologous to the procephalon is shifted anterior by one segment in chelicerates (Damen et al., 1998; Telford and Thomas, 1998).

Free-swimming Nauplius larvae occur in various pancrustacean lineages. They comprise the procephalic segments and the anterior gnathocephalic segment, i.e. the mandible (e.g., Williams, 1994). Therefore, these segments are also referred to as naupliar segments (reviewed in Minelli, 2001). Although *Parhyale*, like all amphipods, undergoes direct development, i.e. does not develop into a free-swimming larva, this lay-out is mirrored by the morphology of the *Parhyale* germ rudiment at S10, before trunk segmentation is initiated (Browne et al., 2005; Williams, 1994). In this process, the ectoderm precursors arrange into a

regular grid, beginning anteriorly at the level of the mandibular segment (Browne et al., 2005).

## 2.5 Characterisation of the amphipod *Parhyale hawaiensis*

**Figure 1:**

***Parhyale hawaiensis*.**

Shown are adult wild-type *Parhyale hawaiensis*. The larger animal on top is male, the smaller one below is female. *Parhyale hawaiensis* exhibits sexual dimorphism, e.g. in the subchelate gnathopods, which are segmental appendages of T2 and T3.

### 2.5.1 Ecology and habitat of *Parhyale hawaiensis*

*Parhyale hawaiensis* (Figure 1) is a marine amphipod. It is found in tropical oceans around the world, where it aggregates in dense populations in intertidal and shallow water habitats. *Parhyale hawaiensis* are detritovorous. In their natural ecological environment, they thrive even under rapid changes in temperature, salinity and water quality in general.

Today's laboratory populations derive from specimen that were isolated from the John G. Shedd Aquarium, Chicago, IL (Browne et al., 2005), where they were considered a pest species in water filter systems. These characteristics make *Parhyale* well-suited for rearing under laboratory conditions (Rehm et al., 2009b).

### 2.5.2 General aspects of *Parhyale hawaiensis* development

*Parhyale hawaiensis* develops directly. This means that no free-swimming larva exists; instead, newly hatched *Parhyale* exhibit adult morphology with regard to a complete set of segments and fully developed appendages, organs and cuticle (Figure 98). The duration of Parhyale embryogenesis is relatively short (250h at 26°C). Detailed staging information is available: in detail, 30 stages (S1-S30) have been described (Browne et al., 2005). They explain subsequent and partially overlapping developmental processes of early cleavage, gastrulation, germ band formation and segmentation, germ cell migration, appendage development, body morphogenesis, cuticularisation (see also Havemann et al., 2008), mesoderm differentiation, myogenesis and gut development, CNS development and organogenesis.

Specifically, individual stages of *Parhyale* embryogenesis were defined differently on the base of various features, such as visible early cleavage events (S1-S5), general embryonic morphology (e.g. "soccerball stage" S6 and "rosette stage" S7), emergence of specific morphological traits (e.g. first appearance of head lobes marking the onset of S9, emergence of the dorsal organ at S10 or the proctodeum becoming first visible at S21), developmental and cellular processes (e.g. germ disc condensation during S8, germband row formation at S11 or movement and splitting of the germ cell cluster during S13-S16) and morphogenetic events regarding appendages (e.g. the initiation of limb bud morphogenesis at S15 or the An2 shape change at S22) as well as the gradual folding of the germ band (e.g. the embryo first appearing comma shaped at S18). As a consequence, several stages mark distinct points of time during *Parhyale* development (e.g. S9, S17 or S27), while others cover a continuous phase of *Parhyale* embryogenesis (e.g. S12 or S16). In this work, this staging system will be used with the following adjustment: all stages will be treated as temporal phases with one given stage ending at the onset of the next stage (e.g. considering S8, germ disc condensation, ending as soon as the embryo's head lobes first appear at S9).

### 2.5.3 The *Parhyale hawaiensis* body design

*Parhyale hawaiensis* is a typical representative of the clade amphipoda, in that it exhibits the pivotal synapomorphic character of this group: bidirectional orientation of the

pereopods, i.e. the appendages of the fourth to eighth thoracic segments (T4-T8) in relation to the body axis, with the T4 and T5 pereopods oriented anterior and the T6-T8 pereopods posterior. With other amphipods, *Parhyale* also shares sessile compound eyes, large coxal plates and a lateral compression of the body (Figure 1; Browne et al., 2005).

During embryogenesis, reiterated segments are established in anterior-posterior sequence. Except the pre-antennal segment and the asegmental telson, all segments bear paired appendages. Although *Parhyale* develops directly, naupliar segments emerge earlier than post-naupliar ones. This is reminiscent of Nauplia larvae found in other crustaceans and also likely reflects two different segmentation mechanisms, primary and secondary segmentation (reviewed in Minelli, 2001; Tautz, 2004). From anterior to posterior, the naupliar segments are: preantennal (pAn, carrying labrum and eyes), first antennal (An1), second antennal (An2, corresponding to the intercalary segment in insects, 2.4) and mandibular (Mn). The first antennae are also referred to as the antennules, as compared to the antennae (An2), strictly speaking. During secondary segmentation, the following segments are established (anterior to posterior): first (Mx1) and second maxillary (Mx2, corresponding to the labial segment in insects), followed by eight thoracic segments, the first of which carries the maxillipeds (T1), the second and third carrying subchelate gnathopods (T2, T3) and the remaining carrying pereopods (T4-T8). T2-T8 form the pereon. The appendages of T4 to T8 carry specialised epipods conferring respiratory function, the gills. Posterior to those, six abdominal segments and the telson follow: the anterior three abdominal segments carry pleopods (A1-A3, constituting the pleon), and the posterior three carry uropods (A4-A6, forming the urosome). The telson does not carry paired appendages (Browne et al., 2005). The nature of the gnathocephalic and the T1 appendages reflects the feeding mode of *Parhyale*: food is transported from posterior to anterior, passing all feeding appendages until it reaches the stomodeum. In adult animals, the segments pAn through T1 form the cephalon or head, with the respective segmental appendages concentrated as the morphologically distinct buccal mass (Browne et al., 2005).

*Parhyale* exhibits sexual dimorphism. Male adults have specialised gnathopods (T3) that are used for mating (2.5.4). The endites of females' T2-T5 appendages constitute the brood pouch, into which the fertilised eggs are deposited (Browne et al., 2005).

### 2.5.4 *Parhyale hawaiensis* mating behaviour

Adult *Parhyale* reach sexual maturity ca. 6 weeks after hatching. Mature adults reproduce throughout the year, roughly every 2-3 weeks (Browne et al., 2005). In detail, sexually mature male and female Parhyale hawaiensis enter premating amplexus with the larger male grasping the female with its gnathopods (sexually dimorphic T3 appendages). A mating pair remains in this position over several days, until the female moults. At this moment, the male deposits its sperm into the female's paired oviducts and subsequently releases her. The female sheds her eggs through the oviducts into a ventral marsupium, or brood pouch, fertilizing them in the process (see also Figure 2). After that, the female's cuticle hardens. The brood pouch is composed of the specialized appendage endites of T2-T5. Embryos of one batch usually develop synchronously. Their numbers may range from 1 up to 25 (Browne et al., 2005).

**Figure 2: Ovigerous female *Parhyale hawaiensis* (top) and shed cuticle (bottom).** Until hatching, *Parhyale* eggs develop in the mother's brood pouch.

### 2.5.5 Early developmental processes of *Parhyale hawaiensis*

In *Parhyale*, the intial cleavages of the zygote are total. Two slightly inequal cleavages and a subsequent distinctly inequal cleavage give rise to four macromeres and four micromeres, each of which exhibit an unambiguous morphology. As shown by cell lineage experiments, each of these eight cells gives rise to specific progeny having invariant cell fates (Gerberding et al., 2002). Importantly, this shows that at this early time of development, anterior-posterior, dorso-ventral and bilateral polarity has been established in *Parhyale*. This mode of early cleavage is specific to amphipods and is believed to have evolved convergently to the mode of holoblastic cleavage found in lophotrochozoans and annelids. Apart from amphipods, several other malacostracan taxa exhibit holoblastic cleavages during the intial phase of embryogenesis as well (e.g., Biffis et al., 2009; Hertzler et al., 1994). These cases are also believed to resemble secondary cleavage evolution (Scholtz and Dohle, 1996). In contrast, *Drosophila* and *Tribolium* embryos undergo intial syncytial cleavages that result in an early syncytial blastoderm (e.g., Lecuit, 2004; Schröder et al., 2008).

#### 2.5.5.1 *Parhyale hawaiensis* development from S1 through S7

As mentioned above, the implementation of axial polarity is preceded by the specification of cell lineage founders within the S4 *Parhyale* embryo. At this stage, the *Parhyale* embryo is composed of four macromeres and four micromeres. From each of these cells, specific germ layers or germ layer fractions encompassing tissues with 'polarised' positional identity derive (Gerberding et al., 2002). In detail, three ectodermal precursors will subsequently contribute to the developing germ band, one to the left embryonic hemisphere (El), one to the right (Er) and the remaining precursor to the posterior part of the embryo (Ep). As a consequence, the establishment of anterior-posterior polarity starts conceptually within the S4 embryo due to the specification of the Ep macromere. This polarity is maintained during subsequent cleavages leading to the 'soccerball stage' (S6) and to gastrulation, as derivatives of the three ectodermal founders retain the relative position to each other within the embryo (Gerberding et al., 2002). During early gastrulation ('rosette stage', S7) mesodermal and germ line cells cluster at the anterior of the condensing germ disc, indicating

the anterior-posterior polarity of the embryo and marking the site where head structures will be formed (Browne et al., 2005; Gerberding et al., 2002).

#### 2.5.5.2 *Parhyale hawaiensis* germ disc condensation during S8

As a result of the completion of gastrulation, the germ disc aggregates at the anterior ventral region of the egg during S8 (Browne et al., 2005). It covers slightly more than a third of the embryo's surface and is characterised by evenly distributed, large cells that contain coin-shaped nuclei. Around the periphery of the germ disc, single cells of similar morphology are scattered. The anterior-posterior polarity of the germ disc is reflected by its relative position within the embryo, with the posterior end expanding farther ventrally than the anterior end does dorsally (Browne et al., 2005).

#### 2.5.5.3 Emergence of head lobes in *Parhyale hawaiensis* at S9

As embryogenesis proceeds to stage 9, the germ disc elongates along the anterior-posterior axis and acquires a distinct tripartite morphology that comprises two emerging head lobes and the post-naupliar, posterior region from which trunk segmentation originates. The head lobes represent the developing left and right head hemispheres that encompass the prospective pre-antennal region of the embryo as well as the first and second antennal segments. The site of the future mandibular segment, which is considered the most posterior naupliar segment in crustaceans (reviewed in Minelli, 2001), shows no hemispheric division and represents the base from which both head lobes expand towards anterior. The emergence of the extra-embryonic dorsal organ, composed of cells found anterior of the developing embryonic head, marks the onset of stage 10 of *Parhyale* embryogenesis (Browne et al., 2005).

### 2.5.6 Technical repertoire of the genetic system *Parhyale hawaiensis*

Several characteristics make *Parhyale hawaiensis* amenable to embryological and molecular techniques (Rehm et al., 2009b). Batches of synchronously developing embryos can be obtained from ovigerous females shortly after fertilisation, i.e. before they undergo first cleavages. They can be incubated outside the brood pouch until they reach a desired age. In addition, the embryos are relatively large (size) and therefore accessible for injection-based techniques (Rehm et al., 2009d). The fact that from S6, the pigmented yolk fraction of an embryo is separated from the opaque germ rudiment enables fast and reliable staging of an individual embryo (Browne et al., 2005).

Specifically developed and adjusted histological techniques (whole mount *in situ* hybridisation and antibody staining) allow for visualisation of transcript and protein expression (MM; e.g. Browne et al., 2006; Prpic and Telford, 2008; Rehm et al., 2009a; for a technical overview, see Rehm et al., 2009c). *pMinos*-based stable transgenesis can be applied to *Parhyale*. *Parhyale*-specific genetic drivers, in particular an *hsp70b*-derived heat shock element as well as a muscle-specific enhancer element, have been isolated and can be used for transgenesis of expression constructs (Pavlopoulos and Averof, 2005; Pavlopoulos et al., 2009). *Parhyale* embryogenesis starts with a series of total and unequal cleavages. These lead to four micromeres and four macromeres, each of which found invariant cell lineages (Gerberding et al., 2002). This particular feature can be used for clonal analyses (e.g., Extavour, 2005; Price et al., 2010). Adult *Parhyale* are only weakly pigmented and therefore transparent, enabling easy detection of fluorescent dyes and markers (Browne et al., 2005; Pavlopoulos and Averof, 2005). The amphipod body plan of *Parhyale* (2.5.3) opens various possibilities to study morphological phenotypes.

However, although attempts to apply siRNA and morpholinos for performing loss-of-function experiments have provided preliminary results (Liubicich et al., 2009; Ozhan-Kizil et al., 2009) and injection of capped mRNA has been successful in causing ectopic expression in *Parhyale* (Pavlopoulos et al., 2009), functional techniques had not been available at the beginning of this work. Still, the development of versatile, robust functional techniques is far from being completed.

## 2.6 Conserved genetic factors of head development

Comprehensive analyses of larval cuticular phenotypes generated through a large-scale genetic mutagenesis screen in *Drosophila melanogaster* (Nüsslein-Volhard and Wieschaus, 1980) provided first candidates for the molecular and genetic mechanisms underlying head development (St Johnston and Nüsslein-Volhard, 1992). Homologs of several of these candidate head development genes have been identified in the majority of well-studied bilaterian species and, importantly, confer similar functions (Acampora et al., 2000a; Finkelstein and Boncinelli, 1994; Hanson, 2001; Kozmik, 2008). These findings strengthen the idea that an anterior located head exhibiting several key features, such as condensed brain, a basic repertoire of sensory organs and maybe specialised feeding organs surrounding the stomodeum, was already present in the bilaterian ancestor, and, in addition, that principal elements of the genetic network governing bilaterian head development may be represented by these conserved high-level hierarchic key players.

However, the individual evolutionary paths extant species have followed during their diversification and adaption imply unknown levels of plasticity with regard to the genetic regulation mechanism within this network, and therefore bear the risk of bias and misinterpretation when performing comparative studies in head development (Scholtz and Edgecombe, 2006). Even within closely related species of a monophylic clade, conserved players of head development may show a discrete functional diversity, such as being involved in different hierarchic levels of head development, covering additional roles in development or having lost ancestral functions. This problem can be addressed by examining the situation of these factors in an outgroup species.

In this work, several of these conserved players in head development were identified and analysed in the amphipod *Parhyale hawaiensis*, which, accordingly, represents a pancrustacean outgroup species for comparative studies with insects. In the following, I will introduce them and characterise their shared and divergent functions in detail.

### 2.6.1 *orthodenticle* as a conserved player in anterior head and brain development

The recognition that *Drosophila melanogaster orthodenticle (otd)* (Cohen and Jürgens, 1990) and its mouse homologs, especially *Otx2* (Simeone et al., 1992; Simeone et al., 2002), both share a comparable, anterior region of activity and highly similar loss-of-function phenotypes, in particular the deletion of the anterior part of the head, encompassing the protocerebrum or the forebrain, respectively (Kammermeier and Reichert, 2001), has strengthened the hypothesis of a conserved anterior head and/or brain patterning mechanism in bilaterians. *otd* genes encode paired-class homeodomain transcription factors (Treisman et al., 1992). Recent phylogenomic evidence suggests that they were already present in the cnidarian-bilaterian ancestor (Ryan et al., 2006). Among extant bilaterian species, the number of *otd* paralogs may vary. For example, single *otd* representatives have been identified from *Drosophila melanogaster* (Finkelstein and Perrimon, 1990; Finkelstein et al., 1990) and *Platynereis dumerilii* (Pdu-otx, Steinmetz et al., 2011), while two *otd* paralogs are known from *Tribolium castaneum* (Tc'otd1 and Tc'otd2, Schinko et al., 2008; Schröder, 2003) and four from mouse (Otx1, Otx2, Otx3, Crx, Chen et al., 1997; Simeone et al., 2002; Zhang et al., 2002).

In accordance with the initial findings from *Drosophila* and mouse, the conserved role of *otd* genes in bilaterians is now, that data from many more species has been gathered, impressively illustrated by a common anterior expression domain during early patterning in embryogenesis, that excludes just the anterior end of the embryo (Acampora et al., 1998; Finkelstein and Boncinelli, 1994) as well as, correspondingly, highly conserved activity in anterior regions of the brain. In bilaterian species exhibiting ectodermal segmentation, such as annelids and arthropods, the early anterior domain of *otd* expression typically lies within the anterior-most presumptive segment or segmental region (e.g., Finkelstein and Boncinelli, 1994; Steinmetz et al., 2011). Similar to *Drosophila*, for example, *Tc'otd1* expression arises during blastoderm stage within the prospective pre-antennal or ocular segmental region (Schinko et al., 2008). In mouse, the regionalisation of the anterior neural plate is dependent on *Otx2* activity (Acampora et al., 2000a; Kurokawa et al., 2004a; Kurokawa et al., 2004b). Nested expression domains of *Otx1* and *Otx2* are found in the developing forebrain, again excluding the most rostral part of the forebrain (Simeone et al., 1992). In contrast to these aspects of extreme conservation, a detailed comparison of *otd/otx* genes can also be used to

exemplify the varying extents of plasticity of expression patterns and functions of homologous genes in different bilaterian species. In this regard, *otd* expression varies even within closely related higher dipterans, ranging from maternal and anterior localized *otd* expression (*Anastrepha suspensa otd*), maternal, but evenly distributed expression (*Ceratitis capitata otd*) or strictly zygotic expression in *Drosophila* (Schetelig et al., 2008). Interestingly, *Drosophila otd* is under control of the evolutionarily novel anterior determinant *bicoid* (bcd, Finkelstein and Perrimon, 1990; McGregor, 2005). In the parasitic wasp *Nasonia vitripennis*, *otd1* is localised both anterior and posterior within the blastodermal embryo, providing patterning information for both poles (Lynch et al., 2006). Apart from these drastic differences, *otd* genes show also smaller variations in expression and function. E.g., in contrast to *Tc'otd1*, whose expression is restricted to the pre-antennal/ocular segment during germ band stages, *Drosophila otd* expression reaches into the anterior part of the antennal segment (Cohen and Jürgens, 1990).

To summarise and conclude this overview, *otd/otx* genes share conserved functions in anterior determination, cephalisation, segmentation of the anterior head and regionalisation of the anterior brain as well as differentiation of neurons and sensory organs, in particular photoreceptors (e.g., Ranade et al., 2008). At least the involvement of *otd/otx* genes in early anterior patterning and segmentation as well as anterior brain regionalisation appear to represent highly conserved developmental processes in bilaterians. However, even within insects, *otd* genes exhibit in part drastic variations in expression and function. To elucidate the putative default situation of insect *otd* genes, the *otd* repertoire of the outgroup pancrustacean *Parhyale hawaiensis* was examined in detail and compared to the insect data in this work.

### 2.6.2 *paired*-class homeodomain transcription factors and head development

The DNA-binding properties of *otd/Otx* genes are mediated by a *paired*-class homeodomain (Finkelstein et al., 1990; Treisman et al., 1992). This protein domain is characteristic for a large family of transcription factors constituted by several other gene families, aside from *otd/Otx* genes (Galliot et al., 1999; Treisman et al., 1992). In many species, specific representatives of these groups of paired-class homeodomain encoding genes are also involved in head development (for an overview, see Holland and Takahashi, 2005). As with *otd/otx* genes, there is increasing evidence from phylogenetic analyses that they have

been part of the genetic repertoire of the bilaterian ancestor (Ryan et al., 2006). This suggests that at least several of the genes might exhibit conserved roles in head development. Typically, members of one of these groups are characterised by unique protein motifs and domains found in addition to the *paired*-class homeodomain (see below). Due to the close sequence similarity of *paired*-class homeoboxes in general, homologs of several of these genes were isolated along the *Paryhale otd* genes in this work. For these reasons, I want to characterise some of the additional *paired*-class homeodomain encoding gene families in more detail, in particular the *paired*-box (PAX) group of genes, the *aristaless* group of genes and the LIM domain encoding group of genes.

Several members of the *paired/pax* subgroup of paired-like homeodomain transcription factors convey important and conserved functions in head and brain development. In *Drosophila* and mouse, *eyeless/pax6* genes are involved in neuronal specification and eye development (e.g., Hanson, 2001). In vertebrates, *Pax2/5/8* genes are part of the isthmic organiser, also referred to as midbrain-hindbrain-organiser (MHO, reviewed in Raible and Brand, 2004). In this work, however, no *Parhyale* homologs of these genes were identified. Instead, *Parhyale* homologs of *Pax3/7/paired* genes which are involved in segmentation (reviewed in Noll, 1993) were isolated. One of these homologs, *Parhyale pairberry1* (*Ph pby1*), was used to analyse early specification processes of head as opposed to trunk in *Parhyale*.

In *Drosophila melanogaster*, the *aristaless* subgroup of *paired*-like homeobox genes includes *aristaless* (*al*), *homeobrain* (*hbn*), *goosecoid* (*gsc*), *munster* (*mu*), *orthopedia* (*otp*), *reverse polarity* (*repo*) and *Retinal Homeobox* (*Rx*) (Mazza et al., 2010; Schneitz et al., 1993; Walldorf et al., 2000). The proteins encoded by these genes share several characteristic motifs and properties, although to varying extents: An N-terminal transcriptional repression domain (octapeptide/GEH) is encoded by *hbn*, *Rx* and *gsc*. The genes *al*, *otp*, *mu* and *Rx* encode a common OAR/aristaless domain in the C-termini of the proteins (Walldorf et al., 2000). Specifically, the *Parhyale* homologs of *hbn* and *al* were isolated and analysed in this work. *Drosophila* and *Tribolium hbn* share expression in the anterior head of the embryo, beginning at blastoderm stage and remaining locally restricted during embryogenesis (Walldorf et al., 2000; Sebastian Kittelmann, personal communication). While *Drosophila al* is required for the patterning of the imaginal discs of antennae and legs but has no known function in head or brain development (Schneitz et al., 1993), the vertebrate *al* homolog *arx* is expressed in neuromeric patterns in various parts of the developing brain (El-Hodiri et al., 2003; Miura et

al., 1997). In humans, several hereditary diseases exhibiting phenotypes of intellectual disability have been linked to mutations in the *arx* gene (e.g., Shoubridge et al., 2010). Importantly, neuronal expression of an *aristaless* homologue has also been described in an ecdysozoan species, the nematode *C. elegans* (Melkman and Sengupta, 2005).

From the LIM domain encoding subgroup of *paired*-class homeobox genes, the *Parhyale* homolog of *arrowhead/Lhx6/8* was cloned and analysed in this work. In mouse, Lhx6/8 genes are important for early forebrain patterning and neuronal differentiation (Grigoriou et al., 1998; Manabe et al., 2007; Manabe et al., 2008; Zhao et al., 2003). Although the *Drosophila* homolog *arrowhead* shows head and brain expression, the only reported function to date is in the specification of imaginal discs (Curtiss and Heilig, 1995, 1997).

### 2.6.3 *optix/six3* and *unplugged/gbx* genes: conserved interaction partners of *orthodenticle*

As described above, *otd/Otx* genes may represent ancestral anterior regionalisation factors. In *Platynereis dumerilii*, the region anterior to the *otd* domain expresses *optix/six3* genes (Steinmetz et al., 2010), while posterior of the *otd*-positive region in the developing head, genes of the *unpg/gbx* group are active (Steinmetz et al., 2011). In many bilaterian species, expression of *optix/six3* and *unpg/gbx* genes are reminiscent of this situation (Lichtneckert and Reichert, 2005; Oliver et al., 1995; Steinmetz et al., 2010). Both *optix/six3* and *unpg/gbx* genes have been shown to interact with *otd* on a genetic level (Lichtneckert and Reichert, 2005; Posnien, 2009). While *optix/six3* requirement for patterning the anterior-most region of the embryonic head and the rostral part of the brain, respectively, is currently discussed as an ancestral feature in bilaterian species (Steinmetz et al., 2010), involvement of vertebrate *gbx* genes in establishing the isthmic or midbrain-hindbrain organiser (MHO; reviewed in Raible and Brand, 2004) is considered to be only partially corresponded by the situation in insects, where *unpg/gbx* and *otd/otx* interact in the deutocerebrum (reviewed in Urbach, 2007). In the following, I want to characterise bilaterian *optix/six3* and *unpg/gbx* functions in more detail.

*Six/so* genes represent a subgroup of homeobox genes that encode both a *Six* domain and a homeodomain. They fall into three groups: *sine oculis/six1* genes, *optix/six3* genes and *six4* genes (Kawakami et al., 2000). They are present in bilaterian and cnidarian species

(Kawakami et al., 2000; Ryan et al., 2006; Seo et al., 1999; Stierwald et al., 2004). Recent phylogenetic studies support previous suggestions that this gene family was already present in the metazoan ancestor (Bebenek et al., 2004; Holland and Takahashi, 2005; Ryan et al., 2006). All three groups are considered to be involved in head and brain development (Kawakami et al., 2000; Seo et al., 1999). *optix/six3* genes, in particular, are necessary for the patterning of anterior-most head and brain structures (Steinmetz et al., 2010). In many species, *six3* activity arises very early in embryogenesis during anterior regionalisation of the head in a region anterior and adjacent to the *otd*-positive domain (Oliver et al., 1995; Posnien, 2009; Steinmetz et al., 2010). In *Tribolium castaneum*, *Tc'six3* loss-of function results in the deletion of the entire anterior-median area, including the labrum and the anterior-most area of the protocerebrum as well as anterior and dorsal head cuticular structures (Posnien, 2009). There is evidence that *Tc'six3* and *Tc'otd1* are part of mutual negative regulation system (Posnien, 2009). Although expression of the *Drosophila melanogaster* homolog *optix* is highly similar to *Tc'six3*, aspects of *Drosophila optix* loss-of-function have so far not been addressed comprehensively (Kenyon et al., 2005). In mammals, *six3* homologs are involved in the patterning of the most rostral parts of the brain (Lavado et al., 2008).

Within the group of homeodomain transcription factors, proteins encoded by homologs of *unpg/gbx* genes have been characterised as members of the *Antennapedia*/*Hox*-related class (ANTP/Hox-related; Galliot et al., 1999; Holland and Takahashi, 2005). Comprehensive phylogenetic analyses of homeobox sequences including recent data from the diploblastic sea anemone *Nematostella vectensis* suggest that *unpg/gbx* genes were already part of the genetic repertoire of the cnidarian-bilaterian ancestor (Ryan et al., 2006). Important functions of these genes in head and brain patterning as well as neurogenesis have been identified in various bilaterian species, suggesting that *unpg/gbx* genes are highly conserved head development genes (Burroughs-Garcia et al., 2011; Chiang et al., 1995; Hirth et al., 2003; Lichtneckert and Reichert, 2005; Urbach and Technau, 2003). In many diverse species, expression of *unpg/gbx* arises in a domain that is posterior and adjacent to *otd/otx* expression (Hirth et al., 2003; Lichtneckert and Reichert, 2005; Steinmetz et al., 2011). In vertebrates, *otx* and *gbx* genes form a genetic interface at the midbrain-hindbrain boundary that gives rise to the isthmic organizer (Raible and Brand, 2004; Wurst and Bally-Cuif, 2001). In addition, *otx* and *gbx* regulate each other negatively in vertebrates (Acampora et al., 1997; Millet et al., 1999). A putatively homologous area involving *otd/otx* and *unpg/gbx* exists in the insect brain, more precisely at the deutocerebral/tritocerebral boundary (DTB). In *Drosophila*, mutual negative

regulation of *otd* and *unpg* at this location has been shown as well (Hirth et al., 2003). However, this area does not convey organiser function (Urbach, 2007).

### 2.6.4 *knirps* genes as novel players in arthropod head development

Transcription factors encoded by the *knirps* (*kni*) gene family constitute a subgroup within the nuclear hormone receptors. Members of this family are uniquely characterised by a shared protein domain encoded by the *kni*-box (Nauber et al., 1988; Rothe et al., 1989). Because for none of the *kni*-encoded hormone receptors that have been analysed so far, ligands were found, this group has also been referred to as "orphan receptors" (Arnosti et al., 1996; Laudet et al., 1992). To date, detailed studies have been done on the *Drosophila kni* repertoire, consisting of *Drosophila knirps* (kni, Nauber et al., 1988), *knirps-related* (knrl, Oro et al., 1988) and *eagle* (eg, Rothe et al., 1989). *Drosophila melanogaster kni* is a zygotically expressed gap gene. Its function is necessary for the establishment of posterior segments. Loss of *kni* function results in the deletion of adjacent abdominal segments, specifically the A2–A6 segments (Nauber et al., 1988). Although it is co-expressed with *knrl* in a broad anterior domain covering the ventral and lateral parts of the embryonic head during blastoderm stage, loss-of-function does not impede head regionalisation or cause head gap phenotypes. Instead, disruption of both *kni* and *knrl* function in the anterior part of the *Drosophila* embryo leads to severe defects in the differentiation and projection of the labral neuronal elements and the stomatogastric nervous system (Gonzalez-Gaitan et al., 1994). In addition, *Drosophila kni* and *knrl* share redundant functions in tracheal placode development (Chen et al., 1998). Specifically, they mediate DPP signalling and coordinate opposing DPP and EGF signals. These functions are required for the establishment, maintenance and morphogenesis of individual tracheal branches (Myat et al., 2005). *Drosophila kni* and *knrl* also organise the second wing vein (Lunde et al., 1998). *Drosophila eg* expression is restricted to the late embryonic gonads (Rothe et al., 1989).

*Tribolium* possesses two *kni* family genes, *Tc'kni* and *Tc'eg* (e.g., Xu et al., 2010). *Tc'kni* is the single ortholog of both *Drosophila kni* and *knrl* (Cerny et al., 2008), suggesting that the *Drosophila* genes derive from a gene locus duplication event specific for the lineage leading to higher dipterans. Like *Drosophila kni* and *knrl, Tc'kni* shows early zygotic expression in an anterior gap-like domain during blastoderm stage. In contrast to *Drosophila*

*kni* and *knrl*, loss of *Tc'kni* function leads to the deletion of the antennal and mandibular segments. This is a much more severe effect than even the combined loss of *Drosophila kni* and *knrl* function in head development. Posterior *Tc'kni* expression is restricted to the prospective first abdominal segment. However, RNAi-mediated loss of *Tc'kni* function results in only mild abdominal defects that manifest posterior to its expression domain. *Tc'kni* is expressed later during *Tribolium* development in the distal tip of the labrum and, similar to *Drosophila kni* and *knrl*, in the putative tracheal placodes and branches (Cerny et al., 2008). *Tc'eg* is maternally expressed. It is localised anterior in the embryo. However, RNAi-based loss-of-function of *Tc'eg* does not result in early anterior patterning defects (Bucher et al., 2005).

Two *kni* paralogs have been isolated from the milk weed bug *Oncopeltus fasciatus*. Early anterior expression has also been found for *Oncopeltus kni1*, as well as additional expression in central and posterior stripes of the embryo. However, no function of *Oncopeltus kni* genes has been discovered so far (Jonas Schwirz, personal communication; Ben-David and Chipman, 2010)

Aside from *Drosophila*, *Tribolium* and *Oncopeltus*, *kni* genes have been isolated from or are present in the genomes of various other arthropod species, more precisely several insects (*Ceratitis capitata*, diptera, Jonas Schwirz, personal communication; *Nasonia vitripennis* and *Apis mellifera*, hymenoptera, 3.4.6, A6.2), crustaceans (*Daphnia pulex*, cladocera, branchipoda, http://wfleabase.org/; species of the *Caligus* genus, copepoda, maxillopoda, 3.4.6, A6.2), a myriapod (*Strigamia maritima*, Carlo Brena, personal communication) and a chelicerate species (*Ixodes scapularis*, acari, arachnida, 3.4.6, A6.2). So far, *kni* genes have not been found in species outside the clade of arthropods. In particular, no sequences similar to *kni*-box sequences can be found in the sequenced genomes of non-arthropods (3.4.6). Therefore, it is highly probable this group of genes has evolved within arthropods.

Based on known expression data of *kni* genes in insects, a conserved role in anterior and head patterning in arthropods has been suggested (Cerny et al., 2008). However, the functional range of *kni* genes varies considerably among these species, as described above. In addition, *kni* genes appear to conveying pleiotropic functions in arthropod development and to be involved in many diverse developmental processes. To find out whether *kni* genes share conserved functions in arthropod head development, a close examination of the situation in an outgroup, such as *Parhyale hawaiensis*, will prove extremely useful.

## 2.7 Aims of this work

To identify conserved elements of insect head development, which might derive from an ancestral molecular and genetic mechanism common to all bilaterian species, the aim of this work was to clone and analyse *Parhyale hawaiensis* homologs of known conserved genetic regulators of head development. In detail, I wanted to identify *Parhyale homologs* of the conserved anterior head gene *orthodenticle (otd)* along with other representatives of *paired*-class homeodomain encoding factors, as well as homologs of the suggested conserved *otd* interaction partners *optix/six3* and *unplugged/gbx* and compare my findings to the situation in insects.

In addition, I sought to isolate *Parhyale* representatives of the *knirps* gene family, which appear to be an evolutionary novelty of arthropods and convey pleiotropic functions in the insects *Drosophila melanogaster* and *Tribolium castaneum.* Here, my aim was to examine whether the divergent roles of insect *knirps* genes in head development might share or derive from an ancestral, conserved function.

As a prerequisite on which I could base my comparative molecular and genetic studies, I needed to further examine and describe the embryologic and morphological implications of *Parhyale hawaiensis* head development in detail.

*Parhyale hawaiensis* represents a promising emerging model system for malacostracans and arthropods in general. Many embryologic, genetic and molecular techniques had been established. However, reliable functional techniques had not been available at the beginning of this work. Therefore, the final aim of my work was to test the efficacy of various functional approaches in *Parhyale*. In particular, I focused on developing loss-of-function techniques based on RNA interference.

# 3 Results

## 3.1 Hallmarks of head morphogenesis in *Parhyale hawaiensis*

For the following morphological analyses of the *Parhyale* embryo, the S1-S30 staging nomenclature is used (Browne et al., 2005). Elements of early anterior-posterior axial polarisation manifest during stages 4-7 of *Parhyale* embryogenesis (2.5.5.1; Browne et al., 2005; Gerberding et al., 2002). The first morphologically visible head-related structures are the embryonic head lobes that emerge at stage 9 (2.5.5.3; Browne et al., 2005).

In this work, the development of head-related structures in *Parhyale* has been analysed in more detail, starting from stage 11 through stage 23 of embryogenesis. Since *Parhyale* embryos of later stages than S23 are no longer accessible for WMISH (5.4.3), head development from S24 until hatching (S30) has not been examined further. The nature and relative positions of prospective segments, fields and anlagen were identified and determined on the base of previous analyses of *Parhyale* development (Browne et al., 2005) and as the result of a comparative, comprehensive view of *Parhyale* embryogenesis in its entirety.

### 3.1.1 Establishment and regionalisation of the embryonic head lobes (S11 and S12)

Following the progress of embryogenesis through stages 9-12, the former even distribution of cells within the germ disc increasingly gives way to a distinct topology of cell density and cell clustering within each head lobe. Located medially within each head lobe and roughly halfway between the anterior end of the embryo and the region of the future mandibular segment to the posterior, densely clustered cells flag the future site of the An1 appendage field (Figure 3, yellow asterisks). The ectodermal cells of the anterior head segments have begun to asymmetrically cleave, some of their progeny likely being neuroblasts (Browne et al, 2005).

At stage 12, the area between both head lobes contains only few cells and is visible morphologically as a cleft (Figure 3, red arrowheads). The transition from the anterior end of the embryo to the extra-embryonic periphery does not appear sharp. Instead, this border

remains blurry through stages 8-12; the transition is characterised by a continuous decrease in cell density (Figure 3 B1, B2). Strikingly, many cells appear arranged and aligned in ray-like columns that radiate into the extra-embryonic periphery (Figure 3 B1, blue arrow). This observation may reflect the cleavage pattern of these cells. Similar to the region representing the future An1 segment, the central part of the developmental field that lies at the site of the future mandibular segment also harbours a relatively high number of densely clustered cells. At stage 12, this area appears triangular-shaped (Figure 3 C1, C2, green asterisk). Descendants of the mesodermal founder cells form the bilateral midgut anlagen that directly abut the head lobes laterally. At stage 12, they can be easily identified as prominent disc-shaped structures. The midgut anlagen extend from the level of the mandibular developmental field to the posterior part of the pre-antennal head region (Figure 3 A2, lateral of the An1 appendage fields; Browne et al., 2005).

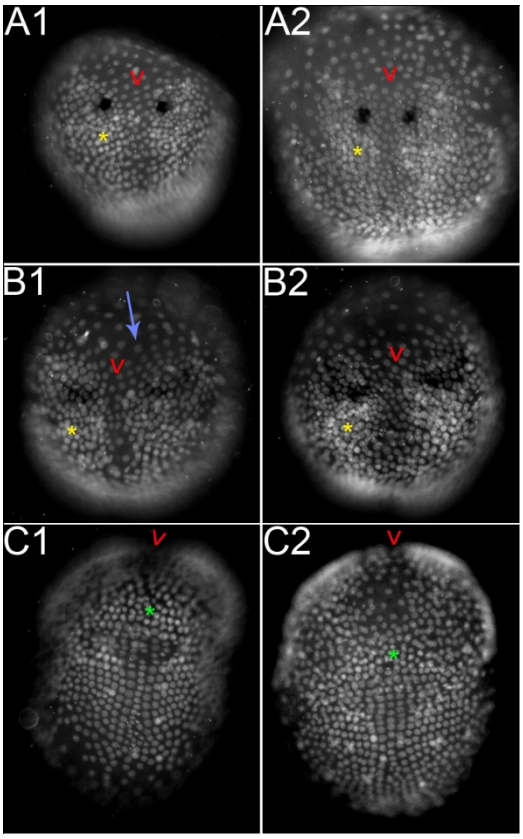



**Figure 3:** ***Parhyale* head development during stage 12.** Shown are grey-scale images of nuclear labelled embryos (SYTOX®). The images to the left (#1) depict early stage 12 embryos. The images to the right show late stage 12 embryos. In all images, anterior is up. Red arrowheads mark the anterior end of the cleft that separates both head lobes. **A** Anterior view of an early stage 12 (A1) and a late stage 12 (A2) embryo. The future sites of the An1 anlagen appear as dense clusters of cells (yellow asterisks indicate the right embryonic An1 appendage fields). The embryos had previously been subjected to *Ph six3* WMISH (3.3.1.7). **B** Anterior view of an early stage 12 (A1) and a late stage 12 (A2) embryo. A blue arrow marks cells of the anterior periphery of the left head lobe in B1 that are arranged in ray-like columns. Like in A, yellow asterisks indicate the right embryonic An1 appendage fields. These embryos had previously been subjected to *Ph otd1* WMISH (3.2.1; Browne et al., 2006). **C** Ventral views of the embryos shown in B. A green asterisk marks the centre of the triangular developmental field at the site of the future mandibular segment. The posterior end of the medial cleft directly borders the anterior of this field (C2).

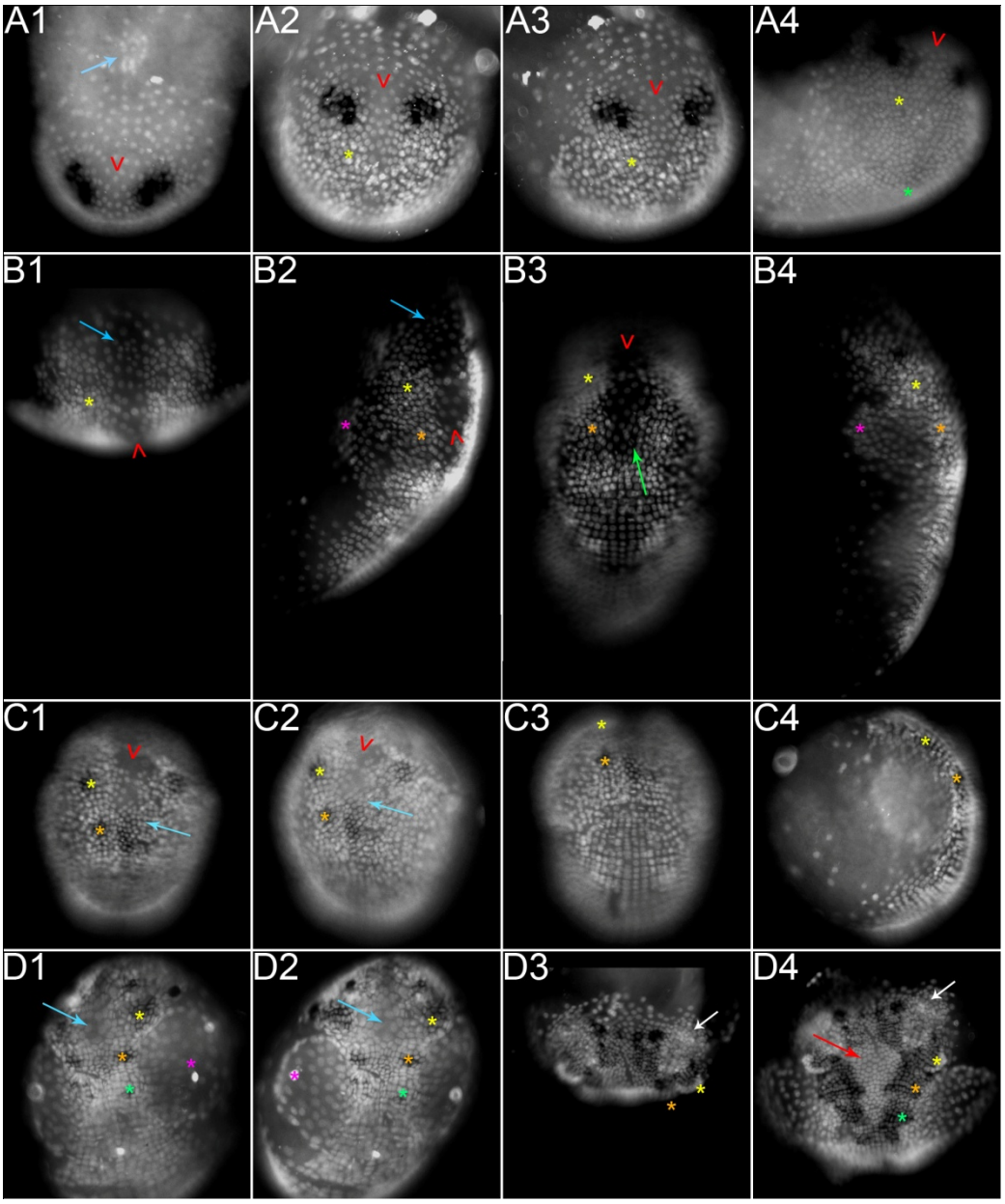
### 3.1.2 Condensation and fusion of the head lobes (S13-15)

While the general morphology of the embryonic head observed during stage 12 is maintained during stage 13, the number of cells that each head lobe comprises increases. As a consequence, the embryonic head reaches its largest expansion on the egg's surface (Figure 4 A1-A4). Directly following from that point of embryogenesis, the head lobes start to condense and fuse. At the onset of stage 14, each individual head lobe has narrowed (Figure 4 B2, B3). Densely packed groups of cells flag the appendage anlagen of An1 and An2 (Figure 4 B1-B4). Bordering the Mn and An2 segments laterally, the midgut anlagen have emerged as distinguishable, ovoid structures (Figure 4 B2, B4, pink asterisks; Browne et al., 2005). The medial cleft that separates the head lobes is still visible as a region of low cell density (Figure 4 B1-B3, red arrowheads). It directly abuts the anterior of the future Mn segment and reaches until the anterior end of the embryo. Within the posterior part of the cleft, at the level of the An2 segment, cells undergo mitoses (Figure 4 B3, green arrow). At the same time, a cluster of cells has emerged at the anterior end of the cleft (Figure 4 B3, blue arrow). In the middle of stage 14, cells cover the medial region of the former separated head lobes at the level of the An2 segment (Figure 4 C1, C2, blue arrows). At the onset of stage 15, the fusion of the head lobes is almost completed. Only a small patch located medially to the An1 appendage anlagen still exhibits lower cell density (Figure 4 D1, D2, blue arrows). The An1, An2 and Mn segments have condensed further. The appendages of these segments start to project distally. In contrast to the condensing head segments, the lateral midgut anlagen have expanded. Their periphery is composed of closely arranged cells (Figure 4 D1, D2, pink asterisks). Therefore, they delineate sharply from the ectodermal segments of the embryo. In embryos that have completed stage 15, the pre-antennal hemispheres of the embryonic head project laterally. The centres of each of these projections show a high number of densely clustered cells that are arranged in multiple layers (Figure 4 D3, D4, white arrows). The medial cleft between the head lobes has been replaced by a single-layered, triangular sheet of cells (Figure 4 D4, red arrow). This sheet of cells is bordered laterally by the emerging appendage anlagen of the An1, An2 and Mn segments, which are clearly visible as proximal-distal outgrowths (Figure 4 D4, yellow, orange and green asterisks indicate the An1, An2 and Mn appendage anlagen, respectively). The anterior end of the embryo has acquired the shape of an almost horizontal border line that separates densely clustered embryonic cells from few scattered extra-embryonic cells, with the lateral anterior-most parts of the pre-antennal hemispheres bulging

slightly more anterior than the medial region of the pre-antennal head (Figure 4 D3). The presence of Distal-less protein directly anterior to the An1 segment in a confined area medially within the pre-antennal head indicates the onset of the organisation of the labral developmental field (Browne et al., 2005).

**Figure 4 (next page):** ***Parhyale* head development during stages 13 to 15.** Shown are grey-scale images of nuclear labelled embryos (SYTOX®). In all images, anterior is up. **A** Stage 13 *Parhyale* embryo. Red arrowheads mark the anterior end of the medial cleft in all images of this embryo. The embryo had been subjected to *Ph hbn* WMISH (3.2.2.9). **A1** Dorsal view. A blue arrow indicates the dorsal organ. The dorsal organ emerges at stage 10 of *Parhyale* embryogenesis (Browne et al., 2005). A2 Anterior view. The appendage field of the An1 segment is indicated by a yellow asterisk. A3 Anterior view; the embryo is tilted to the left. Directly bordering the appendage field of the An1 segment (yellow asterisk) laterally, the midgut anlagen have increased in size as compared to stage 12 (Browne et al., 2005). At this stage, the embryonic head reaches its largest expansion on the egg's surface. A4 Ventral view; the embryo is tilted to the left. The appendage field of the An1 segment is indicated by a yellow asterisk and the site of the future Mn segment is marked by a green asterisk. **B** *Parhyale* embryo at the onset of stage 14. Red arrowheads mark the medial cleft (B1-3). The appendage field of the An1 segment is indicated by a yellow asterisk (B1-4) and that of the An2 segment by an orange asterisk (B2-4). The lateral, ovoid midgut anlagen are marked by a pink asterisk (B2, B4). At the anterior end within the medial cleft, a cluster of cells has emerged (blue arrows, B1, B2). Within the posterior end of the medial cleft, cells undergo mitoses (green arrow, B3). The embryo had been subjected to *Ph six3* WMISH (3.3.1.7). B1 Anterior view. B2 Lateral view; the embryo is tilted to the right. B3 Ventral view. B4 Lateral view. **C** *Parhyale* embryo in the middle of stage 14. At the level of the An2 segment, cells cover the medial region of the previously separated head lobes (blue arrows, C1, C2). The remaining medial cleft is marked by red arrowheads (C1, C2). The appendage field of the An1 segment is indicated by a yellow asterisk (C1-4) and that of the An2 segment by an orange asterisk (C1-4). The embryo had been subjected to *Ph kni1* WMISH (3.4.7). C1 Anterior view. C2 Anterior view, the embryo is tilted slightly to the right. C3 Ventral view. C4 Lateral view, sagittal focal plane. **D** *Parhyale* embryo at the onset (D1, D2) and at the end (D3, D4) of stage 15. D1 Tilted ventral and D2 ventral view of a *Parhyale* embryo at the onset of stage 15. Except a restricted medial area at the level of the An1 appendage anlagen (blue arrows), the fusion of the former head lobes is complete. The appendages of the An1 (yellow asterisks), An2 (orange asterisks) and Mn (green asterisks) segments start to project distally. In D1, the left appendage anlagen are marked. In D2, the right ones are. The left (D1) and right (D2) lateral midgut anlagen are marked by pink asterisks. The embryo had been subjected to *Ph kni1* WMISH (3.4.7). D3 Dorsal and D4 ventral view of a *Parhyale* embryo that has completed stage 15. The pre-antennal hemispheres of the embryonic head project laterally. In their central parts, they show densely clustered cells that are arranged in multiple layers (white arrows). Medially, a single-layered, triangular sheet of cells has emerged (red arrow). It is bordered laterally by the appendage anlagen of the An1 (yellow asterisks), An2 (orange asterisks) and Mn segments (green asterisk, D4). The embryo had been subjected to *Ph six4* WMISH (3.3.1.8).

**Figure 4:** ***Parhyale*** **head development during stages 13 to 15.**



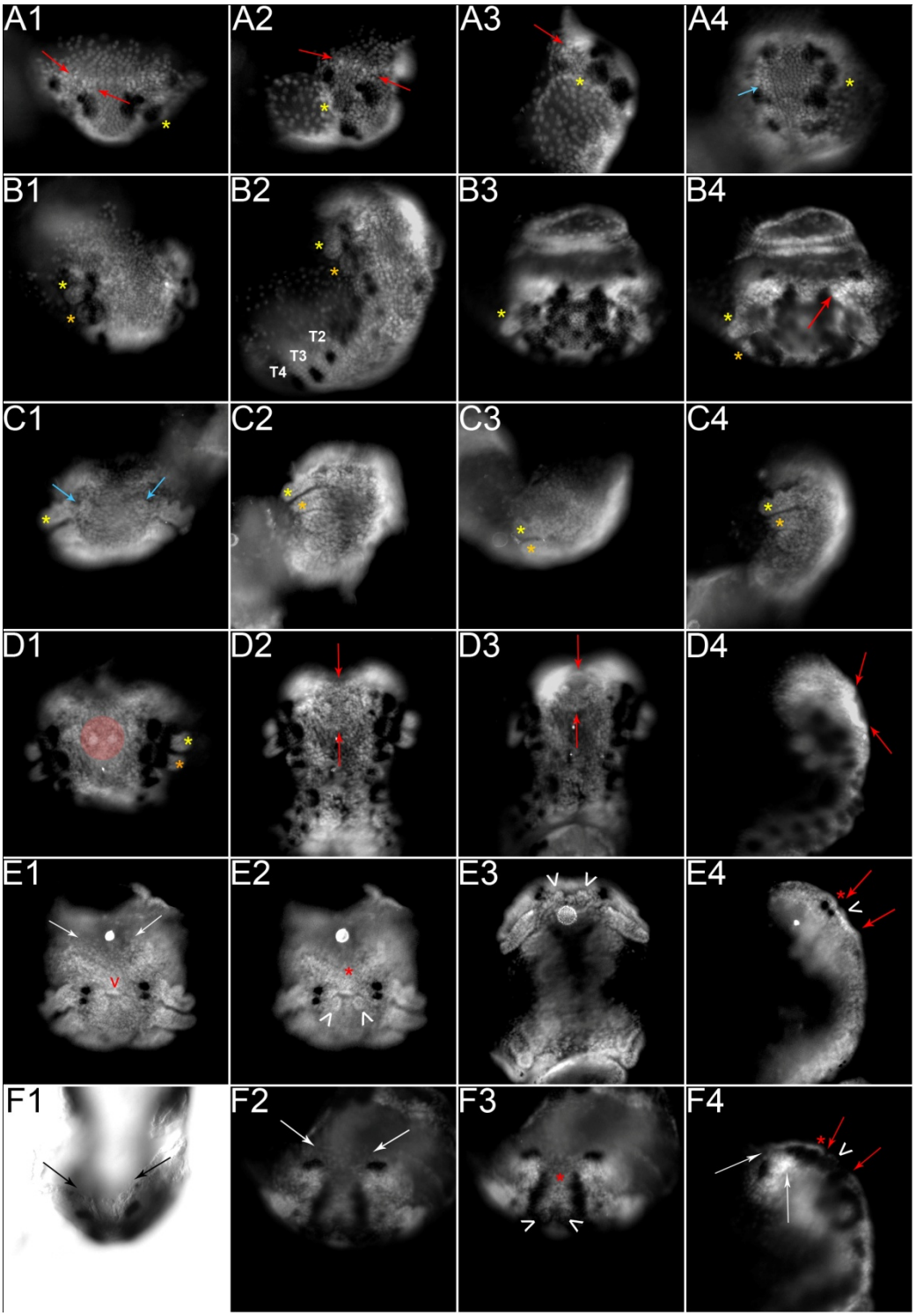
### 3.1.3 Establishment of lateral pre-antennal hemispheres, the stomodeum and the labrum (S16-20)

During stages 16, 17 and 18, the pre-antennal, laterally projected hemispheres undergo significant cell proliferation. The cells become smaller, more densely packed and have point-shaped nuclei. Arched ridges reaching from the medial area to the lateral hemispheres show the highest density of nuclei, indicating multiple layers of cells (Figure 5 A1, A2, A3, B4, red arrows). In contrast, the periphery of an individual pre-antennal hemisphere appears to be composed of fewer cells (Figure 5 A1, A2, B4, areas adjacent to red arrows). The An1 and An2 appendages elongate and produce additional articles (Figure 5 A2, A4, B2, B4, C2, C4). Their development is advanced in comparison to the thoracic appendage anlagen (Figure 5 B2, see also Figures 74 and 75). At the onset of stage 18, two small opposing medial outgrowths, situated between the An1 and An2 segments, mark the onset of stomodeum formation. (Figure 5 C1, blue arrows; compare to A4, blue arrow). During stage 18, the stomodeum begins to invaginate. At the beginning of stage 19, the two opposing outgrowths are clearly visible on lateral positions within the deepening stomodeal opening. Subsequently, they will be referred to as “lateral stomodeal projections” (Figure 5 D1, stomodeal field is highlighted red, lateral stomodeal projections are highlighted white; D2-D4, red arrows indicate the anterior and posterior boundaries of the stomodeal field). The anterior ridge of the stomodeum becomes first visible as a short horizontal stripe with a high number of densely clustered nuclei (Figure 5 E1, red arrowhead), similar to the outlines of the developing An1, An2 and Mn appendages. As the oesophagus invaginates during the stages 19 and 20, the stomodeal field itself becomes smaller (Figure 6). Anterior of the stomodeum, the labrum becomes visible at stage 19 (Figure 5 E2, F3, red asterisks). It begins to protrude distally, tending towards the posterior (Figure 6). At the end of stage 20, the labrum is visibly composed of two lobes that begin to grow over the distinct stomodeal opening (Figure 6 A3, A4, C3, C4). At stage 20, just posterior to the labrum, the brink of the stomodeum is visible as a raised medial ridge at the level of the An2 segment (Figure 6 C4; Browne et al., 2005). During stage 19, the lateral portions of the pre-antennal hemispheres expand. As a consequence, they appear wedge-shaped (Figure 5 E1, E2, F2, F4) and extend in part to the dorsal face of the egg (Figure 5 F1). As a consequence, the anterior borderline appears now more V-shaped, as the expanding preantennal hemispheres conclude an angle of about 100-120° (Figure 5 E1, F2, white arrows; F1, black arrows). The delineation between embryonic

tissue and the extra-embryonic dorsal region of the egg is clearly cut (Figure 5 E1, F1). During stage 19 the paragnaths become visible just medial to the mandibles (Figure 6 C1-C3; Browne et al., 2005).

**Figure 5 (next page):** ***Parhyale* head development from stage 16 to the onset of stage 20.** Shown are grey-scale images of nuclear labelled embryos (SYTOX®) and one grey-scale DIC image (F1). In all images, anterior is up. **A** Early stage 17 *Parhyale* embryo in dorsal (A1), tilted anterior (A2), lateral (A3) and anterior (A4) view. Red arrows indicate the arched ridges that show the highest density of nuclei within the pre-antennal hemispheres. Yellow asterisks mark the An1 appendage anlagen. The blue arrow (A4) points towards a cluster of cell that will subsequently form one of the lateral oesophageal projections (see also C1). The embryo had been subjected to *Ph kni1* WMISH (3.4.7). **B** Stage 17 (B1, B2) and late stage 17 (B3, B4) *Parhyale* embryos. Yellow and orange asterisks mark the elongating An1 and An2 appendage anlagen, respectively. B1 Tilted anterior and B2 tilted ventral view of a stage 17 embryo. In B2, the thoracic appendage anlagen of T2, T3 and T4 are labelled (see also al1S17). The embryo had been subjected to *Ph al1* WMISH (see R). B3 Anterior view, focal plane on the medial region at the level of the An1 segment, and B4 Anterior view, focal plane on the pre-antennal ridges of a late stage 17 embryo. A red arrow (B4) indicates the left pre-antennal ridge (see also A1-A3). The embryo had been subjected to *Ph kni1* WMISH (3.4.7). **C** *Parhyale* embryo at the onset of stage 18. Yellow and orange asterisks mark the An1 and An2 appendages, respectively. C1 Anterior view. Blue arrows point to two small opposing medial outgrowths that will subsequently be recognised as lateral oesophageal projections (see also E2). C2 Tilted ventral view. C3 Tilted anterior view. C4 Lateral view. The embryo had not been subjected to WMISH. **D** *Parhyale* embryo at the onset of stage 19. The stomodeum forms. The embryo had been subjected to *Ph al1* WMISH (3.2.2.10). D1 Anterior view. The stomodeal field is highlighted red. The lateral oesophageal projections are highlighted white. Yellow and orange asterisks mark the An1 and An2 appendages, respectively. D2 Ventral view. D3 Ventral view, alternative focal plane. D4 Lateral view, sagittal focal plane. Red arrows (D2-D4) indicate the upper and lower boundaries of the invaginating stomodeal field. **E** *Parhyale* embryo at stage 19. The embryo had been subjected to *Ph gbx* WMISH (3.3.2.4). E1 Anterior view. As a consequence of the lateral expansion of the pre-antennal hemispheres, the anterior borderline appears V-shaped with an angle of about 100-120° (white arrows). The anterior ridge of the stomodeum is indicated by a red arrowhead. E2 Anterior view, focal plane on the lateral oesophageal projections (white arrowhead). Anterior of the stomodeum, the outgrowing labrum is recognisable as a dense cluster of nuclei (red asterisk). E3 Ventral view. White arrowheads mark the lateral oesophageal projections. E4 Lateral view, sagittal focal plane. The position of the labrum is indicated by a red asterisk. Red arrows mark the anterior and posterior brinks of the invaginating stomodeum. A white arrowhead points at the lateral oesophageal projections. **F** *Parhyale* embryo at the onset of stage 20. The embryo had been subjected to *Ph kni1* WMISH (3.4.7). F1 DIC image of the embryo in dorsal view. The anterior end of the embryonic head delineates sharply from the dorsal, extra-embryonic region. The pre-antennal hemispheres conclude an angle of about 100-120° (black arrows). F2 Anterior view, focal plane on the anterior end of the embryonic head. The pre-antennal hemispheres conclude an angle of about 100-120° (white arrows). F3 Anterior view, focal plane on the labrum (indicated by a red asterisk). The white arrowheads point towards the lateral oesophageal projections. F4 Lateral view, sagittal focal plane. The pre-antennal hemispheres have expanded during stages 18 and 19 and now appear wedge-shaped (white arrows). The position of the labrum is indicated by a red asterisk. Red arrows mark the anterior and posterior brinks of the invaginating stomodeum. A white arrowhead points at the lateral oesophageal projections.

**Figure 5:** ***Parhyale*** **head development from stage 16 to the onset of stage 20.**



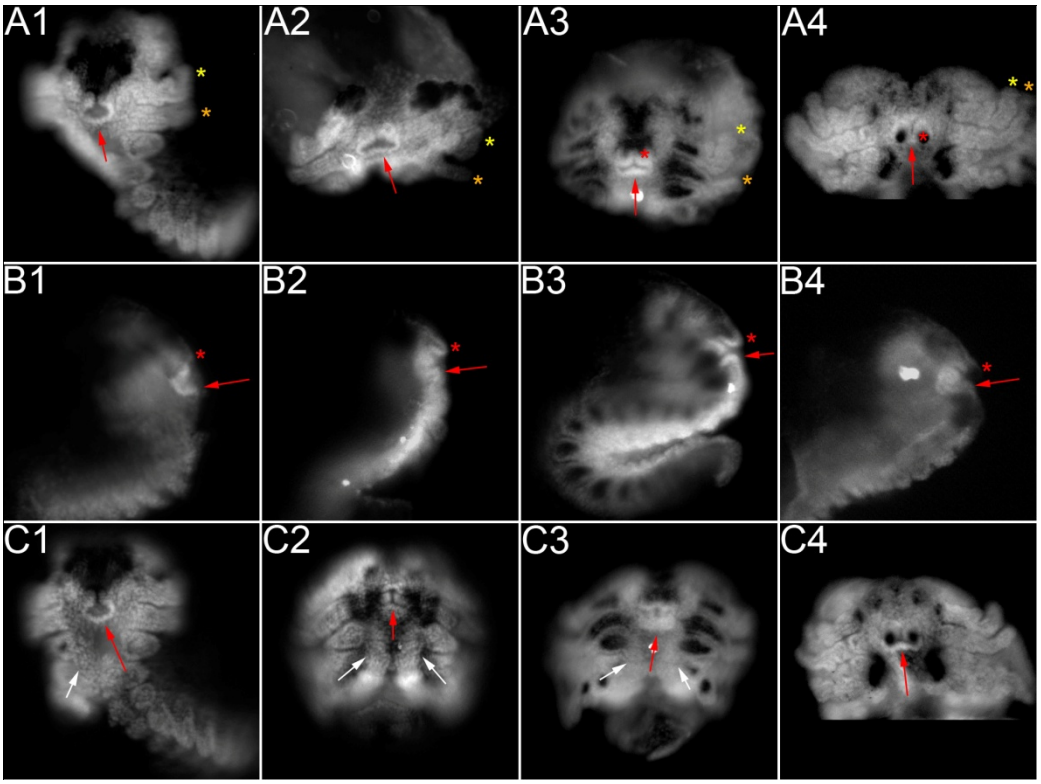



**Figure 6: Development of stomodeum and labrum in *Parhyale* during stage 20.** Shown are grey-scale images of nuclear labelled embryos (SYTOX®). In all images, anterior is up. Shown is the gradual invagination of the stomodeum along with the elongation of the labrum from the onset of stage 20 (#1) through early stage 20 (#2) and the mid of stage 20 (#3) until the end of stage 20 (#4). A red arrow indicates the posterior stomodeal ridge. The embryos had been subjected to *Ph six3* WMISH (A1, B1, C1; 3.3.1.7), to *Ph hbn* WMISH (A2, B2; 3.2.2.9), to *Ph kni1* WMISH (C2; 3.4.7), to *Ph six4* WMISH (A3, B3, C3; 3.3.1.8) and to *Ph awh* WMISH (A4, B4, C4; 3.2.2.4). **A** Tilted anterior (A1) and anterior (A2-A4) views. Yellow and orange asterisks mark the distal articles of the An1 and An2 appendages, respectively. Red asterisks mark the left lobe of the bi-lobed labrum in A3 and A4. **B** Lateral views, sagittal focal plane. Red asterisks mark the distal tip of the labrum. **C** Ventral views, focal plane on the Mn segments. White arrows indicate the emerging paragnaths (C1, C2, C3).

### 3.1.4 Three-dimensional head character morphogenesis (S21-23)

Already at the end of stage 19, the head region has moved dorsally and appears cup-shaped, occupying the anterior face of the egg (Figure 5 F4; Browne et al., 2005). As the *Parhyale* embryo progresses through stages S20 to S22, the head develops an increasingly complex, three-dimensional morphology.

The pre-antennal head lobes expand in size and develop into three-dimensional oval hemispheres that fold dorso-laterally from the anterior, medial region of the embryonic head (Figure 7 A3, A4, B3, B4, C1-C3). Around stage 21, bilateral cellular strands that connect the medullar hemispheres to the anterior An1 segment gain shape. At this position, the future hemi-ellipsoid bodies form (Figure 7 A3, C1; Browne et al., 2006). At the end of stage 22, the head has started to close dorsally. In nuclear stainings, well-defined morphological structures have become visible in the proto- and deutocerebral derivates of the embryonic brain. From anterior to posterior, the visible protocerebral neuromeres are: the large, laterally und dorsally extending medullae (md), the medially situated hemi-ellipsoid bodies (hb) (Figure 7 C1; Browne et al., 2006), and, grouped circumferentially around the stomodeum, the medial protocerebral commissure (pcc) (Figure 7 C2; Browne et al., 2006) as well as the lateral lobes (ll) (Figure 7 C1; Browne et al., 2006). In dorsal view, the deutocerebral neuromeres, i.e. the antennal lobes (al) can be recognised laterally of the stomodeal opening (st) (Figure 7 C4; Browne et al., 2006).

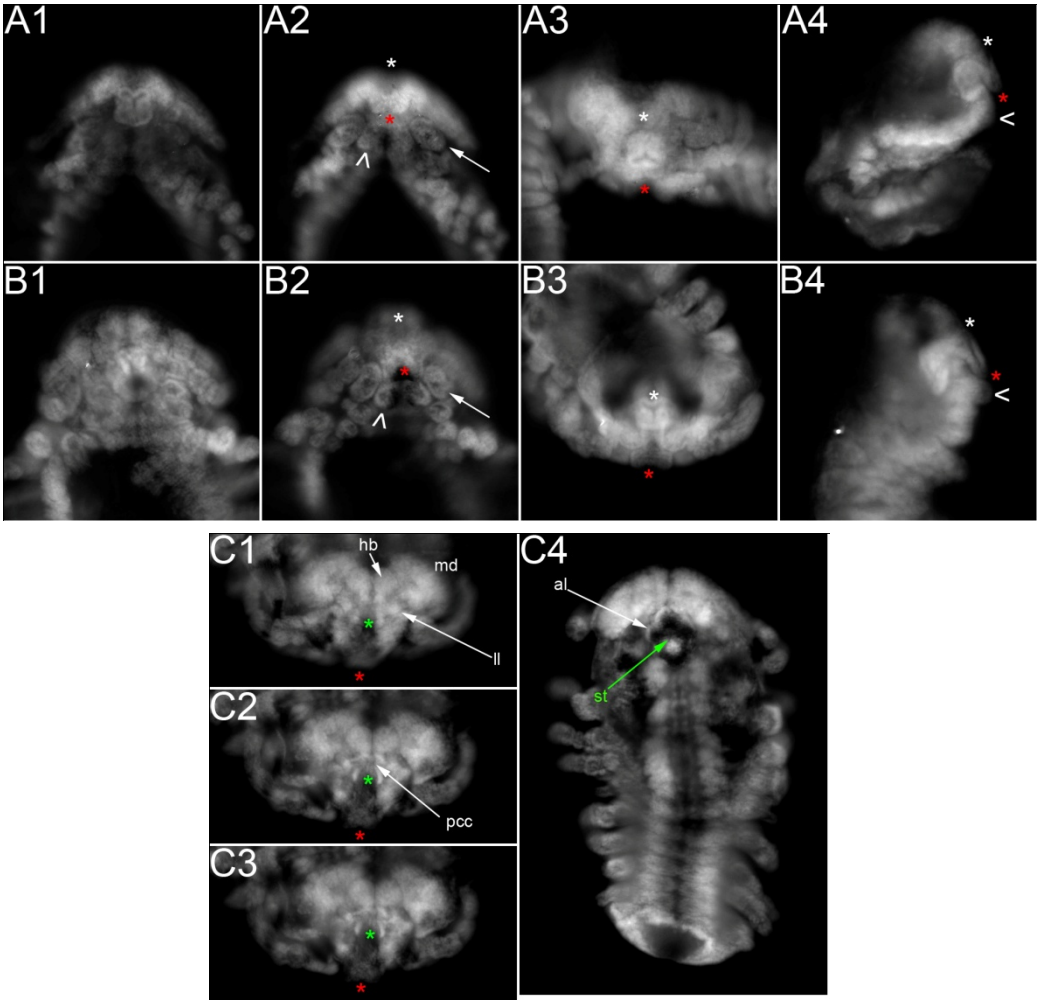
The bag-like oesophageal invagination still found in S21 embryos (Figure 7 A4) develops a more complex structure. The anterior part of it elongates inwards the thickening frontal head region and forms the oesophagus and foregut. At stage 22, the foregut resembles an arched tube that lies underneath the proximal part of the labrum (Figure 7 B4). From the posterior, basal part of the invagination, bilateral sack-like cavities emerge (Figure 7 compare A4 to B4). They are surrounded by dense, proliferating tissue. Based on their appearance, these structures might resemble gland primordia. When viewing directly upon the stomodeal opening of stage 21 to 23 embryos, the lumen of the oesophagus appears clover-shaped (Figure 7 A1, A3, B1, B3).

The labrum projects to the posterior and elongates, eventually covering the stomodeal opening (Figure 7 A2, A4). At stage S22, the labrum has fused and completed its posterior extension. Its distal tip directly abuts the paragnaths, with the mandibles prominently protruding laterally on both sides of the labrum (Figure 7 B2, B4; Browne et al., 2005).

At the onset of stage 20, the second antennae have undergone rapid elongation (Figure 8 A1-A3; Browne et al., 2005). They are composed of five proximal elements forming the peduncle and one distal flagella element (Figure 8 A3; Browne et al., 2005). Until stage 23, the first and second antennae complete the establishment of their entire sets of articles. Their length increases and their tips become pointy as individual antennal articles elongate and acquire distinct morphologies (Figure 8 A1-A4; Browne et al., 2005). By stage 23, the second antennae begin to conform to their final cylindrical uniramous shape (Figure 8 B1-B4; Browne et al., 2005).

**Figure 7 (next page):** ***Parhyale* head development at stages 21 and 22.** Shown are grey-scale images of nuclear labelled embryos (SYTOX®). In all images, anterior is up. White asterisks mark the proximal base of the labrum (A2- A4, B2-B4) and red asterisks its distal tip (A2- A4, B2-B4, C1-C3). White arrows indicate the (left) mandibles (A2, B2). White arrowheads mark the (right) paragnaths (A2, B2, A4, B4). **A** Stage 21 embryo. A1 Ventral view, focal plane on the stomodeum. A2 Ventral view, focal plane on the labrum. A3 Anterior view, focal plane on stomodeum. A4 Lateral view, sagittal focal plane. This embryo had been subjected to *Ph kni2* WMISH (3.4.8). **B** Stage 22 embryo. B1 Ventral view, focal plane on the stomodeum. B2 Ventral view, focal plane on the labrum. B3 Anterior view, focal plane on stomodeum. B4 Lateral view, sagittal focal plane. This embryo had been subjected to *Ph otd1* WMISH (3.2.1.5). **C** Late stage 22 embryo. The embryo had been subjected to *Ph kni1* WMISH (3.4.7). At this stage, several neuromeres can be distinguished in the head of nuclear labelled embryos. C1-C3 depict different focal planes of an anterior view of the embryo. Green asterisks mark the site of the stomodeum. Indicated are: medulla (md), hemi-ellipsoid bodies (hb), lateral lobes (ll) and protocerebral commissure (pcc). C4 Dorsal view, focal plane on the neuromeres surrounding the stomodeum (st, green arrow). On both sides of the stomodeum, the deutocerebral antennal lobes (al) can be recognised. In this image, the left al is marked (see also Browne et al., 2006).

**Figure 7:** ***Parhyale*** **head development at stages 21 and 22.**



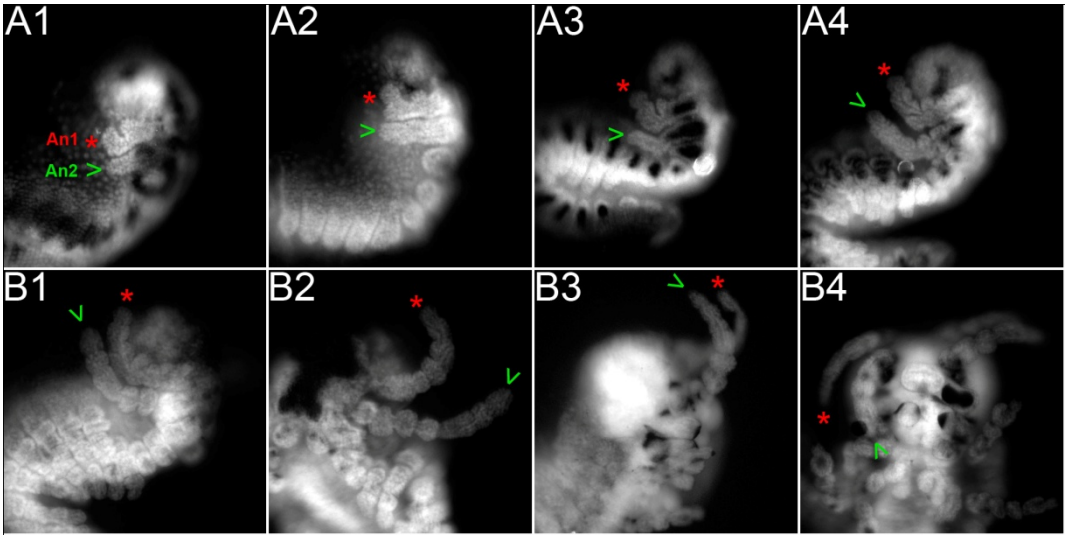



**Figure 8: Development of the An1 and An2 appendages (i.e. the first and second antennae) in *Parhyale* from stage 19 to 23.** Shown are grey-scale images of nuclear labelled embryos (SYTOX®). In all images, anterior is up. Also, the tips of the first antennae are marked by a red asterisk and the tips of the second antennae by a green arrowhead. **A** Gradual development and elongation of the first and second antennae from stage 19 until stage 22. All images depict the embryos in lateral orientation, ventral is to the right. **A1** S19 embryo. The first and second antennae are labelled An1 and An2, respectively. The embryo had been subjected to *Ph kni1* WMISH (3.4.7). **A2** Embryo of early S20. The embryo had been subjected to *Ph six3* WMISH (3.3.1.7). **A3** Embryo of late S20. The embryo had been subjected to *Ph six4* WMISH (3.3.1.8). **A4** S22 embryo. The embryo had been subjected to *Ph six4* WMISH (3.3.1.8). **B** Morphogenesis of the first and second antennae during S23. **B1, 2** Early S23 embryo. The embryo had been subjected to *Ph six3* WMISH (3.3.1.7). **B1** Lateral view, ventral is to the left. **B2** Tilted ventral view. Shown are the An1 and An2 appendages of the left body side. **B3, 4** Late S23 embryo. The embryo had been subjected to *Ph kni2* WMISH (3.4.8). **B3** Lateral view. Ventral is to the right. **B4** Ventral view.

## 3.2 *Parhyale orthodenticle* and other *paired*-class homeobox genes involved in head development

*Parhyale* homologues of representatives from various groups and families of transcription factors with known functions in head development in other species were isolated and characterized in this work. In many arthropod and vertebrate species, *orthodenticle* genes play major roles in the developmental processes mentioned above (Acampora et al., 2000b; Cohen and Jürgens, 1990; Finkelstein and Boncinelli, 1994; Finkelstein et al., 1990; Schinko et al., 2008; Simeone et al., 2002; Steinmetz et al., 2011). For this reason, the first candidate genes to be isolated and characterised in this work were the putative Parhyale homologs of *orthodenticle* (*otd*). To this end, repeated PCR screens on a collection of *Parhyale* cDNA covering all stages of embryogenesis (varied-stringency degenerate PCR, 5.2.3.1) were conducted.

Aside from novel *Parhyale hawaiensis otd* sequences, several sequence fragments with close resemblance to additional *paired*-class homeobox genes were recovered. From those, promising candidates for the study of the molecular processes underlying anterior-posterior polarization, head regionalization, head segmentation and the development of brain structures and neuronal organs in *Parhyale hawaiensis* were chosen, examined in more detail and compared with findings from insects and other arthropods. A set of PCR screens, which were adapted to target paired-class homeobox genes in general (depletive varied-stringency degenerate PCR, 5.2.3.2), was conducted to support the isolation and confirmation of these additional candidates. Complete cDNA isolation and phylogenetic characterisation of all these genes resulted in the identification of *Parhyale otd1* and *otd2*, *pairberry1* (*Ph pby1*), *homeobrain* (*Ph hbn*), of two *aristaless* genes (*Ph al1* and *Ph al2*) and of the *Parhyale arrowhead* homolog (*Ph awh*). Subsequently, the expression profiles of these genes were visualized via WMISH (5.4.3).

### 3.2.1 *Parhyale hawaiensis orthodenticle* genes

In arthropods as well as vertebrates, transcription factors encoded by *orthodenticle (otd)* genes play major roles in early anterior polarity, head and brain regionalisation, neurogenesis

and the development of head and brain organs (Acampora et al., 2000b; Cohen and Jürgens, 1990; Finkelstein and Boncinelli, 1994; Finkelstein et al., 1990; Schinko et al., 2008; Simeone et al., 2002; Steinmetz et al., 2011). Despite the conserved roles of *otd* and *otx* genes, their expression pattern and, accordingly, the functions of OTD/Otx proteins are modulated and individually adjusted to specific developmental needs even within closely related species(Schetelig et al., 2008; Schinko et al., 2008). In order to elucidate the roles that *otd* genes play in *Parhyale* head development, I cloned and analysed *Parhyale hawaiensis otd* genes in this work.

#### 3.2.1.1 Isolation and characterisation of *Parhyale hawaiensis orthodenticle* genes

Because sufficient genomic sequence data is not currently available for *Parhyale hawaiensis*, the isolation of putative *otd* sequences was approached by performing varied-stringency degenerate PCR screens (5.2.3.1). To this end, degenerate primers based on comparative amino acid alignments of homeodomains (HD) of described Orthodenticle proteins were used (5.2.4). From repeated varied-stringency degenerate PCR screens on a collection of *Parhyale* cDNA covering all stages of embryogenesis (5.2.2) and a supplemental PCR screen targeted at *paired*-class homeobox sequences (depletive lowered-stringency PCR screen, 5.2.3.2), two *Parhyale hawaiensis otd* sequences, *Parhyale hawaiensis orthodenticle1* (*Ph otd1*) and *orthodenticle2* (*Ph otd2*) were recovered. The isolated fragments contained 109 bp of specific *Ph otd1* and *Ph otd2* sequence and were sufficient for unambiguous homology identification (5.3.3.1; A2.1; A2.2).

In detail, five primary sequences were recovered from the approaches described above. BLAST evaluation revealed that they represent novel *otd/otx* sequences. They fall into two groups, *Ph otd1* and *Ph otd2*, with three sequences representing *Ph otd1* and two sequences representing *Ph otd2*. Individual sequences of one group differ from any sequence of the other in at least 12.8% nucleotide positions, supporting that the recovered homeobox fragments actually derive from two different, albeit closely related genes. The sequence fragments of the *Ph otd1* group cover several polymorphic sites (see below), which makes them differ from each other in single nucleotides only. They are therefore considered identical. The two recovered *Ph otd2* homeobox fragments are completely identical with regard to their

nucleotide sequence. Since each of the five recovered sequences derives from a different priming event, they are considered independent.

Based on this primary sequence information, specific oligonucleotides were designed and used to prime *Ph otd1* and *Ph otd2* RACE reactions (5.2.2.3). Consequently, the remaining mRNA transcript sequences both 5' and 3' of the initial fragments covering the HD were obtained.

#### 3.2.1.2 Isolation of *Ph otd1*

From single-step 5' RACE, three independent *Ph otd1* 5' cDNA fragment clones were obtained. Their sequences show the same transcription start. The fraction of sequence they cover contains several polymorphic nucleotide exchanges (see below). If these are not taken into account, the nucleotide variation between any two 5' RACE sequences is not exceeding 1.5%. For this reason, the sequences of the recovered *Ph otd1* 5' cDNA fragment clones are considered to belong to one gene locus (A2.1; A5).

*Ph otd1* 3' cDNA sequence was fully recovered by single-step 3' RACE. Three independent clones covering identical (less than 1% nucleotide variation), full 3' *Ph otd1* transcript sequences were obtained. They are of the same length, showing only minor differences in poly(A) tailing.

Sequence information provided by all recovered 5' and 3' *Ph otd1* RACE clones was sufficient for assembling the complete *Ph otd1* cDNA sequence *in silico*. In order to verify the consistency of the established *Ph otd1* cDNA sequence, coherent *Ph otd1* ORF and cDNA fragments were isolated via long-distance PCR reactions (5.2.3.4) performed on *Parhyale* cDNA collections (5.2.2). Five independent *Ph otd1* cDNA sequences and six independent *Ph otd1* ORF sequences were obtained (A2.1). They are consistent with the findings from 5' and 3' RACE (A2.1). Notably, three cDNA clones, Ph_otd1_cDNA27, Ph_otd1_cDNA29 and Ph_otd1_cDNA34 show insertions at the same position 5' of the homeobox that translate in frame into various additional N-terminal protein elements, putatively representing *Ph otd1* transcript isoforms (Ph_otd1_cDNA27: 15 bp insertion corresponding to 5 additional amino acids; Ph_otd1_cDNA29 and Ph_otd1_cDNA34: 216 bp insertions corresponding to 72 additional amino acids). If neither these insertions nor the polymorphic nucleotide exchanges found within the *Ph otd1* cDNA (see below) are taken into account, the nucleotide variation

between any two of the recovered *Ph otd1* ORF and cDNA sequences is not exceeding 1.5% (A2.1; A5). These findings suggest that all recovered *Ph otd1* sequences derive from transcripts of different alleles of the same gene (5.3.3.2).

#### 3.2.1.3 Characterisation of *Ph otd1*

The comparison of all recovered *Ph otd1* sequences led to the identification of 127 sites of nucleotide exchange within the complete *Ph otd1* cDNA. 93 of those occur uniquely and are therefore considered sporadic (5.3.3.2). The majority of these (56) are found in the ORF where they account for 36 amino acid changes (missense alterations). The remaining 20 nucleotide exchanges found within the ORF do not alter the translated amino acid sequence (silent exchanges). These findings go well with the expected randomness of artificial nucleotide exchanges. 34 nucleotide exchanges are present in more than one sequence and therefore considered polymorphic. Of the 20 polymorphic nucleotide exchanges found within the ORF, 18 do not alter the amino acid sequence (silent exchanges). The remaining two lead to amino acid exchanges outside the homeodomain. These observations suggest that these nucleotide exchanges represent naturally occurring polymorphism of wild type *Ph otd1* alleles (5.3.3.2).

Based on these findings, the sequence of the clone Ph_otd1_cDNA25 was chosen as the source of the *Ph otd1* reference sequence (*Ph otd1_ref*). It was extended 5' by adding 30 bp derived from clone Ph_otd1_5R01 and 3' by adding 57 bp derived from clone Ph_otd1_3R12 (A2.1.1). Sporadic and putatively artificial nucleotide exchanges were substituted with the 'consensus' nucleotides that are present in all other obtained sequences (1439G>A,1502G>A,1519 G>A, 1885T>-, 1949 A>G). An alignment of *Ph otd1_ref* with all recovered *Ph otd1* sequences confirms its statistical consensus relevance (A2.1.2). Within the fraction of sequence they share, all recovered Ph otd1 sequences show less than 1.2% difference from *Ph otd1_ref*. *Ph otd1_ref* has been used for phylogenetic studies.

The *Ph otd1* transcript is 2.0 kb in length and encodes a protein of 396 amino acids. The Parhyale OTD1 protein has a HD (60 amino acids) that is highly similar to HD of described OTD and Otx proteins (Figures 9, 10). However, it lacks a conserved SIWSPASI motif found C-terminally within several described OTD and Otx proteins of other species (Browne et al., 2006).

**Figure 9: Schematic view of the *Ph otd1* transcript.** The length of the *Ph otd1* transcript is 2036 bases. Shown are: 5' UTR in grey (length 255 bases, nucleotide positions 1-255), ORF in black (length 1191 bases, encoding 396 amino acids; nucleotide positions 256-1446) and the 3' UTR in grey (length 590 bases, nucleotide positions 1447-2036). The ORF region encoding the HD is depicted in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 517-696).

```
0001                    AGTTGTGCGAACGTAAGCTTGTCGAATACGTTGTGTAGGATACCTCTCAAAACTTGTGTGCGGACTTCCATTAGT
0076          GATTCTTCCATTTAAAACCGGACATGCTAATTGATAGTCGGATTAGCAGCTGCAGTAATCTCCCTTTGCGCCTGTGCCAGTGAACGTACC
0166          TATTCCAGAAACTCCAACTCTGAAGGGTATTGTGATATACCCAGTGTTCTTCAGCGCCTCTTTCGAGACATCTTTGATCATCCACCACCG

0001          M  H  M  N  G  D  P  M  A  Y  H  G  S  F  V  T  Q  S  Q  T  L  G  N  T  M  A  W  R  T  A
0256          ATGCACATGAACGGAGATCCCATGGCCTACCATGGTTCATTTGTGACTCAGTCACAAACTCTGGGTAACACGATGGCCTGGCGGACCGCC

0031           M  S  N  A  T  A  G  L  P  M  G  M  A  T  H  H  P  H  S  M  V  D  P  T  L  H  M  S  Y  H
0346          ATGTCGAACGCCACAGCTGGGTTGCCCATGGGTATGGCAACCCATCATCCCCACTCCATGGTGGACCCTACGCTGCACATGAGCTACCAT

0061           S  T  A  E  S  R  D  H  L  L  G  M  S  S  F  P  S  L  R  M  P  P  Q  P  G  R  K  Q  R  R
0436          TCCACCGCAGAATCCAGGGATCATCTACTTGGGATGAGTTCATTTCCTTCCCTTCGAATGCCACCACAACCAGGCCGCAAGCAACGCAGG

0091           E  R  T  T  F  T  R  A  Q  L  D  V  L  E  T  L  F  G  K  T  R  Y  P  D  I  F  M  R  E  E
0526          GAGCGCACCACCTTCACTAGAGCTCAGCTCGACGTACTAGAAACTTTGTTCGGCAAGACGCGGTATCCCGACATCTTCATGAGAGAGGAA

0121           V  A  I  K  I  N  L  P  E  S  R  V  Q  V  W  F  K  N  R  R  A  K  C  R  Q  Q  N  T  N  N
0616          GTTGCGATCAAAATCAACCTACCAGAGAGTAGAGTGCAGGTTTGGTTCAAGAACCGCCGAGCGAAGTGTCGTCAGCAGAACACCAACAAC

0151           N  N  N  S  N  S  N  N  S  S  S  N  N  N  S  S  S  S  S  S  G  G  T  A  T  S  N  T  A  S
0706          AATAACAACAGCAACAGTAACAACAGCAGCAGCAACAATAACAGTTCCAGCAGTTCCTCAGGTGGCACCGCTACCAGCAACACAGCCAGC

0181           T  V  G  S  V  A  S  G  T  S  T  T  D  S  T  R  A  A  A  P  R  A  K  K  S  T  A  K  P  S
0796          ACTGTCGGCAGTGTGGCTTCTGGCACCAGCACGACGGACAGCACTCGGGCCGCCGCTCCTAGGGCCAAGAAATCCACAGCCAAACCAAGC

0211           T  P  P  V  A  T  A  T  N  T  T  S  S  A  V  P  S  T  V  K  S  E  P  R  T  Q  S  P  A  A
0886          ACCCCACCTGTAGCTACCGCCACCAACACTACATCTTCTGCTGTGCCATCCACGGTCAAATCTGAACCCAGGACCCAATCCCCCGCAGCA

0241           Y  K  P  P  T  T  S  P  A  S  A  H  T  P  P  N  T  T  P  T  P  T  Q  P  Y  A  G  F  Q  G
0976          TACAAACCCCCAACCACCTCTCCTGCATCAGCACATACCCCACCCAACACCACACCAACCCCTACCCAACCCTATGCTGGGTTCCAGGGA

0271           A  A  H  T  L  S  T  N  V  V  D  P  N  A  Y  S  S  S  L  W  Q  S  T  A  L  R  S  D  L  G
1066          GCTGCTCATACCCTTTCGACGAACGTGGTCGACCCCAACGCCTATTCTTCCAGCCTATGGCAATCGACAGCTTTACGGTCGGACCTGGGT

0301           T  P  N  C  M  Q  T  T  G  H  Y  T  P  M  P  P  H  A  K  T  P  P  S  W  N  Q  S  Y  T  P
1156          ACACCGAACTGTATGCAGACTACGGGACACTACACGCCGATGCCTCCTCATGCCAAGACCCCCCCTTCATGGAACCAGAGCTACACCCCC

0331           S  Y  Y  G  N  I  E  Y  L  H  Q  P  N  H  F  N  T  S  H  M  T  N  F  S  Q  M  A  V  G  H
1246          TCATACTACGGTAACATTGAGTACCTCCACCAGCCAAATCATTTCAACACGAGCCACATGACCAACTTTTCTCAGATGGCCGTCGGTCAT

0361           H  N  M  N  S  Q  M  T  A  A  Q  Q  Y  R  A  A  P  T  K  Q  E  C  M  D  Y  G  G  T  A  A
1336          CATAATATGAATTCCCAAATGACAGCAGCCCAGCAGTACAGAGCGGCTCCGACCAAACAGGAATGTATGGATTACGGAGGTACCGCAGCC

0391           E  S  F  H  N  L  &                                                                    0396
1426          GAAAGCTTTCACAATCTTTGAACGGCATTCTTATAGATACGCCCTTCAAATATTTTGTTTATTACTTGCTGTTTAGAGTTAGCGGATTGC
1516          AATAGGGATTGCTTGGTGTGTTGAAGGTTTAGAGTTTCTGTTGTGGAGACGCTCCGTTGTTTGTTACGATAAGGCTGAACAAGAAGCAGT
1606          TATAGCATCGCCATGACTTTGTGCTGCTCCAAGTTCAGCAGCAGTATTGCGCTGTAAAACAATAAAAGGTTAAGGCGAGGATAGAGGAGC
1696          ATACCTGCGGAAAGATGAGTACTTGTGTGAACTGCTGTGGCAATTCTGATAACAATCGCTGAACTGGTAATGACCTTTGATTAAATGATG
1786          TCCATATTGAGAAGTTTTCCATGTTCGTTTTTTTGTTATCTACCGGAATGCATGGATGAATTTTTGCATAAAACACTTTTCATTTTTTTT
1876          TTTTTTTTGAGAAGGTAAGATGTTGAGTATGTGAATTGGCATTATTACATTGATACATTTGTTTCGTATAAGTGGACACGAAATTATACT
1966          TTCTTTTTTTTTTCCACTGTTACGATATGTTTGACCTTTTCTACTTGCGAGAAAAAAAAAAAAAAAAAAAA                2036
```

**Figure 10: *Ph otd1* cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph otd1* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The HD is shown in blue. The nucleotide sequence that encodes this domain is shown in blue as well. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

### 3.2.1.4 Isolation and characterisation of *Ph otd2*

In contrast to *Ph otd1*, RACE reactions revealed a significantly more complex organisation of the *Ph otd2* paralog: five unique 5' and three unique 3' sequence fragments were recovered. Several of these share identical fractions of sequence while differing in the remaining sequence (Figure 11). This indicates the existence of multiple alternative splice variants. The true number of alternative *Ph otd2* transcripts that can be produced is unknown. For more details on the isolation and characterisation of *Ph otd2*, see (Browne et al., 2006).

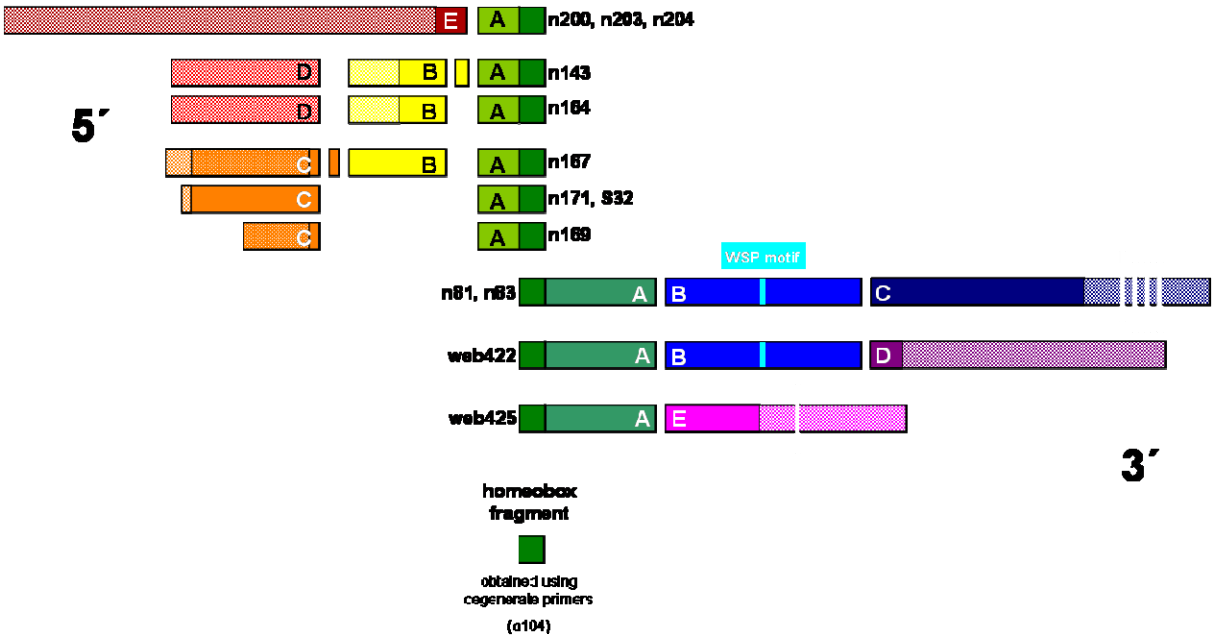



**Figure 11: Schematic overview over 5' and 3' RACE clones of *Parhyale hawaiensis otd2* and their putative splicing structure.** A schematic alignment of the nucleic acid sequences of 5' RACE clones (top of panel, left direction), 3' RACE clones (middle of panel, right direction) and the short *otd2* homeobox fragment obtained by PCR using degenerate primers (5.2.3.1, bottom of panel). Individual putative splicing units are characterized by respective colours 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, as with n171 and S32) is comprised of a different combination of respective splicing units. Coding sequence within the obtained RACE clones was determined by 5'-most methionine codon (5' RACE clones) or first 3' stop codon (3' RACE clones, respectively). Coding sequence is in solid colour, non-coding sequence is hatched in the respective colour. A WSP-motif present in translated 3'B splicing units and found in vertebrate *otx* orthologs is indicated in cyan. For more detailed information about individual RACE clones, see A1 and Browne et. al, (2006). All putative splicing units are displayed in their correct relative scale. White perpendicular strokes (clones n81/n83 3'C and clone web425 3'E) substitute for 500 bp each. 5'C of n169 appears to be truncated as compared to 5'C of n171/S32. It is thus characterized by an individual hatching since all of it appears to be open reading frame. A similar assumption applies to 5'C of n167. Here, 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, A1; Browne et al., 2006).

#### 3.2.1.5 Phylogeny and Expression of *Ph otd1 and Ph otd2*

Detailed analyses of the phylogeny as well as the expression profiles of *Ph otd1* and *Ph otd2* have already been published in "Expression of *otd* orthologs in the amphipod crustacean, *Parhyale hawaiensis*" (Browne et al., 2006, Appendix). This publication is the result of collaboration between all listed authors.

My contribution to this collaborative approach was the isolation of *Ph otd1* 5' and 3' RACE sequences independently and in parallel to those isolated by W. E. Browne as well as the isolation and verification of complete and coherent *Ph otd1* cDNA sequences. The sequence information gathered hereby was used to confirm the identity of *Ph otd1*. In addition, I isolated the majority of 5' and 3' RACE *Ph otd2* sequences. I put up an overview and a model for putative splicing units of *Ph otd2* to provide an explanation for the large number of transcript isoforms that were found for this gene (Browne et al., 2006; see also 3.2.1.4).

Because *Ph otd1* expression at stage 22 has not been presented in Browne et al. (2006), I want to give a detailed analysis of it in this work. In stage 22 embryos, the labrum has completed its posterior extension. Its lobes directly abut the paragnaths anterior, with the mandibles prominently protruding laterally on both sides of the labrum (Figure 12 D1-D3, E1-E3; 3.1.4; Browne et al., 2005). *Ph otd1* expression is found bilaterally in the medullar areas of the pre-antennal head. Here, *Ph otd1* expression appears strongest in the proximal parts of the medullar hemispheres (Figure 12 A1-A3, C1-C3). Anterior, this region will give rise to the medulla terminalis, and posterior, to the medulla interna (see also 3.1.4; Browne et al., 2006). The medial area of the pre-antennal head, where the hemi-ellipsoid bodies form, is free of *Ph otd1* expression (Figure 12 A1-A3, B1-B3). *Ph otd1* expressing cells appear as projections from the proximal medullar expression domains directed towards the bases of the An1 appendages (Figure 12 A2, A3). Small groups of *Ph otd1*-positive cells can be found to both sides of the stomodeum (Figure 12 D2). Strong *Ph otd1* expression is visible in the medial, distal part of the labrum and in medial cells of the paragnaths (Figure 12 D1-D3, E1-E3, F1-F3). *Ph otd1*-positive cells can be found at the bases of the mandibles (Figure 12 B2, D2, E2) as well as the ventral midline within the mandibular segment (Figure 12 B1-B3). *Ph otd1* is expressed in the ectoderm of the developing proctodeum and hindgut (Figure 12 E1-E3).

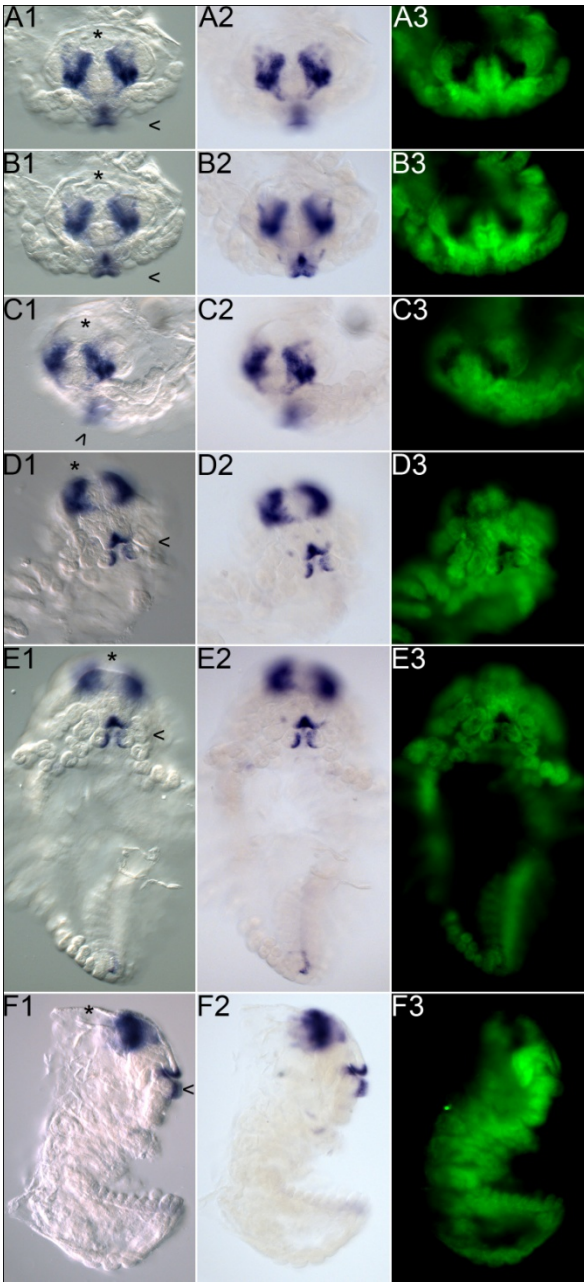



**Figure 12: Expression of *Ph otd1* at stage S22**. The embryo is mounted in anterior orientation, focal plane on the bases of the An1 appendages (A1-A3), in anterior orientation, focal plane on the labrum (B1-B3), in anterior orientation, tilted to the right (C1-C3), in ventral orientation, tilted to the left (D1-D3), in ventral orientation (E1-E3) and in lateral orientation, sagittal focal plane (F1-F3) in order to facilitate better understanding of the three-dimensional pattern of *Ph otd1* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-F1), bright field images in the middle panel (A2-F2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–F3; nuclei are green, expression of *Ph otd1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-F1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** Anterior view, focal plane on the bases of the An1 appendages. **B** Anterior view, focal plane on the labrum. **C** Tilted anterior view. **D** Tilted ventral view. **E** Ventral view. **F** Lateral view, sagittal focal plane. Ventral is to the right.

### 3.2.2 *Parhyale hawaiensis* genes of the *aristaless* group

In *Drosophila*, genes of the *aristaless* (*al*) group encode transcription factors that are involved in various aspects of development, including head and brain patterning and neurogenesis (Davis et al., 2003; Goriely et al., 1999; Hahn and Jäckle, 1996; Simeone et al., 1994; Walldorf et al., 2000; Xiong et al., 1994). Several *Parhyale* representatives of this group were isolated from the PCR screen targeting *paired*-class homeobox gene sequences. They were examined for their roles in *Parhyale* head development.

#### 3.2.2.1 Isolation and characterisation of *Parhyale hawaiensis aristaless* group genes

Varied-stringency degenerate PCR screens (5.2.3.1) targeting *Parhyale paired-class* homeobox genes were subjected to several rounds of known-sequence fragment depletion (5.2.3.2) in order to specifically enrich novel paired-type homeobox sequences. From this approach, three *Parhyale aristaless*-like genes were recovered. The isolated fragments contained 109 bp of specific sequence that was sufficient for unambiguous homology identification (5.3.3.1).

In detail, nine sequences were recovered. BLAST evaluation (5.3.3.1) revealed that they represented three different, novel *aristaless*-like genes. Since each sequence derives from a different priming event, they are considered independent.

Three of those sequences were identified as *Parhyale homeobrain* (*Ph hbn*), based on their high sequence similarity to *Drosophila hbn* (blastx; taxid: 7227; A6.1, table A5; A6.2, table A6; 5.3.3.1). Four sequences resemble *Parhyale aristaless1* (*Ph al1*), showing the highest sequence similarity to *Drosophila aristaless* (blastx; taxid: 7227; A6.1, table A5; A6.2, table A6; 5.3.3.1). The remaining two sequences were identified as *Parhyale aristaless2* (*Ph al2*) since they also show the highest sequence similarity to *Drosophila aristaless* (blastx; taxid: 7227; A6.1, table A5; A6.2, table A6; 5.3.3.1), but at a lower BLAST score than *Ph al1*. The two sequences differ from each other in two nucleotide positions that were, however, identified as polymorphic (5.3.3.2).

A direct comparison of all recovered *Ph al1* and *Ph al2* homeobox fragments reveals a nucleotide sequence difference of at least 23.9%, strongly supporting that they actually derive

from two different genes. The nucleotide sequence of *Ph hbn* essentially differs from those of *Ph al1* and *Ph al2 (*35.8% and 39.4%, respectively).
Based on this primary sequence information, specific oligonucleotides were designed and used to prime *Ph hbn, Ph al1* and *Ph al2* RACE reactions (5.2.3.3, A4.1, table A3). Consequently, the remaining mRNA transcript sequences both 5' and 3' of the initial fragments covering the HD were obtained.

#### 3.2.2.2 Isolation of *Ph hbn*

From single-step 5' RACE, supplemented by two-step 5' RACE, six independent *Ph hbn* 5' cDNA fragment clones were obtained. Within the sequence they share, they show minor nucleotide variation, with differences between any two of them not exceeding 2% (polymorphic sites not considered, see below). These differences are found predominantly within the 5' ends of the sequences. All *Ph hbn* 5' cDNA fragment clones show identical transcription starts. These findings suggest that the recovered *Ph hbn* 5' cDNA fragment clones derive from transcripts of the same gene (5.3.3.3).

From an initial round of two-step 3' RACE, three independent *Ph hbn* 3' cDNA fragment clones were recovered. They cover a consistent ORF fragment, but show premature poly(A)-tailing within the ORF. For this reason, several rounds of supplementary 3' extension RACE were performed, yielding seven additional independent *Ph hbn* 3' cDNA fragment clones. Although they expand 3' to varying extend, the sequence information provided by all of them was sufficient to reconcile the complete 3' fraction of the *Ph hbn* cDNA. Within the sequence they share, they show minor nucleotide variation (less than 1.3%). Given the fact that premature poly(A) tailing was observed in several of the *Ph hbn* 3' cDNA fragment clones, poly(A) tailing varies in general considerably between individual clones.

Sequence information provided by all recovered 5' and 3' *Ph hbn* RACE clones was sufficient for assembling the complete *Ph hbn* cDNA sequence *in silico*. In order to verify the consistency of the established *Ph hbn* cDNA sequence, coherent *Ph hbn* cDNA and, independently, fragments covering only the ORF were isolated via long-distance PCR reactions (5.2.3.4), performed on *Parhyale* cDNA collections (5.2.2). Three independent *Ph hbn* cDNA sequences and three independent *Ph hbn* ORF sequences were obtained (A2.3). They are consistent with the findings from 5' and 3' RACE (A2.3). The nucleotide variation

between any two of these does not exceed 1%, supporting that these sequences derive from transcripts of the same gene (5.3.3.3, A2.3, A3, table A1).

#### 3.2.2.3 Characterisation of *Ph hbn*

The comparison of all recovered *Ph hbn* sequences led to the identification of 44 sites of nucleotide exchange within the complete *Ph hbn* cDNA (see also A3, table A1). 34 of those occur uniquely and are therefore considered sporadic (5.3.3.2). They are found predominantly in the ORF (21) where they account for 13 amino acid changes (missense alterations) and one frame shift. The remaining seven nucleotide exchanges found within the ORF do not alter the translated amino acid sequence (silent exchanges). These findings go well with the expected randomness of artificial nucleotide exchanges. Ten nucleotide exchanges are present in more than one sequence and therefore considered polymorphic. Of the four polymorphic nucleotide exchanges found within the ORF, three do not alter the amino acid sequence. The remaining one leads to a change of the respective amino acid residue (T415>S415) outside the conserved HD. These observations suggest that the polymorphic nucleotide exchanges represent transcript variations resulting from different wild type *Ph hbn* alleles (5.3.3.2).

Based on these findings, the sequence of the clone Ph_hbn_cDNA01 was chosen as the source of the *Ph hbn* reference sequence (*Ph hbn_ref*). It was extended 5' by adding 60 bp derived from clone Ph_hbn_5Rn25 and 3' by adding 32 bp derived from clone Ph_hbn_3RHH357 (A2.3.1). An Alignment of *Ph hbn_ref* with all recovered *Ph hbn* sequences confirms its statistical consensus relevance (A2.3.2). If polymorphic nucleotide exchanges are not taken into account, all recovered Ph hbn sequences show a maximum of 1.1% difference from *Ph hbn_ref*, confirming that they in fact derive from transcripts of *Ph hbn*. *Ph hbn_ref* has been used for phylogenetic studies.

The *Ph hbn* transcript is 2.7 kb in length and encodes a protein of 493 amino acids. The *Parhyale* HBN protein has a HD (60 amino acids) that his highly similar to the HD of *Drosophila* HBN (3.2.2.8). Accordingly, glutamate is found at position 50 of the *Parhyale* HBN HD. N-terminally, an amino acid motif is present that shows high similarity with octapeptide/GEH domains found in *Drosophila* HBN and other closely related proteins (Figures 13, 14).

**Figure 13: Schematic view of *Ph hbn* transcript.** The length of the *Ph hbn* transcript is 2651 bases. Shown are: 5' UTR in grey (length 171 bases, nucleotide positions 1-171), ORF in black (length 1482 bases, encoding 493 amino acids; nucleotide positions 172-1653) and the 3' UTR in grey (length 998 bases, nucleotide positions 1654-2651). The ORF region encoding the octapeptide/GEH domain is depicted in orange (length 24 bases, encoding 8 amino acids; nucleotide positions 517-540).The ORF region encoding the HD is depicted in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 694-873).

```
0001                GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAGTCGAGGAGATACTTACCAAAGTGTAAAAATGAGGAAAAATTGAA
0082        CTTGCAAAATTATGTAAAATGAAACTCCAACGTTTCTACACTTCCTGAAGTTTTTTGACTGGTGACATAAACTCTAGACCGAAATCTACC

0001        M  S  V  N  S  T  I  A  S  V  I  S  T  P  P  G  S  K  S  T  T  A  D  H  L  A  R  D  Q  T
0172        ATGAGTGTCAATAGCACCATTGCCAGCGTCATATCAACCCCACCAGGATCGAAATCTACGACAGCTGACCATCTAGCGAGAGATCAAACC

0031         C  S  K  I  M  R  S  P  S  H  Q  L  T  V  K  T  I  P  F  S  T  V  S  D  C  P  T  V  G  N
0262        TGTAGCAAAATTATGCGAAGCCCTTCGCACCAGTTAACGGTGAAAACAATTCCTTTTTCTACTGTCAGTGACTGTCCGACAGTAGGAAAT

0061         T  V  T  V  D  M  S  P  S  N  C  T  M  A  G  D  T  V  N  V  P  A  S  E  V  T  I  S  Q  I
0352        ACAGTCACGGTAGACATGTCTCCCAGTAACTGTACAATGGCAGGTGATACAGTTAATGTACCAGCCTCAGAAGTTACAATAAGTCAAATT

0091         S  S  M  V  E  P  V  P  T  V  I  C  P  S  I  P  N  A  S  Q  K  T  P  Q  I  Y  T  I  E  N
0442        TCGTCCATGGTGGAGCCCGTGCCAACGGTCATCTGCCCGTCAATTCCTAATGCTAGTCAGAAAACCCCACAAATTTACACCATAGAAAAT

0121         I  L  A  L  K  T  S  S  S  H  N  K  H  I  D  S  D  V  N  D  A  G  C  P  S  E  S  P  S  P
0532        ATACTCGCCCTCAAGACCAGCTCATCTCATAATAAACATATCGACAGCGACGTCAATGATGCTGGCTGCCCGTCCGAATCCCCCTCGCCG

0151         P  K  C  P  P  S  P  S  S  D  H  A  V  E  G  T  D  L  E  K  P  R  K  Q  V  R  R  S  R  T
0622        CCCAAGTGTCCCCCGTCTCCTTCCTCCGACCACGCCGTGGAGGGAACGGACTTGGAGAAACCCCGCAAGCAAGTTCGGCGATCCCGTACA

0181         T  F  T  T  F  Q  L  H  Q  L  E  R  A  F  E  K  T  Q  Y  P  D  V  F  T  R  E  E  L  A  L
0712        ACGTTCACAACGTTCCAACTCCACCAGCTGGAGCGGGCGTTCGAGAAAACCCAGTATCCAGACGTCTTCACCCGAGAGGAGTTGGCACTC

0211         N  L  D  L  S  E  A  R  V  Q  V  W  F  Q  N  R  R  A  K  W  R  K  R  E  K  A  M  G  R  E
0802        AACCTTGACCTATCCGAGGCACGCGTGCAGGTGTGGTTCCAGAACCGACGAGCCAAATGGCGCAAGAGAGAGAAGGCGATGGGCAGGGAG

0241         S  P  S  P  Y  I  A  G  A  H  N  D  A  I  I  S  S  L  P  D  I  V  G  I  P  I  T  S  S  A
0892        TCGCCCTCGCCGTACATTGCTGGCGCTCACAACGATGCAATCATATCCAGTCTGCCAGATATTGTCGGTATACCAATAACATCGTCTGCT

0271         P  H  P  P  S  A  S  S  A  A  M  A  A  A  A  A  A  S  T  A  D  F  F  K  L  Q  A  F  H  A
0982        CCTCATCCTCCGTCGGCGTCATCTGCAGCTATGGCTGCAGCTGCCGCGGCTTCCACTGCAGATTTCTTTAAGCTGCAGGCGTTCCACGCT

0301         L  Q  L  N  T  F  L  P  H  L  Q  P  K  S  V  V  H  P  S  H  P  L  Q  S  L  F  Q  P  Y  L
1072        CTGCAACTCAATACCTTCCTGCCTCATCTGCAGCCCAAATCGGTTGTCCATCCCAGTCATCCACTCCAGTCTTTATTTCAGCCATACCTT

0331         L  P  S  S  L  P  I  I  P  N  H  P  P  I  F  P  A  L  P  I  S  P  M  G  S  P  D  V  V  S
1162        CTTCCATCTTCCCTCCCCATAATCCCCAACCATCCACCTATCTTTCCAGCTCTCCCAATATCCCCAATGGGGTCTCCAGATGTGGTCTCC

0361         T  P  S  T  S  P  N  S  G  L  A  T  T  A  P  Q  F  P  S  C  S  T  S  S  D  L  P  N  V  S
1252        ACCCCATCCACCTCCCCAAATTCCGGTTTAGCCACAACGGCACCTCAATTTCCCAGCTGCAGTACTTCCTCTGATCTTCCGAACGTTTCC

0391         A  H  N  V  A  V  I  N  P  L  T  S  L  A  H  M  T  A  A  P  S  M  D  L  T  T  T  S  V  A
1342        GCACACAATGTTGCCGTTATCAACCCATTAACTTCCTTAGCCCACATGACTGCAGCACCTTCCATGGATCTGACCACCACCTCGGTCGCG

0421         S  T  T  P  L  L  Y  T  A  D  S  R  G  Q  S  V  D  L  L  R  L  K  A  R  Q  H  Q  V  L  L
1432        TCTACCACCCCTCTACTATACACCGCCGACTCACGAGGGCAATCCGTGGATCTGCTGCGGCTCAAGGCGCGCCAGCACCAGGTCCTGCTG

0451         E  H  L  S  S  A  C  G  S  G  G  A  G  V  A  G  G  V  V  G  T  E  S  G  P  G  G  V  P  G
1522        GAGCATTTATCCTCCGCCTGTGGTTCCGGGGGTGCAGGGGTGGCCGGTGGTGTCGTTGGGACTGAATCAGGGCCGGGGGGAGTCCCTGGG

0481         T  L  K  I  D  P  Q  L  P  S  Q  L  S  #                                                  0493
1612        ACTCTAAAGATTGACCCCCAACTGCCTTCGCAATTGTCCTAACTGAAGTTCCGCCGACCCCCCATAACTCCGATTGTTGTGACGTGTCAG
1702        GCGCAGACATGGAACAAGAGCTGATCTTTCGGGTTTTCCTTTCGTTCATGCACCCTGTCAGCTGCAACACAATCTTTGTCAGACGTTTTC
1792        ACTCTATCAACCACACGGCACGTACTGTCTTCCAGTTCTGCCCCGATCTCATCCCTCTGTCAGTCGTGCTTCAGCGTCCAGTCGATCGTT
1882        TGCAGAGTCTGTTCGCATTCTTGTACAATATGCAGTCAACATTTTAACGATATTTACTAGAACTGCGCAACCCAACCATATCCGAAACAT
1972        TTCCTTTCCGCCAGAAGTACACAAGCCAGTTATTTTGGGCAAAATACGCCTGTTAGCTACACATCGAGCCAGTTAATTACGGTGTAAAAT
2062        GGTTGCAATGTCTGCCAGTTTACGATTCAGTCAGCGAGCTATGTTACGGTGCTCGTGCTCATCCGGCAACAAATAAGTTTGACCATCTAC
2152        TAGTTTGTCTCTCAATGTAAAACCTTTGTGACGATGTCTAGCAGCTTGTTTCTCATCCAGCTAAATTATAACTCATCCAACGACCAGTGT
2242        TACATAAATCGCATGCGTCGGGTTTTCCTTTATCGTGCCAGCTATGAAACGCTGTCTGCCGCTCATCCAGTCAGATAAATTGAACCAGTT
2332        GACAGACTTGCACCCATCTAGCCCGCCATTTTGAAGGCTCTAATACGAAGCAGATGCTACCTCTTCTCACAATGAAAGGGTTCGTTTTTT
2422        AGCGCTTAGCAGCAACCGCTTGTGATCTAACGAGCGTTGGAGCGTTGCCACCCATGTTGTACCGAGGGATGGTGTGATAGCACGGTACCA
2512        ATCAATGGTATATGGTACTAATATAGCGCATAACACCACATAAATTATTAGAGTACACTGTTTTGCACTAACTTCGTCTTGCTTATCGGC
2601        AAAGATTTGTTGTAAGAGATCGTGATCAAGGGATAAACCCCATGACATAC                                         2651
```

**Figure 14: *Ph hbn* cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph hbn* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and

above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The octapeptide/GEH domain is shown in pink and the HD in blue. The nucleotide sequences that encode these domains are shown in the respective colours. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

#### 3.2.2.4 Isolation of *Ph al1*

From two-step 5' RACE, five independent *Ph al1* 5' cDNA fragment clones were obtained. Within the sequence they share, several polymorphic sites of nucleotide exchange have been identified (see below). If these are not taken into account, the nucleotide variation between any two of them is well below 1%. All *Ph al1* 5' cDNA fragment clones show identical transcription starts. These findings strongly support that they derive from transcripts of the same gene.

*Ph al1* 3' cDNA sequence was fully recovered by two-step 3' RACE. Three independent *Ph al*1 3' cDNA fragment clones were obtained. The nucleotide variation between any two of their sequences does not exceed 1.7%, with most nucleotide exchanges found within the sequence of clone Ph_al1_3Rn7 (A2.4, A3, table A1). All three cover the same fraction of *Ph al1* 3' cDNA and show only minor differences in poly(A) tailing. These observations suggest that they derive from transcripts of the same gene.

Sequence information provided by all recovered 5' and 3' *Ph al1* RACE clones was sufficient for assembling the complete *Ph al1* cDNA sequence in silico. In order to verify the consistency of the established *Ph al1* cDNA sequence, coherent *Ph al1* cDNA and, independently, fragments covering only the ORF were isolated via long-distance PCR reactions (5.2.3.4), performed on Parhyale cDNA collections (5.2.2). Three independent *Ph al1* cDNA sequences and six independent *Ph al1* ORF sequences were obtained (A2.4.1). They are consistent with the findings from 5' and 3' RACE (A2.4.2). Apart from polymorphic nucleotide exchanges (3.2.2.5), the sequences of any two of these vary in less than 1% of nucleotide positions. This strongly suggests that these sequences derive from transcripts of different alleles of the same gene (5.3.3.2, see also A3, table A1).

### 3.2.2.5 Characterisation of *Ph al1*

The comparison of all recovered *Ph al1* sequences led to the identification of 76 sites of nucleotide exchange within the complete *Ph al1* cDNA (see also A3, table A1). 39 of those occur uniquely and are therefore considered sporadic (5.3.3.2). Half (15) of those found within the ORF (30) alter the amino acid sequence (missense alterations). The other half of them does not change the amino acid sequence (silent exchanges). These findings suggest that these nucleotide exchanges are artificial. 37 nucleotide exchanges are present in more than one sequence and therefore considered polymorphic. All of them are found within the ORF. Eleven of these cause amino acid exchanges outside the conserved homeodomain. It is important to note that these amino acid variations appear to be linked, meaning that within individual *Ph al1* cDNA and ORF clones, either all of the amino acid alternatives are found or none, in this case resembling the *Ph al1* reference sequence (see below, A2.4.2, A3, table A1). The differences in the two resulting proteins are minor, however, arguing against the existence of two closely related *Ph al1* paralogs. In two of the recovered *Ph al1* 5' RACE clones, an additional adenine is present within a stretch of six adenines. This causes a frame shift and would lead to a truncated AL1 protein *in vivo*. Notably, none of the *Ph al1* cDNA and ORF clones shows this single nucleotide insertion. The remaining 25 nucleotide exchanges in the ORF that were identified as polymorphic do not alter the amino acid sequence. These observations suggest that at least two rather diverse *Ph al1* alleles exist.

Based on these findings, the sequence of the clone 'Ph_al1_cDNA01 was chosen as the source of the *Ph al1* reference sequence (*Ph al1_ref*). It was extended 5' by adding 35 bp derived from clone Ph_al1_5Rn2 and 3' by adding 62 bp derived from clone Ph_al1_3Rn8. One sporadic nucleotide exchange was altered in order to represent the statistically prevalent nucleotide residue (1048 T>C, A2.4.1). An Alignment of *Ph al1_ref* with all recovered *Ph al1* sequences confirms its statistical consensus relevance (A2.4.2). *Ph al1_ref* has been used for phylogenetic studies.

The *Ph al1* transcript is 1.8 kb in length and encodes a protein of 510 amino acids. The *Parhyale* AL1 protein has a HD (60 amino acids) that his highly similar to the HD of Aristaless proteins of other species (3.2.2.8). Accordingly, glutamate is found at position 50 of the *Parhyale* AL1 HD. C-terminally, a sequence motif is present that is highly similar to OAR/*aristaless* domains found in *Drosophila* AL and other closely related proteins (Figures 15, 16).

**Figure 15: Schematic view of *Ph al1* transcript.** The length of the *Ph al1* transcript is 1849 bases. Shown are: 5' UTR in grey (length 169 bases, nucleotide positions 1-169), ORF in black (length 1533 bases, encoding 510 amino acids; nucleotide positions 170-1702) and the 3' UTR in grey (length 147 bases, nucleotide positions 1703-1849). The ORF region encoding the HD is depicted in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 791-970). The ORF region encoding the OAR is depicted in orange (length 63 bases, encoding 21 amino acids; nucleotide positions 1529-1591).

```
0001                GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAATTTGAATAATATGAAACTAATATATCAAATTAACTGCTATTT
0080          TAATTCATTAATAGTCTTCTACTTACTTCTGTTTACTGCTGTATAAAAAAATTATAAGCAGTCAAAATTGAATCATTCAAGTGAAGAAAT

0001           M  V  N  F  S  E  L  A  V  Q  G  A  A  S  Q  V  R  P  V  F  G  A  C  R  T  S  P  I  T  S
0170          ATGGTGAATTTCAGCGAACTCGCTGTGCAGGGCGCCGCCTCTCAGGTACGTCCCGTCTTCGGCGCCTGCAGGACCTCCCCCATCACCAGT

0031           D  P  A  T  G  P  R  L  H  P  S  S  T  D  H  S  I  N  S  I  V  V  Y  N  N  H  S  N  D  N
0260          GATCCAGCGACGGGTCCTCGACTCCACCCTTCAAGCACAGACCATTCCATCAACAGCATCGTCGTCTACAACAATCACAGCAACGACAAT

0061           L  N  N  N  L  T  G  D  G  K  R  S  S  P  S  F  K  P  A  N  E  Q  E  T  A  T  Q  T  E  V
0350          TTAAATAACAACTTAACCGGCGACGGTAAAAGATCTTCTCCAAGTTTCAAACCTGCAAACGAACAAGAAACCGCCACCCAAACCGAGGTG

0091           Q  T  S  I  H  K  P  N  N  K  R  N  L  Q  S  F  H  L  H  H  I  L  D  Q  P  P  P  T  P  H
0440          CAAACTTCAATACACAAACCAAATAACAAACGAAATTTACAGAGTTTTCATCTGCACCATATTCTTGACCAACCTCCACCTACCCCGCAC

0121           I  N  V  V  D  D  D  D  Q  E  D  S  S  I  R  I  K  K  N  E  I  D  Y  D  S  N  E  E  S  Q
0530          ATCAATGTAGTCGATGATGACGATCAAGAAGACTCAAGCATAAGAATAAAAAAGAATGAAATTGATTACGACAGTAATGAGGAATCGCAA

0151           T  R  S  L  A  S  K  S  T  E  F  S  S  Q  S  S  S  K  E  L  Q  K  D  A  E  D  C  T  E  D
0620          ACGAGATCCCTTGCGTCAAAATCCACTGAATTTTCATCTCAAAGCAGTTCAAAGGAATTGCAGAAAGACGCAGAAGATTGTACAGAAGAT

0181           D  A  E  I  D  I  N  E  S  T  A  E  D  V  D  D  V  G  D  E  V  T  P  N  K  K  K  Q  R  R
0710          GACGCCGAAATTGATATCAATGAAAGCACCGCAGAAGACGTCGATGACGTAGGAGATGAAGTCACCCCTAACAAGAAAAAACAACGTCGA

0211           Y  R  T  T  F  T  S  F  Q  L  E  E  L  E  K  A  F  S  R  T  H  Y  P  D  V  F  T  R  E  E
0800          TATAGAACAACGTTCACGTCATTCCAGTTGGAAGAACTGGAGAAAGCGTTCTCCAGAACACATTACCCAGATGTCTTCACAAGGGAAGAA

0241           L  A  M  K  I  G  L  T  E  A  R  I  Q  V  W  F  Q  N  R  R  A  K  W  R  K  Q  E  K  V  G
0890          CTGGCAATGAAAATAGGTCTCACGGAAGCCCGTATCCAGGTATGGTTCCAAAATCGGCGAGCAAAATGGCGCAAGCAAGAGAAGGTCGGA

0271           P  S  H  H  L  A  G  P  F  P  S  S  S  S  V  P  L  H  P  P  P  L  P  S  S  F  I  G  T  T
0980          CCGAGCCATCACTTGGCGGGACCTTTCCCATCGTCTTCTTCGGTCCCACTTCATCCTCCTCCACTTCCTTCCTCTTTTATTGGCACTACG

0301           Y  T  S  P  F  A  P  P  P  H  P  P  T  G  A  P  H  S  S  G  K  T  M  D  F  P  I  L  P  G
1070          TACACTTCACCGTTCGCTCCTCCACCTCATCCTCCTACCGGTGCTCCGCACTCTTCCGGGAAGACGATGGACTTCCCGATACTTCCCGGG

0331           V  P  R  L  P  G  Y  L  T  G  L  V  P  S  C  L  A  S  L  G  G  L  S  S  L  G  G  L  G  A
1160          GTGCCTCGACTACCTGGCTACCTCACCGGTCTTGTGCCGAGCTGCCTGGCCAGCTTGGGAGGTCTCTCCAGTCTAGGCGGGTTGGGCGCT

0361           G  A  L  R  T  D  F  N  S  H  I  R  A  D  L  S  S  H  L  R  N  D  L  G  S  P  L  R  G  D
1250          GGTGCACTAAGAACGGACTTCAATAGCCACATTCGCGCAGATTTGAGTAGTCACCTCAGAAACGATCTGGGGAGCCCACTGAGAGGAGAT

0391           L  A  P  R  L  P  L  L  P  Q  T  L  P  Q  S  Y  P  P  V  F  Q  Q  I  L  A  G  L  R  P  P
1340          TTAGCTCCACGACTTCCCCTACTTCCCCAAACTCTGCCTCAGTCCTACCCCCCGGTGTTCCAGCAGATTCTGGCTGGTCTTCGTCCCCCG

0421           S  I  P  E  L  N  D  Y  S  S  I  L  A  S  S  A  A  L  S  T  A  S  D  V  G  P  N  N  S  D
1430          TCAATCCCCGAACTCAACGACTATAGCTCAATCTTGGCGAGCTCGGCAGCTCTCTCGACGGCATCGGATGTTGGACCCAATAATTCTGAC

0451           S  K  E  Q  E  R  R  A  S  S  I  A  E  L  R  Q  R  A  R  E  H  E  K  R  L  A  S  L  S  R
1520          AGCAAAGAACAAGAACGACGGGCTTCTTCTATAGCGGAGCTGCGACAAAGAGCCAGAGAGCATGAGAAAAGACTGGCGTCCCTCAGCCGG

0481           K  N  K  D  S  G  G  T  D  G  D  D  R  S  P  T  R  R  S  P  R  C  S  P  K  R  E  I  K  T      0510
1610          AAGAACAAAGATTCTGGAGGGACAGACGGAGACGACAGGTCTCCTACACGGAGATCCCCGCGATGTTCACCCAAAAGAGAGATCAAAACA

   &
1700          TGAATAAAATGGTAACAATTGTCAGAGAACAACATAAAACATCGATGAAAAATATTCTCGAAAAGGAGTCATATATGTAAATAGTAATGC
1790          TGGTTAGACTAAACATCAAAGAAAACACTTCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAA                              1849
```

**Figure 16: *Ph al1* cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph al1* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The

putative start and stop codons are shown in green and red, respectively. The HD is shown in blue and the OAR in orange. The nucleotide sequences that encode these domains are shown in the respective colours. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

#### 3.2.2.6 Isolation of *Ph al2*

From single-step 5' RACE, supplemented by two-step 5' RACE, seven independent *Ph al2* 5' cDNA fragment clones were obtained. Except clone Ph_al2_5Rg1 which appears 5' truncated, they all show identical transcription starts. A variation in the fraction of sequence they share manifests as a 15 bp sequence insertion (A2.5). In addition to that, several polymorphic nucleotide exchanges have been identified within the *Ph al2* cDNA fraction they cover. If these are not taken into account, the nucleotide variation between any two of them is well below 1%. These findings suggest that they derive from transcripts of the same gene.

In order to fully recover the *Ph al2* 3' cDNA sequence, 3' extension RACE was employed to overcome artificial premature poly(A) tailing present in the single clone obtained from an initial round of two-step 3' RACE. Two additional, independent *Ph al2* 3' cDNA fragment clones were obtained. Their sequences expand 3' to different extends. Within the fraction of sequence they share, the nucleotide variation is well below 1%, suggesting that they derive from transcripts of the same gene.

Sequence information provided by all recovered 5' and 3' *Ph al2* RACE clones was sufficient for assembling the complete *Ph al2* cDNA sequence in silico. In order to verify the consistency of the established *Ph al2* cDNA sequence, coherent *Ph al2* ORF fragments were isolated via long-distance PCR reactions (5.2.3.4), performed on Parhyale cDNA collections (5.2.2). Three independent *Ph al2* ORF sequences were obtained (A2.5.1). They are consistent with the findings from 5' and 3' RACE (A2.5.2) except that all three lack a nucleotide triplet encoding a threonine residue (T85') found in all *Ph al2* 5' RACE clones. Apart from polymorphic nucleotide exchanges (3.2.2.7), the sequences of any two of these vary in a maximum of 1.2% of nucleotide positions. This suggests that these sequences derive from transcripts of different alleles of the same gene. (5.3.3.2, see also A3, table A1).

#### 3.2.2.7 Characterisation of *Ph al2*

The comparison of all recovered *Ph al2* sequences led to the identification of 78 sites of nucleotide exchange within the complete *Ph al2* cDNA (see also A3, table A1). 34 of those occur uniquely and are therefore considered sporadic (5.3.3.2). Of the 32 sporadic nucleotide exchanges found within the ORF, eight alter the amino acid sequence (missense alterations), two lead to premature translation stops (nonsense alterations) and two manifest as single nucleotide insertions or deletions that lead to frame shifts with regard to the translated amino acid sequence. The remaining 20 of them do not change the amino acid sequence (silent exchanges). The high number of silent nucleotide exchanges suggests that at least some of these might not be artificial, but occur naturally in existing *Ph al2* alleles. 44 nucleotide exchanges are present in more than one sequence and therefore considered polymorphic. 35 of them are found within the ORF. Six of these cause amino acid exchanges outside the conserved homeodomain. As mentioned above, all three recovered *Ph al2* ORF sequences lack a nucleotide triplet encoding a threonine residue (T85') that is found in all *Ph al2* 5' RACE clones. A single nucleotide insertion is also found more than once, causing a frame shift. The remaining 20 polymorphic nucleotide exchanges found within the ORF do not alter the amino acid sequence. It is important to note that similar to the situation observed for *Ph al1*, several of the amino acid variations appear to be linked (see also A3, table A1). The differences in the resulting protein variants are minor, however, arguing against the existence of two closely related *Ph al2* paralogs. These observations suggest that several rather diverse *Ph al2* alleles exist (5.3.3.2).

Based on these findings, the sequence of the clone Ph_al2_ORF05 was chosen as the source of the *Ph al2* reference sequence (*Ph al2_ref*). It was extended 5' by adding 593 bp derived from clone Ph_al2_5Rn11 and 3' by adding 294 bp derived from clone Ph_al2_3Rfwx05. Several sporadic nucleotide exchanges were altered in order to represent the statistically prevalent nucleotide residues (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, A2.5.1). An Alignment of *Ph al2_ref* with all recovered *Ph al2* sequences confirms its statistical consensus relevance (A2.5.2). Ph al2_ref has been used for phylogenetic studies.

The *Ph al2* transcript is 2.6 kb in length and encodes a protein of 586 amino acids. The *Parhyale* AL2 protein has a HD (60 amino acids) that his highly similar to the HD of

Aristaless proteins of other species (3.2.2.8). Accordingly, glutamate is found at position 50 of the *Parhyale* AL2 HD. C-terminally, a sequence motif is present that is highly similar to OAR/*aristaless* domains found in *Drosophila* AL and other closely related proteins (Figures 17, 18).

**Figure 17: Schematic view of *Ph al2* transcript.** The length of the *Ph al2* transcript is 2594 bases. Shown are: 5' UTR in grey (length 566 bases, nucleotide positions 1-566), ORF in black (length 1761 bases, encoding 586 amino acids; nucleotide positions 557-2327) and the 3' UTR in grey (length 567 bases, nucleotide positions 2328-2594). The ORF region encoding the HD is depicted in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 1464-1643). The ORF region encoding the OAR is depicted in orange (length 63 bases, encoding 21 amino acids; nucleotide positions 2226-2288).

```
0001                                                                                                GAGTTTACCACCTGGCTTCGTTCTGG
0027          AAGCACTGCGCGTATTGCTCCATTGAACTGCAAGGCAATTATCACCTGGGCAGTGTGTCTTGTTTCGCAGTTATTTGGTACCTTACGGGG
0117          TTGAATTAAAAACAAACTTTTAATATAGGTCTTACCCTCAGGAACAGAACTGAAATATTAATTGCTGTGCTCATACTGAATTTTTTTCTA
0207          AAAGTGTATGTGTTGAAGCTTGATCTAAAATTTCAATGTGACGAGACTGGTATTTTTCGAAAGTTAGCTACTTGAGATAAACGATTTTCT
0297          TCTTTTGTGGTGTCAGTGTCACACTGTACATGGGATCTTAAATTATTCTTCGGAACTGCGCGAGAAAAGTCCTCTGAATTTATCAACAGT
0387          CTGGAGACCCTCATCCGAAACCCAGATGACTACAAGATTAACAGCCTCAGCTGCACCTACACAGGACTACTGAATTCTTCACTGGAGAGG
0477          TTTGGCATACAGTGACTTCACTCAAACTTCTATTTCTGAAAGTAAAATTCTCTTCTTACCTGCAATAGATTGAGCAAGTTACTTAAAACA
0001          M  G  I  L  E  E  C  S  V  S  S  Q  S  R  T  N  P  L  A  L  D  M  C  I  K  R  A  S  P  T
0567          ATGGGAATTCTGGAAGAGTGCAGCGTCTCCTCTCAGTCGCGGACAAACCCCCTGGCTCTGGACATGTGCATCAAGAGGGCCTCACCAACC

0031           M  L  A  H  S  D  A  V  A  S  S  S  S  S  S  L  M  P  V  D  P  E  M  T  S  T  T  V  S  D
0657          ATGCTCGCTCATAGCGATGCCGTAGCCAGCAGTTCCAGTTCCAGCCTGATGCCTGTAGACCCTGAGATGACGAGCACTACAGTGTCTGAC

0061           L  T  S  G  V  I  S  H  N  I  H  S  I  A  N  L  S  S  Q  H  F  G  Q  Q  G  G  R  T  S  P
0747          CTCACCTCGGGTGTCATTTCCCACAACATTCACAGCATTGCAAACCTCAGCTCGCAACATTTCGGCCAACAGGGTGGTCGAACATCTCCC

0091           R  G  R  S  S  P  K  S  P  K  S  P  K  N  V  L  I  P  K  P  T  S  E  T  I  S  R  A  Q  S
0837          AGAGGTAGAAGCTCACCAAAGTCCCCCAAGTCGCCGAAAAACGTGTTGATTCCCAAACCAACGAGCGAAACAATTTCCAGAGCACAGTCC

0121           I  L  S  I  H  N  L  T  K  Q  L  S  P  P  L  N  K  K  L  H  D  T  L  V  Q  M  V  T  C  A
0927          ATCCTCAGCATTCACAACCTGACAAAGCAGCTGTCTCCCCCGCTAAACAAAAAACTCCATGACACCCTGGTGCAAATGGTTACTTGTGCC

0151           Q  S  M  P  S  T  T  A  S  T  S  T  S  S  P  H  Q  N  L  H  S  R  R  A  L  S  P  P  F  S
1017          CAGTCGATGCCATCTACCACGGCTTCGACATCAACTTCATCTCCTCACCAAAATCTCCATAGTCGGCGAGCTCTCTCACCTCCGTTCAGT

0181           P  A  N  A  G  D  E  P  Q  S  L  A  H  M  K  L  P  M  L  H  Q  G  S  G  G  P  A  H  R  N
1107          CCAGCAAATGCTGGTGACGAACCACAGTCCTTAGCTCACATGAAGCTGCCTATGCTTCATCAAGGGAGCGGCGGACCCGCGCACAGGAAT

0211           T  K  K  N  R  S  A  Q  P  D  H  D  L  E  L  E  V  D  V  V  E  T  D  S  E  L  E  T  E  S
1197          ACGAAGAAAAACAGGTCGGCCCAGCCAGACCACGACCTGGAATTGGAAGTTGACGTCGTCGAAACAGACTCTGAATTGGAAACCGAATCA

0241           S  S  T  V  S  K  M  E  E  D  P  S  V  V  A  D  G  S  S  S  H  L  E  G  C  D  G  G  E  E
1287          TCCTCAACTGTTTCAAAAATGGAAGAAGATCCTTCGGTTGTTGCCGATGGATCTTCTTCACATCTGGAAGGGTGTGATGGAGGGGAAGAA

0271           V  A  D  E  N  G  G  E  D  G  E  D  C  G  E  G  G  S  G  G  D  T  A  G  P  T  K  R  K  Q
1377          GTAGCAGATGAGAACGGAGGAGAAGATGGAGAAGACTGTGGAGAGGGTGGGAGCGGAGGCGACACAGCTGGTCCTACCAAAAGGAAGCAA

0301           R  R  Y  R  T  T  F  T  S  Y  Q  L  E  E  L  E  K  V  F  S  R  T  H  Y  P  D  V  F  T  R
1467          AGGAGATATCGAACTACCTTCACTTCATACCAACTCGAAGAATTAGAGAAAGTGTTTTCCAGGACTCATTATCCGGATGTGTTTACCAGG

0331           E  E  L  A  M  K  I  G  L  T  E  A  R  I  Q  V  W  F  Q  N  R  R  A  K  W  R  K  Q  E  K
1557          GAGGAGTTGGCAATGAAGATAGGCCTGACGGAGGCACGCATACAGGTCTGGTTCCAGAACCGACGTGCAAAGTGGCGCAAGCAAGAAAAA

0361           V  G  P  S  S  H  P  Y  G  G  A  F  P  P  T  S  P  V  P  L  H  P  P  P  L  L  P  S  F  G
1647          GTTGGACCCTCAAGTCACCCGTACGGAGGAGCCTTCCCACCGACGTCCCCCGTTCCTCTACATCCTCCACCTCTTCTCCCATCATTTGGA

0391           S  T  G  G  A  Y  P  H  P  F  P  P  H  P  S  I  R  G  G  K  P  L  E  T  P  R  L  S  P  Y
1737          AGCACCGGCGGTGCTTACCCTCACCCGTTCCCACCCCACCCCTCCATACGGGGAGGCAAGCCCCTGGAAACCCCTCGTCTGTCCCCGTAC

0421           L  P  S  P  P  P  S  L  L  H  A  S  Y  L  A  G  G  L  P  G  L  A  T  P  T  G  M  A  S  L
1827          CTCCCCTCACCGCCGCCATCTTTGCTACACGCATCGTATCTCGCAGGAGGTCTGCCGGGGCTCGCGACCCCCACCGGCATGGCTTCTCTG

0451           L  L  R  E  L  P  S  A  Y  P  R  P  A  L  P  A  L  L  P  S  Q  T  P  S  F  T  T  V  L  A
1917          CTGCTCAGGGAGCTTCCCTCAGCTTATCCTCGACCAGCGCTACCTGCTCTCTTACCCTCACAAACGCCATCTTTTACAACCGTGTTAGCG

0481           Q  L  S  H  R  S  K  I  D  P  S  A  Y  Y  L  P  P  Y  L  A  H  L  T  A  L  P  P  S  T  L
2007          CAACTGTCACACCGATCCAAAATAGACCCATCAGCTTATTATCTACCACCCTACTTGGCCCATCTCACTGCTCTACCCCCGTCAACTCTC

0511           P  S  S  L  S  P  A  L  T  P  P  S  P  S  S  T  F  P  S  P  I  P  S  T  S  S  P  P  I  N
2097          CCCTCTTCTTTGTCCCCTGCCCTCACCCCACCTTCACCCTCGTCCACCTTTCCCTCTCCTATCCCTTCGACATCTTCTCCACCGATAAAT

0541           G  L  G  L  A  T  P  P  G  L  S  V  E  A  E  R  R  Q  A  S  L  H  D  L  R  L  K  A  R  N
2187          GGCCTGGGACTTGCGACTCCCCCAGGCCTCAGTGTAGAGGCAGAGCGGCGACAGGCGTCTCTCCACGACCTCAGACTTAAGGCACGAAAT

0571           Q  E  V  R  L  E  M  L  R  K  A  G  E  I  T  A  &                                          0586
2277          CAGGAAGTGCGACTTGAGATGCTACGAAAAGCTGGCGAAATTACAGCGTGAGGTAGAAGTAGTTTACTGGGTTCTTCAACGTAACTTTAG
2367          GGAACAACCATAAAGAATCAATACCATAGTTGCACCATAATATCAGTTATAAATTTTTCTGTAATGGTATATGCTATACTTTGTTTCCAA
2457          CTCCAGCAATATCATACGGAAAAATGATTGCCAATAGCTCGTCTACGATATTTAAGCCTTTAATATTCGTTTTTAATGATTCATTCCGCT
2547          CTCTGCTCAAATCATATCATTAGAAAAAAAAAAAAAAAAAAAAAAAAAA                                          2594
```

**Figure 18:** ***Ph al2*** **cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph al2* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The HD is shown in blue and the OAR in orange. The nucleotide sequences that encode these domains are shown in the respective colours. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

#### 3.2.2.8 Phylogeny of the *Parhyale hawaiensis aristaless* group genes *Ph hbn, Ph al1* and *Ph al2*

BLAST searches (Altschul et al., 1997) using the complete amino acid sequence of *Ph* HBN as query against the protein databases *of Drosophila melanogaster* (blastp, taxid 7227) and mouse (blastp, taxid 10090) revealed highest sequence similarity to *Drosophila* HOMEOBRAIN (*Dm* HBN) and murine Aristaless related homeobox (*Mm* ARX), respectively. For *Ph* AL1 and *Ph* AL2, analogous BLAST searches yielded *Drosophila* ARISTALESS (*Dm* AL) for both *Parhyale* proteins and again *Mm* ARX, also for both *Parhyale* proteins. Independently, BLAST searches were performed using the complete amino acid sequences of *Ph* HBN, *Ph* AL1 and *Ph* AL2 as queries against the complete translated nucleotide database (tblastn). From these searches, a putative protein sequence from the Acorn worm *Saccoglossus kowalevskii* (Hemichordata, Enteropneusta; PREDICTED: Saccoglossus kowalevskii predicted protein-like (LOC100366437), mRNA) showed highest sequence similarity to *Ph* HBN. A BLAST search using this putative *Saccoglossus kowalevskii* amino acid sequence as query against the protein database *of Drosophila melanogaster* (blastp, taxid 7227) revealed highest sequence similarity to *Dm* HBN. A putative protein sequence from the mosquito *Anopheles gambiae* (Insecta, Diptera; Anopheles gambiae str. PEST AGAP007985-PA (AgaP_AGAP007985) mRNA, partial cds) showed highest sequence similarity to *Ph* AL1. A BLAST search using this putative *Anopheles gambiae* amino acid sequence as query against the protein database *of Drosophila melanogaster* (blastp, taxid 7227) revealed highest sequence similarity to *Dm* AL. A putative protein sequence from the fruit fly *Drosophila mojavensis* (Drosophila mojavensis GI21624 (Dmoj\GI21624), mRNA) showed highest sequence similarity to *Ph* AL2. A BLAST search using this putative *Drosophila mojavensis* amino acid sequence as query against the protein database *of Drosophila melanogaster* (blastp, taxid 7227) revealed highest sequence similarity to *Dm* AL.

In *Drosophila*, HBN and AL have been described as representatives of a subgroup of paired-like homeodomain transcription factors. Additional members of this group are the gene products of *orthopedia* (OTP), *munster* (MU), *goosecoid* (GSC), *reverse polarity* (REPO) and *Retinal Homeobox* (RX). These proteins share several characteristic motifs and properties to a varying extent: At position 50 of the homeodomain, they possess a glutamine residue, with the exception of GSC (lysine). Further, an N-terminal transcriptional repression domain

(octapeptide/GEH) is present in HBN, RX and GSC. The genes *al*, *otp*, *mu* and *Rx* encode a common OAR/*aristaless* domain in the C-termini of the proteins. In addition, the majority of these genes reside in close chromosomal proximity in *Drosophila*, where they form two gene clusters (Walldorf et al., 2000). Representatives of this subgroup of *paired*-class homeodomain transcription factors have been found in other insect and arthropod species as well as in vertebrates, e.g. mouse (Mazza et al., 2010; Meijlink et al., 1999) . However, no distinct murine orthologs of *hbn* and *al* have been identified so far. In order to confirm the identities of *Ph* HBN, *Ph* AL1 and *Ph* AL2 as orthologs of *Drosophila* HBN and AL, respectively, as they have been suggested by BLAST searches and the presence of protein motifs that share high sequence similarities to HBN-specific octapeptides and OAR domains, the phylogenetic relation of the *Parhyale* proteins to HBN- and AL-like transcription factors of *Drosophila* and Mouse was calculated (Mr Bayes, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The phylogenetic calculation was based on an alignment of *Ph* HBN, *Ph* AL1 and *Ph* AL2 together with the other protein sequences of choice, using the CLUSTALW2 algorithm (Larkin et al., 2007). Specifically, described representatives of the *aristaless* subgroup of transcription factors from *Drosophila* and their putative murine orthologs were used as well as additional *Drosophila* and mouse proteins with high sequence similarity to *Ph* HBN, *Ph* AL1 and *Ph* AL2, as obtained from the BLAST searches described earlier (A6.2, table A6). From the later, all those BLAST hits were excluded that clearly belong to differing *paired*-class homeodomain factor subgroups, e. g. proteins that possess *paired*-box domains. The protein sequences included in the alignment vary strongly with regard to sequence length, the relative position of the HD and the presence of additional protein motifs. For this reason, only the part of the alignment that covers the HD was used for calculating the phylogeny (A6.3). Mus musculus *paired related homeobox protein-like 1* (*Prrxl1*) was set as outgroup since preliminary phylogenetic calculations suggested this factor as least related to *Ph* HBN, *Ph* AL1 and *Ph* AL2.

The resulting phylogram confirms the identity of *Ph* HBN as distinct ortholog of *Dm* HBN. Further, it suggests that *Ph* AL1 and *Ph* AL2 are co-orthologous to *Dm* AL since all three factors cluster distinctly in a phylogenetic subdivision of the tree (Figure 19). However, the low credibility value of the phylogenetic relation of *Ph* AL1 and *Ph* AL2 (51%) does not resolve the relation of all three proteins unambiguously. In addition, the phylogram confirms previous assumptions that it has yet been impossible to identify murine orthologs of

HBN and AL, while other *Drosophila aristaless*-like transcription factors such as OTP, REPO, RX and GSC have murine orthologs (Figure 19).

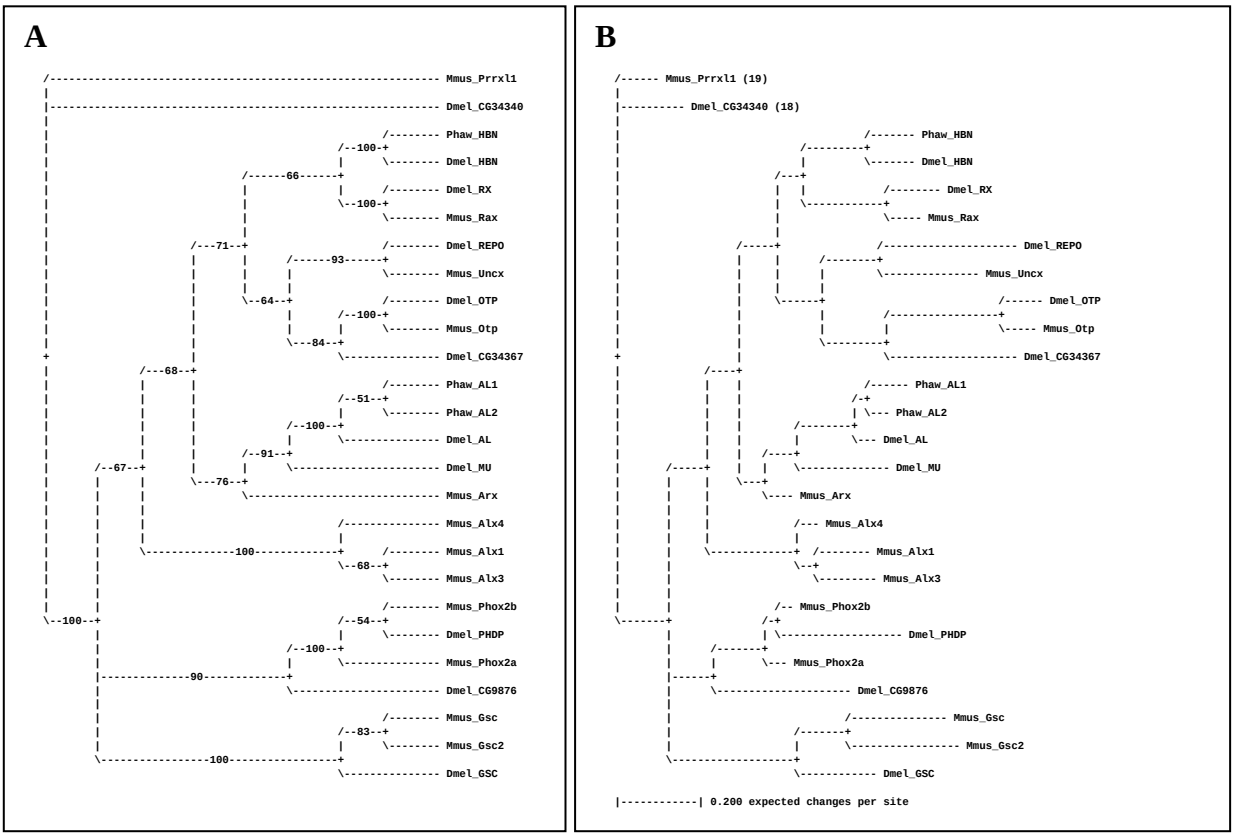



**Figure 19 (next page): Phylogenetic analysis of *Parhyale* HBN, AL1 and AL2 and *aristaless*-like *paired*-class homeodomain transcription factors from *Drosophila* and mouse.** Shown are: *Parhyale hawaiensis* AL1 (Phaw_AL1), *Parhyale hawaiensis* AL2 (Phaw_AL2), *Parhyale hawaiensis* HBN (Phaw_HBN), *Drosophila melanogaster* AL (Dmel_AL), *Drosophila melanogaster* GSC (Dmel_GSC), *Drosophila melanogaster* MU (Dmel_MU), *Drosophila melanogaster* HBN (Dmel_HBN), *Drosophila melanogaster* OTP (Dmel_OTP), *Drosophila melanogaster* Putative homeodomain protein (Dmel_PHDP), *Drosophila melanogaster* REPO (Dmel_REPO), *Drosophila melanogaster* RX (Dmel_RX), *Drosophila melanogaster* gene product of CG9876 (Dmel_CG9876), *Drosophila melanogaster* gene product of CG34340 (Dmel_CG34340), *Drosophila melanogaster* gene product of CG34367 (Dmel_CG34367), *Mus musculus* gene product of ALX homeobox 1 (Mmus_Alx1), *Mus musculus* gene product of aristaless-like homeobox 3 (Mmus_Alx3), *Mus musculus* gene product of aristaless-like homeobox 4 (Mmus_Alx4), *Mus musculus* gene product of aristaless related homeobox (Mmus_Arx), *Mus musculus* gene product of goosecoid homeobox (Mmus_Gsc), *Mus musculus* gene product of goosecoid homeobox 2 (Mmus_Gsc2), *Mus musculus* Orthopedia homolog (Mmus_Otp), *Mus musculus* gene product of paired-like homeobox 2a (Mmus_Phox2a), *Mus musculus* gene product of paired-like homeobox 2b (Mmus_Phox2b), *Mus musculus* gene product of paired related homeobox protein-like 1 (Mmus_Prrxl1), *Mus musculus* gene product of retina and anterior neural fold homeobox (Mmus_Rax) and *Mus musculus* gene product of UNC homeobox (Mmus_Uncx). Mmus_Prrxl1 was set as outgroup. For reference regarding the non-*Parhyale* protein sequences, see A6, tables A5 and A6. **A** Phylogram showing Clade credibility values. **B** Phylogram based on average branch lengths. Phylogenetic analysis was done using Mr Bayes (Huelsenbeck and Ronquist, 2001.3.2; 1,000,000 generations; Ronquist and Huelsenbeck, 2003).

#### 3.2.2.9 Expression of *Ph hbn*

Expression of *Ph hbn* is first detectable in stage 8 embryos. During this stage, the germ disc forms as a result of gastrulation (Browne et al., 2005). *Ph hbn* is expressed symmetrically in groups of cells that roughly form hemispheric arches at the anterior periphery of the germ disc (Figure 20 A1, A2). Interestingly, the area of the embryo in which the *Ph hbn*-positive cells are located exhibits a lower cell density than the centre of the germ disc (Figure 20 A2). No expression of *Ph hbn* was found in embryos of earlier stages.

At stage 9, the head lobes appear as morphologically distinct anterior structures (Figure 21 A3, B3, C3; Browne et al., 2005). *Ph hbn* is expressed bilaterally in groups of cells that constitute major anterior parts of the head lobes (Figure 21 A1-A3, B1-B3, C1-C3, D1-D3).

In the stage 10 embryo, the dorsal organ has formed and is visible as a prominent ring of thickened cells along the anterodorsal midline surface of the developing egg (Browne et al., 2005). *Ph hbn* is expressed bilaterally in clusters of 6-7 cells located in an anterior, slightly medial area of the head lobes (Figure 22 A1-A3, B1-B3).

Stage 12 is characterised by the initial phase of germ band elongation. Both head lobes have increased in size and are visibly separated by a medial cleft (Figure 23 A2, B2, C2, D2; 3.1.1; Browne et al., 2005). *Ph hbn* is expressed bilaterally in arch-like domains that cover single rows of cells in the pre-antennal part of the head (Figure 23 A1, A2). At the end of stage 12, these arch-like domains have broadened and cover additional cells to the anterior (Figure 23 C1, C2).

Stage 13 embryos undergo trunk segmentation and germ band elongation (Figure 24 D1-D3, E1-E3; Browne et al., 2005). The embryonic head is composed of two head lobes that clearly cover the pre-antennal part as well as the An1 and An2 segments. The head lobes are still separated by a medial cleft that is characterised by a lower density of cells. The head lobes have reached their largest expansion on the egg's surface (Figure 24, compare A3, B3 and C3; 3.1.2; Browne et al., 2005). *Ph hbn* is expressed bilaterally in arch-like domains that cover several rows of cells in the pre-antennal part of the head (Figure 24 A1-A3, B1-B3). The medial region between the morphologically distinct head lobes is free of *Ph hbn* expression (Figure 24 A1-A3, C1-C3).

In stage 20 embryos, the labrum is composed of two lobes that begin to grow over the distinct stomodeal opening (Figure 25, compare A3, B3 and D3; 3.1.3, Browne et al., 2005). *Ph hbn* expression covers the majority of the pre-antennal head (Figure 25 A1-A3). In each

hemisphere, a prominent anterior *Ph hbn* expression domain is bordered by several lateral clusters of *Ph hbn*-positive cells (Figure 25 B1-B3). The medial region of the head including the stomodeal opening does not show distinct *Ph hbn* expression (Figure 25 A1-A3, B1-B3, D1-D3).

In stage 22 embryos, the stomodeum has internalised into the developing embryonic head at the level of the An2 segment. It is covered by the labrum that has completed its elongation and directly abuts the paragnaths (Figure 26, compare C3 and D3; 3.1.4; Browne et al., 2005). Expression of *Ph hbn* remains strong in the pre-antennal part of the embryonic head (Figure 26 A1-A3). Similar to stage 19 embryos, the *Ph hbn* expression profile of one hemisphere is composed of a prominent apical domain that sharply borders the medial region of the embryonic head. *Ph hbn* expressing cells extend laterally at the anterior most, peripheral lining of the embryonic head, forming an arch-shaped expression domain (Figure 26 A1-A3, B1-B3, D1-D3, F1-F3). Additional groups of cells expressing *Ph hbn* are found throughout the pre-antennal area of the head (Figure 26 A1-A3, B1-B3, F1-F3). Small, dot-like expression domains are found on the level of the An2 appendage bases (Figure 26 A2, B2, E2). During Parhyale embryogenesis, *Ph hbn* is expressed exclusively in the developing head (Figure 25 C1-C3, D1-D3, E1-E3; Figure 26 C1-C3, D1-D3, F1-F3). No *Ph hbn* expression was found within the medial region of the head, encompassing the stomodeal field and the labrum (Figure 25 A1-A3, D1-D3; Figure 26 A1-A3, D1-D3, E1-E3).

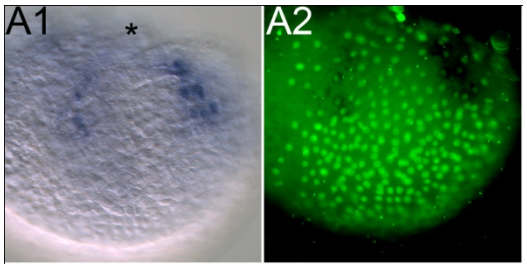



**Figure 20: Expression of *Ph hbn* at stage S8**. The embryo is mounted in ventral orientation. In the images, anterior is up. Shown is a DIC image to the left (A1) and nuclear labelling of the same embryo (SYTOX®) to the right (A2; nuclei are green, expression of *Ph hbn* is detectable as it quenches the nuclear labelling). In the DIC image (A1), the anterior end of the embryo is indicated by a black asterisk.

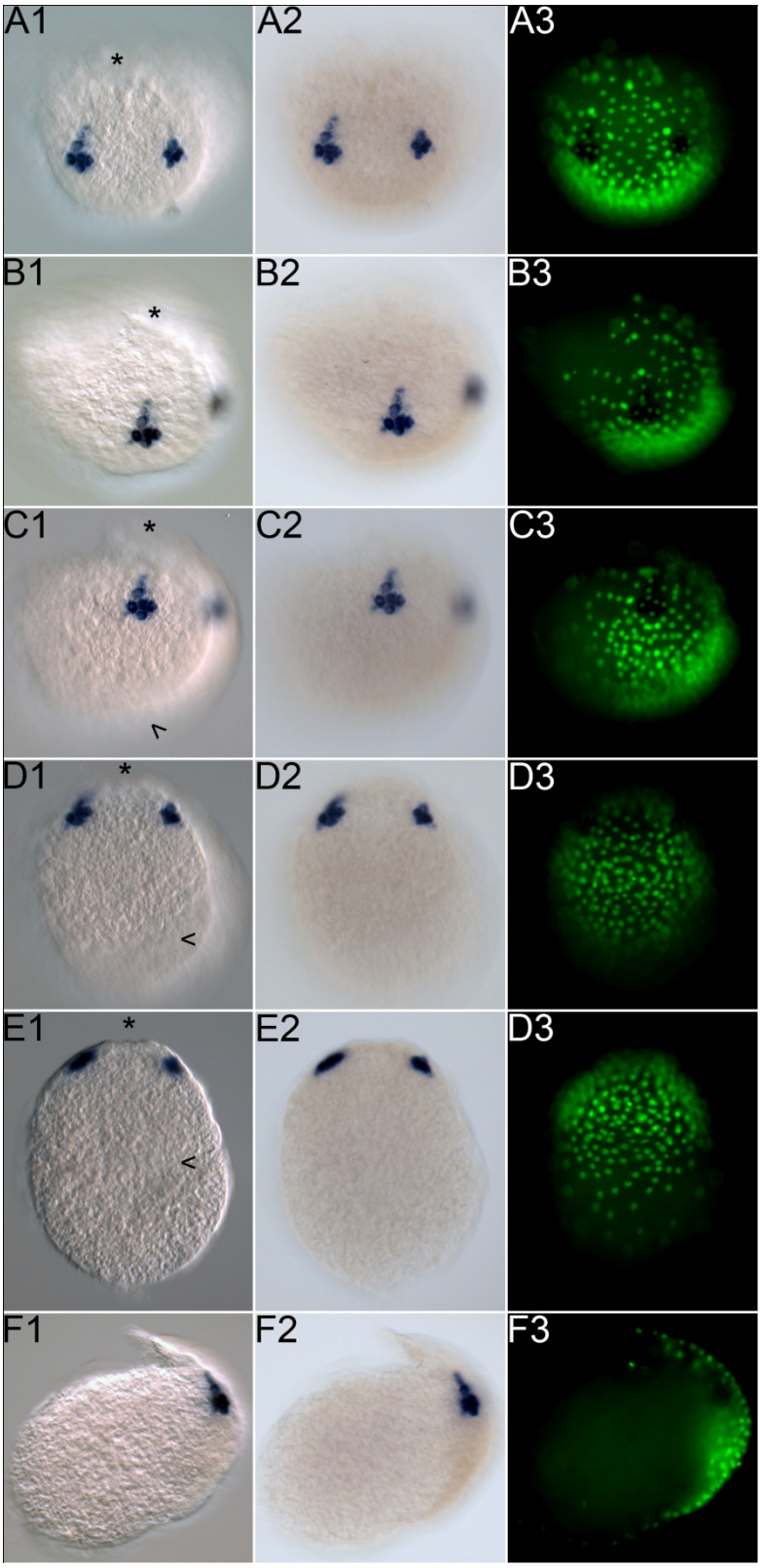



**Figure 21: Expression of *Ph hbn* at stage S9**. The embryo is mounted in anterior (A1-A3), tilted anterior (B1-B3), tilted ventral (C1-C3), ventral (D1-D3), posterior (E1-E3) and lateral, sagittal focal plane (F1-F3, ventral is to the right) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph hbn* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-F1), bright field images in the middle panel (A2-F2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–F3; nuclei are green, expression of *Ph hbn* is detectable as it quenches the nuclear labelling). In all DIC images (A1-F1), the anterior end of the embryo is indicated by a black asterisk. Where appropriate, the position of the future Mn segment is marked by a black arrowhead.

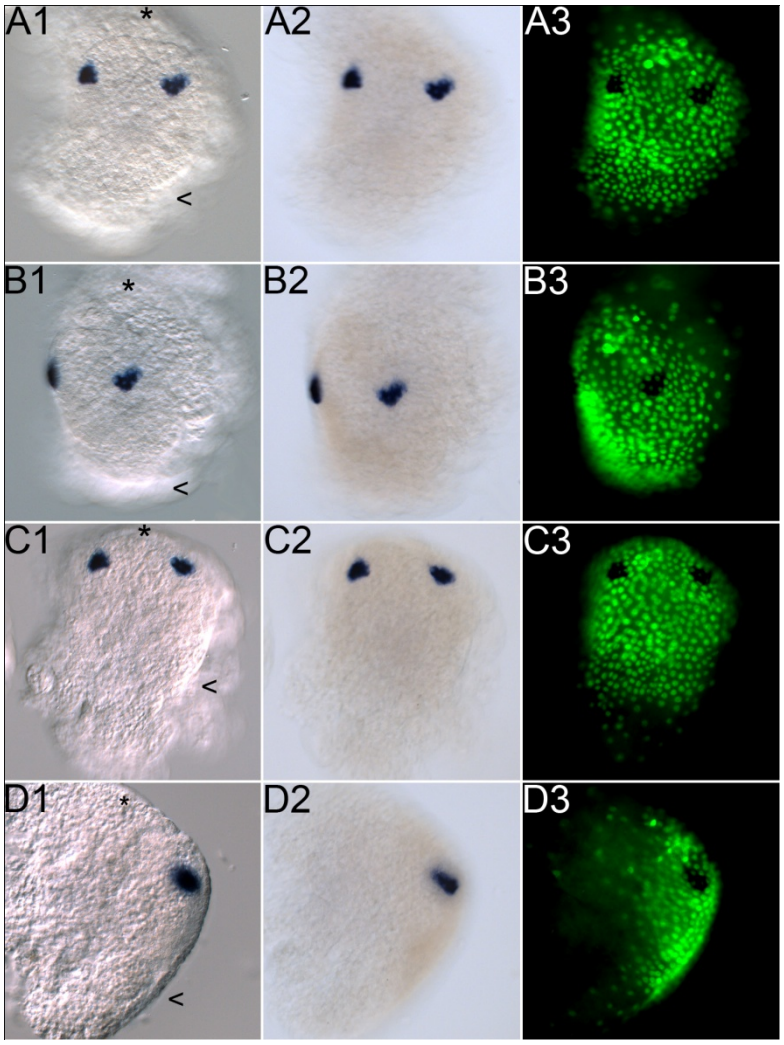



**Figure 22: Expression of *Ph hbn* at stage S10**. The embryo is mounted in frontal (A1-A3), tilted anterior (B1-B3), ventral (C1-C3) and lateral (D1-D3, ventral is to the right) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph hbn* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2-D2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – D3; nuclei are green, expression of *Ph hbn* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the position of the future Mn segment is marked by a black arrowhead. *Ph hbn* is expressed bilaterally in clusters of 6-7 cells located in a slightly medial, apical area of the head lobes.

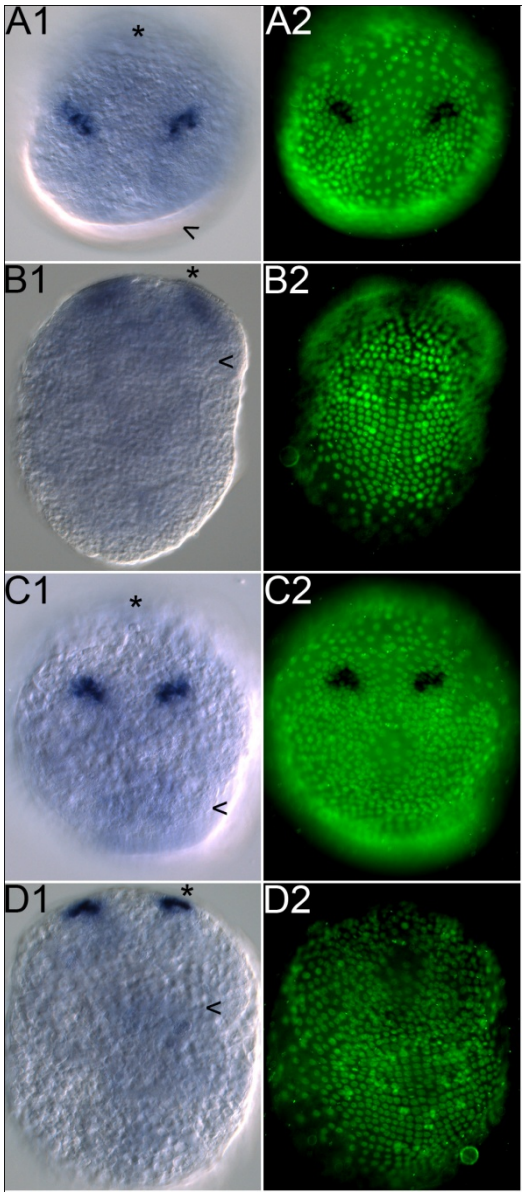



**Figure 23: Expression of *Ph hbn* at stage S12**. Shown are two embryos, one at the onset of stage 12 (A1, A2, B1, B2) and one during late stage 12 (C1, C2, D1, D2). The embryos are mounted in anterior (A1, A2, C1, C2) and ventral (B1, B2, D1, D2) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph hbn* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-D1) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A2-D2; nuclei are green, expression of *Ph hbn* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the position of the future Mn segment is marked by a black arrowhead.

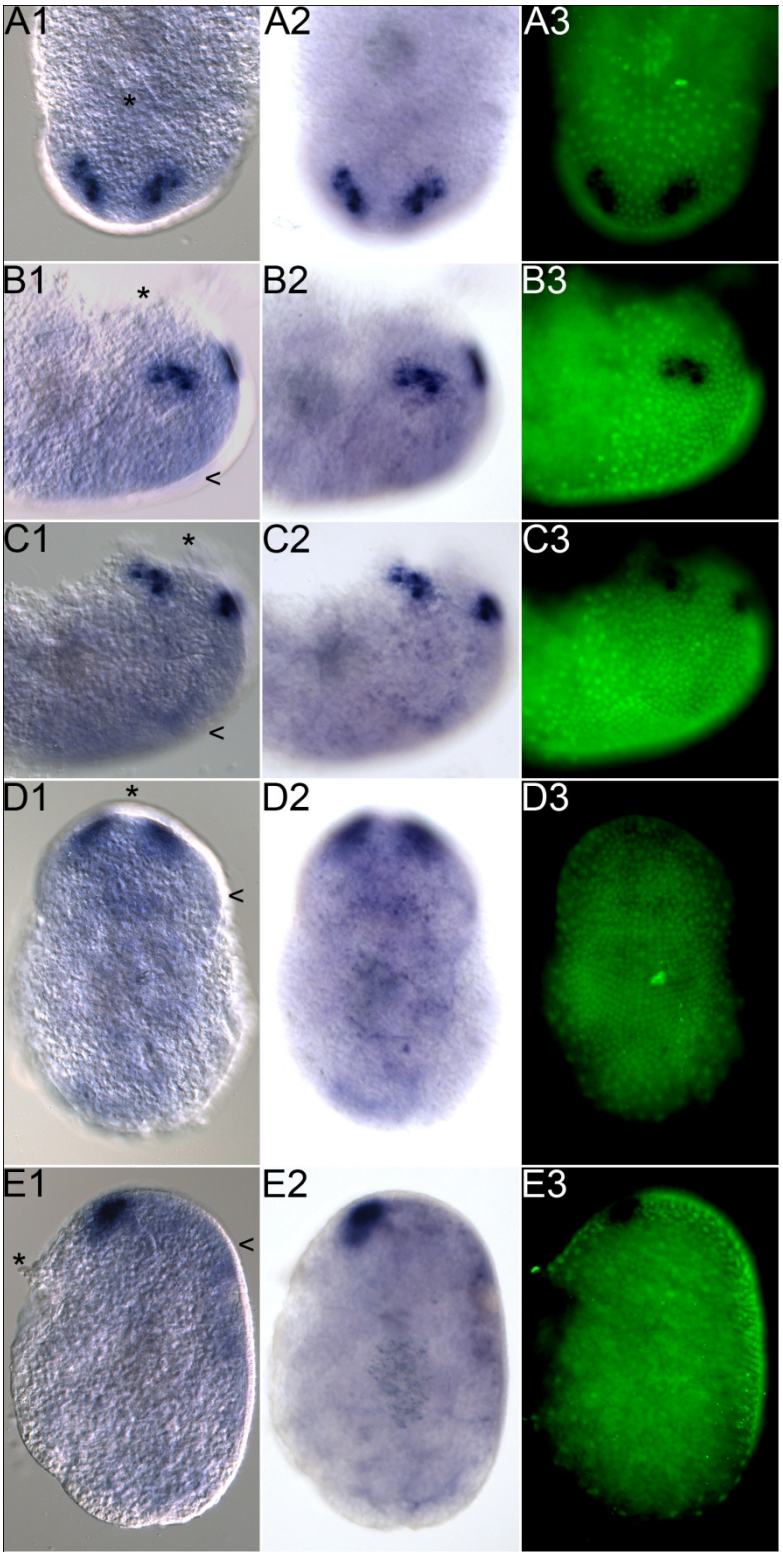



**Figure 24: Expression of *Ph hbn* at stage S13**. The embryo is mounted in frontal (A1-A3), tilted anterior (B1-B3), tilted ventral (C1-C3), ventral (D1-D3) and lateral (E1-E3, ventral is to the right) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph hbn* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – E3; nuclei are green, expression of *Ph hbn* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A**, **B** *Ph hbn* is expressed bilaterally in arch-like domains that cover several rows of cells in the pre-antennal part of the head. **C** The medial region between the head lobes is free of *Ph hbn* expression.

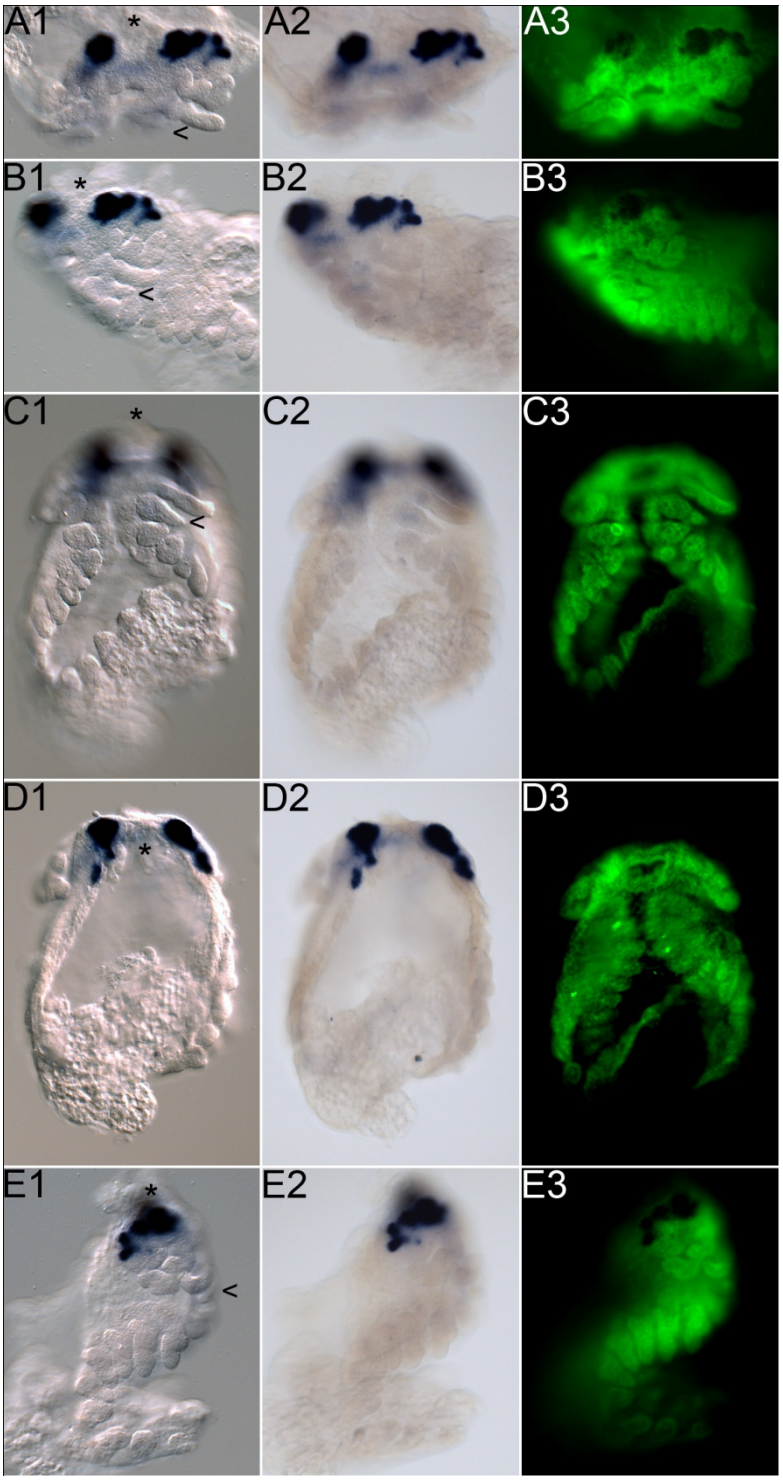



**Figure 25**: **Expression of *Ph hbn* at stage S20**. The embryo is mounted in frontal (A1-A3), tilted anterior (B1-B3), ventral (C1-C3), dorsal (D1-D3) and lateral (E1-E3, ventral is to the right) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph hbn* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – E3; nuclei are green, expression of *Ph hbn* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A**, **B** *Ph hbn* expression covers the majority of the pre-antennal head. In each hemisphere, a prominent anterior *Ph hbn* expression domain is bordered by several lateral clusters of *Ph hbn*-positive cells. **C** The embryo is centred up the Mn, Mx1 and Mx2 segments. **D** No distinct *Ph hbn* expression is detectable within the medial region of the head.

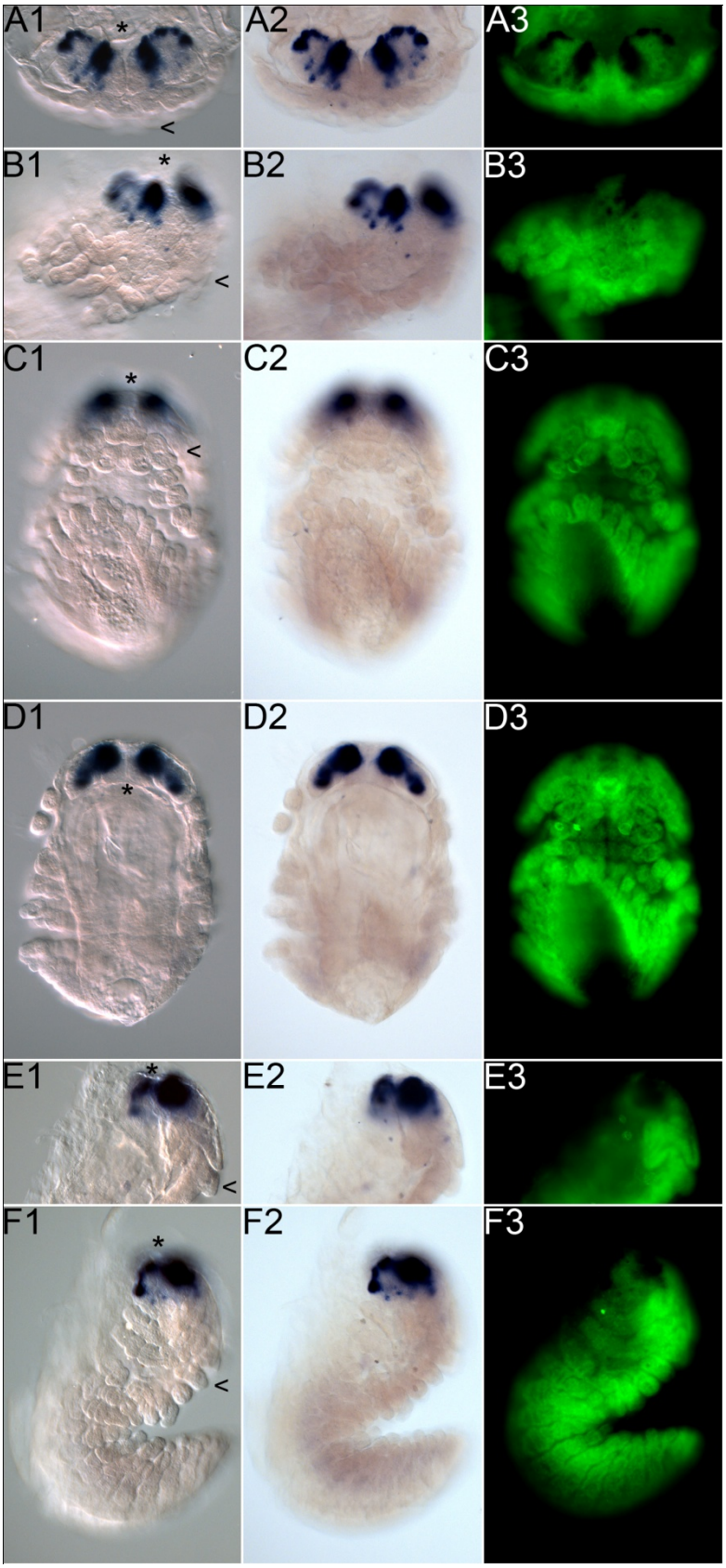



**Figure 26: Expression of *Ph hbn* at stage S22**. The embryo is mounted in frontal (A1-A3), tilted anterior (B1-B3), ventral (C1-C3), dorsal (D1-D3) and lateral (E1-E3, ventral is to the right) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph hbn* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – E3; nuclei are green, expression of *Ph hbn* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph hbn* expression covers the majority of the pre-antennal head. **B** In each hemisphere, a prominent apical *Ph hbn* expression domain is bordered by several lateral clusters of *Ph hbn*-positive cells. **C** The embryo is centred up the Mn, Mx1 and Mx2 segments. **D** No distinct *Ph hbn* expression is detectable within the medial region of the head.

#### 3.2.2.10 Expression of *Ph al1* and *Ph al2*

Expression of *Ph al1* is first detectable in stage 17 embryos. (Figure 27 A1-A3, B1-B3; 3.1.3; Browne et al., 2005). *Ph al1* is expressed in the An1, An2 and Mn appendages (Figure 27 B1-B3) as well as in the appendage fields of the segments Mx1, Mx2 (Figure 27 C1-C3) and T1-T5 (Figure 27 D1-D3). The expression pattern of *Ph al1* within the An1 and An2 appendages shows a complex topology, with small clusters of cells that strongly express *Ph al1* nested within the appendage that shows relatively low *Ph al1* expression as a whole (Figure 27 A1-A3; B1-B3). At the base of each Mn appendage bud, two groups of cells - one located anterior, one posterior - express *Ph al1* (Figure 27 B1-B3). In the remaining appendage fields that show *Ph al1* expression, it is found in all cells (Figure 27 C1-C3). Several cells of the lateral midgut anlagen appear positive for *Ph al1* (Figure 27 B1-B3, D1-D3), but further confirmation is required. No expression of *Ph al1* was found in embryos of earlier stages.

In stage 19 embryos, the stomodeum has begun to invaginate. The two stomodeal projections are clearly visible on lateral positions within the deepening stomodeal opening (Figure 28 C3; 3.1.3, Browne et al., 2005). Expression of *Ph al1* is still found exclusively within the appendages and appendage fields. The expression pattern of *Ph al1* within the An1 and An2 appendages is similar. It consists of two parallel rings of strong *Ph al1* expression in the proximal half of the appendage (Figure 28 A1-A3, B1-B3, E3). In the Mn, Mx1 and Mx2 appendages, *Ph al1* is expressed in two proximal-distal bands, one of them at the anterior of the appendage, and the other located posterior (Figure 28 B1-B3, C1-C3, D1-D3). The appendage buds of T1 to T4 show matching expression domains of *Ph al1*, encompassing two proximal and one distal group of cells (Figure 28 C1-C3, D1-D3). Strong *Ph al1* expression is found in the proximal halves of the T5 and T6 appendage buds (Figure 28 D1-D3). In the T7, T8, A1 and A2 segments, *Ph al1* expression covers the entire appendage field (Figure 28 D1-D3). Since at this stage of embryogenesis, the development of the posterior appendages lags behind that of more anterior ones, the *Ph al1* expression profile of the T1 to A2 appendages might reflect a temporal expression sequence of *Ph al1* as well.

In stage 22 embryos, the stomodeum has internalised into the developing embryonic head at the level of the An2 segment. It is covered by the labrum that has completed its elongation and directly abuts the paragnaths (Figure 29 compare A3 and C3; 3.1.4, Browne et al., 2005). *Ph al1* is expressed in all developing segmental appendages. In individual

appendages the expression profile of *Ph al1* is very complex, but morphologically similar appendages appear to have similar *Ph al1* expression. In the An1 and An2 appendages, *Ph al1* expression is found predominantly in the proximal articles (Figure 29 A1-A3, B1-B3, C1-C3). In the mandibles, paragnaths, Mx1 and Mx2 appendages, *Ph al1* is strongly expressed in posterior regions (Figure 29 A1-A3, B1-B3, C1-C3). T1 to T3 appendages show manifold expression domains within all articles (Figure 29 B1-B3, C1-C3, D1-D3). Within the segments T4 to T8, *Ph al1* expression is visible mainly in the proximal parts of the appendages and in adjacent cells of the segments themselves (Figure 29 D1-D3, E1-E3). In contrast, the pleopod and uropod primordia of segments A1 to A6, express *Ph al1* strongly within the distal appendage articles (Figure 29 C1-C3, D1-D3). On the dorsal side of the labrum, dot-like *Ph al1* expression can be found. Similar dot-like expression domains are visible bilaterally on the dorsal parts of the thoracic and abdominal segments (Figure 29 A2 – labral *Ph al1* expression, F1, F2). During *Parhyale* embryogenesis, no *Ph al1* expression was found within the head apart from the appendages.

For *Ph al2,* no specific *in situ* pattern could be recovered via WMISH (5.4.3) in embryos of stages 8 to 22 (not shown).

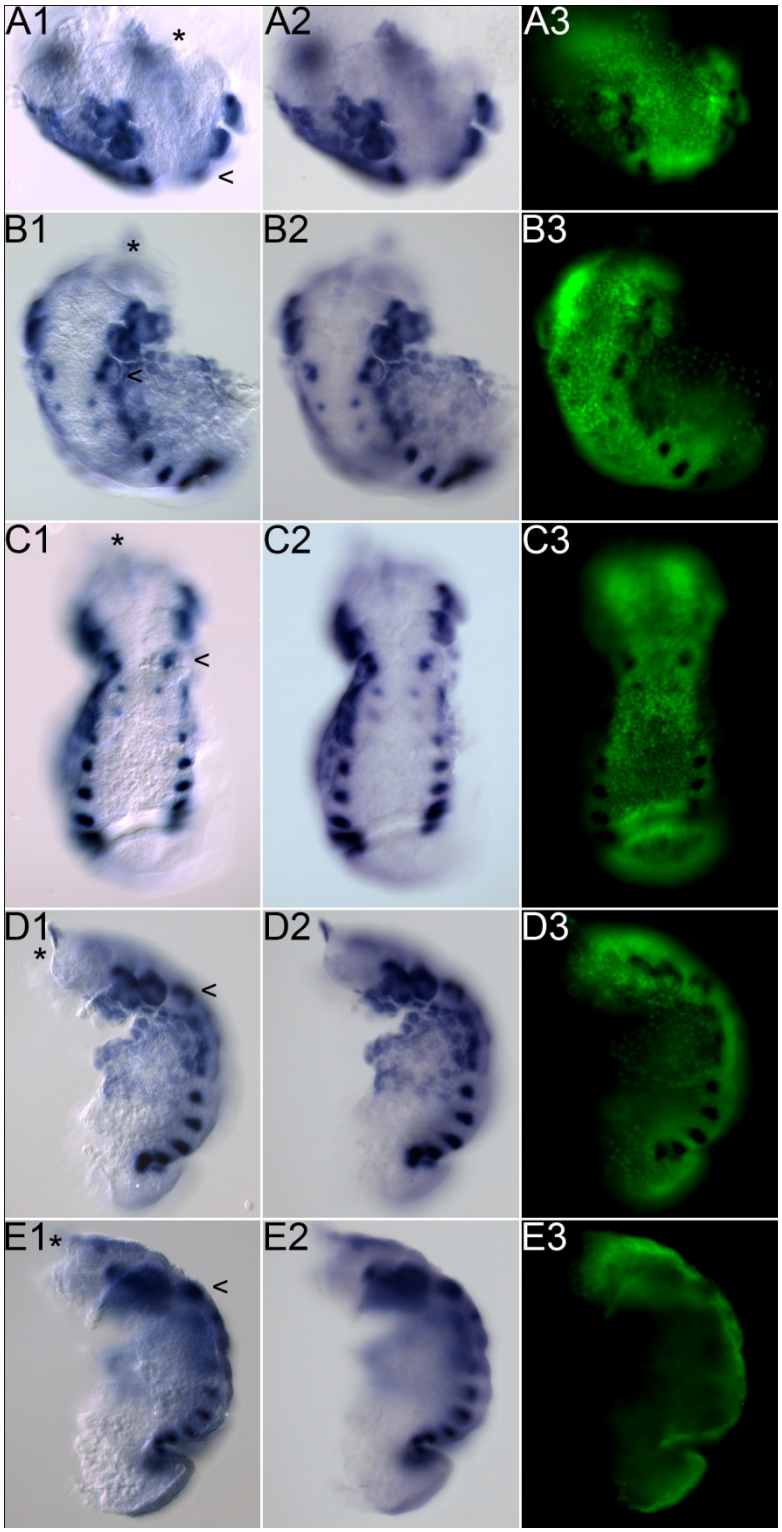



**Figure 27: Expression of *Ph al1* at stage S17**. The embryo is mounted in frontal (A1-A3), tilted anterior (B1-B3), ventral (C1-C3) and lateral (D1-D3 focal plane on An2 appendage; E1-E3 sagittal focal plane; ventral is to the right) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph al1* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – E3; nuclei are green, expression of *Ph al1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** The An1 and An2 appendages show a complex topology of *Ph al1* expression. **B** *Ph al1* expression is found in two groups of cells at the base of each Mn appendage bud. **C** The embryo is centred up the Mn, Mx1 and Mx2 segments. *Ph al1* expression is found bilaterally in the appendage fields of Mx1, Mx2 and T1-T5.

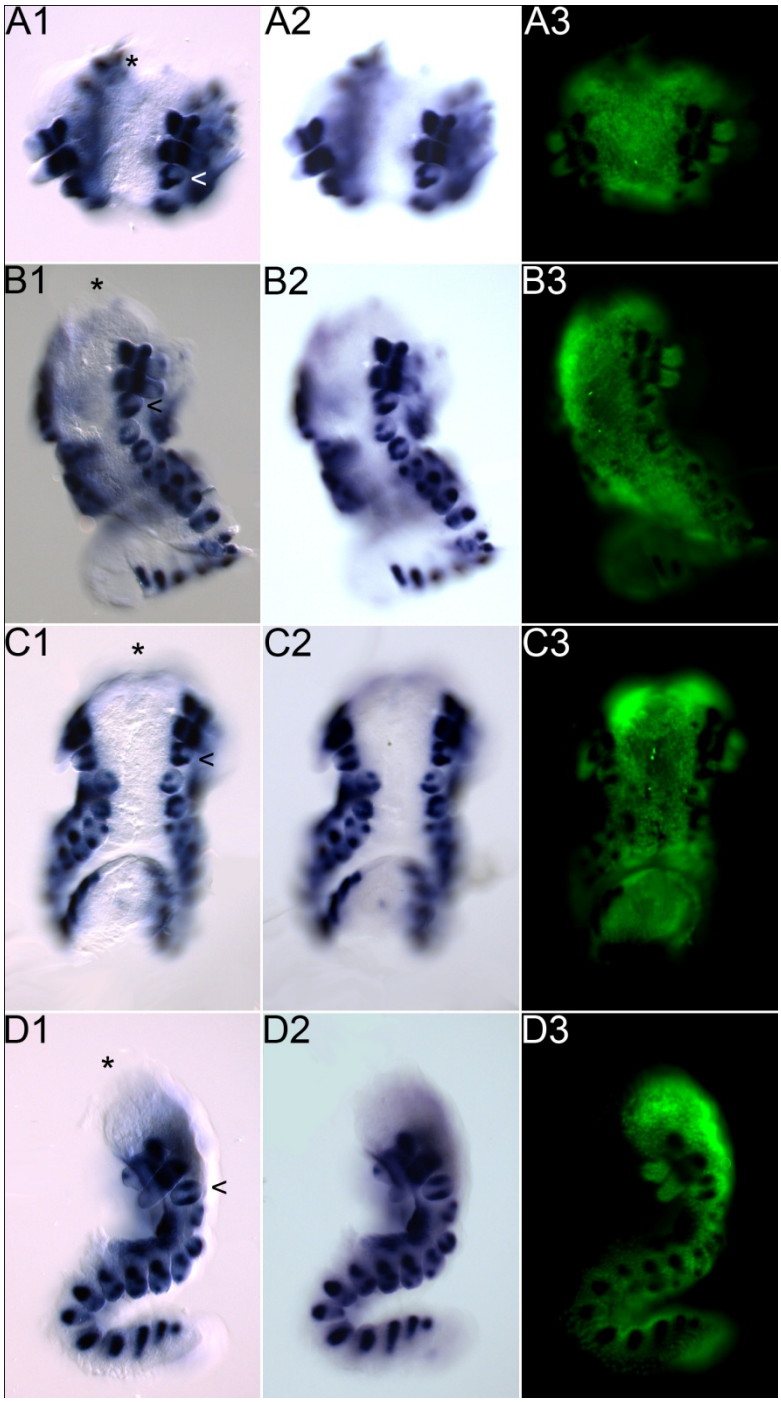



**Figure 28: Expression of *Ph al1* at stage S19**. The embryo is mounted in frontal (A1-A3), in tilted anterior (B1-B3), in ventral, native coiled posture (C1-C3) and lateral (D1-D3, ventral is to the right) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph al1* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2-D2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – D3; nuclei are green, expression of *Ph al1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph al1* is strongly expressed in two parallel rings in the proximal half of the An1 and An2 appendages. **B** In the Mn, Mx1 and Mx2 appendages, *Ph al1* is expressed in two proximal-distal bands, one of them at the anterior of the appendage, and the other located posterior. **C** The embryo is mounted in its native, coiled posture. The Mx1 segment is in the centre. In the Mn, Mx1 and Mx2 appendages, *Ph al1* is expressed in two proximal-distal bands, one of them at the anterior of the appendage, and the other located posterior. **D** Ventral is to the right. Within the T1 to T4 appendage buds, *Ph al1* is expressed in two proximal and one distal group of cells. Strong *Ph al1* expression is found in the proximal halves of the T5 and T6 appendage buds. In the T7, T8, A1 and A2 segments, *Ph al1* expression covers the entire appendage field.

**Figure 29:** **Expression of *Ph al1* at stage S22**.

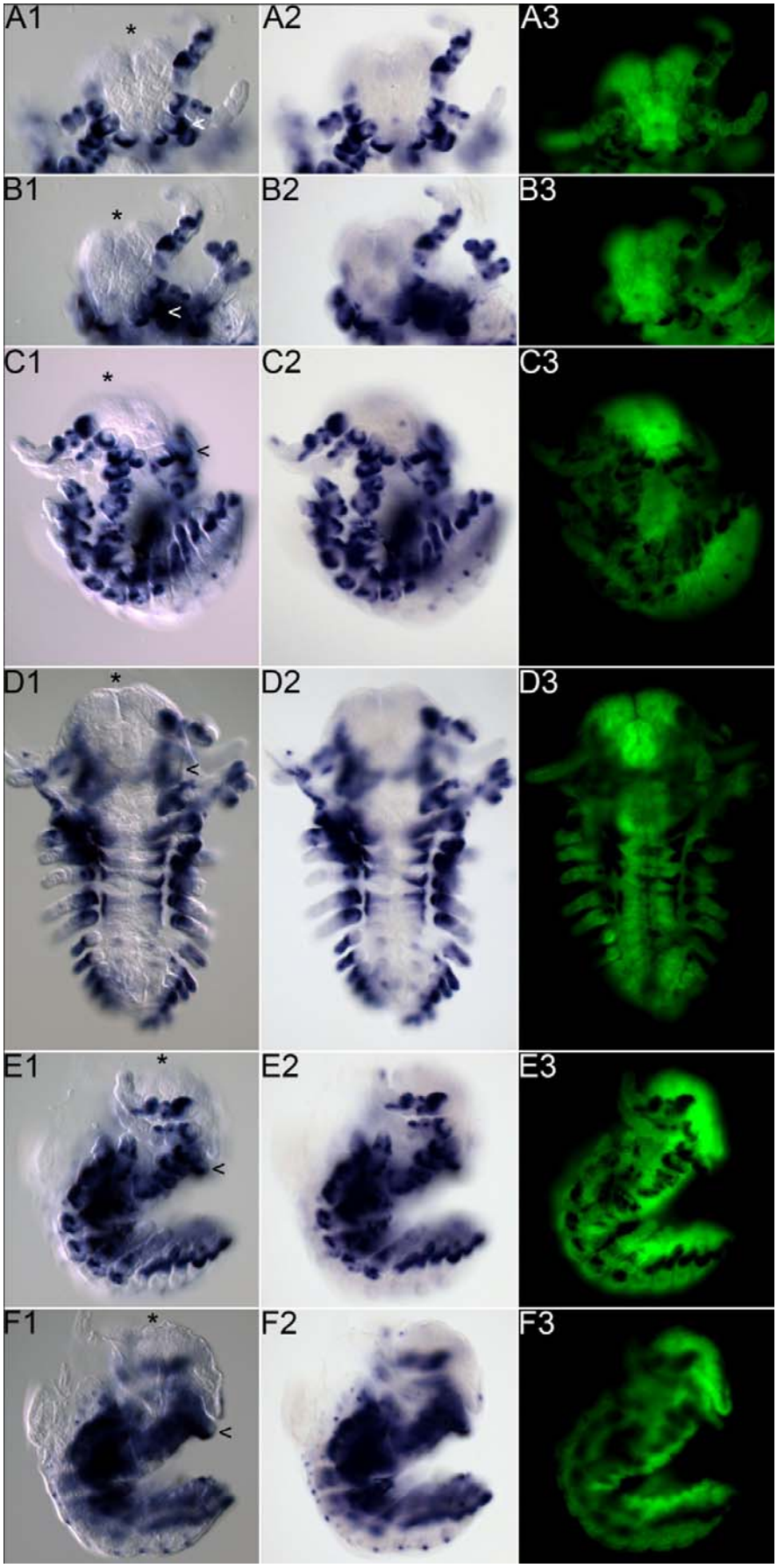


**Figure 29 (previous page): Expression of *Ph al1* at stage S22**. The embryo is mounted in frontal (A1-A3), in tilted anterior (B1-B3), in semi-lateral, native coiled posture (C1-C3), in ventral orientation, dissected flat, focal plane on T4-T8 (D1-D3), in lateral orientation, focus on the segmental appendages (E1-E3, ventral is to the right) and lateral orientation, sagittal focal plane (F1-F3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph al1* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-F1), bright field images in the middle panel (A2-F2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – F3; nuclei are green, expression of *Ph al1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-F1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** In the An1 and An2 appendages, *Ph al1* is expressed predominantly in the proximal articles, in a complex topology. During mounting, the left An1 appendage has been torn off the embryo. **B** To the right, the distal parts of the T1 to T3 appendages can be seen. They show manifold expression domains within all articles. **C** In the mandibles, paragnaths, Mx1 and Mx2 appendages, *Ph al1* is strongly expressed in posterior regions. **D** The embryo is dissected flat, focal plane is on T4-T8. Within these segments, *Ph al1* is mainly expressed in the proximal parts of the appendages and in adjacent cells of the segments themselves. The pleopod and uropod primordia of segments A1 to A6 express *Ph al1* strongly within the distal appendage articles. **F** Dot-like *Ph al1* expression is found on the dorsal side of the labrum, as well as bilaterally on the dorsal parts of the thoracic and abdominal segments.

### 3.2.3 *Parhyale hawaiensis pax*-box genes

Members of the well described and highly conserved *paired*-box homeobox family of genes encode transcription factors that play important roles in early development, segmentation, head and brain patterning and the establishment of sense organs (Kozmik, 2008; Noll, 1993). Two different representatives of these paired-box homeobox genes were isolated from the performed low-stringency degenerate PCR screens targeted at *paired*-type homeobox sequences (5.2.3.1).

#### 3.2.3.1 Isolation and characterisation of *Parhyale hawaiensis pax-box* genes

Because sufficient genomic sequence data is not currently available for *Parhyale hawaiensis*, the isolation of gene-of-interest sequences was approached by performing varied-stringency degenerate PCR screens (5.2.3.1). From repeated low-stringency degenerate PCR screens targeted at paired-type homeobox sequences, homeobox fragments of two different putative *Parhyale hawaiensis pax3/7/paired* paralogs, *Parhyale hawaiensis pairberry1* and *pairberry2*, were recovered. The isolated fragments contained 109 bp of specific *Ph pby1* and *Ph pby2* sequence and were sufficient for unambiguous homology identification (5.3.3.1, A2.6).

In detail, six sequences were recovered, representing two novel *paired*-type homeobox genes. Five of them (*Ph pby1*) show the highest similarity to murine *pax7* and *Drosophila paired*, and the remaining single sequence (*Ph pby2*) shows highest similarity to murine *pax3* and again, to *Drosophila paired* (BLAST evaluation via blastx; taxid: 7227 and taxid: 10090; 5.3.3.1; A6.1, table A5; A6.2, table A6). The fact that both groups of sequences differ from each other in more than 30% of nucleotides comprehensibly supports that they derive from different genes (5.3.3.3). Within the group of *Ph pby1*, any two individual sequences may vary considerably (up to 5% of nucleotide positions). However, two polymorphic sites of nucleotide exchange were identified within the sequence they cover (3.2.3.3). For this reason, their nucleotide variations are considered minor, supporting the assumption that they derive from the same gene. Since each sequence derives from a different priming event, all are considered independent.

Based on this primary sequence information, specific oligonucleotides were designed and used to prime *Ph pby1* and *Ph pby2* RACE reactions (5.2.3.3, A4.2, table A3). Consequently, the remaining mRNA transcript sequences both 5' and 3' of the initial fragments covering the homeobox were obtained.

*Ph pby2* has been characterised in detail by another research group. It appears to be involved in *Parhyale* segmentation (Ron Parchem, personal communication). In this work, *Ph pby2* has not been examined further.

#### 3.2.3.2 Isolation of *Ph pby1*

From repeated two-step 5' RACE reactions, three independent *Ph pby1* 5' cDNA fragment clones were obtained. Within the sequence they cover, several sites of polymorphic nucleotide exchange were identified (3.2.3.3). If these are not taken into account, the nucleotide variation within the sequence of any two of the *Ph pby1* 5' cDNA fragment clones does not exceed 1%. All three of them show the same transcription start. These observations suggest that they derive from transcripts of different alleles of the same gene.

*Ph pby1* 3' cDNA sequence was recovered by two-step 3' RACE. Six independent *Ph pby1* 3' cDNA fragment clones were obtained. They fall into two groups that represent distinct *Ph pby1* isoforms. All six *Ph pby1* 3' RACE clones share a fraction of 333 bp and accordingly, exhibit different 3' ends. Importantly, both isoforms translate into proteins with identical conserved domains that show differences only in the C-termini.

Three *Ph pby1* 3' cDNA fragment clones represent the longer isoform A (*Ph pby1*A). Within the sequence they cover, several polymorphic sites of nucleotide exchange were identified (see below), resulting in an overall considerable nucleotide variation between any two of the sequences. If these polymorphic sites are not taken into account, any two individual *Ph pby1* 3' RACE/ isoform A sequences differ in less than 2%, suggesting that they all derive from the same *Ph pby1* transcript isoform. Additionally, they cover the same fraction of sequence und show only minor variations in poly(A) tailing. The three remaining *Ph pby1* 3' cDNA fragment clones represent the shorter isoform B (*Ph pby1B*). They cover the same fraction of sequence and show only minor (less than 1%) nucleotide variation. Similar to the sequences representing *Ph pby1* isoform A, they show minor differences in poly(A) tailing. Again, this suggests that they derive from the same transcript isoform.

Sequence information provided by all recovered 5' and 3' *Ph pby1* RACE clones was used to assemble the complete *Ph pby1* cDNA sequence of both isoform A and B in silico. In order to verify the consistency of the established *Ph pby1* cDNA sequences, coherent *Ph pby1A* and *Ph pby1B* ORF fragments were isolated via long-distance PCR reactions (5.2.3.4), performed on *Parhyale* cDNA collections (5.2.2). Three independent *Ph pby1A* and three independent *Ph pby1B* ORF sequences were obtained (A2.6.1). They are consistent with the respective findings from 5' and 3' RACE (A2.6.2). If polymorphic sites of nucleotide exchange are not taken into account, any two individual ORF sequences of one isoform vary in less than 1%, strongly supporting that they derive from the same transcript isoforms (see also A3, table A1).

#### 3.2.3.3 Characterisation of *Ph pby1*

The comparison of all recovered *Ph pby1* sequences led to the identification of 103 sites of nucleotide exchange within both *Ph pby1* cDNA isoforms (see also A3, table A1). 83 of those occur uniquely and are therefore considered sporadic (5.3.3.2). Within the fraction of sequence shared by both *Ph pby1A* and *Ph pby1B*, 40 sporadic nucleotide exchanges are found: seven in the 5' UTR and 33 in the ORF, where they result in a majority of 18 amino acid alterations (missense exchanges). The remaining 15 do not change the amino acid sequence (silent exchanges). Within the fraction of sequence specific for *Ph pby1A*, 22 sporadic nucleotide exchanges are found within the 3' UTR and twelve are found within the ORF, where they account for ten amino acid exchanges. Within the fraction of sequence specific for *Ph pby1B*, nine sporadic nucleotide exchanges are found, all of them within the 3' UTR. Taken together, these findings go well with the expected random occurrence of artificial nucleotide exchanges. 20 nucleotide exchanges are present in more than one sequence and therefore considered polymorphic. Eleven are found within the shared fraction of the ORF, where they lead to one amino acid exchange outside of conserved protein domains. The remaining ten polymorphic nucleotide exchanges do not alter the amino acid sequence. Within the fraction of sequence specific for *Ph pby1A*, nine additional polymorphic nucleotide exchanges were identified, four of which cause amino acid sequence alterations. Again, these occur outside conserved protein domains within the C-terminal part of the PBY1A protein.

Within the fraction of sequence specific for *Ph pby1B*, no additional polymorphic nucleotide exchanges were identified.

These findings support the fact that at least two specific *Ph pby1* isoforms exist, probably the result of alternative splicing events. The differences between both isoforms lie within the C-terminal parts of the respective proteins and the 3' UTRs. Because no differences between the expression patterns of both isoforms could be identified (3.2.3.5), the question, whether the isoforms correlate to specific, alternative functions of the PBY1 protein during *Parhyale* development remains to be answered.

Based on these considerations, the sequence of the clone Ph_pby1_ORF16 was chosen as the source of the *Ph pby1A* reference sequence (*Ph pby1A*_ref). It was extended 5' by adding 129 bp derived from clone Ph_pby1_5R0101 and 3' by adding 421 bp derived from clone Ph_pby1_3R2219 (A2.6.1). One sporadic and putatively artificial nucleotide exchange was substituted with the 'consensus' nucleotide that is present in all other obtained sequences (1597C>T). An Alignment of *Ph pby1A*_ref with all recovered *Ph pby1* sequences that are relevant for isoform A shows that *Ph pby1A*_ref represents the statistically most common transcript of isoform A (A2.6.2).

The sequence of the clone Ph_pby1_ORF01 was chosen as the source of the *Ph pby1B* reference sequence (*Ph pby1B*_ref). It was extended 5' by adding 129 bp derived from clone Ph_pby1_5R0101 and 3' by adding 530 bp derived from clone Ph_pby1_3Rf316 (A2.6.1). Sporadic and putatively artificial nucleotide exchanges were substituted with the 'consensus' nucleotides that were present in all other obtained sequences (433C>T, 1273T>C). An Alignment of *Ph pby1B*_ref with all recovered *Ph pby1* sequences that are relevant for isoform B shows that *Ph pby1B*_ref represents the statistically most common transcript of isoform B (A2.6.2). *Ph pby1A*_ref and *Ph pby1B*_ref have been used for phylogenetic studies.

The *Ph pby1A* transcript is 2.1 kb in length and encodes a protein of 550 amino acids. The *Ph pby1B* transcript is 1.8 kb in length and encodes a protein of 396 amino acids. Both *Parhyale* PBY1A and PBY1B proteins have an N-terminal PAX domain (PAXD, 132 amino acids) and a HD (60 amino acids) that are highly similar to the corresponding domains of described PAIRED and Pax3/7 proteins (Figures 30, 31, 32, 33).

**Figure 30: Schematic view of the *Ph pby1* transcript, isoform A.** The length of the *Ph pby1* transcript, isoform A, is 2149 bases. Shown are: 5' UTR in grey (length 99 bases, nucleotide positions 1-99), ORF in black (length 1653 bases, encoding 550 amino acids; nucleotide positions 100-1752) and the 3' UTR in grey (length 397 bases, nucleotide positions 1753-2149). The ORF region encoding the PAX is depicted in orange (length 396 bases, encoding 132 amino acids; nucleotide positions 250-645) and the region encoding the HD in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 727-906). The fraction of transcript that both *Ph pby1* transcript isoforms share is framed with a red line (nucleotide positions 1-1174).

**Figure 31:Schematic view of the *Ph pby1* transcript, isoform B.** The length of the *Ph pby1* transcript, isoform B, is 1791 bases. Shown are: 5' UTR in grey (length 99 bases, nucleotide positions 1-99), ORF in black (length 1191 bases, encoding 396 amino acids; nucleotide positions 100-1290) and the 3' UTR in grey (length 501 bases, nucleotide positions 1291-1791). The ORF region encoding the PAX is depicted in orange (length 396 bases, encoding 132 amino acids; nucleotide positions 250-645) and the region encoding the HD in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 727-906). The fraction of transcript that both *Ph pby1* transcript isoforms share is framed with a red line (nucleotide positions 1-1174).

```
0001                                                                                                GGATTTTAC
0010   AAAGAACCCCTTTACCAACTCTTAACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCA

0001   M  E  C  M  Y  S  F  R  P  Q  P  F  G  P  M  G  V  S  V  P  F  P  S  A  S  A  F  P  P  T
0100   ATGGAGTGCATGTATAGCTTTAGGCCTCAGCCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACC

0031   A  P  L  P  S  T  A  N  L  D  H  E  N  N  D  F  D  N  A  H  A  N  S  R  V  N  Q  L  G  G
0190   GCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGCCGCGTCAACCAGCTGGGGGGC

0061   K  F  V  N  G  R  P  L  P  S  N  V  R  R  E  I  I  A  L  A  A  Q  G  I  R  P  C  N  I  S
0280   AAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCC

0091   K  A  L  K  V  S  H  G  C  V  S  K  I  L  N  R  Y  Q  E  T  G  S  I  K  P  G  V  L  A  G
0370   AAAGCCCTCAAAGTGTCCCACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGG

0121   V  A  R  S  L  P  P  R  L  D  E  R  I  V  Q  R  I  W  S  F  N  N  C  H  G  V  M  T  L  E
0460   GTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCATTCAATAATTGCCATGGCGTCATGACTCTGGAA

0151   Q  L  R  E  R  L  V  A  E  F  T  G  T  P  V  N  V  P  P  L  P  V  L  G  K  G  L  R  P  Y
0550   CAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTAC

0181   R  D  Q  L  P  F  L  E  C  T  N  T  D  S  D  S  E  E  D  T  E  P  G  L  T  I  R  R  K  Q
0640   AGGGACCAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAG

0211   R  R  S  R  T  S  F  T  A  I  Q  L  E  I  L  E  Q  E  F  C  S  N  Q  Y  P  D  I  T  T  R
0730   AGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGC

0241   E  H  L  A  E  R  T  G  L  T  E  A  R  V  Q  V  W  F  S  N  R  R  A  R  Q  R  K  A  S  N
0820   GAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCTAGCAAC

0271   S  P  E  E  A  A  Q  R  P  A  N  N  R  N  G  G  S  P  V  A  A  I  S  P  F  D  P  A  V  P
0910   AGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCT

0301   G  P  S  H  G  L  D  G  G  Y  G  M  F  N  R  Y  G  L  Q  A  P  A  S  D  A  V  V  P  P  V
1000   GGACCCAGTCACGGTCTGGACGGCGGCTATGGAATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTG

0331   K  I  E  H  S  P  S  S  S  C  E  D  S  S  S  S  P  N  Q  P  S  Y  H  P  P  A  Y  Q  G  T
1090   AAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCGCCTGCATACCAAGGAACA

0361   S  L  L  Y  P  R  Q  Q  V  D  G  F  G  F  N  Y  N  T  P  Y  N  P  A  V  H  P  G  Y  P  F
1180   TCGCTGCTGTACCCACGGCAGCAGGTTGACGGCTTTGGCTTCAACTACAACACGCCGTACAACCCGGCAGTTCATCCTGGCTACCCATTT

0391   G  D  P  G  Y  P  G  Q  M  Y  L  G  H  V  A  A  Q  Q  A  F  Q  A  P  G  R  S  G  Q  S  E
1270   GGGGACCCTGGCTACCCAGGACAGATGTACTTGGGTCACGTGGCGGCCCAGCAGGCATTCCAGGCCCCGGGGAGGTCGGGTCAGAGTGAA

0421   V  Q  T  P  M  S  G  D  R  P  S  A  G  Q  E  L  N  Q  R  H  R  S  R  D  S  L  Q  F  P  G
1360   GTCCAGACTCCCATGAGTGGTGATCGTCCCAGCGCTGGTCAGGAGCTAAACCAGCGCCACCGCTCCCGTGACAGCCTCCAGTTCCCTGGG

0451   P  Q  S  T  Y  G  Y  P  Q  P  L  S  N  N  T  S  D  V  T  A  T  R  L  S  P  P  C  T  S  Y
1450   CCCCAGAGCACTTACGGTTACCCTCAGCCCCTTAGCAACAACACTAGTGACGTGACAGCAACTAGGCTGTCGCCTCCCTGCACCAGTTAC

0481   A  G  A  S  E  Q  Q  Q  Q  G  P  V  T  T  S  A  A  N  F  W  Q  P  S  N  L  K  P  E  P  G
1540   GCTGGAGCCTCAGAGCAGCAGCAGCAGGGTCCTGTTACCACCTCAGCAGCCAACTTCTGGCAGCCCAGCAACCTAAAGCCCGAGCCTGGC

0511   T  S  S  L  G  S  S  S  F  A  T  A  A  N  G  G  S  F  C  E  N  Q  T  G  T  S  S  S  A  D
1630   ACATCCTCACTTGGAAGCTCGAGCTTCGCCACTGCTGCCAATGGCGGATCTTTCTGTGAGAACCAAACGGGAACTTCTTCGAGCGCCGAC

0541   Q  Q  R  P  M  S  V  G  L  A  &                                                          0550
1720   CAGCAGCGACCGATGTCCGTGGGCCTCGCTTGAGGCCTTGCCCTCAGCTACAGCATGGTAGTCCTGTACTGTGTCATGCTACACCGTACT
1810   CTGCAGTACTGTAACACCAGTAATACTGCGCGTCGTAGCGCTGTAACTTACAGAACTGTGTTACATAGTACCATAGCCTGCACAAGTTGG
1900   ACCCGTTGTTGTAGACCCCAGCTCATTGTACAGGAGTGTACTGTAGTCTTGCTGAAGATTCTCAGAAGAATTTTTGCTTCCTCACGCAAG
1990   GTGAAGTTATTTAATTGGTGTTGTTGCTTCTAGCTGTGGACGGCGAAGATTGCTCTGCAGGGTGCTTCTATTTATGTTGTAAGCTTAGTT
2080   TATAAGTGAAATCCACTTACGATGGTTGGTCAAATGGGTCCAGGAAAAAGAAAAAAAAAAAAAAAAAAAA                   2149
```

**Figure 32:** ***Ph pby1*** **(isoform A) cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph pby1* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The PAXD is shown in orange and the HD in blue. The nucleotide sequences that encode these domains and motifs are shown in the respective colours. In the C-terminus of the protein sequence, the first amino acid that is specific to this isoform is highlighted in pink. The 3' fraction of the cDNA that is specific to this isoform and differs from isoform B is printed italic. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

```
0001                                                                                                GGATTTTAC
0010        AAAGAACCCCTTTACCAACTCTTAACTGAGCACCTATAGTTACAGGCTAACTATCACCTGACAGTGTCCCTAGTTCTATATTATACGTCA

0001        M  E  C  M  Y  S  F  R  P  Q  P  F  G  P  M  G  V  S  V  P  F  P  S  A  S  A  F  P  P  T
0100        ATGGAGTGCATGTATAGCTTTAGGCCTCAGCCCTTTGGGCCAATGGGTGTGTCTGTTCCATTCCCGTCGGCTTCGGCATTCCCGCCGACC

0031         A  P  L  P  S  T  A  N  L  D  H  E  N  N  D  F  D  N  A  H  A  N  S  R  V  N  Q  L  G  G
0190        GCTCCCTTGCCATCCACCGCCAACTTAGATCATGAGAACAATGACTTCGATAACGCACATGCAAACAGCCGCGTCAACCAGCTGGGGGGC

0061         K  F  V  N  G  R  P  L  P  S  N  V  R  R  E  I  I  A  L  A  A  Q  G  I  R  P  C  N  I  S
0280        AAGTTCGTCAACGGCCGCCCCCTCCCCAGCAACGTCCGGCGGGAGATCATCGCCTTGGCAGCCCAGGGCATCCGTCCCTGCAATATCTCC

0091         K  A  L  K  V  S  H  G  C  V  S  K  I  L  N  R  Y  Q  E  T  G  S  I  K  P  G  V  L  A  G
0370        AAAGCCCTCAAAGTGTCCCACGGCTGCGTCTCCAAGATACTTAATCGCTACCAGGAGACGGGGTCCATCAAGCCAGGCGTTCTTGCGGGG

0121         V  A  R  S  L  P  P  R  L  D  E  R  I  V  Q  R  I  W  S  F  N  N  C  H  G  V  M  T  L  E
0460        GTCGCTCGCTCCCTGCCTCCCCGTCTAGACGAGAGGATTGTGCAGCGTATCTGGTCATTCAATAATTGCCATGGCGTCATGACTCTGGAA

0151         Q  L  R  E  R  L  V  A  E  F  T  G  T  P  V  N  V  P  P  L  P  V  L  G  K  G  L  R  P  Y
0550        CAGCTGCGGGAGAGATTGGTTGCTGAATTCACGGGAACGCCCGTCAATGTGCCACCTCTCCCAGTGCTCGGCAAAGGTCTGCGCCCCTAC

0181         R  D  Q  L  P  F  L  E  C  T  N  T  D  S  D  S  E  E  D  T  E  P  G  L  T  I  R  R  K  Q
0640        AGGGACCAGCTGCCCTTCTTAGAGTGCACCAACACAGACTCGGACAGCGAGGAGGACACAGAGCCTGGCCTGACCATCCGCCGCAAGCAG

0211         R  R  S  R  T  S  F  T  A  I  Q  L  E  I  L  E  Q  E  F  C  S  N  Q  Y  P  D  I  T  T  R
0730        AGGCGGTCGCGCACTTCCTTCACTGCAATTCAGCTGGAAATTCTGGAGCAGGAGTTTTGCAGCAACCAATACCCGGACATTACGACTCGC

0241         E  H  L  A  E  R  T  G  L  T  E  A  R  V  Q  V  W  F  S  N  R  R  A  R  Q  R  K  A  S  N
0820        GAGCACCTGGCAGAGAGGACTGGGCTCACTGAGGCTAGAGTGCAGGTTTGGTTCAGCAATAGACGCGCTCGCCAGAGAAAGGCTAGCAAC

0271         S  P  E  E  A  A  Q  R  P  A  N  N  R  N  G  G  S  P  V  A  A  I  S  P  F  D  P  A  V  P
0910        AGCCCGGAAGAGGCCGCTCAGCGCCCTGCCAACAACAGAAACGGCGGATCCCCAGTAGCTGCCATCAGTCCATTCGACCCAGCAGTACCT

0301         G  P  S  H  G  L  D  G  G  Y  G  M  F  N  R  Y  G  L  Q  A  P  A  S  D  A  V  V  P  P  V
1000        GGACCCAGTCACGGTCTGGACGGCGGCTATGGAATGTTCAACAGATACGGTCTGCAGGCTCCTGCAAGCGATGCTGTTGTCCCCCCTGTG

0331         K  I  E  H  S  P  S  S  S  C  E  D  S  S  S  S  P  N  Q  P  S  Y  H  P  P  A  Y  Q  D  G
1090        AAGATCGAGCATAGCCCCTCTTCTTCCTGCGAGGACTCCTCCAGCAGCCCAAACCAGCCCTCGTACCACCCGCCTGCATACCAAGATGGT

0361         P  G  F  L  F  P  M  Q  G  Q  A  A  V  L  N  M  S  F  S  D  G  D  E  P  E  N  Y  S  G  N
1180        CCTGGATTCCTCTTTCCAATGCAAGGACAAGCTGCTGTGCTGAACATGTCTTTCTCGGATGGAGACGAACCTGAGAATTATAGCGGGAAT

0391         P  P  E  A  S  Y  @                                                                      0396
1270        CCCCCGGAAGCTTCTATTAGTTTTGTTTTACACGTCAACTTCCTCGCCACTTGGTGTTTCATATTGTAATGGCTCTAAATAAGGGAGTT
1360        GTTTTATGTTCTTCTAGCATTTCACTAAATTAAGCATGATTTGCTCCCCATAGAAGTAGCAGTTTTCATATGATGTTTAGAGCGTGGGAA
1450        CTGGGAACTACTGCCCCAGTAGTGCCAAGTCAGAAGTTGAATGGCTGAGCGCCACAATACGATTTAATAGATTCAGATGAACCCTTAGCG
1540        AACATCATTGGTTGTATTTTATCGACTTCACTCAATCACTTTATAATATGATGCTACTTTTCTAAGCTGCCTAGTTTTAAACTTTTGACT
1630        ACATCACTTCGTACTTACGCATTAACTTAACTTTGAGGACAGCCAGAGTATCCTTCGTTTAAATTTTGAATATTATGGACTACTGTACAA
1720        GTGTGTTAATAGTACTGTAAATAAAGTAGTGGGAGCAATGACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA                1791
```

**Figure 33:** ***Ph pby1*** **(isoform B) cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph pby1* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The PAXD is shown in orange and the HD in blue. The nucleotide sequences that encode these domains and motifs are shown in the respective colours. In the C-terminus of the protein sequence, the first amino acid that is specific to this isoform is highlighted in pink. The 3' fraction of the cDNA that is specific to this isoform and differs from isoform A is printed italic. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

#### 3.2.3.4 Phylogeny of *Ph pby1A* and *Ph pby1B*

BLAST searches (Altschul et al., 1997) using the complete amino acid sequence of *Ph* PBY1A as query against the protein databases of *Drosophila melanogaster* (blastp, taxid 7227) and mouse (blastp, taxid 10090) revealed highest sequence similarity to *Drosophila* GOOSEBERRY (*Dm* GSB) and the murine gene product of *paired box gene 7* (*Mm* Pax7), respectively. For *Ph* PBY1B, an analogous BLAST search yielded identical results. Independently, a BLAST search was performed using the complete amino acid sequences of *Ph* PBY1A and *Ph* PBY1B as query against the complete translated nucleotide database (tblastn). From this search, a predicted protein sequence from the pea aphid *Acyrthosiphon pisum* (PREDICTED: *Acyrthosiphon pisum* similar to AGAP010358-PA (LOC100163852), mRNA; NCBI Reference Sequence: XM_001942939.1) showed highest sequence similarity to *Ph* PBY1A. A BLAST search using this predicted *Acyrthosiphon pisum* amino acid sequence as query against the protein database of *Drosophila melanogaster* (blastp, taxid 7227) revealed highest sequence similarity to *Dm* GSB. Another predicted protein sequence from the pea aphid *Acyrthosiphon pisum* (PREDICTED: *Acyrthosiphon pisum* similar to *paired* (LOC100161564), mRNA; NCBI Reference Sequence: XM_001948983.1) showed highest sequence similarity to *Ph* PBY1B. A BLAST search of this protein sequence against the protein database of *Drosophila melanogaster* (blastp, taxid 7227) revealed highest sequence similarity to *Dm* PRD. Both *Ph* PBY1 isoforms share the major part of the protein sequence encompassing the N-terminal PAXD and the HD. In order to find indications whether additional protein functions were present in the differing C-terminal parts of both isoforms, additional BLAST searches (tblastn) using only the unique fractions of both protein isoforms were performed. For *Ph* PBY1A, no specific results that would indicate additional protein domains or functions were retrieved (first BLAST hit/low BLAST score: common marmoset: PREDICTED: *Callithrix jacchus* D-3-phosphoglycerate dehydrogenase-like (LOC100412268), miscRNA). For *Ph* PBY1B, no results exhibiting any significant sequence similarity were obtained.

Within the family of paired-class homeodomain transcription factors, proteins that possess a PAIRED/Pax domain and an adjacent HD (PAX-HD proteins) constitute a conserved subfamily of homeodomain transcription factors (Noll, 1993). Members of this group are present in species as divergent as *Drosophila melanogaster* and mouse. From previous BLAST searches, representatives of the PRD/Pax3/7 subgroup of PAX-HD proteins

showed the highest sequence similarity to both isoforms of *Ph* PBY1. In order to substantiate the identity of *Ph* PBY1 and to elucidate its homology to known PRD/Pax3/7 proteins of other species, the phylogenetic relation of the *Parhyale* protein isoforms *Ph* PBY1A and *Ph* PBY1B as well as *Ph* PBY2 to known PAX-HD proteins of *Drosophila* and mouse was calculated (Mr Bayes, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The phylogenetic calculation was based on an alignment of *Ph* PBY1A, *Ph* PBY1B and *Ph* PBY2 together with the sequences of the PAX-HD proteins *Dm* GSB, *Drosophila* GOOSEBERRY-NEURO (*Dm* GSB-N), *Dm* PRD, *Drosophila* EYELESS (*Dm* EY), its paralog *Drosophila* TWIN OF EYELESS (*Dm* TOY) and *Drosophila* SHAVEN (*Dm* SV, PAXD with high similarity to the PAXD of *Ph* PBY1) as well as their murine homologues Pax3, Pax7, Pax6 and Pax2 (homologue of *Dm* SV). The protein sequence alignment was done using the CLUSTALW2 algorithm (Larkin et al., 2007). Specifically, the full-length protein sequences of all included factors were used for calculating the phylogeny (A6, table A6).

In the resulting phylogram, the PRD/Pax3/7 proteins from *Drosophila* and mouse are grouped together with *Ph* PBY1A, *Ph* PBY1B and *Ph* PBY2 (Figure 34). These factors constitute a subdivision that is clearly separated from the group of Pax6/EY proteins and from the Pax2/SV proteins that represent the outgroup of the phylogram. Within the group of PRD/Pax3/7, *Ph* PBY2 appears as the ortholog of *Dm* GSB. Like murine Pax3 and Pax7, however, the isoforms of *Ph* PBY1 are not shown to have unambiguous *Drosophila* orthologs. Instead, they appear in a divergent branch within the group of PRD/Pax3/7 proteins (Figure 34). Accordingly, both *Dm* PRD isoforms included in the phylogenetic calculation are grouped together with *Dm* GSB-N, forming a subdivision that is parallel to the *Ph* PBY2/*Dm* GSB branch (Figure 34). These findings suggest that *Ph* PBY1 is a PRD/Pax3/7 representative that is specific to the lineage leading to *Parhyale*. Importantly, since genomic sequence data of *Parhyale* was not available in this work, additional *Parhyale prd/pax3/7* genes might exist. Therefore, it cannot be excluded that more conserved and unambiguous orthologs of *Dm* PRD and *Dm* GSB-N are present in *Parhyale*.

A

```
/------------------------------------------------------------ Mmus_Pax2
|
|------------------------------------------------------------ Dmel_SV
|
|                                                   /--------- Dmel_TOY
|                                         /---89---+
+                                         |         \--------- Dmel_EY
|        /-----------------100----------------+
|        |                                 \------------------ Mmus_Pax6
|        |
|        |                                           /--------- Phaw_PBY1A
|        |        /-----------------100----------------+
\---100--+        |                                 \--------- Phaw_PBY1B
         |        |
         |        |                                 /--------- Dmel_PRDA
         |        |                         /---100--+
         |        |                         |         \--------- Dmel_PRDB
         \---92---+                 /---100--+
                  |                 |        \------------------ Dmel_GSB-N
                  |        /---100---+
                  |        |         |                 /--------- Phaw_PBY2
                  |        |         \-------100-------+
                  \---100--+                           \--------- Dmel_GSB
                           |
                           |                           /--------- Mmus_Pax7
                           \------------100------------+
                                                       \--------- Mmus_Pax3
```

B

```
/------------ Mmus_Pax2
|
|--------------- Dmel_SV
|
|                                        /-------- Dmel_TOY
|                                     /--+
+                                     |  \---------------- Dmel_EY
|                        /--------------+
|                        |             \----- Mmus_Pax6
|                        |
|                        |                                     / Phaw_PBY1A
|                        |   /----------------------------------+
\------------------------+   |                                 \- Phaw_PBY1B
                         |   |
                         |   |                              / Dmel_PRDA
                         |   |               /------------+
                         |   |               |            \ Dmel_PRDB
                         \---+          /----+
                             |          |    \----------- Dmel_GSB-N
                             |      /---+
                             |      |   |    /---------- Phaw_PBY2
                             |      |   \----+
                             \------+        \----------- Dmel_GSB
                                    |
                                    |            /---- Mmus_Pax7
                                    \------------+
                                                 \----- Mmus_Pax3

|--------------| 0.500 expected changes per site
```

**Figure 34: Phylogenetic analysis of *Parhyale* PBY1 and PBY2 and PRD/Pax3/7, EY/Pax6 and SV/Pax2 proteins from Drosophila and mouse.** Shown are: *Parhyale hawaiensis* PBY1, isoform A (Phaw_PBY1A), *Parhyale hawaiensis* PBY1, isoform B (Phaw_PBY1B), *Parhyale hawaiensis* PBY2 (Phaw_PBY2), *Drosophila melanogaster* EYELESS (Dmel_EY), *Drosophila melanogaster* GOOSEBERRY (Dmel_GSB), *Drosophila melanogaster* GOOSEBERRY-NEURO (Dmel_GSB-N), *Drosophila melanogaster* PAIRED, isoform A (Dmel_PRDA), *Drosophila melanogaster* PAIRED, isoform B (Dmel_PRDB), *Drosophila melanogaster* SHAVEN (Dmel_SV), *Drosophila melanogaster* TWIN OF EYELESS (Dmel_TOY), *Mus musculus* product of *paired box gene 2* (Mmus_Pax2), *Mus musculus* product of *paired box gene 3* (Mmus_Pax3), *Mus musculus* product of *paired box gene 6* (Mmus_Pax6) and *Mus musculus* product of *paired box gene 7* (Mmus_Pax7). Mmus_Pax2 was set as outgroup. For reference regarding the non-*Parhyale* protein sequences, see A6, tables A5 and A6. **A** Phylogram showing clade credibility values. **B** Phylogram based on average branch lengths. Phylogenetic Analysis was done using Mr Bayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; 5.3.2; 100,000 generations, sample frequency=10).

#### 3.2.3.5 Expression of *Ph pby1*

Expression of *Ph pby1* is first detectable in embryos that have exited gastrulation and have started the process of germ disc condensation (onset of stage 8, Browne et al., 2005). Weak *Ph pby1* expression is present in scattered, peripheral cells of the condensing germ disc (Figure 35 A).

Subsequently, still at early stage 8, strong *Ph pby1* expression has arisen abruptly in all cells of the germ disc, with bilateral clusters of cells in the posterior quadrants of the germ disc exhibiting the strongest *Ph pby1* signal Figure 35 B). From this point of development, the expression profile of *Ph pby1* is highly dynamic. As stage 8 progresses, *Ph pby1* expression recedes from the anterior-most cells of the germ disc. In addition, the level of *Ph pby1* expression declines in those cells that are located in the centre of the germ disc (Figure 35 C). At the end of stage 8, *Ph pby1* is strongly expressed in a bilateral, posterior domain that extends over the entire width of the germ disc and is interrupted by a small group of medial cells that show weak *Ph pby1* expression (Figure 35 D). The *Ph pby1* expression domain covers several rows of cells from anterior to posterior. Anterior of this domain, weak *Ph pby1* expression is found in several groups of cells (Figure 35 D).

At stage 9, *Ph pby1* is expressed exclusively in the bilateral posterior domain (Figure 35 E). *Ph pby1* expression remains unaltered as embryogenesis proceeds through stage 10 (not shown). At stage 11, the formation of the germ band begins. Initial ectodermal and mesodermal cell rows form (Browne et al., 2005; Figure 35 F2). Until the onset of stage 12, the *Ph pby1* expression domain covers exactly the two anterior-most rows of cells of the germ band (Figure 35 F, G, H). The domain remains interrupted by few medial cells that form the anterior end of the emerging ventral midline.

From stage 12, the regular alignment of germ band cells into a well-defined grid of cell rows proceeds in a co-linear spatial and temporal wave from anterior to posterior and from centre to periphery (Browne et al., 2005). *Ph pby1* expression from this stage has been studied in detail by another research group and is therefore not presented here. Their preliminary findings suggest that during trunk segmentation, *Ph pby1* spearheads the segmental cell row formation as it is expressed in the horizon of cells that are in the very moment of grid-like row alignment. Accordingly, *Ph pby1* expression gradually moves posterior (Ron Parchem, personal communication).

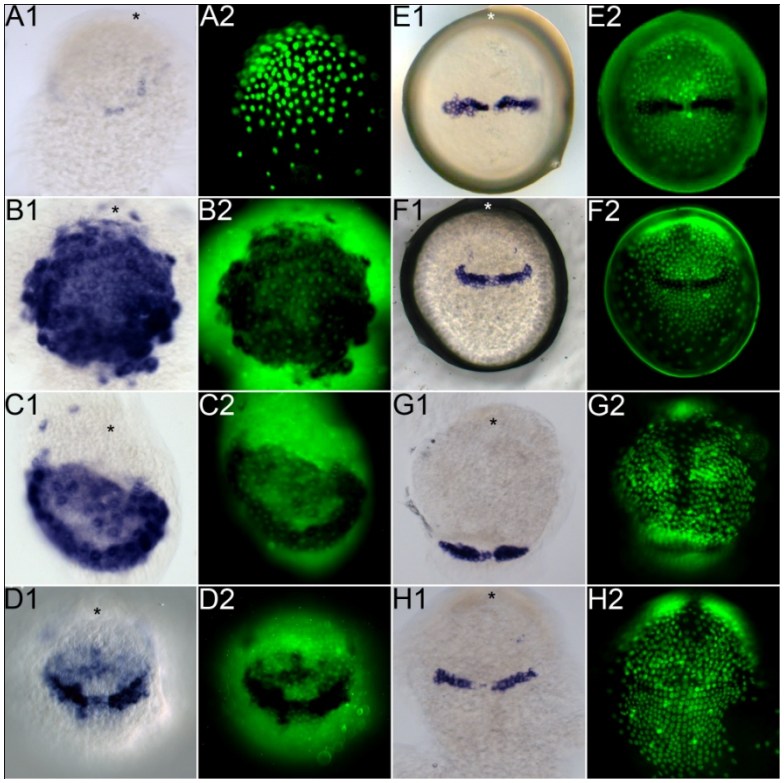



**Figure 35: Expression of *Ph pby1* at S8 to S12**. The embryo is mounted in ventral orientation (B, C, D, E, F, H), in ventral orientation, tilted to the right (A) and in anterior orientation (G). In all images, anterior is up. Shown are bright field (A, B, C, E, G, H) and DIC images (D, F) to the left (#1) and nuclear labelling of the same embryos (SYTOX®) to the right (#2; nuclei are green, expression of *Ph pby1* is detectable as it quenches the nuclear labelling). In all DIC and bright field images (#1), the anterior end of the embryo is indicated by a black asterisk. **A** Expression of *Ph pby1* at the onset of stage S8. **B** Expression of *Ph pby1* at early stage S8. **C** Expression of *Ph pby1* during the mid of stage S8. **D** Expression of *Ph pby1* at the end of stage S8. **E** Expression of *Ph pby1* at stage S9. **F** Expression of *Ph pby1* at stage S11. **G** Expression of *Ph pby1* at the onset of stage S12. Anterior view of the embryo. **H** Expression of *Ph pby1* at the onset of stage S12. Ventral view of the same embryo as in G.

#### 3.2.3.6 Expression of *Ph otd1* and *Ph pby1* at stages 8 and 9

Expression of both *Ph otd1* and *Ph pby1* arises in embryos of stage 8 (Figure 36 A; Figure 35 A, B; Browne et al., 2006). In order to study the expression dynamics of both *Parhyale* genes in relation to each other, a WMISH approach to visualise the expression of both *Ph otd1* and *Ph pby1* in a single embryo by using two alternative staining techniques has been pursued (5.4.4).

*Ph otd1* expression is first detectable in embryos that are in mid-stage 8. It is found in few cells that cluster at the anterior end of the germ disc (Figure 36 A). At this point of development, *Ph pby1* expression is most prominent in a posterior crescent-shaped domain, while *Ph pby1* signal in cells at the centre of the germ disc is of very low intensity. In the anterior-most section of the germ disc, no *Ph pby1* expression is found (Figure 36 A3, see also Figure 35 C).

At the end of stage 8, *Ph otd1* is expressed in a broad domain that covers approximately the anterior third of the germ disc. Within this domain, *Ph otd1* expression appears strongest in lateral clusters of cells on both sides of the anterior germ disc (Figure 36 B). *Ph pby1* expression is found in a horizontal stripe within the posterior third of the germ disc (Figure 36 B3, see also Figure 35 D). The central part of the germ disc does not show *Ph otd1* or *Ph pby1* signal. In nuclear labelling images, the cells within this region appear multi-layered (Figure 36 B3).

At stage 9, the developing bi-lobed head of the embryo has become distinctly visible. Both head lobes are separated by a relatively anuclear cleft (Figure 36 C3, Browne et al., 2005). *Ph otd1* is expressed in the majority of cells of each head lobe (Figure 36 C, Browne et al., 2006). In addition, novel *Ph otd1* expression has arisen in a cluster of cells located medially at the posterior of the embryo (Figure 36 C, Browne et al., 2006). Directly abutting posterior, *Ph pby1* is expressed in a bilateral, horizontal domain covering several cell rows from anterior to posterior (Figure 36 C3, see also Figure 35 E).

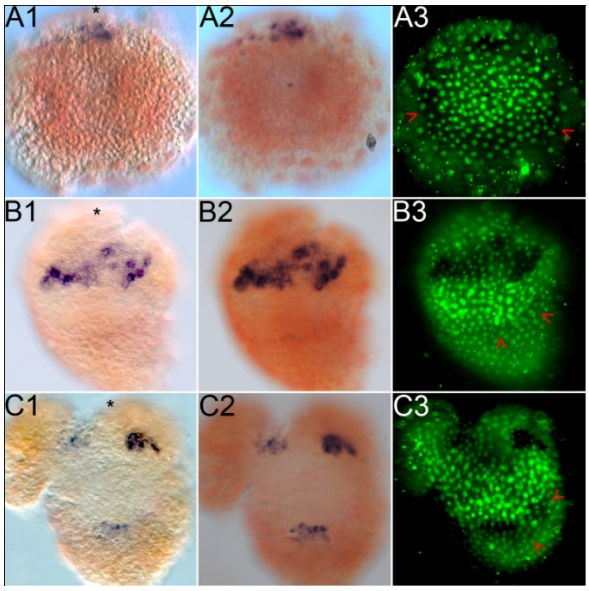



**Figure 36: Expression of *Ph otd1* and *Ph pby1* at S8 and S9.** The embryos have been subjected to WMISH using probes for both *Ph otd1* and *Ph pby1* in combination with two alternative staining techniques (MM). *Ph otd1* expression is detectable as blue staining (using a Digoxigenin-labelled probe). *Ph pby1* expression is detectable as red staining (using a Fluorescein-labelled probe). Please note that in DIC and bright field images, the unspecific orange signal caused during visualisation of the Fluorescein-labelled *Ph pby1* probe masks in part the specific red signal indicating *Ph pby1* expression. The specific Fluorescein-derived staining is visible in the images showing nuclear labelling of the embryos since it quenches the SYTOX® signal (red arrowheads). In contrast, the unspecific orange background does not quench the nuclear labelling. All embryos are mounted in ventral orientation. In all images, anterior is up. Shown are DIC images to the left, bright field images in the centre panel and nuclear labelling of the same embryos to the right (SYTOX®, nuclei are green, expression of *Ph otd1* is detectable as it strongly quenches the nuclear labelling and expression of *Ph pby1* is detectable through a different, slighter quenching effect, red arrowheads). In all DIC images, the anterior end of the embryo is indicated by a black asterisk. **A** Expression of *Ph otd1* (blue signal) and *Ph pby1* (red signal) during the mid of stage S8. **B** Expression of *Ph otd1* (blue signal) and *Ph pby1* (red signal) at the end of stage S8. . **C** Expression of *Ph otd1* (blue signal) and *Ph pby1* (red signal) at stage S9.

### 3.2.4 *Parhyale hawaiensis arrowhead*

Representatives of the LIM homeobox family of genes have been described to play important roles in head and brain development of various species (Curtiss and Heilig, 1997; Hobert and Westphal, 2000). For this reason, I cloned and analysed the putative *Parhyale hawaiensis arrowhead* (*Ph awh*) gene in this work. *Ph awh* belongs to the LIM8/*awh* subgroup of LIM homeobox genes and was identified during the homeobox screen (3.2).

#### 3.2.4.1 Isolation of *Ph awh*

Because sufficient genomic sequence data is not currently available for *Parhyale hawaiensis*, the isolation of gene-of-interest sequences was approached by performing varied-stringency degenerate PCR screens (5.2.3.1). From repeated low-stringency degenerate PCR screens targeted at paired-type homeobox sequences, homeobox fragments of the putative *Parhyale hawaiensis awh* ortholog, a LIM-box containing homeobox gene, were recovered. The isolated fragments contained 109 bp of specific *Ph awh* sequence and were sufficient for unambiguous homology identification (5.3.3.1, A2.7).

In detail, two sequences were recovered. BLAST evaluation revealed that they represent a novel *awh/LIM* sequence. Both obtained sequences are considered identical because the nucleotide variation between them is less than 1%. Since each sequence derives from a different priming event, they are considered independent.

Based on this primary sequence information, specific oligonucleotides were designed and used to prime *Ph awh* RACE reactions (5.2.3.3, A4.2, table A3). Consequently, the remaining mRNA transcript sequences both 5' and 3' of the initial fragments covering the homeobox were obtained.

In detail, *Ph awh* 5' cDNA sequence was recovered by two-step 5' RACE. Seven independent *Ph awh* 5' cDNA fragment clones were obtained. Their nucleotide sequence varies considerably (see also A3, table A1), with any two individual sequences typically differing in more than 3% of nucleotides. A comparison of all obtained *Ph awh* sequences reveals a high frequency of polymorphic nucleotide positions within the fraction of sequence covered by the *Ph awh* 5' cDNA fragment clones (3.2.4.2). These account for most of the relative nucleotide differences between individual *Ph awh* 5' cDNA sequences (see also A3, table A1). Additionally, all recovered *Ph awh* clones show significant variations in the fraction of *Ph awh* 5' sequence that they cover, represented by sequence insertions or deletions, respectively. The alignment of all *Ph awh* 5' sequences suggests a putative common transcription start. Taken together, these findings indicate the occurrence of various *Ph awh* transcript isoforms that derive from different alleles of the same gene. The observed sequence variations do not exceed 10% difference in nucleotides between two individual sequences, which does not support the existence of *Ph awh* paralogs.

*Ph awh* 3' cDNA sequence was recovered by two-step 3' RACE. Seven independent *Ph awh* 3' cDNA fragment clones were obtained. Similar to the observations made for the

*Ph awh* 5' RACE results, the recovered *Ph awh* 3' cDNA sequences vary considerably with regard to nucleotide positions (up to 4% variation), most of which represent polymorphic sites (3.2.4.2). They also show differences concerning the fraction of sequence they cover (relative insertions and deletions). These findings corroborate the suggested occurrence of various *Ph awh* transcript isoforms that derive from different alleles of the same gene. All recovered *Ph awh* 3' cDNA sequences show minor differences in poly(A) tailing.

Sequence information provided by all recovered 5' and 3' *Ph awh* RACE clones was used to assemble the complete *Ph awh* cDNA sequence *in silico*. In order to verify the consistency of the established *Ph awh* cDNA sequence, coherent *Ph awh* ORF fragments were isolated via long-distance PCR reactions (5.2.3.4), performed on *Parhyale* cDNA collections (5.2.2). Three independent *Ph awh* ORF sequences were obtained (A2.7.1). They are consistent with the findings from 5' and 3' RACE (A2.7.2) and considered identical, since the nucleotide variation between any two of these is not exceeding 1% (A3, table A1). Importantly, despite repeated attempts a complete consistent *Ph awh* cDNA sequence clone including 5' and 3' UTRs has so far not been recovered.

#### 3.2.4.2 Characterisation of *Ph awh*

The comparison of all recovered *Ph awh* sequences led to the identification of 145 sites of nucleotide exchange within the complete *Ph awh* cDNA. 54 of those occur uniquely and are therefore considered sporadic (5.3.3.2). They are evenly distributed within the 5' UTR (24), ORF (14) and 3' UTR (16). In the ORF, they account for five amino acid changes (missense alterations) and one premature stop (nonsense alterations). The remaining eight nucleotide exchanges found within the ORF do not alter the translated amino acid sequence (silent exchanges). This finding suggests that these nucleotide exchanges occur randomly and are therefore artificial. 91 nucleotide exchanges are present in more than one sequence, hence being considered polymorphic. They are found predominantly in the 5' UTR (41). Of the 25 polymorphic nucleotide exchanges found within the ORF, the majority (21) do not alter the translated amino acid sequence. The remaining four polymorphic nucleotide exchanges lead to three amino acid exchanges in the N-terminal part of the protein, outside conserved protein domains (see also A2.7.2).

Most insertions and deletions found throughout all recovered *Ph awh* cDNA sequences lie within the 5' and 3' UTRs. One deletion, however, leads to a loss of 16 amino acids in the N-terminal part of the protein. It is found in two independent 5' RACE sequences (Ph_awh_5Rn122 and Ph_awh_5Rn150) and does not lead to a frame shift with regard to the remaining ORF, suggesting an alternative splicing event.

These findings support the fact that all recovered *Ph awh* sequences derive from transcripts of the same gene. Due to the frequent occurrence of polymorphic nucleotide exchanges and the strong variation in the fraction of sequence that is covered by any individual sequence clone it is probable that different *Ph awh* alleles exist and that *Ph awh* transcripts are spliced alternatively.

Based on these considerations, the sequence of the clone Ph_awh_ORF21 was chosen as the source of the *Ph awh* reference sequence (*Ph awh_ref*). It was extended 5' by adding 1210 bp derived from clone Ph_awh_5RCR02 and 3' by adding 1298 bp derived from clone Ph_awh_3Rn21 (A2.7.1). An alignment of *Ph awh_ref* with all recovered *Ph awh* sequences shows that *Ph awh_ref* represents the statistically most common transcript isoform. However, individual recovered *Ph awh* sequences may show more than 3% difference from *Ph awh_ref*. *Ph awh_ref* has been used for phylogenetic studies.

The *Ph awh* transcript is 3.6 kb in length and encodes a protein of 372 amino acids. The *Parhyale* AWH protein has two N-terminal LIM (LIM1, 53 amino acids and LIM2, 54 amino acids) and a HD (60 amino acids) that are highly similar to the corresponding domains of described AWH and LIM6/8 proteins (Figures 37, 38; 3.2.4.3).

**Figure 37: Schematic view of *Ph awh* transcript.** The length of the *Ph awh* transcript is 3567 bases. Shown are: 5' UTR in grey (length 1181 bases, nucleotide positions 1-1181), ORF in black (length 1119 bases, encoding 372 amino acids; nucleotide positions 1182-2300) and the 3' UTR in grey (length 1267 bases, nucleotide positions 2301-3567). The ORF region encoding the LIM1 is depicted in orange (length 159 bases, encoding 53 amino acids; nucleotide positions 1530-1688) and the region encoding the LIM2 in bright orange (length 162 bases, encoding 54 amino acids; nucleotide positions 1713-1874). The ORF region encoding the HD is depicted in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 1947-2126).

```
0001                                                                                                GAATTTTCACA
0012   CTCCCCTTTACCGAGTTTTACTGTTACTCATGAATTATTTCTCAGAGCACTTAAAAGAAAATGACATGACTTCAGAATTTAATGCATGTG
0102   GAGGGTGCATTTTCCAGAAAATATTTGCTGATGTGTGAGAAAATATTCCTGCAGGAAAATGTTACTCAAAAGGATTGTTAAATTTAGCTG
0192   GGAAGAAGCAACAAACAAAAACAAGCAACAAGCTTGTGACGGAACTGCTGTTAGTATTGTAGGTGATATTGAGTTCCTCGCCGCTTTTTG
0282   ACCAGAATATTAATTTGTTAATCAGATAGAAAATATCACTGCAATAGTTGTTACAGAAGTTTACTGAAATGGACTGTGCAAACTTGAAGT
0372   AAACTTCGGCGGCGTGTGTTGTATGGACGTGCGGACCTCGTTCTGAACTGTAACTGTTGTACATAGAAACTAACTAGTTGTAGGACTGGC
0462   TGTGTGGGGGGTAAAGTGTTAAGTGTTGGCGAAGAACAATTGAAAAAAGAGGTTAATACGGTAAGTAGCGTTTCAAGCTACGGATATGAT
0552   AGAACACAACTTCTGGATAAAGAACTACCCGCGAAATTCGGCATGGTTCTTGGTATCCATGGATGACGAGAATCCAGAAGACAGTGGAAT
0642   ATTGCTCCAATAAGTTTAGTTTCTGCCAGTAGAATTGCAGAAATTGATTATATTTTAATTATCAGCAGAACTGTGCACTACTGTAGCTGA
0732   ACTAATAGAGATTTTCATATCCTTATGTAAGTGTTCTTAAGAACAATAGAGAACAGAAGAACCACTGGAAAAACTAATATATATAACTCA
0822   GTGAACAAGGATACCGGCAGCACCCGTTATTAAGGACTATTTGTGAAATAAGACAAGGTAAGGTTCTTACTAATAGCAAGCCGAACCAGA
0912   ACCAGTAACGGAAGAAAATTAGTTAAAGTAGAACAGGTTCTCAGCGCTAATATACCACGGAACAACAAACGAAAGAGGAATCTGTGAATA
1002   GAAATGAAAACTGATATTACGGACTGTCAAGGAAATGATCTAAAGTTCTTAGAAGACAACTAAACAAAAAGAGAAACTGCGAGTGAACCA
1092   GTTAGGCAAAACGTCAGAACCAATAATTAGAACTATTAGTGAAGTACTGCCAGGTTCTAAAGCGCAACAGACAACAGCAGAACTAACAGG

0001   M  M  A  S  A  T  V  L  L  P  M  N  R  I  K  A  E  R  N  S  P  V  S  S  P  E  A  S  S  G
1182   ATGATGGCCAGCGCAACAGTTCTGTTGCCAATGAACAGAATCAAAGCAGAGCGCAACAGTCCGGTGTCGAGCCCTGAGGCCTCCAGTGGT

0031    N  T  G  D  G  D  P  S  G  P  P  C  C  M  G  G  S  V  K  L  T  V  S  G  L  Q  H  T  A  R
1272   AACACAGGTGATGGGGACCCCTCGGGACCCCCCTGCTGTATGGGGGGCAGCGTTAAGCTGACTGTAAGCGGCCTACAACACACGGCGAGA

0061    E  R  R  L  G  C  L  P  L  S  V  N  N  N  N  N  S  S  T  T  N  N  N  N  S  N  S  A  S  A
1362   GAGCGGAGGCTGGGGTGTTTGCCTCTGAGCGTAAACAACAACAACAACAGTTCCACCACAAACAACAACAATAGCAACAGTGCCAGTGCT

0091    A  P  H  G  V  E  G  G  G  G  V  R  G  L  S  G  G  G  C  G  A  P  T  T  E  S  C  A  G  C
1452   GCCCCACACGGGGTGGAGGGAGGTGGGGGTGTGCGGGGGCTCTCTGGAGGAGGGTGCGGCGCCCCCACCACAGAGAGCTGTGCCGGCTGT

0121    A  R  I  I  T  D  R  F  L  L  R  V  N  C  L  S  W  H  Q  S  C  L  R  C  C  V  C  Q  L  A
1542   GCCAGGATCATCACTGACAGGTTCCTGCTGCGTGTCAACTGCCTCTCCTGGCACCAGAGCTGCCTGCGCTGCTGCGTGTGCCAGCTGGCC

0151    L  D  R  Q  P  S  C  F  I  R  D  Q  N  I  Y  C  K  S  D  Y  S  R  N  F  G  A  R  C  A  K
1632   CTGGACCGCCAGCCGTCCTGCTTCATACGCGACCAAAACATTTACTGCAAGTCCGACTACTCTCGAAATTTTGGGGCCCGCTGCGCCAAA

0181    C  C  R  A  I  G  S  T  D  W  V  R  R  A  R  D  R  V  Y  H  L  A  C  F  A  C  D  A  C  K
1722   TGCTGCCGGGCCATCGGGTCGACGGACTGGGTCCGGAGGGCGAGGGATCGGGTCTACCATCTCGCCTGTTTCGCCTGCGACGCCTGTAAG

0211    R  Q  L  S  T  G  E  E  F  A  L  H  E  S  R  V  L  C  K  Q  H  Y  L  E  A  V  E  G  G  A
1812   AGACAACTGAGCACAGGAGAGGAGTTCGCCTTGCATGAGAGCAGAGTATTGTGTAAACAGCATTACTTGGAGGCCGTGGAGGGAGGAGCA

0241    T  S  N  D  E  N  T  D  G  E  G  G  Q  R  K  S  K  R  V  R  T  T  F  T  D  E  Q  L  Q  V
1902   ACTAGTAATGATGAGAACACGGACGGTGAGGGCGGCCAGCGCAAGAGTAAGCGGGTCCGGACGACCTTCACGGACGAGCAGCTGCAGGTG

0271    L  Q  A  N  F  Q  I  D  S  N  P  D  S  Q  D  L  E  R  I  A  Q  L  T  G  L  S  K  R  V  T
1992   CTCCAGGCTAACTTCCAGATCGACTCGAACCCTGACAGCCAGGACCTGGAGCGCATCGCACAGCTCACGGGGCTGAGCAAGCGGGTGACG

0301    Q  V  W  F  Q  N  S  R  A  R  Q  K  K  H  M  T  S  K  K  H  H  Q  G  L  E  R  G  L  H  D
2082   CAGGTGTGGTTCCAAAACAGCAGAGCGCGGCAGAAGAAGCACATGACTTCTAAAAAACATCACCAAGGTCTGGAGCGCGGCCTGCACGAC

0331    I  T  G  S  L  A  R  T  I  D  L  H  F  M  Y  S  F  R  P  Q  S  A  E  P  E  D  L  S  S  P
2172   ATCACAGGCAGCTTAGCCCGTACAATCGACCTGCACTTCATGTATTCCTTCCGTCCCCAGAGTGCAGAGCCAGAGGACCTGTCGTCTCCC

0361    E  N  S  L  M  T  L  M  P  G  E  A  &                                                      0372
2262   GAGAACTCGCTTATGACGCTCATGCCTGGAGAAGCATGACGGTACGACCCCTCCACCTCCCCTCCACCTCCTCCTCCACCCACCCTCCAC
2352   TGACGGCGCTGTGTCACCCGTTGTGGCACTCCTGGCGAGAGTCGCACTTTGTGACACCAACAGACTGCTCTGTGTCACCAGTTGTGACAC
2442   TAACCGGCACACGCAATGAAGCAAACATTTCCAGAGAAGTTGACCTTAGAGCAGAATGGTATTAACGAGTTGGGAATGTAACGTCAAGCT
2532   GGCATTGATTTTCAATTTGAAGAAATGAAAATATAACATACTTGGAAGTCATATATCTGAAACTAAACAAATGACAGCAATCTTGACTAC
2622   CAGTCGGACAATTACACGAGACAATCTCAGCTACCAGGCTTACGATTACACGAGTCATCTCAAGGGAAGCAGGATTTTTTAGAGGTCGGA
2712   TGAGATGGGACTCCGTGTGTCGCTACTGAAGTTGGTTAAATTCTTCATGAGACCCAGTGGATCCGTAAGGTCAGGAGTCAAATGGATCTG
2802   AGATCCAGTGGATCAGTTACACAGAGTTCAATGACAGCGACTAGCAATAACTAGTTTCCAGACTAGTTACTTCTGCTAGTTCCAAACATA
2892   TTATGGGTAGTGACAAGAATTTTCGAATAATCATTGCGGCGAATTCGGCCCACTCGTGGCTGTCGAGGAAAGCCAGTCAGCTTGGCTTCT
2982   CCCGGTGGCAGAGAGCAGAGACTCTGGCACCGTCCGAGTGCCAGTAAGTCAAGCTTGAAGCTCAATCTAAAGCGCGACTCGAACGCCGGC
3072   GATGGAAACTGCGGAAACGTTAAAGCTAAGTGGGATTAATTTAATCGATGACAGTAGGCGAATTGCCCACGAGTGTGTCTTATTCAGCAT
3162   ACTATAGATAAGCATGCGTATGATCTAACGCAGCACTCTAGGGAAGGTTGCGTATGGTCAAACAGGGCACCCTAGGGAACGTTGCGTATG
3252   GTTGCAGCACACTTCTGCGTATGTTCTGAGTCGCGCGGCCGCAGATAAGTGCAGGGCGGAACCACGCAGCTTTGCTCAAGTTGGACAGAT
3342   TCAGCTCTGTAGGAATGGAAGAATTACACAACAGTACGCGGGCGATGAGTTGCAGCAAATACGACACCGTGGTCATCTGGAAATGTTCTG
3432   CACAGCTTCTACATCAAGAGTACCCTAGGTGGGGGCACGGACCGTCAGAGAGCGCGTGATACCATTGTAGGGGATTCATGATAAATATAA
3522   TTGGTACACTCAGTACCGGTAAAAAAAAAAAAAAAAAAAAAAAAAAA                                            3567
```

**Figure 38:** ***Ph awh* cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph awh* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The LIM1 domain is shown in orange, the LIM2 domain in bright orange and the HD in blue. The nucleotide sequences that encode these domains are shown in the respective colours. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

### 3.2.4.3 Phylogeny of *Ph awh*

BLAST searches (Altschul et al., 1997) using the complete amino acid sequence of *Parhyale* AWH as query against the protein databases *of Drosophila melanogaster* (blastp, taxid 7227) and mouse (blastp, taxid 10090) revealed highest sequence similarity to *Drosophila* AWH and murine LIM homeobox protein 6 (Lhx6), respectively. *Mus musculus* LIM homeobox protein 8 is found as the second-most similar protein within the mouse proteome. Other proteins that exhibit the same domain structure (LIM LIM HD) were found with significantly lower BLAST score. Independently, a BLAST search was performed using the complete amino acid sequence of *Parhyale* AWH as query against the complete translated nucleotide database (tblastn). From this search, the putative *Tribolium castaneum* AWH (PREDICTED: *Tribolium castaneum* similar to GA10520-PA (LOC659841), mRNA; NCBI Reference Sequence: XM_966109.1) showed highest sequence similarity to *Parhyale* AWH.

Within the family of homeodomain transcription factors, proteins that possess two consecutive LIM domains and an adjacent HD (LIM-HD proteins) constitute a conserved subfamily of transcription factors that divides further into six groups (Hobert & Westphal, 2000). Members of each LIM-HD protein group are present in species as divergent as mouse, *Drosophila melanogaster* and *Caenorhabditis elegans*. The identity of *Ph* AWH as a representative of the Lhx6/8 group within the family of LIM-HD proteins was confirmed by calculating the phylogenetic relation of *Ph* AWH to all known individual LIM-HD proteins present within the mouse and Drosophila proteomes (Mr Bayes, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The phylogenetic calculation was based on an alignment of *Ph* AWH with Drosophila and mouse LIM-HD proteins using the CLUSTALW2 algorithm (Larkin et al., 2007). Specifically, a coherent fragment covering both LIM domains and the HD was used for calculating the phylogeny (A6.3). In the resulting phylogram, *Ph* AWH is grouped together with *Dm* AWH and murine Lhx6 and Lhx8 proteins into one clade with absolute credibility (Figure 39). Accordingly, this confirms *Ph* AWH as the ortholog to *Dm* AWH. Interestingly, both of the two murine LIM-HD proteins Lhx6 and Lhx8 are identified as equally orthologous to *Ph* AWH. This strongly suggests that they represent paralogs derived from a duplication event that occurred within the lineage leading to vertebrates and mammalians.

In addition, an alignment of *Ph* AWH with Lhx6/8 and AWH proteins of different species was constructed using the CLUSTALW2 algorithm (Larkin et al., 2007) and used for

phylogenetic calculation (Mr Bayes, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The resulting phylogram perfectly reconciles the phylogenetic relations of the species included (Figure 40, see also A6.1, table A5), suggesting a high level of conservation of Lhx6/8/AWH proteins. The fact that both Lhx6 and Lhx8 proteins are present in zebrafish (*Danio rerio*), corroborates the suggestion that these proteins represent paralogs derived from a duplication event that occurred within the lineage leading to vertebrates and mammalians and that zebrafish Lhx6 and Lhx8 proteins as well as mouse Lhx6 and Lhx8 proteins are co-orthologous to *Ph* AWH.

**Figure 39 (next page): Phylogenetic analysis of *Parhyale* AWH and LIM-HD proteins from *Drosophila* and mouse.** Shown are: *Parhyale hawaiensis* AWH (Phaw_AWH), *Drosophila melanogaster* AWH (Dmel_AWH), *Drosophila melanogaster* LIM1 (Dmel_LIM1), *Drosophila melanogaster* LIM3 (Dmel_LIM3), *Drosophila melanogaster* TAILUP/ISLET (Dmel_TUP), *Drosophila melanogaster* APTEROUS (Dmel_AP), *Drosophila melanogaster* gene product of CG32105 (Dmel_CG32105), *Drosophila melanogaster* gene product of CG4328 (Dmel_CG4328), *Mus musculus* LIM homeobox protein 1 (Mmus_Lhx1), *Mus musculus* LIM homeobox protein 2 (Mmus_Lhx2), *Mus musculus* LIM homeobox protein 3 (Mmus_Lhx3), *Mus musculus* LIM homeobox protein 4 (Mmus_Lhx4), *Mus musculus* LIM homeobox protein 5 (Mmus_Lhx5), *Mus musculus* LIM homeobox protein 6 (Mmus_Lhx6), *Mus musculus* LIM homeobox protein 8 (Mmus_Lhx8), *Mus musculus* LIM homeobox protein 9 (Mmus_Lhx9), *Mus musculus* LIM homeobox transcription factor 1 alpha (Mmus_Lmx1a), *Mus musculus* LIM homeobox transcription factor 1 beta (Mmus_Lmx1b), *Mus musculus* ISL1 transcription factor (Mmus_Isl1) and *Mus musculus* insulin related protein 2 (Mmus_Isl2). Dmel_AP was set as outgroup. For reference regarding the non-*Parhyale* protein sequences, see A6, tables A5 and A6. **A** Phylogram showing Clade credibility values. **B** Phylogram based on average branch lengths. Phylogenetic Analysis was done using Mr Bayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; 5.3.2; 100,000 generations).

**A**

```
/------------------------------------------------------------ Dmel_AP
|
|                                                   /--------- Phaw_AWH
|                                          /--100--+
|                                          |        \--------- Dmel_AWH
|        /----------------100---------------+
|        |                                 |        /--------- Mmus_Lhx6
|        |                                 \--100--+
|        |                                          \--------- Mmus_Lhx8
|        |
|        |                                          /--------- Mmus_Lhx4
|        |                                 /--100--+
|        |                                 |        \--------- Mmus_Lhx3
|        |                        /---100--+
|        |                        |        \----------------- Dmel_LIM3
|        |                        |
|---100--+                /---100--+                /--------- Mmus_Lhx5
|        |                |       |        /--100--+
|        |                |       |        |        \--------- Mmus_Lhx1
+        |                |       \---100--+
|        |                |                \----------------- Dmel_LIM1
|        |       /---100--+
|        |       |        |                          /--------- Mmus_Lmx1a
|        |       |        |                 /--100--+
|        |       |        |                 |        \--------- Mmus_Lmx1b
|        |       |        \-------100-------+
|        |       |                          |        /--------- Dmel_CG32105
|        \--100--+                          \---99--+
|                |                                   \--------- Dmel_CG4328
|                |
|                |                                   /--------- Mmus_Isl2
|                |                          /---88--+
|                |                          |        \--------- Mmus_Isl1
|                \------------100-----------+
|                                           \----------------- Dmel_TUP
|
|                                                   /--------- Mmus_Lhx9
\------------------------100------------------------+
                                                    \--------- Mmus_Lhx2
```

**B**

```
/---------- Dmel_AP (7)
|
|                        /----- Phaw_AWH
|                  /-----+
|                  |     \--------- Dmel_AWH
|            /------+
|            |      |             /---- Mmus_Lhx6
|            |      \-------------+
|            |                    \----- Mmus_Lhx8
|            |
|            |                              /--- Mmus_Lhx4
|            |                         /---+
|            |                         |   \--- Mmus_Lhx3
|            |                  /-------+
|            |                  |       \------- Dmel_LIM3
|            |                  |
|------------+           /------+            /--- Mmus_Lhx5
|            |           |      |       /----+
|            |           |      |       |    \--- Mmus_Lhx1
+            |           |      \-------+
|            |           |              \------------ Dmel_LIM1
|            |     /-----+
|            |     |     |                     /----- Mmus_Lmx1a
|            |     |     |               /-----+
|            |     |     |               |     \---- Mmus_Lmx1b
|            |     |     \---------------+
|            |     |                     |    /-------------- Dmel_CG32105
|            \-----+                     \----+
|                  |                          \---------------- Dmel_CG4328
|                  |
|                  |                          /----- Mmus_Isl2
|                  |                    /----+
|                  |                    |    \---- Mmus_Isl1
|                  \-------------------+
|                                       \-------- Dmel_TUP
|
|              /--- Mmus_Lhx9
\--------------+
               \-- Mmus_Lhx2

|----------------| 0.500 expected changes per site
```

**Figure 39:** **Phylogenetic analysis of *Parhyale* AWH and LIM-HD proteins from *Drosophila* and mouse.**

**A**

```
/------------------------------------------------------------ Mmus_Lhx8
|
|------------------------------------------------------------ Drer_Lhx8a
|
|                                                     /------- Dmel_AWH
|                                             /--82--+
|                                             |      \------- Dpse_AWH
|                                     /--100-+
|                                     |      \-------------- Dvir_AWH
|                             /--100-+
+                             |      \--------------------- Agam_AWH
|                         /--98--+
|                         |      \---------------------------- Tcas_AWH
|                   /--100-+
|                   |      \---------------------------------- Amel_AWH
|             /--100-+
|             |      \---------------------------------------- Phaw_AWH
|      /--100-+
|      |      \------------------------------------------------ Cele_LIM-4
\--100-+
       |                                             /------- Mmus_Lhx6
       \------------------------100----------------------+
                                                     \------- Drer_Lhx6
```

B

```
/----- Mmus_Lhx8
|
|----- Drer_Lhx8a
|
|                                                         / Dmel_AWH
|                                                        /+
|                                                        |\ Dpse_AWH
|                                                     /---+
|                                                     |   \- Dvir_AWH
|                                                 /---+
+                                                 |   \----------- Agam_AWH
|                                             /--+
|                                             |  \----- Tcas_AWH
|                                     /-----+
|                                     |     \---------------- Amel_AWH
|                             /--------+
|                             |        \------------ Phaw_AWH
|     /--------------------------+
|     |                       \----------------------- Cele_LIM-4
\-----+
      |    /--- Mmus_Lhx6
      \----+
           \-- Drer_Lhx6

|--------| 0.200 expected changes per site
```

**Figure 40: Phylogenetic analysis of *Parhyale* AWH and Lhx6/8 and AWH proteins from other species.** Shown are: *Parhyale hawaiensis* AWH (Phaw_AWH), *Drosophila melanogaster* AWH (Dmel_AWH), *Drosophila virilis* AWH (Dvir_AWH), *Drosophila pseudoobscura* AWH (Dpse_AWH), *Anopheles gambiae* AWH (Agam_AWH), *Tribolium castaneum* AWH (Tcas_AWH), *Apis mellifera* AWH (Amel_AWH), *Caenorhabditis elegans* LIM-4 (Cele_LIM-4), *Danio rerio* Lhx6 (*Drer*_Lhx6), *Danio rerio* Lhx8 (*Drer*_Lhx8), *Mus musculus* LIM homeobox protein 6 (Mmus_Lhx6) and *Mus musculus* LIM homeobox protein 8 (Mmus_Lhx8). Mmus_Lhx8 was set as outgroup. For reference regarding the non-*Parhyale* protein sequences, see see A6, tables A5 and A6. **A** Phylogram showing Clade credibility values. **B** Phylogram based on average branch lengths. Phylogenetic Analysis was done using Mr Bayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; 5.3.2; 100,000 generations).

#### 3.2.4.4 Expression of *Ph awh*

Expression of *Ph awh* is first detectable in stage 17 embryos. The elongating germ band has begun to fold inwards toward the egg interior at segment T5. This posterior flexure is visible as a distinct furrow in the germ band (Figure 41 B1-B3, D1-D3; 3.1.3; Browne et al., 2005). *Ph awh* is expressed bilaterally in clusters of cells that lie medially and slightly anterior to the base of the An1 appendages (Figure 41 A1-A3, C1-C3). Additionally, starting from the Mn segment and extending until the posterior flexure, dispersed *Ph awh* expression is visible medially in two anterior-posterior stretches, abutting the very central row of cells of the ventral midline (Figure 41 B1-B3, D1-D3, E1-E3). No expression of *Ph awh* was found in embryos of earlier stages.

In stage 19 embryos, the bi-lobed labrum begins to grow over the stomodeal opening (Figure 42 compare A1 and B1, compare A3 and B3; 3.1.3; Browne et al., 2005). *Ph awh* expression is visible in the pre-antennal part of the embryonic head in broad, symmetric domains that appear to be composed of smaller clusters of *Ph awh* positive cells (Figure 42 C1-C3). Additionally, *Ph awh* expression is found at the base of the An1 appendages (Figure 42 A1-A3; C1-C3). *Ph awh* is strongly expressed in the tips of both labral lobes (Figure 42 B1-B3, F1-F3). Strong *Ph awh* expression is also apparent at the bases of the Mn appendage buds, reaching into the anterior part of the Mx1 segment (Figure 42 B1-B3; D1-D3). From there, extending until the posterior end of the embryo, dispersed *Ph awh* expression remains in two mediolateral stretches that lie at the bases of the developing segmental appendages (Figure 42 C1-C3; D1-D3; E1-E3).

In stage 22 embryos, the stomodeum is entirely covered by the labrum that has completed its elongation and directly abuts the paragnaths (Figure 43 C1-C3; 3.1.4, Browne et al., 2005). *Ph awh* is expressed strongly in almost the entire medial part of the pre-antennal head (Figure 43 A1-A3; B1-B3). In the lateral pre-antennal hemispheres, three clusters of *Ph awh*-positive cells can be found in proximity to the medial expression domains (Figure 43 B2, upper red arrowhead). Additionally, a small group of cells expressing *Ph awh* is visible in the more peripheral part of the lateral pre-antennal hemispheres (Figure 43 B2, lower red arrowhead). In the tips of both labral lobes, strong *Ph awh* expression persists (Figure 43 A1', A2', C1-C3). The bases of the Mn appendages as well as the paragnaths show strong *Ph awh* expression (Figure 43 C1-C3). In all segments posterior to Mn, *Ph awh* expression is found at the base of the appendages. In thoracic and abdominal segments, *Ph awh*-positive cells line

the distal periphery of the coxal plates, where the coxae of the appendages connect (Figure 43 B1-B3, C3, D2, E1, E2). In stage 22, the ventral side of the thoracic and abdominal segments is intensively stained. However, such staining can also be observed in WMISH with some other probes, questioning whether this indicates specific *Ph awh* expression or whether it resembles unspecific staining caused by freshly shed cuticle that was not completely removed during dissection (compare Figure 43 B1 and Figure 78). The same considerations apply for the unilateral staining found in the left An1 and An2 appendages (Figure 43 D1).

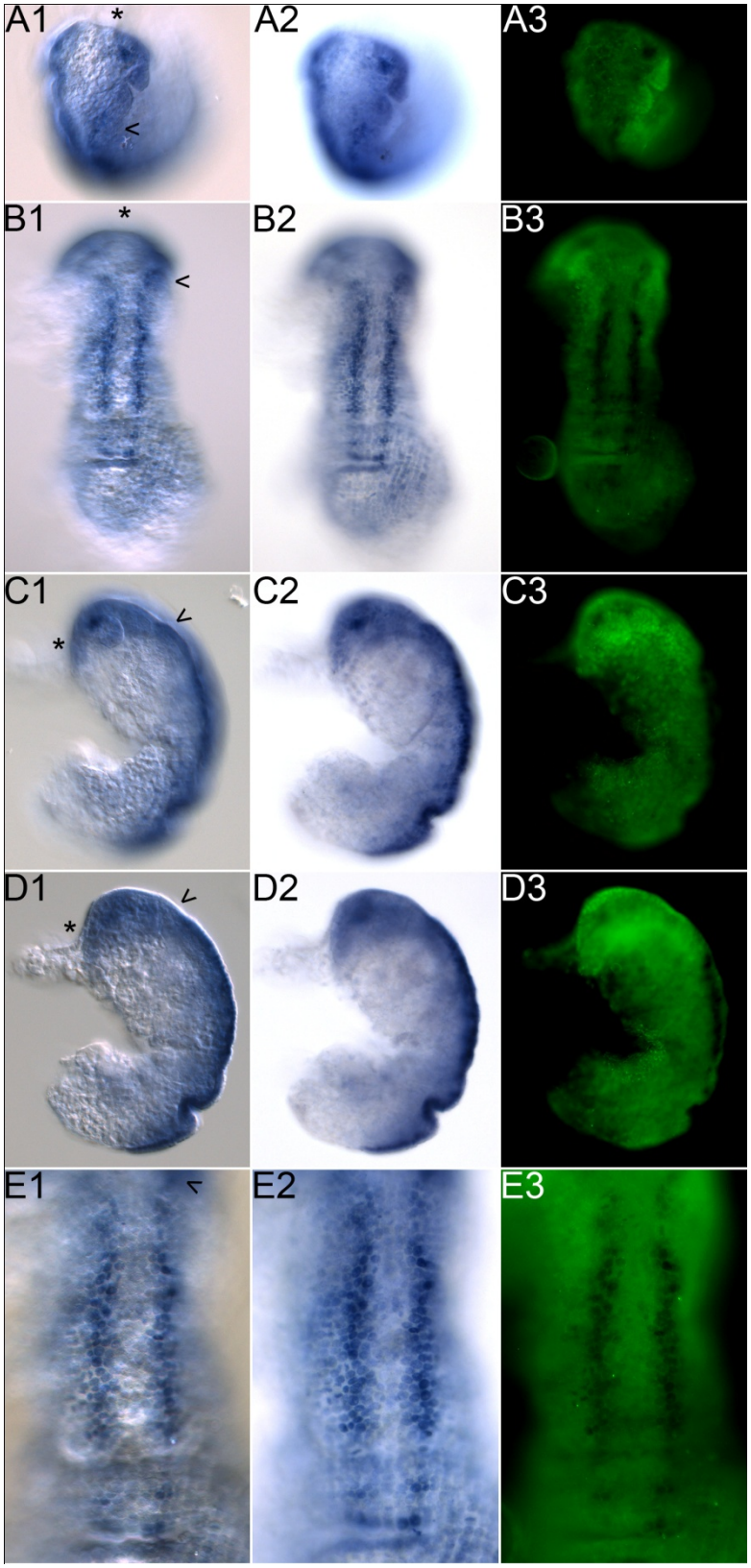



**Figure 41: Expression of *Ph awh* at stage S17**. The embryo is mounted in anterior, slightly tilted orientation (A1-A3), in ventral orientation (B1-B3), in lateral orientation, focal plane on the An1 and An2 appendages, (C1-C3, ventral is to the right) and lateral orientation, sagittal focal plane (D1-D3, ventral is to the right) and in ventral orientation, showing in detail a thoracic fraction of the germ band at higher magnification (x20; E1-E3) in order to facilitate better understanding of the three-dimensional pattern of *Ph awh* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – E3; nuclei are green, expression of *Ph awh* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph awh* is expressed bilaterally in clusters of cells that lie medially and slightly anterior to the base of the An1 appendages. **B** Starting from the Mn segment and extending until the posterior flexure, dispersed *Ph awh* expression is visible medially in two anterior-posterior stretches. **C** The focal plane is set on the An1 and An2 appendages. **E** Shown is in detail a thoracic fraction of the germ band at higher magnification (x20).

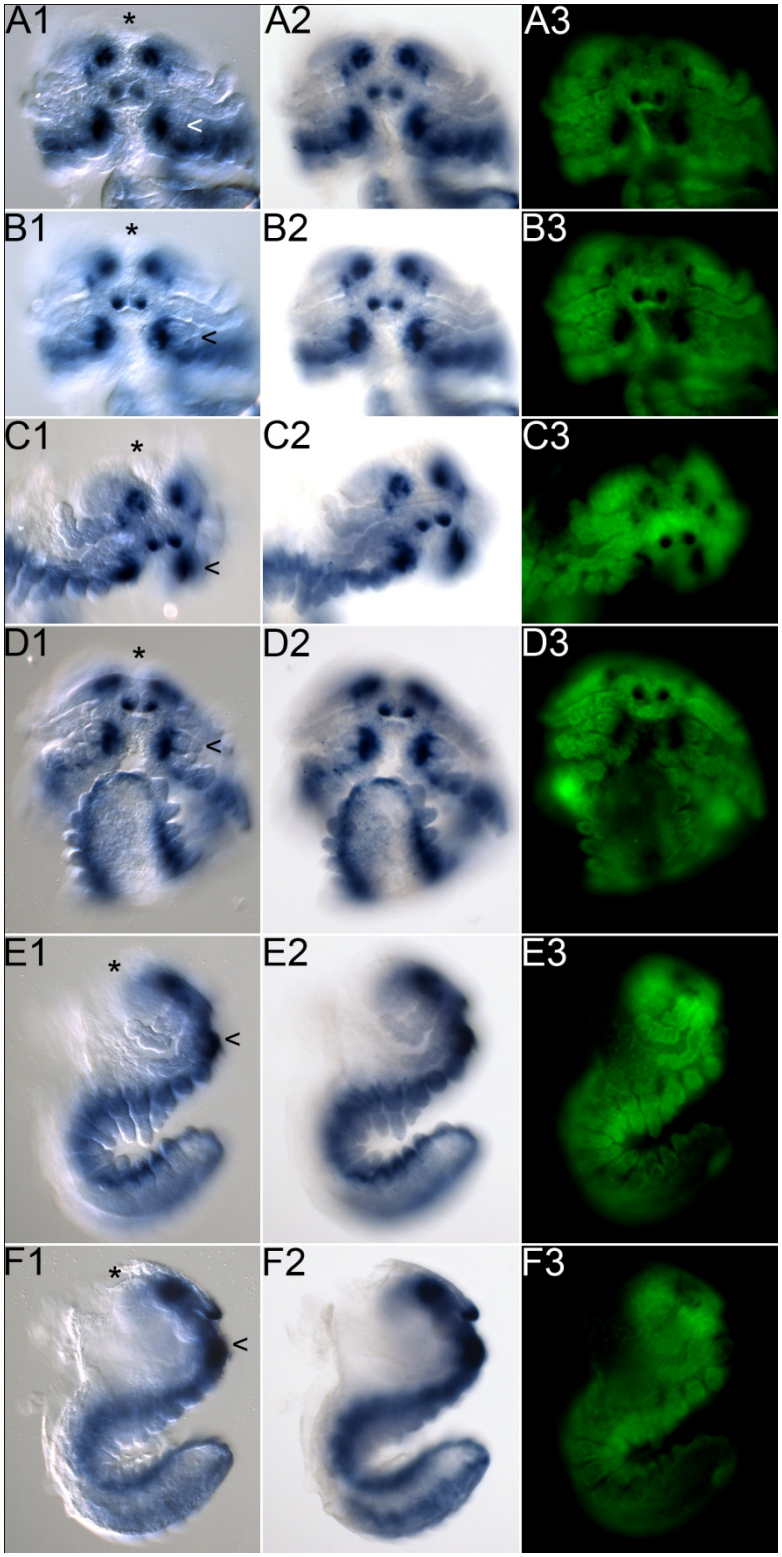



**Figure 42: Expression of *Ph awh* at stage S19**. The embryo is mounted in frontal orientation, focus set on the stomodeal opening (A1-A3), in frontal orientation, focus set on the labral lobes (B1-B3), in tilted anterior orientation (C1-C3), in ventral, native coiled posture (D1-D3), in lateral orientation, focal plane on the An1 and An2 appendages (E1-E3, ventral is to the right) and lateral orientation, sagittal focal plane (F1-F3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph awh* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-F1), bright field images in the middle panel (A2-F2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – F3; nuclei are green, expression of *Ph awh* is detectable as it quenches the nuclear labelling). In all DIC images (A1-F1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** The focal plane is set on the stomodeal opening. *Ph awh* expression is found at the base of the An1 appendages. **B** The focal plane is set on the labral lobes. *Ph awh* is strongly expressed in the tips of both labral lobes. **C** *Ph awh* expression is visible in the pre-antennal part of the embryonic head in broad, symmetric domains that appear to be composed of smaller clusters of *Ph awh* positive cells. **D** The embryo is mounted in its native, coiled posture. *Ph awh* is strongly expressed at the bases of the Mn appendage buds, reaching into the anterior part of the Mx1 segment. **E** The focal plane is set on the An1 and An2 appendages. From the Mn segment, extending until the posterior end of the embryo, *Ph awh* expression is visible in medial stretches that lie at the bases of the developing segmental appendages.

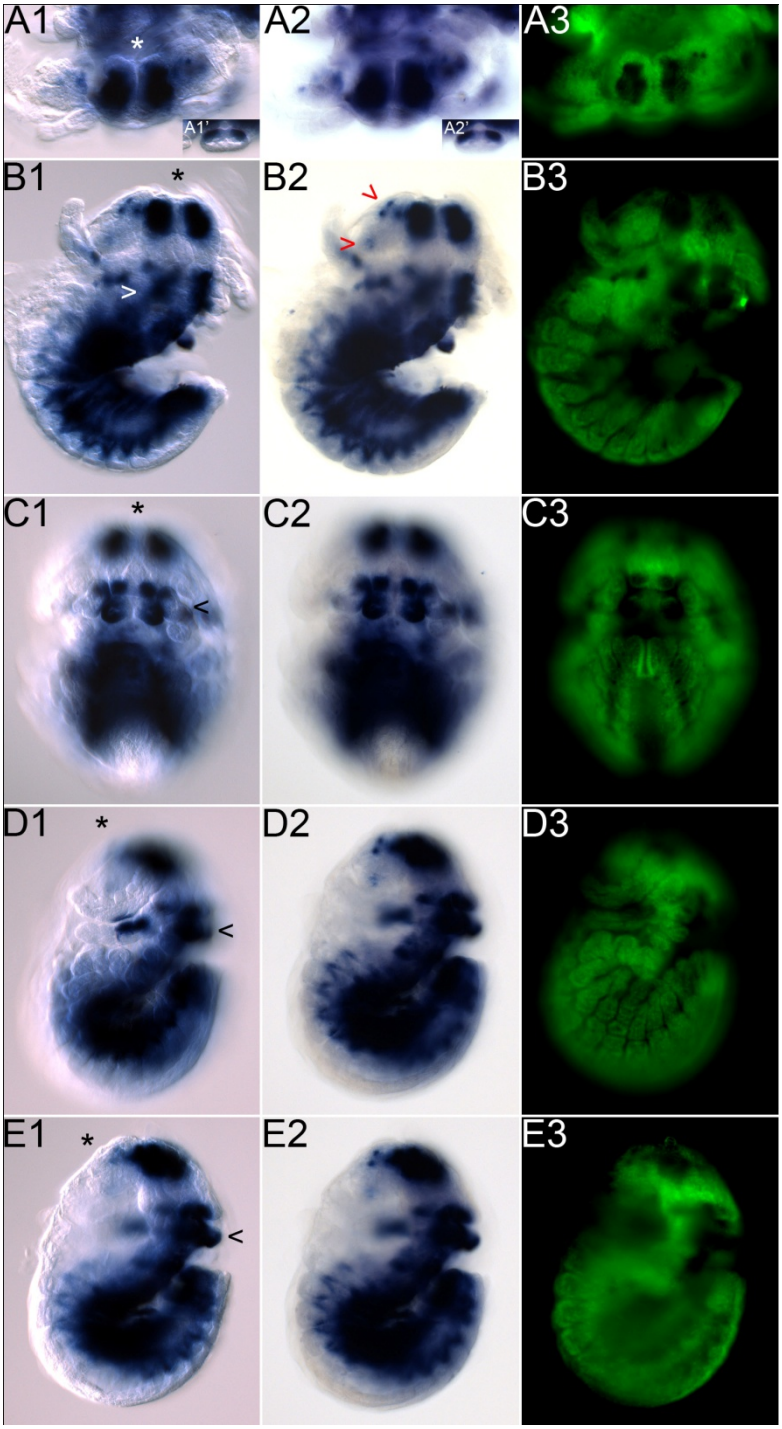



**Figure 43: Expression of *Ph awh* at stage S22**. The embryo is mounted in frontal orientation (A1-A3; in A1' and A2', focal plane is on the labrum), in tilted anterior orientation (B1-B3), in ventral, native coiled posture (C1-C3), in lateral orientation, focal plane on the An1 and An2 appendages (D1-D3, ventral is to the right) and in lateral orientation, sagittal focal plane (E1-E3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph awh* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – E3; nuclei are green, expression of *Ph awh* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph awh* is expressed strongly in almost the entire medial part of the pre-antennal head. In the tips of both labral lobes, strong *Ph awh* expression persists (A1', A2'). **B** In the lateral pre-antennal hemispheres, several clusters of *Ph awh*-positive cells can be found in proximity to the medial expression domains (B2, upper red arrowhead) and in a more peripheral area (B2, lower red arrowhead). **C** Ventral view. The embryo is mounted in its native, coiled posture. The bases of the Mn appendages as well as the paragnaths show strong *Ph awh* expression. **D** The focal plane is set on the An1 and An2 appendages. **E** In all segments posterior to Mn, *Ph awh* expression is found at the base of the appendages. In thoracic and abdominal segments, *Ph awh*-positive cells line the distal periphery of the coxal plates, where the coxae of the appendages connect.

## 3.3 *Parhyale hawaiensis sine oculis/Six* and *unplugged/gbx* genes

In many species, activity of *sine oculis/Six* (*so/Six*) genes arises early in development in a region anterior and directly adjacent to *otd/Otx* (e.g., Steinmetz et al., 2010). Genetic interaction and mutual regulation has been shown *Tribolium six3* and *otd1* (Posnien, 2009). These findings suggest that the combination of these genes might be an ancestral element of a conserved bilaterian mechanism of early anterior patterning and regionalisation. Similarly, activity of *unplugged/gbx* (*unpg/gbx*) genes arises posterior of *otd/Otx*-positive regions in the developing anterior head and brain regions of even distantly related species (Hirth et al., 2003; Raible and Brand, 2004; Steinmetz et al., 2011). Interaction of Otx2 and Gbx2 has been shown to be pivotal for establishing the midbrain/ hindbrain boundary in vertebrates (Raible and Brand, 2004). To address the question whether these genes might be ancestral bilaterian "neighbours" to *otd/Otx* genes with regard to patterning along the anterior-posterior axis, I cloned and analysed the *Parhyale* orthologs of *so/Six* and *unpg/gbx* genes in this work.

### 3.3.1 *Parhyale hawaiensis sine oculis/Six* genes

Since representatives of the *so/Six* class family of genes appear highly conserved throughout the animal phylogeny and share important functions in anterior head and brain development in many different species (Kawakami et al., 2000; Lavado et al., 2008; Oliver et al., 1995; Posnien et al., 2011; Seo et al., 1999; Steinmetz et al., 2010), I cloned and analysed *Parhyale hawaiensis six3* and *six4* as candidates for studies on *Parhyale* head development in this work.

#### 3.3.1.1 Isolation and characterisation of *Parhyale hawaiensis so/Six* genes

Because sufficient genomic sequence data is not currently available for *Parhyale hawaiensis*, the isolation of putative *so/Six* sequences was approached by performing varied-stringency degenerate PCR screens (5.2.3.1). To this end, degenerate primers based on comparative amino acid alignments of both Six domains (SD) and homeodomains (HD) of described Optix/Six3 proteins were used (5.2.4). Two *Parhyale hawaiensis so/Six* sequences,

*Parhyale hawaiensis optix/six3* (*Ph six3*) and *myotonix/six4* (*Ph six4*) were recovered from repeated varied-stringency degenerate PCR screens on a collection of *Parhyale* cDNA covering all stages of embryogenesis (5.2.2). The isolated fragments contained 217 bp of specific *Ph six3* and *Ph six4* sequence and were sufficient for unambiguous homology identification (5.3.3.1, see also A2.8, A2.9).

In detail, six sequences were recovered from repeated varied-stringency degenerate PCR screens. BLAST evaluation (5.3.3.1) revealed that three of them represented a novel *six3* sequence fragment and the remaining three s*ix4*, respectively. Within each group, obtained sequences are identical (less than 1% nucleotide variation, polymorphic nucleotide exchanges not considered). Since each sequence derives from a different priming event, they are considered independent.

Based on this primary sequence information, specific oligonucleotides were designed and used to prime *Ph six3* and *Ph six4* RACE reactions (5.2.3.3, A4.2, table A3). Consequently, the remaining mRNA transcript sequences both 5' and 3' of the initial fragments covering SD and HD were obtained.

#### 3.3.1.2 Isolation of *Ph six3*

From two-step 5' RACE, five independent *Ph six3* 5' cDNA fragment clones were obtained. The fraction of sequence they share is identical, whereas their 5' starting positions differ significantly. In order to determine the full *Ph six3* 5' cDNA sequence and, consequently, identify the actual transcription start of *Ph six3*, supplemental two-step 5' extension RACE was performed. Six additional, independent *Ph six3* 5' cDNA fragment clones were recovered, eventually providing complete 5' sequence information. Apart from the differences in length, all recovered *Ph six3* 5' RACE clones share sequence fractions that are considered identical (less than 1% nucleotide variation or less than 3 nucleotide exchanges, respectively, polymorphic nucleotide exchanges not considered; 5.3.3.2, see also A3, table A1).

*Ph six3* 3' cDNA sequence was completely recovered by two-step 3' RACE. Three independent clones covering identical, full 3' *Ph six3* transcript sequences were obtained. They are of same length, showing only minor differences in poly(A) tailing. They show only

minor nucleotide variations (less than 1% nucleotide variation, polymorphic nucleotide exchanges not considered).

Sequence information provided by all recovered 5' and 3' *Ph six3* RACE clones was sufficient for assembling the complete *Ph six3* cDNA sequence *in silico*. In order to verify the consistency of the established *Ph six3* cDNA sequence, coherent *Ph six3* ORF and cDNA fragments were isolated via long-distance PCR reactions (5.2.3.4), performed on *Parhyale* cDNA pools (5.2.2). Two independent *Ph six3* cDNA sequences and three independent *Ph six3* ORF sequences were obtained (A2.8.1). They are consistent with the findings from 5' and 3' RACE (A2.8.2) and are considered identical, since they vary in less than 1% of nucleotide positions (5.3.3.3, see also A3, table A1).

#### 3.3.1.3 Characterisation of *Ph six3*

The comparison of all recovered *Ph six3* sequences led to the identification of 66 sites of nucleotide exchange within the complete *Ph six3* cDNA, corresponding to 1.8% nucleotide variation. 42 of those occur uniquely and are therefore considered sporadic (5.3.3.2). They are found predominantly in the ORF (30) where they account for 20 amino acid changes (missense alterations) and one premature stop (nonsense alteration). Nine nucleotide exchanges found within the ORF do not alter the translated amino acid sequence (silent exchanges). These findings are concordant with the expected randomness of artificial nucleotide exchanges. The remaining 24 nucleotide exchanges are found in more than one independent sequence and are therefore considered polymorphic. They are predominant in the 3' UTR (16). In the ORF, the remaining eight polymorphic nucleotide exchanges do not alter the translated amino acid sequence. These findings corroborate the assumption that they represent naturally occurring polymorphisms of wild type *Ph six3* alleles (5.3.3.2).

Based on these findings, the sequence of the clone Ph_six3_cDNA09 was chosen as the source of the *Ph six3* reference sequence (*Ph six3*_ref). Sporadic and putatively artificial nucleotide exchanges were substituted with the 'consensus' nucleotides that are present in all other obtained sequences (754G>A, 2080G>A, 2886G>A). The sequence was extended 5' by adding 25 bp derived from clone Ph_six3_5R02 and 3' by adding 58 bp derived from clone Ph_six3_3R08 (A2.8.1). An Alignment of *Ph six3*_ref with all recovered *Ph six3* sequences

confirms its statistical consensus relevance (A2.8.2). *Ph six3*_ref has been used for phylogenetic studies.

The *Ph six3* transcript is 3.7 kb in length and encodes a protein of 543 amino acids. The *Parhyale* Six3 protein has a typical N-terminal Six domain (SD, 119 amino acids) with closest resemblance to Six3 SD of other species. Directly adjacent to the SD, *Parhyale* Six3 has a homeodomain (HD, 60 amino acids) with high sequence similarity to the HD of described Six3 proteins of other species (Figure 44, Figure 45, 3.3.1.6).

**Figure 44: Schematic view of *Ph six3* transcript.** The length of the *Ph six3* transcript is 3655 bases. Shown are: 5' UTR in grey (length 77 bases, nucleotide positions 1-77), ORF in black (length 1632 bases, encoding 543 amino acids; nucleotide positions 78-1709) and the 3' UTR in grey (length 1946 bases, nucleotide positions 1710-3655). The ORF region encoding the SD is depicted in orange (length 357 bases, encoding 119 amino acids; nucleotide positions 195-551). The ORF region encoding the HD is depicted in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 552-731).

**Figure 45 (next page): *Ph six3* cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph six3* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The SD is shown in orange and the HD in blue. The nucleotide sequences that encode these domains are shown in the respective colours. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

```
0001                GGAGTGCTGCAAGCGCCACCGACACGCTGGCTTCGAGCACCTCTCTGCCCGACCTTCTGCGCGCCCTGCAGCCCAAC

0001      M  A  L  L  A  G  L  S  A  Y  S  A  A  L  L  G  P  A  P  H  G  G  A  P  G  P  V  L  P  S
0078      ATGGCGCTCCTGGCGGGGCTGTCCGCGTACTCCGCCGCTCTCCTGGGCCCCGCACCCCACGGAGGGGCCCCAGGGCCCGTGTTGCCTTCC

0031       S  N  V  S  L  P  S  L  S  F  T  A  T  Q  V  A  A  V  C  E  T  L  E  E  S  G  D  I  E  R
0168      TCAAACGTGTCTTTACCTTCCCTATCCTTCACCGCGACACAAGTGGCAGCCGTGTGTGAAACATTAGAAGAAAGCGGTGACATTGAAAGG

0061       L  G  R  F  L  W  S  L  P  V  A  H  P  H  L  K  D  L  N  K  H  E  A  V  I  R  A  R  A  L
0258      TTAGGTAGATTTTTGTGGTCCCTCCCAGTAGCTCATCCTCATTTGAAAGACCTGAACAAGCATGAAGCAGTGATACGTGCCCGTGCCCTT

0091       V  A  F  H  L  G  N  F  R  E  L  Y  S  L  I  E  C  N  R  F  T  R  A  S  Y  P  R  L  Q  A
0348      GTTGCTTTCCACTTAGGTAATTTCCGTGAACTTTATTCCTTAATTGAGTGCAACAGATTCACCCGCGCGTCGTACCCTCGGTTGCAGGCC

0121       L  W  L  E  S  H  Y  Q  E  A  E  R  L  R  G  R  P  L  G  P  V  D  K  Y  R  V  R  K  K  F
0438      CTCTGGCTCGAGTCCCACTACCAGGAGGCAGAGCGCCTCAGAGGGAGGCCTTTGGGACCCGTCGACAAGTACCGCGTCAGGAAAAAGTTC

0151       P  F  P  R  T  I  W  D  G  E  Q  K  S  H  C  F  K  E  R  T  R  S  L  L  R  E  A  Y  L  Q
0528      CCGTTTCCGCGCACGATATGGGATGGGGAGCAGAAGTCCCACTGTTTCAAGGAACGCACGCGGTCCCTCCTGAGGGAGGCCTACCTGCAG

0181       D  P  Y  P  N  P  S  R  K  R  E  L  A  Q  A  T  T  L  T  P  T  Q  V  G  N  W  F  K  N  R
0618      GATCCGTACCCAAACCCCAGCAGGAAGCGGGAACTGGCTCAGGCCACGACCCTCACCCCCACGCAGGTCGGCAACTGGTTCAAGAACAGA

0211       R  Q  R  D  R  A  A  A  I  K  N  R  L  H  Q  Q  S  S  S  I  S  S  S  A  K  L  E  T  G  P
0708      AGACAGCGGGATAGAGCCGCCGCTATCAAGAACAGACTTCACCAGCAGAGTTCGTCAATATCGTCCTCTGCCAAACTGGAGACGGGGCCG

0241       R  S  P  G  A  D  D  D  E  D  D  L  I  N  P  G  S  P  T  P  F  E  D  D  V  S  E  D  E  V
0798      CGGTCGCCAGGAGCCGACGATGACGAGGACGACCTCATCAACCCAGGCTCTCCGACGCCGTTCGAAGACGACGTCTCAGAAGACGAAGTA

0271       S  L  G  Q  Q  S  P  L  S  K  L  G  Q  P  P  L  A  A  F  D  G  V  I  T  G  D  P  P  Q  H
0888      AGTCTTGGGCAGCAGTCTCCGCTCTCCAAACTGGGCCAGCCGCCCCTGGCTGCCTTCGATGGAGTTATCACAGGAGATCCTCCTCAACAC

0301       G  G  S  K  G  P  L  D  D  L  S  K  D  A  D  T  A  L  K  C  S  S  A  G  S  S  F  S  L  S
0978      GGGGGCTCCAAGGGACCGTTGGATGATTTAAGCAAAGACGCAGATACTGCCTTGAAGTGCAGCTCTGCGGGGAGTTCGTTTTCCCTGAGT

0331       E  G  A  S  S  D  T  S  N  S  A  F  S  S  P  Y  P  P  T  S  L  A  A  G  T  S  W  S  T  K
1068      GAAGGTGCCTCGTCTGACACTTCCAACAGCGCCTTCAGCAGCCCTTACCCTCCCACCAGTCTAGCAGCTGGCACTTCTTGGAGCACCAAA

0361       S  S  S  Q  E  I  F  L  P  A  S  P  V  P  L  P  P  D  R  L  F  R  I  N  P  D  G  D  H  S
1158      TCTTCCTCACAAGAAATTTTTCTGCCCGCCTCTCCAGTTCCTCTTCCTCCAGATCGCCTGTTTCGAATAAACCCCGACGGGGACCACAGT

0391       A  L  R  L  P  L  P  G  A  L  F  S  P  F  A  S  A  L  P  P  R  L  P  L  A  F  P  P  H  L
1248      GCCTTAAGGTTGCCTCTCCCAGGTGCACTATTTTCTCCCTTTGCATCAGCGCTGCCCCCCAGACTGCCTCTAGCGTTCCCTCCGCACCTC

0421       S  T  W  S  Q  Y  A  P  F  P  T  I  S  T  P  G  D  S  T  Y  H  R  S  F  P  A  P  L  S  I
1338      AGCACCTGGTCCCAATACGCTCCATTCCCCACCATCAGCACCCCTGGGGACAGTACCTACCACAGATCCTTCCCAGCACCTCTGTCCATC

0451       D  S  L  I  S  P  A  R  N  K  R  E  A  P  P  R  S  P  S  P  H  T  V  T  S  K  T  E  P  E
1428      GACTCGCTCATTAGCCCGGCACGAAATAAGAGAGAAGCTCCTCCTCGCTCTCCTTCCCCCCACACCGTTACGTCCAAGACGGAACCTGAG

0481       D  P  G  I  G  A  D  L  T  T  A  Q  G  D  G  P  S  L  R  S  P  T  Q  S  P  V  A  K  R  L
1518      GACCCTGGCATTGGTGCTGACCTGACCACCGCGCAGGGTGACGGGCCCTCGCTGCGGTCCCCCACACAGTCCCCAGTGGCCAAACGGCTC

0511       H  T  S  P  G  D  T  S  I  S  D  Q  Q  P  R  P  H  I  F  Q  P  I  K  L  E  F  A  P  S  T
1608      CACACAAGTCCAGGCGACACCAGCATCAGTGACCAGCAGCCTCGGCCCCACATTTTCCAGCCCATTAAACTAGAGTTCGCTCCCTCTACC

0541       E  T  S  &                                                                                0543
1698      GAGACCTCATGATTACATCTCAACATTTTTATTGTCTCCTATTAGCTCTCGATATTTATGATGTACATTAATTGTGTATACTTAAAGTTA
1788      TAGAATTATTAGACTAGCAGGAAATATTTCCTTAATTCTCTTGATCTTTATACATTTACAGTCATTATTTTTTTCCCTTGAATTTATTAG
1878      CGCAGGATGAAGAACTTGCTGTGGTATTTTTCTTAGTGAAATGTACAGGGAAGATAAAAATGAACAAATTGTCTCGAGCCTAAATATGAG
1968      AGTCTACGGAGCATTTAGCTCCAGACAGCATTACCCATATCTAATTGCATCTACCTTATTTATCAAAGGCTTATATTTCACTTTTTCTTA
2058      GTAACTGGAAACAATCTTCAGTAGGCAATCGAGAATATTTTTGTTAGTTAGTATTAAAACTGCAGGCTTCCATGTCTCTGAGAGGGTATG
2148      AAAGAACCTAAAGTATAGGGAATTTATTCCTGTGTGACAAAATTATTATCTCATGACATGAATCTTATGATCAGTTTATGTAACGCTTTG
2238      TGTACACACACGTGTTGTTATTCTCTCGGTTGAACCGCAGGTGGTGGGTGCATTATTTTCTTTTTTCCACACGACCGAAGTTTGAGTCGG
2328      TTGCTTGGCCTCTCTGCCTCATGCAGAGGTCACAGCATTAAGTAGATTCAGCAGTGGGAGGTATGACAACTTAGGGCTCGTTCCATTTGC
2418      TGGTCGATGTAGAACTCAGCGAAAATCAATCGAGCTTGACTTGTCTGGAACTGCAGATAGAATCTTTGGCTACGTCACTTTGAACCGCGT
2508      TGCAGGTGTTAAGCCCGGTGCCCACATAACTACTTTCGTAGCCGCTAGCAGGCTAGAGACAACTGAACTAATTCGAGGGCGTAGTCAAGG
2598      TCGTGACTTCAGCATGAACTCTGTGAACATCAAGCTATAATTCTGAAAATATCAAGCTTTTCTTACTTAAACTGCTCATAATTGGGAGCG
2688      ATTGTTGAAAGTAGCTACACCATTGCCGCTTGCAGTGGTGTAAAAATTGCAGTAGAGACTGAGATATTCGACAATTTTTTTTAGACAGGT
2778      AAATTTTGCGCAAGCTGAAGCCGAGCCGCGGCTAAGTAAGCGGATTAGTCGAGTTGCTTCTAAGCTGTTAGCAGCAACGAAACTTGTTTC
2868      GTGGGTACCGGGCTTTAGAAACAAGGCGATTATGTTGAGCTGACTGCAAACTCCATTTTGTTGAGTAAGATTAAGTTCTTGTGAAGTCTC
2958      AAAGGTAATCGATTAAGGGTCAACACCTGAAATTTGAATTAACCTATTAAAGTCAGCCCACATGGCTGCAGTTTCGGACAAGTTAAGCCT
3048      CAAAAACCTATTTTGGCCCATCAAACAAACGCGCTCTTAGTGGACTACTGAACTAATCCTAACTTGGCACGCCACGGGTACTGATATTGG
3138      TACAGTCGGCCTGATGAGGCCATAGCAAGCAGTCCGGCAGCAAATGCCACTGTCGCGCGGGGTGCATGTAGCCATGTGACGGTACCGATG
3228      TGGGGCTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGGGGGTAGCCATGTGGCTGCACCGATGTGTTAG
3318      GATATCATGATACATATGTGTGTGCTAGGTCGTCTGATTTGATTGGGTAATTAACTGCCGCGTGATTGATTGATTGCTTGATTGCCGTGT
3408      CAAAGAGGGATAATAATAATGGTGCAAATAATTAATACTTTTTTGTGCAAAAATAGTTCGCCCACAGATAGCTGCTCAGGTGGGCAAAGA
3498      CTGGGTACGAATCAAGTCTCAACTCGGTTGACCAACTCTGAAGTTGTTGCCATGACCAGCACCGGACCTAAACTGACTCTAAAATTCTCT
3588      TCTAACCCCAGCCGAACGTTGCCATTCGGATCCCCCCCACCCAAAAAAAAAAAAAAAAAAAAAAAAAAA                    3655
```

**Figure 45:** ***Ph six3*** **cDNA and derived amino acid sequence.**

#### 3.3.1.4 Isolation of *Ph six4*

From two-step 5' RACE, three independent *Ph six4* 5' cDNA fragment clones were recovered. The sequences of the clones Ph_six4_5R13 and Ph_six4_5R15 show identical transcription starts. The sequence of the clone Ph_six4_5R14 is shorter and appears truncated (A2.9). Within the fraction of sequence they share, the clones Ph_six4_5R13 and Ph_six4_5R14 differ in less than 1% of nucleotides. The sequence of clone Ph_six4_5R15 differs considerably in more than 3% of nucleotides from any of the other two clones' sequences. However, these findings do not support the existence of further *Ph six4* paralogs, suggesting that all *Ph six4* 5' RACE sequences derive from transcripts of different alleles of the same gene.

From single-step 3' RACE, supplemented by two-step 3' RACE, three independent clones covering full 3' *Ph six3* transcript sequences were obtained. Within the sequence they share, they vary in less than 1% of nucleotide positions and are therefore considered identical. They show minor differences in poly(A) tailing.

Sequence information provided by all recovered 5' and 3' *Ph six4* RACE clones was sufficient for assembling the complete *Ph six4* cDNA sequence *in silico*. In order to verify the consistency of the established *Ph six4* cDNA sequence, coherent *Ph six4* ORF fragments were isolated via long-distance PCR reactions (5.2.3.4), performed on *Parhyale* cDNA pools (5.2.2). Three independent *Ph six4* ORF sequences were recovered (A2.9.1). They are consistent with the findings from 5' and 3' RACE (A2.9.2) and vary in less than 1% of nucleotide positions (5.3.3.3, see also A3, table A1). These findings suggest that they derive from transcripts of the same gene.

#### 3.3.1.5 Characterisation of *Ph six4*

The comparison of all recovered *Ph six4* sequences led to the identification of 65 sites of nucleotide exchange within the complete *Ph six4* cDNA, corresponding to 2.4% sequence variation. All of those occur uniquely and are therefore considered sporadic (5.3.3.2). It is important to note that sequence information of more than three independent clones is only available for the *Ph six4* ORF. Therefore, the nature of nucleotide exchanges within the 5' and 3' UTRs remains speculative. Of the 39 nucleotide exchanges found within the *Ph six4* ORF,

18 nucleotide exchanges cause changes in the amino acid sequence (missense alterations) and 21 do not alter the translated amino acid sequence (silent exchanges). These observations do not suggest for any of the nucleotide exchanges within the *Ph six4* cDNA to represent putative allelic polymorphisms. For that reason, all nucleotide exchanges are considered sporadic and artificial with regard to the establishment of a *Ph six4* reference sequence.

Interestingly, the sequence of Ph_six4_5R14 has an addition of three nucleotides, compared to all other *Ph six4* sequences, leading to an additional serine in the amino acid sequence. The serine is found within a region of low complexity of the Six4 protein and adds to an existing stretch of serines. This suggests an allelic variation, similar to findings from gene transcript isoforms of other *Parhyale* genes (see also 4.1.1, 4.1.2).

Based on these considerations, the sequence of the clone Ph_six4_ORF15 was chosen as the source of the *Ph six4* reference sequence (*Ph six4_ref*). It was extended 5' by adding 593 bp derived from clone Ph_six4_5R13 and 3' by adding 569 bp derived from clone Ph_six4_3R24 (A2.9.1). Two nucleotides of the assembled Ph six4 cDNA sequence were changed to represent the statistical consensus (1891C>T, 2692G>A). Alignment of *Ph six4_ref* with all recovered *Ph six4* sequences confirmed its statistical consensus relevance (A2.9.2). *Ph six4_ref* has been used for phylogenetic studies.

The *Ph six4* transcript is 2.7 kb in length and encodes a protein of 504 amino acids. The *Parhyale* Six4 protein has an N-terminal SD (117 amino acids) with closest resemblance to Six4 SD. Directly adjacent to the SD, *Parhyale* Six4 has a HD (60 amino acids) with high sequence similarity to the HD of described Six4 proteins of other species (Figure 46, Figure 47, 3.3.1.6).

**Figure 46: Schematic view of *Ph six4* transcript.** The length of the *Ph six4* transcript is 2718 bases. Shown are: 5' UTR in grey (length 597 bases, nucleotide positions 1-597), ORF in black (length 1515 bases, encoding 504 amino acids; nucleotide positions 598-2112) and the 3' UTR in grey (length 606 bases, nucleotide positions 2113-2718). The ORF region encoding the SD is depicted in orange (length 351 bases, encoding 117 amino acids; nucleotide positions 1234-1584). The ORF region encoding the HD is depicted in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 1585-1764).

```
0001                                                ATTCCCCGGCATGGACGATGCGAGGAAAGTACGCTAGACACCGTAAACTG
0051          ACTCGGCCTGTATTGAATTTGATCTGCTAGTTGATTCGAAAGTGATATGATCCTGTTGCGCTTCTTGTCTGTCTTGCATGAGGTCTTCAT
0141          AGATAAAGTGGCGGAAAATGTGTTGTTGTAATAATGACTCATAGTTTTTGTGGAGTTCAGGGCTTCCGAAGTGTCGTGTTAATTTAGAAG
0231          TGTTTGTGTTGTGCGAGTTTTGAACTTTTGGTGTTACCGTTGGTGTTTGGTGCGAAAGCTATTCGAAGAAAGTGGGCTATGGAAACTGAC
0321          TTTACTGCTATTCAAAAATAGTGCTTATAATAAGCTGAATCCATTTACAACTGTAGATTTTGTTTTGTGGATTTTACGAAAAGTGTTTTT
0411          ATGAACTCAATCGAGACTCGGTAATTAAAATTACTTCGTATTTTATTTGTGTGTGTAGAAATATAATCTGAAACTACCATCTCCGAAGAG
0501          AATTTGAAAACTTTTTAAGATTTGTCTTTCTATGAATAATTTTTACTGAGTTCATGATTGAACTTTGTTTTTTAAATTGTATCAAGAAAA

0001          M  K  T  E  H  P  R  P  S  S  A  E  S  Y  G  S  F  S  I  P  S  P  D  G  L  S  T  P  N  P
0591          ATGAAAACGGAACATCCGCGGCCGTCGTCGGCGGAATCGTACGGAAGCTTCTCAATTCCTTCCCCCGACGGGCTTAGCACTCCTAACCCC

0031           R  L  P  H  Y  S  S  M  N  S  S  S  S  S  V  A  N  F  S  N  L  S  N  F  Q  S  T  A  S  N
0681          CGACTACCGCACTACAGCAGCATGAATTCTAGTTCCTCCAGCGTAGCTAATTTTTCAAACCTAAGCAATTTTCAAAGCACTGCGTCTAAC

0061           L  C  S  R  S  F  K  N  D  T  Y  T  P  P  P  H  I  G  M  G  M  S  P  M  S  S  Q  S  S  V
0771          CTTTGCAGTCGCTCCTTCAAAAACGATACCTACACTCCACCACCTCATATCGGCATGGGTATGTCGCCAATGTCTTCTCAATCTTCTGTA

0091           T  T  P  P  L  S  S  S  P  L  S  P  S  E  P  L  R  N  I  T  N  N  P  H  L  I  H  G  S  A
0861          ACAACGCCGCCTTTATCTTCGTCTCCTTTAAGCCCATCGGAACCGCTAAGAAATATAACAAATAACCCGCATTTGATTCACGGGTCCGCC

0121           A  S  P  F  H  Y  G  A  L  A  A  N  T  S  G  S  T  A  P  C  A  V  S  Q  S  T  S  S  S  C
0951          GCCTCGCCATTCCACTACGGGGCTTTAGCAGCCAACACGTCCGGCAGTACAGCCCCCTGCGCCGTCAGCCAATCCACTTCGTCCAGTTGC

0151           T  T  N  I  S  N  L  S  K  S  P  A  I  K  P  Q  P  T  S  P  H  I  A  T  V  T  P  E  P  H
1041          ACCACTAATATTTCCAATTTATCCAAGTCGCCGGCTATAAAACCACAGCCAACGTCCCCGCATATTGCTACCGTGACACCTGAACCCCAT

0181           H  R  E  S  P  S  L  S  T  P  K  K  T  P  S  S  S  A  S  P  S  S  P  S  S  S  P  V  K  T
1131          CACCGTGAATCACCGTCCCTATCAACCCCGAAGAAAACCCCCTCCTCCTCAGCTTCGCCAAGCTCCCCATCATCGTCCCCAGTAAAAACG

0211           S  Q  F  N  K  E  Q  I  D  C  I  C  E  V  L  I  Q  A  R  D  M  D  K  L  S  K  F  L  N  N
1221          TCCCAGTTCAACAAGGAACAAATAGACTGCATATGCGAGGTTCTCATACAGGCGAGAGATATGGACAAACTCTCTAAATTTTTAAACAAT

0241           L  P  A  S  H  F  A  T  D  A  V  S  E  V  I  L  R  A  K  V  E  V  A  F  A  K  G  N  Y  K
1311          TTACCGGCCTCGCACTTTGCAACCGATGCAGTTTCCGAGGTCATACTAAGGGCCAAAGTTGAAGTTGCTTTTGCGAAGGGCAATTACAAG

0271           E  V  Y  N  I  L  E  A  T  S  F  H  A  N  Y  H  T  H  L  Q  G  M  W  Y  K  A  H  Y  K  E
1401          GAAGTATATAATATCTTAGAGGCGACATCTTTCCATGCCAACTACCACACCCACTTGCAGGGCATGTGGTACAAGGCGCACTATAAAGAA

0301           A  E  R  V  R  Q  R  P  L  G  A  V  D  K  Y  R  I  R  K  K  Y  P  L  P  R  T  I  W  D  G
1491          GCCGAGAGAGTTCGTCAGAGACCTTTGGGTGCCGTCGACAAGTACCGCATACGCAAGAAATACCCGCTGCCGCGCACCATCTGGGACGGG

0331           E  E  T  I  Y  C  F  K  E  K  S  R  M  A  L  K  D  C  Y  R  Q  N  R  Y  P  T  P  D  E  K
1581          GAGGAGACCATCTATTGCTTCAAAGAGAAGTCCAGGATGGCGTTAAAAGACTGCTACCGACAGAACCGCTACCCGACCCCTGATGAGAAG

0361           K  T  L  S  R  Q  T  G  L  T  L  T  Q  V  S  N  W  F  K  N  R  R  Q  R  D  R  N  P  A  P
1671          AAGACTTTGTCCCGCCAGACCGGCCTCACACTCACCCAAGTCTCCAATTGGTTCAAGAACAGGCGGCAGAGGGACCGAAATCCTGCGCCT

0391           R  P  D  L  L  L  G  A  H  M  D  H  G  G  L  A  M  D  H  H  V  H  P  D  M  R  A  M  Y  H
1761          CGACCGGACCTTCTTTTGGGAGCGCACATGGATCACGGCGGTCTGGCCATGGACCATCACGTCCATCCTGACATGAGAGCCATGTACCAC

0421           H  A  K  M  A  A  H  A  Y  D  P  C  K  M  G  T  G  V  H  P  A  Y  D  P  N  N  V  M  C  M
1851          CACGCCAAGATGGCCGCCCATGCGTACGATCCGTGCAAGATGGGCACCGGCGTCCATCCAGCTTATGACCCCAACAACGTCATGTGCATG

0451           S  M  R  G  H  R  A  D  A  I  L  H  H  L  P  N  V  A  M  G  S  Y  P  Y  S  S  Y  D  L  Q
1941          AGCATGAGAGGTCATCGAGCTGATGCCATTCTACACCACCTCCCTAATGTGGCCATGGGTAGCTACCCCTATAGTAGCTACGACCTACAG

0481           S  V  A  A  T  H  G  N  S  H  Q  D  M  R  E  D  K  W  T  A  A  G  V  Q  #                0504
2031          AGCGTGGCGGCGACACACGGCAACTCTCACCAAGACATGCGGGAAGACAAGTGGACGGCCGCGGGAGTGCAATAATTGTTCGCCTCTCAA
2121          ATCTTACGAGCTTCTGACATTTCACGTACCATCGATTGGGACAGTTTCATGGAGACGAACGGATATGAGCTGAAACTTGTTCAGACTTTC
2211          ACTTTTAACCATGCAATTCGCTACTGTGGATCGAGCTTCAAGGGTTAATTCCAAATAGATACCCGCAAAGTTGAGGTTAAAGCCGTCTCA
2301          TGAATAATTTTTGTTTTGAAATATGTCTCTGAAACTCTATCTAAATAATTTTAATGAACGCTTAGTTATTGCTGAAGTTCTGGGTACGTT
2391          AAGGCAAATGAGAGTAGTTTGCCTATTGACACTGTCGCAATCGATATTACTTTCAATTGTGTATATGTTTTATAATGTACTTGTATAGTC
2481          TACCTAGCTATAATTGTCATAGGCAGCAGCTCCTTAGCATAGGATATCGACGTACAACCGAACGTGAATGTGAAAAAGTAGGTGGCCACG
2571          TCACAAGCCTTTTCATTGTGTACTGAGTTTTGTTAGTAATACCTTGCCGTCCATTGTATGCTGAAGAAAATTTAGGGTTTGTTTACGGCT
2661          CGATCACATAGATCTACATGTATCAAAAAAAAAAAAAAAAAAAAAAAAAAAA                                      2711
```

**Figure 47:** ***Ph six4*** **cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph six4* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The SD is shown in orange and the HD in blue. The nucleotide sequences that encode these domains are shown in the respective colours. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

#### 3.3.1.6 Phylogeny of *Parhyale hawaiensis so/Six* genes

BLAST searches (Altschul et al., 1997) using the complete amino acid sequence of *Ph* SIX3 as query against the protein databases *of Drosophila melanogaster* (blastp, taxid 7227) and mouse (blastp, taxid 10090) revealed highest sequence similarity to *Drosophila* OPTIX (*Dm* OPTIX) and murine SIX3 (*Mm* SIX3), respectively. For *Ph* SIX4, an analogous BLAST search yielded *Drosophila* SIX4 (*Dm* SIX4) and murine SIX4 (*Mm* SIX4). Independently, a BLAST search was performed using the complete amino acid sequences of *Parhyale* SIX3 and SIX4 as query against the complete translated nucleotide database (tblastn). From this search, a putative protein sequence from the nematode *Trichinella spiralis* (BAC clone TS195-5H12 from chromosome unknown, complete sequence; GenBank: AC190419.2) showed highest sequence similarity to *Ph* SIX3. A BLAST search using this putative *Trichinella spiralis* amino acid sequence as query against the protein database of *Drosophila melanogaster* (blastp, taxid 7227) revealed highest sequence similarity to *Dm* OPTIX. *Drosophila grimshawi* GH14784 (Dgri\GH14784, mRNA; NCBI Reference Sequence: XM_001984877.1) showed highest sequence similarity to *Ph* SIX4. A BLAST search of this protein sequence against the protein database *of Drosophila melanogaster* (blastp, taxid 7227) revealed highest sequence similarity to *Dm* SIX4.

Within the family of homeodomain transcription factors, proteins that possess a *sine oculis*/SIX domain and an adjacent HD (SIX-HD proteins) constitute a conserved subfamily of homeodomain transcription factors that divides further into three groups (Seo et al., 1999). Members of each of these SIX-HD protein groups are present in species as divergent as mouse, *Drosophila melanogaster* and *Nematostella vectensis* (Kawakami et al., 2000; Seo et al., 1999; Stierwald et al., 2004). In insects, a single factor typically represents each group, namely one ortholog of *Drosophila* SINE OCULIS (*Dm* SO), one of *Dm* OPTIX and one of *Dm* SIX4 (Seo et al., 1999). For this reason, I relate to these SIX-HD subgroups as SO-group, OPTIX/SIX3-group and SIX4-group in this work. However, the number of representatives belonging to any of those groups may vary, especially in non-insect species (Kawakami et al., 2000). In order to confirm the identities of *Ph* SIX3 and *Ph* SIX4 as representatives of the OPTIX and SIX4 groups within the family of SIX-HD proteins, the phylogenetic relation of the *Parhyale* proteins to known SIX-HD members of various other species was calculated (Mr Bayes, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The phylogenetic calculation was based on an alignment of *Ph* SIX3 and

*Ph* SIX4 together with the other protein sequences of choice, using the CLUSTALW2 algorithm (Larkin et al., 2007). Specifically, the full-length protein sequences of all included factors were used for calculating the phylogeny (A6.3). The resulting phylogram confirms the subdivision of the SIX-HD family into three groups. The SO group is well-defined, comprising a sub-group that contains the insect SO proteins, one that contains vertebrate Six1, Six2 and Six9 factors and - as an outgrouped factor to those branches - a diploblastic SO-ortholog (Figure 48). Similarly, all SIX4-factors that have been included in the phylogenetic calculation are grouped comprehensively, with the insect/arthropod SIX4-group containing *Ph* SIX4. Several small additional branches indicate that the vertebrate SIX5 factors, the nematode SIX4-like factors and the cnidarian SIX4 paralogs exhibit increasing levels of phylogenetic divergence (Figure 48). In contrast to the SO- and the SIX4 groups, the OPTIX/SIX3 proteins that were taken into the phylogenetic calculation here appear more divergent as they form several different subgroups: a subgroup that contains all insect OPTIX proteins together with one nematode SIX-HD factor, a subgroup containing vertebrate Six3 factors, a parallel subgroup consisting of vertebrate Six6 factors, and several individual OPTIX/SIX3 proteins of different species. In contrast to *Ph* SIX4, *Ph* SIX3 is not grouped within the insect/arthropod OPTIX subgroup, but instead is positioned remote to all other SIX-HD proteins (Figure 48). From the BLAST searches described above and the structure of the phylogram here, I conclude that *Ph* SIX3 is one factual ortholog of the insect OPTIX proteins, but exhibits a high level of derivation. This might suggest a considerable degree of evolutionary adaptation involving *Ph* SIX3, specific for the *Parhyale* lineage. In this work, no alternative *Parhyale six3/optix* gene was recovered. However, the number of representatives of this SIX-HD subgroup varies outside insect species. In addition, genomic sequence data of *Parhyale* was not available in this work. Therefore, additional *Parhyale* SIX3/OPTIX paralogs that are less derived from the respective insect factors might exist.

**A**

Subtree rooted at node 84:

```
                                                  /----- Cele_H-34
                  /-----------------94-----------------+
                  |                                     \----- Cele_H-33
                  |
                  |                                     /----- Mmus_Six5
                  |    /---------------89--------------+
                  |    |                                \----- Drer_Six4.3
                  |    |
             /-84-+    |                                /----- Mmus_Six4
             |    |    |                          /-100-+
             |    |    |                          |     \----- Drer_Six4.2
             |    |    |          /------100-----+
             |    |    |          |              \----------- Drer_Six4.1
             |    |    |          |
             |    \-96-+          |                        /----- Phaw_SIX4
             |         |          |              /--98-+
             |         |    /--50-+              |     \----- Amel_SIX4
             |         |    |     |    /---100---+
             |         |    |     |    |         \----------- Tcas_SIX4
        /--62-+        |    |     |    |
        |    |         |    |     |    |                    /----- Dmel_SIX4
        |    |         |    |     \-100+         /--97-+
        |    |         \-97-+          |         |     \----- Dpse_SIX4
        |    |              |          |    /-100+
        |    |              |          |    |    \----------- Dvir_SIX4
        |    |              |          \-100+
        |    |              |               \---------------- Agam_SIX4
        |    |              |
        |    |              \------------------------------- Nvec_SIX4.1
        |    |
        |    \-------------------------------------------- Nvec_Six4.2
        |
        |                                                  /----- Dmel_SO
        |                                            /--98-+
        |                                            |     \----- Dpse_SO
        |                                       /-100+
   --100+                                       |    \----------- Dvir_SO
        |                                  /-100+
        |                                  |    \---------------- Agam_SO
        |                             /-100+
        |                             |    |                    /----- Tcas_SO
        |                             |    \------100------+
        |                             |                     \----- Amel_SIX2
        |                             |
        |                             |                     /----- Drer_Six1b
        |                             |                /--99-+
        |                        /--91-+                |     \----- Drer_Six1a
        |                        |     |           /-100+
        |                        |     |           |    \----------- Mmus_Six1
        |                        |     |           |
        |                        |     |    /-97-+          /----- Drer_Six2.1
        |                        |     |    |    |    /-100-+
        \---------95---------+   |     |    |    |    |     \----- Mmus_Six2
                               |     \-58-+    \-100+
                               |          |         \----------- Drer_Zgc_195096
                               |          |
                               |          \--------------------- Drer_Six9
                               |
                               \------------------------------- Nvec_SO
```

Root part of tree:

```
/------------------------------------------------------- Phaw_SIX3
|
|------------------------------------------------------- Nvec_SIX3
|
|                                                  /------ Drer_Six6b
|                                            /--100-+
|                                            |      \------ Drer_Six6a
|     /-----------------100----------------+
|     |                                      \------------- Mmus_Six6
|     |
+     |                                                  /------ Drer_Six3a
|     |                                            /--50--+
|     |                                            |      \------ Mmus_Six3
|     |-----------------100----------------+
|     |                                      \------------- Drer_Six3b
|     |
|     |                                 /------------------- Agam_OPTIX
|     |                                 |
|     |                                 |                  /------ Dmel_OPTIX
\--98-+                          /--97--+     /--56--+
      |                          |     |      |      \------ Dpse_OPTIX
      |                          |     \-100-+
      |                     /--66-+           \------------- Dvir_OPTIX
      |                     |     |
      |                     |     \------------------------- Cele_H-32
      |               /--81--+
      |               |     |                            /------ Amel_OPTIX
      |               |     \-----------100-----------+
      |          /--87-+                               \------ Tcas_OPTIX
      |          |     |
      \--100-+     \-------------------------------------- (84)
             |
             \------------------------------------------- Drer Six7
```

**B**

```
/------------ Phaw_SIX3
|
|- Nvec_SIX3
|
|     /- Drer_Six6b
|    /+
|    |\- Drer_Six6a
|  /-+
|  | \- Mmus_Six6
|  |
|  |   / Drer_Six3a
|  |  /+
|  |  |\- Mmus_Six3
|  |--+
|  |  \--- Drer_Six3b
|  |
|  |              /- Agam_OPTIX
|  |              |
|  |              |              /- Dmel_OPTIX
+  |            /-+             /+
|  |            | |             |\- Dpse_OPTIX
|  |            | \-------------+
|  |         /--+               \-- Dvir_OPTIX
|  |         |  |
|  |         |  \---------------------- Cele_H-32
|  |      /--+
|  |      |  |   /------- Amel_OPTIX
|  |      |  \---+
|  |      |      \--------- Tcas_OPTIX
|  |      |
|  |      |                   /-------------- Cele_H-34
|  |      |               /---+
|  |      |               |   \---------- Cele_H-33
|  |      |               |
|  |      |               |         /------------------------ Mmus_Six5
|  |      |               |    /----+
|  |      |               |    |    \------------------------ Drer_Six4.3
|  |      |               |    |
\--+      |           /---+    |                  /---------- Mmus_Six4
   |      |           |   |    |              /---+
   |      |           |   |    |              |   \-------- Drer_Six4.2
   |      |           |   |    |     /--------+
   |      |           |   |    |     |        \----------- Drer_Six4.1
   |      |           |   |    |     |
   |      |           |   \----+     |            /-------------- Phaw_SIX4
   |      |           |        |     |         /--+
   |   /--+           |        |   /-+         |  \-------- Amel_SIX4
   |   |  |           |        |   | |     /---+
   |   |  |           |        |   | |     |   \-------------- Tcas_SIX4
   |   |  |         /-+        |   | |     |
   |   |  |         | |        |   | |     |               /- Dmel_SIX4
   |   |  |         | |        |   | \-----+              /+
   |   |  |         | |        \---+       |              |\- Dpse_SIX4
   |   |  |         | |            |       |      /-------+
   |   |  |         | |            |       |      |       \- Dvir_SIX4
   |   |  |         | |            |       \------+
   |   |  |         | |            |              \---- Agam_SIX4
   |   |  |         | |            |
   |   |  |         | |            \----- Nvec_SIX4.1
   |   |  |         | |
   |   |  |         | \--------- Nvec_Six4.2
   |   |  |         |
   |   |  |         |                       /-- Dmel_SO
   |   |  |         |                      /+
   |   |  |         |                      |\-- Dpse_SO
   |   |  |         |                 /----+
   |   |  \---------+                 |    \- Dvir_SO
   |   |            |           /------+
   \---+            |           |      \------ Agam_SO
       |            |     /----+
       |            |     |    |     /---- Tcas_SO
       |            |     |    \-----+
       |            |     |          \--------- Amel_SIX2
       |            |     |
       |            |     |      /- Drer_Six1b
       |            |     |     /+
       |            |  /--+     |\--- Drer_Six1a
       |            |  |  |  /--+
       |            |  |  |  |  \-- Mmus_Six1
       |            |  |  |  |
       |            |  |  |/-+   /- Drer_Six2.1
       |            |  |  || |  /+
       |            \--+  || |  |\-- Mmus_Six2
       |               |  \+ \--+
       |               |   |    \--- Drer_Zgc_195096
       |               |   |
       |               |   \---------- Drer_Six9
       |               |
       |               \-- Nvec_SO
       |
       \------- Drer_Six7

|-----------| 0.500 expected changes per site
```

**Figure 48: Phylogenetic analysis of *Parhyale* SIX3 and SIX4 together with SIX-HD proteins from other species.** Shown are: *Parhyale hawaiensis* SIX3 (Phaw_SIX3), *Parhyale hawaiensis* SIX4 (Phaw_SIX4), *Drosophila melanogaster* SINE OCULIS (Dmel_SO), *Drosophila melanogaster* OPTIX (Dmel_OPTIX), *Drosophila melanogaster* SIX4 (Dmel_SIX4), *Drosophila virilis* SINE OCULIS (Dvir_SO), *Drosophila virilis* OPTIX (Dvir_OPTIX), *Drosophila virilis* SIX4 (Dvir_SIX4), *Drosophila pseudoobscura* SINE OCULIS (Dpse_SO), *Drosophila pseudoobscura* OPTIX (Dpse_OPTIX), *Drosophila pseudoobscura* SIX4 (Dpse_ SIX4), *Anopheles gambiae* SINE OCULIS (Agam_SO), *Anopheles gambiae* OPTIX (Agam_OPTIX), *Anopheles gambiae* SIX4 (Agam_SIX4), *Tribolium castaneum* SINE OCULIS (Tcas_SO), *Tribolium castaneum* OPTIX/SIX3 (Tcas_OPTIX), *Tribolium castaneum* SIX4 (Tcas_SIX4),A*pis mellifera* SIX2 (Amel_SIX2), A*pis mellifera* SIX3 (Amel_SIX3), A*pis mellifera* SIX4 (Amel_SIX4), *Caenorhabditis elegans* H-32 (Cele_H-32), *Caenorhabditis elegans* H-33 (Cele_H-33), *Caenorhabditis elegans* H-34 (Cele_H-34), *Danio rerio* Six1a (Drer_Six1a), *Danio rerio* Six1b (Drer_Six1b), *Danio rerio* Six2.1 (Drer_Six2.1), *Danio rerio* Six3a (Drer_Six3a), *Danio rerio* Six3b (Drer_Six3b), *Danio rerio* Six4.1 (Drer_Six4.1), *Danio rerio* Six4.2 (Drer_Six4.2), *Danio rerio* Six4.3 (Drer_Six4.3), *Danio rerio* Six6a (Drer_Six6a), *Danio rerio* Six6b (Drer_Six6b), *Danio rerio* Six7 (Drer_Six7), *Danio rerio* Six9 (Drer_Six9), *Danio rerio* Zgc:195096 (Drer_Zgc_195096), *Mus musculus* Six1 (Mmus_Six1), *Mus musculus* Six2 (Mmus_Six2), *Mus musculus* Six3 (Mmus_Six3), *Mus musculus* Six4 (Mmus_Six4), *Mus musculus* Six5 (Mmus_Six5), *Mus musculus* Six6 (Mmus_Six6), *Nematostella vectensis* SO (Nvec_SO), *Nematostella vectensis* SIX3 (Nvec_SIX3), *Nematostella vectensis* SIX4.1 (Nvec_SIX4.1) and *Nematostella vectensis* SIX4.2 (Nvec_SIX4.2). No outgroup was set. For reference regarding the non-*Parhyale* protein sequences, see A6, tables A5 and A6. **A** Phylogram showing Clade credibility values. **B** Phylogram based on average branch lengths. Phylogenetic Analysis was done using Mr Bayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; 5.3.2; 750,000 generations).

#### 3.3.1.7 Expression of *Ph six3*

Expression of *Ph six3* is first detectable in embryos of stage 11. At this stage, initial ectodermal and mesodermal cell rows form at the posterior end of the germ band (Browne et al., 2005). The anterior of the embryo is clearly composed of two separate head lobes. They cover the future first and second antennal segments as well as the pre-antennal region of the head including the ocular segment (Figure 49 A2; 3.1.1; Browne et al., 2005). *Ph six3* is expressed bilaterally in single, medially located cells within the pre-antennal part of the embryonic head region (Figure 49 A1, A2).

At stage 12, the number of cells that each head lobe comprises has increased (Figure 49 B2). *Ph six3* expression is now found in small groups of cells at the same relative position as during the previous stage (Figure 49 B1, B2).

At the onset of stage 14, the medial cleft that separates both head lobes at the level of the An1 and An2 segments is still visible. The ectodermal grid that characterises the progressing segmentation of the post-mandibular segments extends the full length of the ventral surface of the egg (Figure 50 A3, B3, C3, D3, 3.1.2, Browne et al., 2005). Bilateral *Ph six3* expression remains in small groups of cells that lie medial within each head lobe. This position is just anterior to the An1 appendage anlagen which show a relatively high density of nuclei (Figure 50 A1-A3, B1-B3).

At the onset of stage 16, the fusion of the head lobes is complete (Figure 51 A3, B3; 3.1.2; 3.1.3; Browne et al., 2005). The pre-antennal, laterally projected hemispheres begin to undergo significant cell proliferation. The cells become smaller, more densely packed and have point-shaped nuclei. Arched ridges reaching from the medial area to the lateral hemispheres show the highest density of nuclei, indicating multiple layers of cells (Figure 51 C3, 3.1.3). Like in previous stages, the dorsal organ is visible as a ring of cells located centrally at the dorsal side of the embryo (Figure 51 A3, white arrowhead, Browne et al., 2005). *Ph six3* expression is found in cells that form a coherent medial stripe. The stripe extends from the posterior part of the second antennal segment to the anterior end of the embryo where it branches out laterally to both sides (Figure 51 B1-B3, C1-C3). The anterior-most medial cells are free of *Ph six3* expression (Figure 51 A1-A3, B1-B3). Within the more lateral regions of the pre-antennal hemispheres, very slight *Ph six3* signal can be detected bilaterally in small groups of cells (Figure 51 A1, A2, red arrowheads).

In stage 18 embryos, the ventral flexure of the embryo is fully established (Figure 52 E1-E3, F1-F3, Browne et al., 2005). The stomodeum begins to invaginate. The length of the second element of the An2 appendage has increased relative to the proximal first element (Figure 52 C1, 3.1.3, Browne et al., 2005). *Ph six3* is expressed strongly in a medial, tri-lobed domain that encompasses the medial portion of the future stomodeal field at the border of the An1 and An2 segments and reaches until the anterior brink of the embryo. It entirely covers the labrum anlagen (Figure 52 B1-B3, C1-C3, D1-D3, F1-F3, see also Figure 5). Additionally, *Ph six3* is expressed bilaterally in groups of cells at the lateral periphery as well as in scattered cells throughout the anterior periphery of the pre-antennal lobes (Figure 52 A1-A3, E1-E3). Small clusters of cells at the bases of the An2 appendages also show weak *Ph six3* expression. A very slight *Ph six3* signal is found within the proximal parts of the Mn appendages (Figure 52 C1-C3).

At stage 20, the lower edge of the stomodeum is visible as a raised medial ridge just posterior of the labrum. The An2 appendage is undergoing rapid elongation (Figure 53 B3, C3, 3.1.4, Browne et al., 2005). Strong *Ph six3* expression is maintained. In contrast to stage 18, the posterior part of the tri-lobed medial *Ph six3* expression domain appears slightly condensed, giving the domain a heart-shaped appearance. A population of *Ph six3* expressing cells is found beneath this *Ph six3* expression domain, internally within the embryonic head. These cells are located within the roof of the foregut (Figure 53 B1-B3). Bilateral expression of *Ph six3* in the lateral periphery of the pre-antennal hemispheres has increased slightly as compared to stage 18 (Figure 53 B1-B3). Scattered single cells expressing *Ph six3* are few, but still found in the anterior periphery of the pre-antennal hemispheres (Figure 53 A2, B2). *Ph six3* signal at the base of the An2 appendages has decreased as compared to stage 18. In contrast, the intensity of *Ph six3* signal at the base of the Mn appendages has increased (Figure 53 B1-B3, C1-C3).

In stage 22 embryos, the labrum has completed its posterior extension. Its lobes directly abut the paragnaths anterior, with the mandibles prominently protruding laterally on both sides of the labrum (Figure 54 B3, C3, D1; 3.1.4, Browne et al., 2005). In the medial part of the pre-antennal head, strong *Ph six3* expression is maintained (Figure 54 A1-A3, B1-B3). *Ph six3* expression is found in the labrum, predominantly within the fused labral lobes (Figure 54 B1-B3, C1-C3, D1-D3). Cells that comprise the roof of the foregut continue to express high levels of *Ph six3* (Figure 54 D1, red arrowhead; D2, D3). In the centres of the lateral pre-antennal hemispheres, the future medullae, small dot-like *Ph six3* expression domains are

visible (Figure 54 A1-A3). In embryos of this stage, no expression of *Ph six3* was detectable in the appendages of An1, An2, Mn and Mx1 (Figure 54 C1-C3).

At stage 23, the substructure of the protocerebral neuromere has become morphologically distinct. The medial hemi-ellipsoid bodies and the lateral medullae clearly delineate from each other in nuclear labelled images. Additionally, tergites and coxal plates become visible in thoracic and abdominal segments (Figure 55 A1-A3, D3, 3.1.4, Browne et al., 2005). In the medial pre-antennal head, expression of *Ph six3* remains strong and encompasses the entire hemi-ellipsoid bodies (Figure 55 A1-A3, B1-B3). *Ph six3* is expressed in all cells of the labrum except the most distal ones in the tips of the fused labral lobes (Figure 55 B1-B3). As in stage 22 embryos, expression of *Ph six3* is found in the arched foregut that has expanded farther into the head (Figure 55 E1, red arrowheads, E2, E3). Within the medullae, *Ph six3*-positive cell clusters form ring-shaped expression domains (Figure 55 A1-A3). The Mn appendages, the paragnaths and the proximal appendage articles of the first and second maxillae as well as the maxillipeds (T1) show dot-like *Ph six3* expression domains (Figure 55 B1-B3, C1-C3, D1-D3).

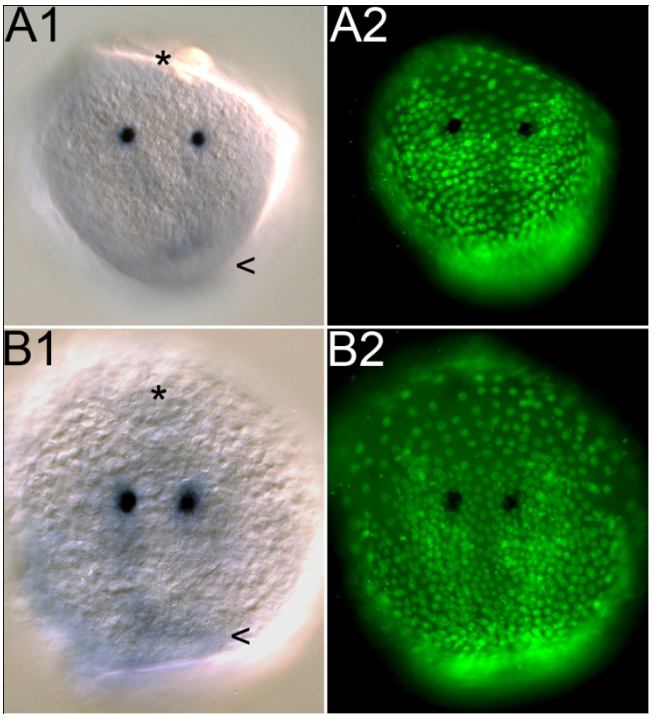



**Figure 49: Expression of *Ph six3* at stages S11 and S12**. An embryo of stage S11 (A1, A2) and an embryo of stage S12 (B1, B2) are mounted in frontal orientation. In all images, anterior is up. Shown are DIC images in the left panel (A1, B2) and nuclear labelling of the same embryos (SYTOX®) in the right panel (A2, B2; nuclei are green, expression of *Ph six3* is detectable as it quenches the nuclear labelling). In all DIC images (A1, B1), the anterior end of the embryo is indicated by a black asterisk and the position of the future Mn segment is marked by a black arrowhead.

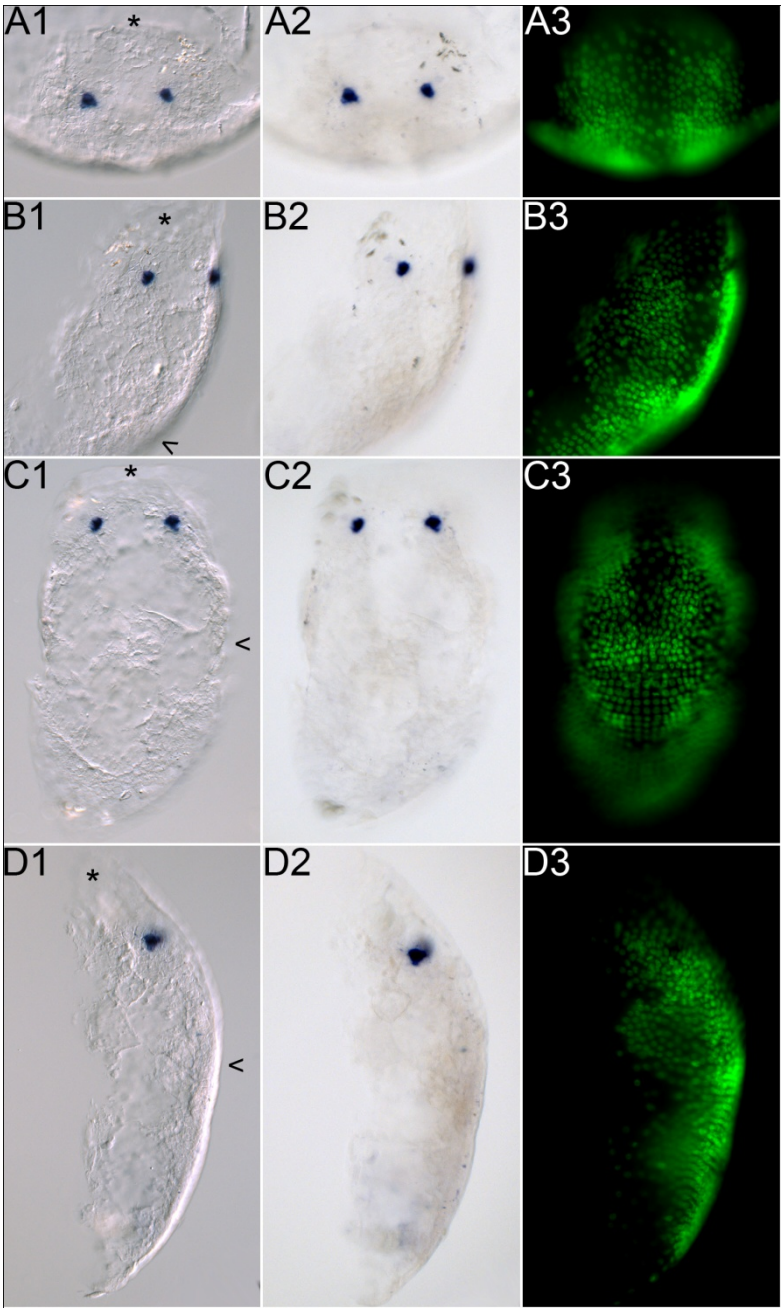
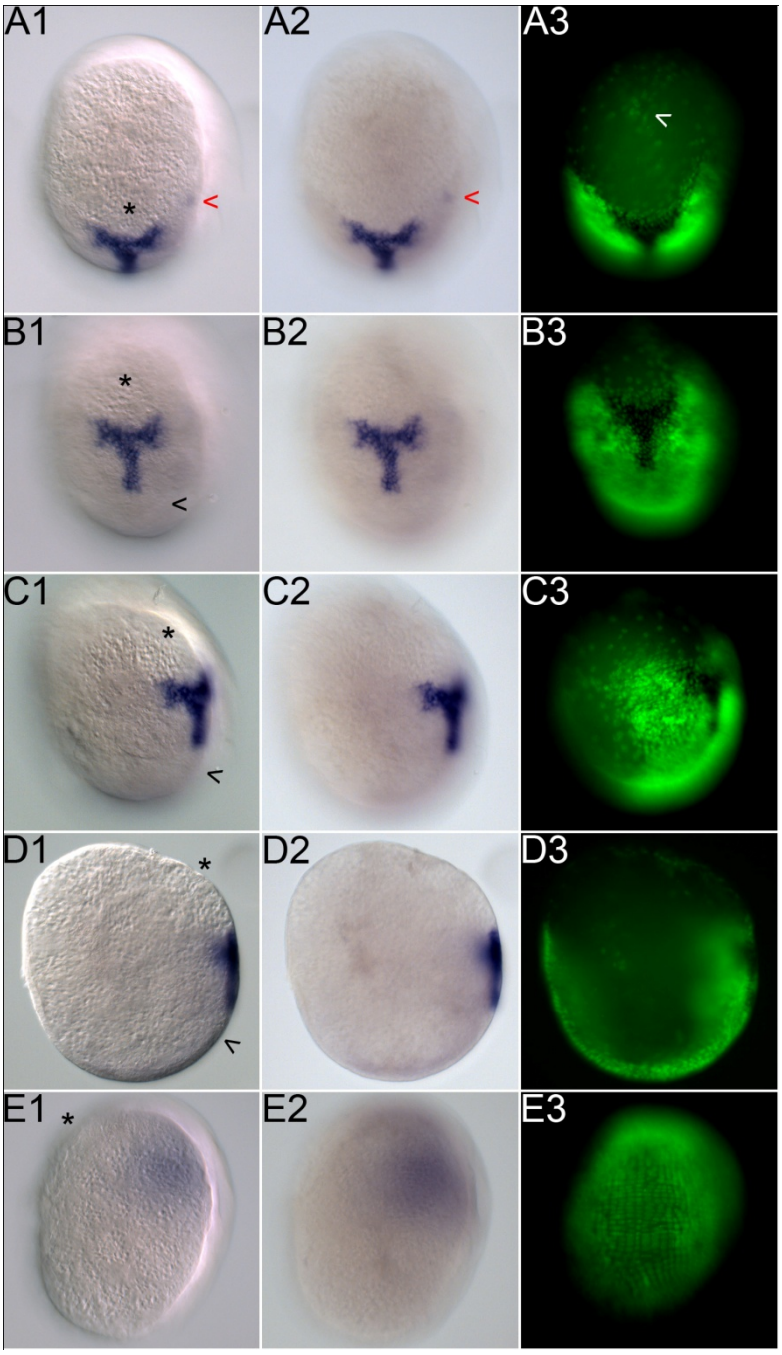



**Figure 50: Expression of *Ph six3* at stage S14**. The embryo is mounted in anterior orientation (A1-A3), anterior, tilted orientation (B1-B3), ventral orientation (C1-C3) and lateral orientation (D1-D3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph six3* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2-D2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–D3; nuclei are green, expression of *Ph six3* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the future position of the Mn segment is marked by a black arrowhead.




**<u>Figure 51:</u> Expression of *Ph six3* at stage S16**. The embryo is mounted in dorsal orientation (A1-A3), anterior orientation (B1-B3), anterior, tilted orientation (C1-C3), lateral orientation, sagittal focal plane (D1-D3, ventral is to the right) and posterior ventral orientation (E1-E3) in order to facilitate better understanding of the three-dimensional pattern of *Ph six3* expression. In all images except A1-A3, anterior is up. In the images A1-A3, the posterior end of the dorsal side of the embryo is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–E3; nuclei are green, expression of *Ph six3* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and, where appropriate, the position of the Mn segment is marked by a black arrowhead. **A** Red arrowheads indicate *Ph six3*-expressing cells in the lateral part of the left pre-antennal hemisphere.

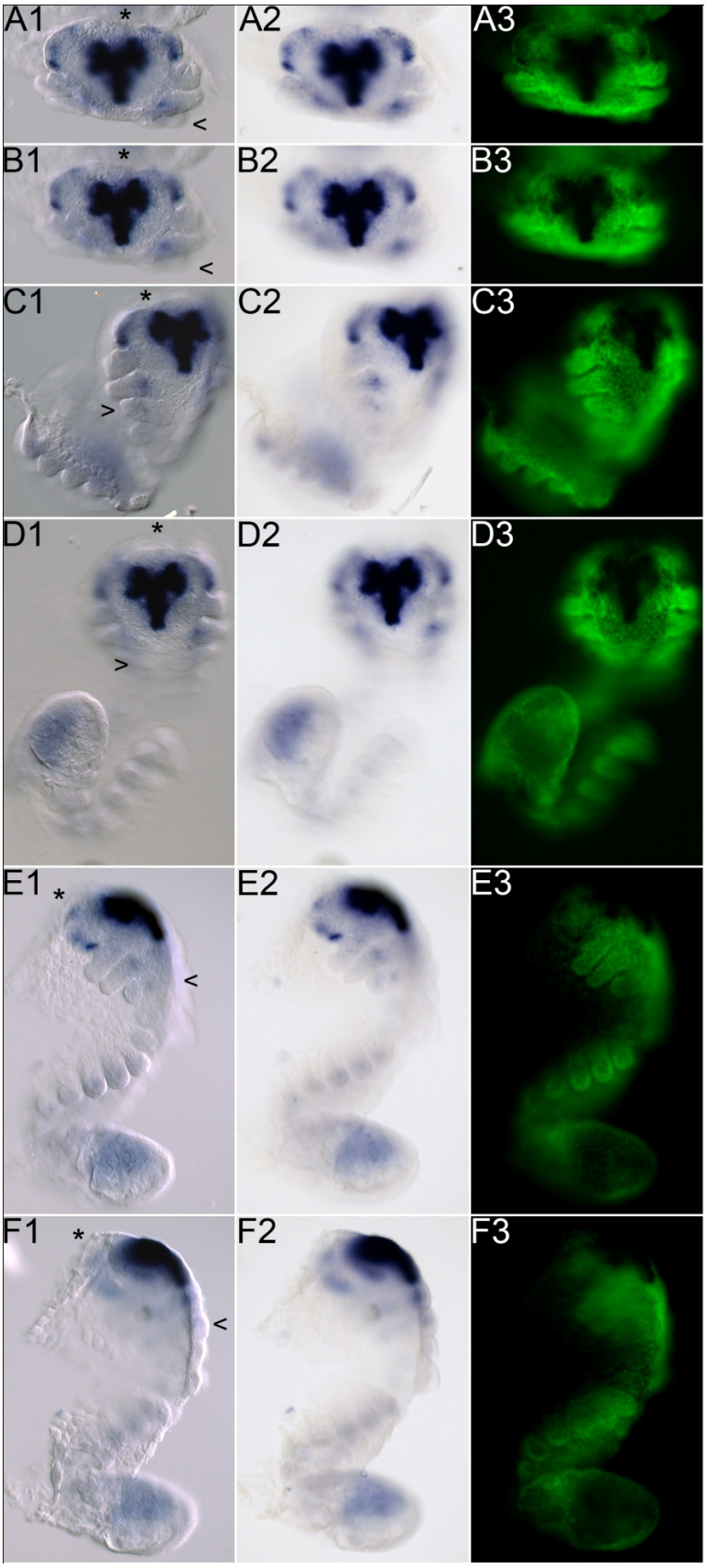



**Figure 52: Expression of *Ph six3* at stage S18**. The embryo is mounted in anterior orientation, focal plane on the lateral parts of the pre-antennal head (A1-A3), in anterior orientation, focal plane on the medial part of the pre-antennal head (B1-B3), in anterior, tilted orientation (C1-C3), in ventral orientation (D1-D3), in lateral orientation, focal plane on the An1 and An2 appendages (E1-E3, ventral is to the right) and in lateral orientation, sagittal focal plane (F1-F3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph six3* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-F1), bright field images in the middle panel (A2-F2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–F3; nuclei are green, expression of *Ph six3* is detectable as it quenches the nuclear labelling). In all DIC images (A1-F1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead.

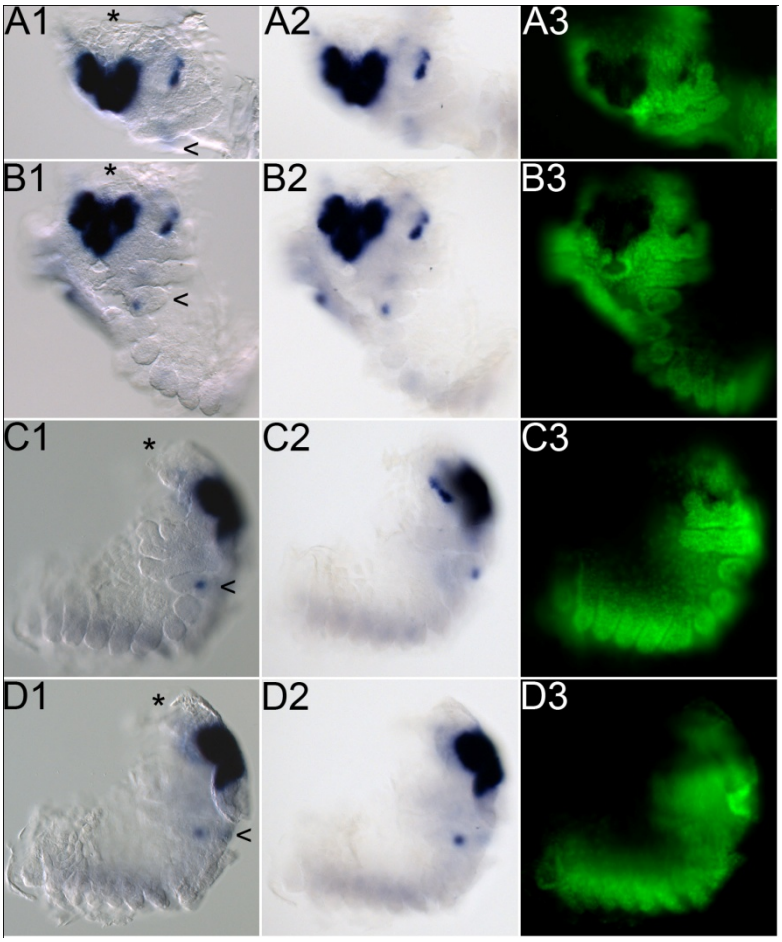



**Figure 53: Expression of *Ph six3* at stage S20**. The embryo is mounted in anterior, slightly tilted orientation (A1-A3), in ventral, tilted orientation (B1-B3), in lateral orientation, focal plane on the An1 and An2 appendages (C1-C3, ventral is to the right) and in lateral orientation, sagittal focal plane (D1-D3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph six3* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2-D2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–D3; nuclei are green, expression of *Ph six3* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead.

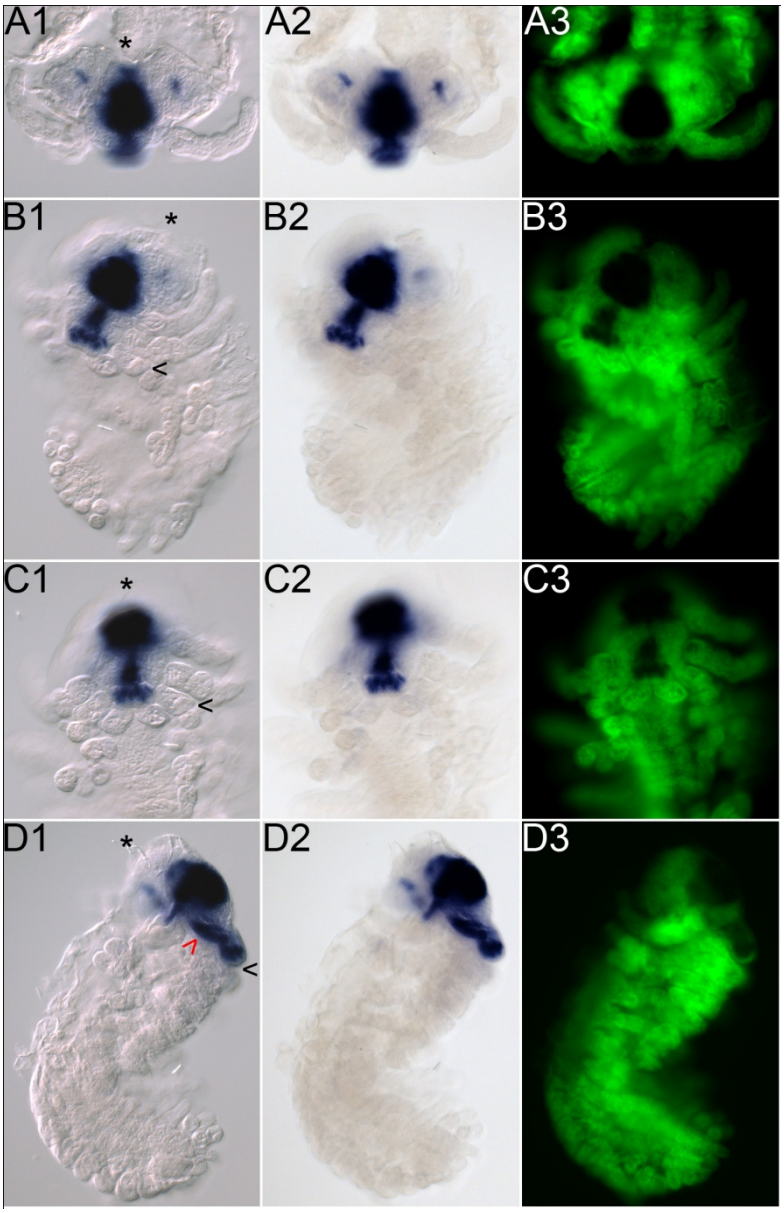



**Figure 54: Expression of *Ph six3* at stage S22**. The embryo is mounted in anterior orientation (A1-A3), in anterior, tilted orientation (B1-B3), in ventral orientation, focal plane on the tip of the labrum and the paragnaths (C1-C3) and in lateral orientation, sagittal focal plane (D1-D3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph six3* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2-D2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–D3; nuclei are green, expression of *Ph six3* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **D1** A red arrowhead marks *Ph six3*-expressing cells in the roof of the foregut.

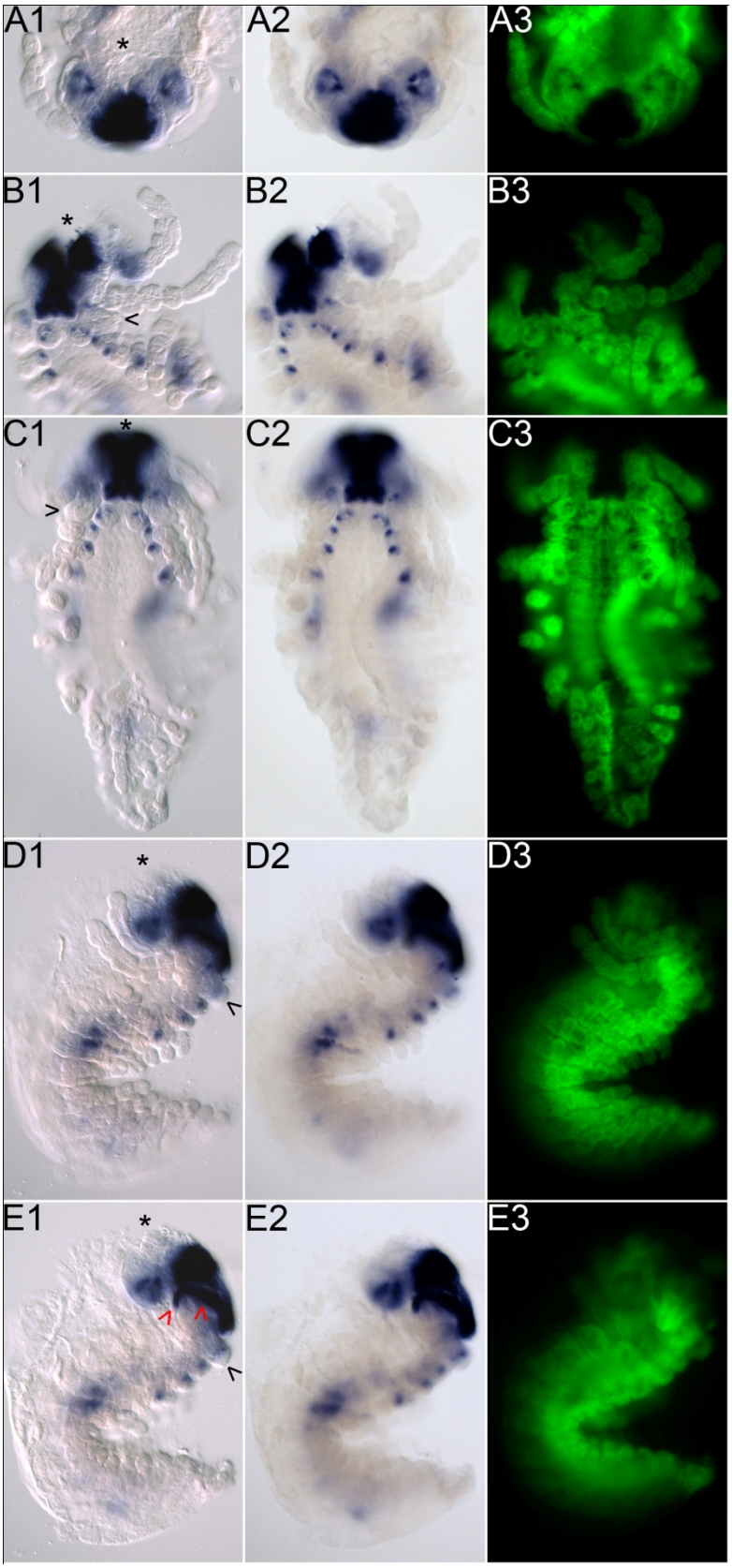



**Figure 55: Expression of *Ph six3* at stage S23**. The embryo is mounted in anterior orientation (A1-A3), in anterior, tilted orientation (B1-B3), in ventral orientation, focal plane on the tip of the labrum (C1-C3), in lateral orientation, focal plane on the An1 and An2 appendages (D1-D3, ventral is to the right) and in lateral orientation, sagittal focal plane (E1-E3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph six3* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–E3; nuclei are green, expression of *Ph six3* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **E1** Red arrowheads mark *Ph six3* signal found in the arched foregut.

#### 3.3.1.8 Expression of *Ph six4*

Expression of *Ph six4* is first detectable in embryos of stage 12. This stage is characterised by the initial elongation of the germ band (Figure 56 C2, Browne et al., 2005). The head lobes cover the pre-antennal head region as well as the An1 and An2 segments and are morphologically separated by a medial cleft (Figure 56 A2, B2, 3.1.1, Browne et al., 2005). *Ph six4* is expressed weakly in cells of the An2 appendage fields (Figure 56 A1, A2). *Ph six4* expression is also found in cells that are grouped medially at the base of the branching point of the head lobes within the future Mn segment (Figure 56 A1, A2; B1, B2).

During stage 14, the medial cleft that separates both head lobes is gradually replaced by cells. The head lobes begin to fuse (Figure 57 A2, B2; 3.1.2; Browne et al., 2005). *Ph six4*-positive cells are scattered on both sides of the closing medial cleft (Figure 57 A1, A2), starting from a bi-lobed expression domain in the Mn segment and extending to the anterior end of the embryo (Figure 57 A1, A2; B1, B2; C1, C2).

In stage 15 embryos, the medial cleft that separates both head lobes during earlier stages has closed. It is replaced by a regular layer of cells. The anterior border of the embryo appears as an almost vertical line (Figure 58 A3, B3; 3.1.3; Browne et al., 2005). Strong *Ph six4* expression is found throughout the head in a bilateral pattern. Two types of *Ph six4* expression domains can be distinguished. The first type is found in the pre-antennal region of the head. Here, the hemispheric topology of *Ph six4* expression encompasses a medially and anterior located small group of cells and, in relation to it, a lateral, larger group of cells (Figure 58 A1-A3, B1-B3, C1-C3). Directly abutting posterior, *Ph six4* is expressed in an almost rectangular sheet of cells, in which two small clusters of cells, aligned in an anterior-posterior orientation, remain free of *Ph six4* expression. A population of cells that is located between both pre-antennal hemispheres is free of *Ph six4* expression (Figure 58 A1-A3, B1-B3). The second bilateral *Ph six4* expression domain covers the segments An1, An2, Mn and Mx1. *Ph six4* is expressed coherently in broad stripes from anterior to posterior, giving this domain a V-shaped appearance. In detail, *Ph six4* expression is found in all cells of the An1, An2 and Mn appendage buds as well as in cells that lie between them (Figure 58 A1-A3, B1-B3). Importantly, the most posterior cells expressing *Ph six4* are found in the Mx1 segment. They are aligned in a regular grid indicating that these cells are undergoing the process of trunk segmentation. This is typical for the elongating germ band during this stage (Figure 58 D1-D3). Cells that are located medially to the *Ph six4* expressing cells of the segments An1 to

Mx1 are free of *Ph six4* expression and form a roughly triangular-shaped cellular sheet (Figure 58 B1-B3). No *Ph six4* expression is found in segments that lie posterior to Mx1 (Figure 58 D1-D3, E1-E3).

In stage 20 embryos, the pre-antennal hemispheres appear oval-shaped and expand dorsally and laterally. The labrum is elongating and begins to cover the stomodeal opening (Figure 59 A1-A3, E3; 3.1.4; Browne et al., 2005). *Ph six4* is expressed in a medial domain that covers the anterior brink of the embryo and extends posterior into the base of the labrum (Figure 59 A1-A3). It directly abuts the anterior limit of the stomodeum. The stomodeum itself is free of *Ph six4* expression (Figure 59 B1-B3). Additionally, *Ph six4* is expressed bilaterally in the central part of the pre-antennal hemispheres within several cell groups of different size (Figure 59 A1-A3, B1-B3). *Ph six4* expression is found within the basal articles of each segmental appendage in all cells that form the core of these articles (Figure 59 B1-B3, C1-C3, D1-D3). Within the Mn appendages, all cells of the appendage core express *Ph six4* (Figure 59 B1-B3, C1-C3).

In stage 22 embryos, the labrum has completed its elongation, covering the stomodeal opening entirely. To the posterior, it directly abuts the paragnaths (Figure 60, compare D3 and E3). The pre-antennal hemispheres have expanded further lateral (Figure 60, A3, B3, C3; 3.1.4, Browne et al., 2005). Expression of *Ph six4* is confined to a distinct stripe in the medial pre-antennal region of the embryonic head, with two oval domains extending into the anterior medial brink of the head (Figure 60 A1-A3). In the centres of the pre-antennal hemispheres, *Ph six4* expression remains largely unaltered as compared to stage 20 embryos. Several clusters of *Ph six4*-expressing cells are visible (Figure 60 C1, red arrowhead in C2, C3). Strong *Ph six4* expression is found at the proximal root of the labrum, with two fine strings of *Ph six4*–positive cells extending into the labral lobes (Figure 60 B1, red arrowhead in B2, B3; E2, F2). Similar to stage 20, *Ph six4* expressing cells are found within the proximal articles of all segmental appendages (Figure 60 C1-C3, E1-E3, G1-G3). In the newly emerged paragnaths, *Ph six4* expression is found in small dot-like domains (Figure 60 E1-E3).

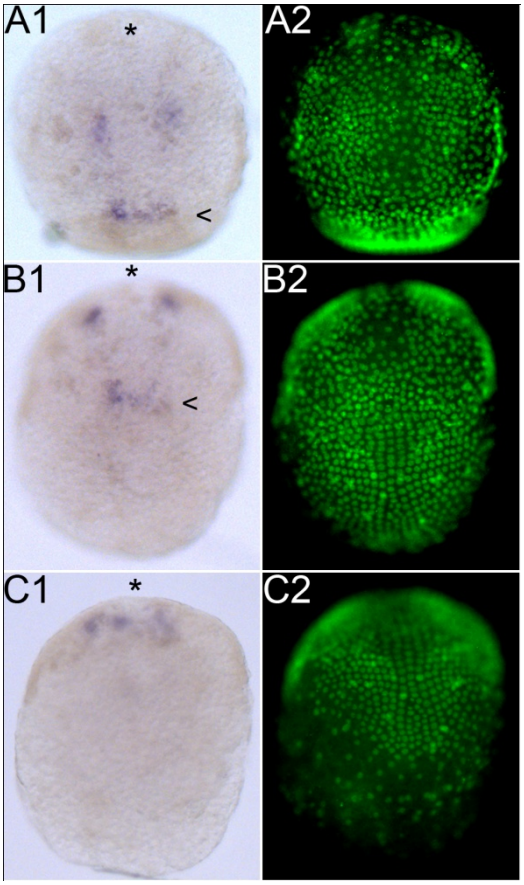



**Figure 56: Expression of *Ph six4* at stage 12**. The embryo is mounted in anterior orientation (A1, 2), ventral orientation (B1, B2) and posterior orientation (C1, C2) in order to facilitate better understanding of the three-dimensional pattern of *Ph six4* expression. In all images, anterior is up. Shown are bright field images in the left panel (A1-C1) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A2–C2; nuclei are green, expression of *Ph six4* is detectable as it quenches the nuclear labelling). In all bright field images (A1-C1), the anterior end of the embryo is indicated by a black asterisk. In the images A1 and B1, the position of the future Mn segment is marked by a black arrowhead.

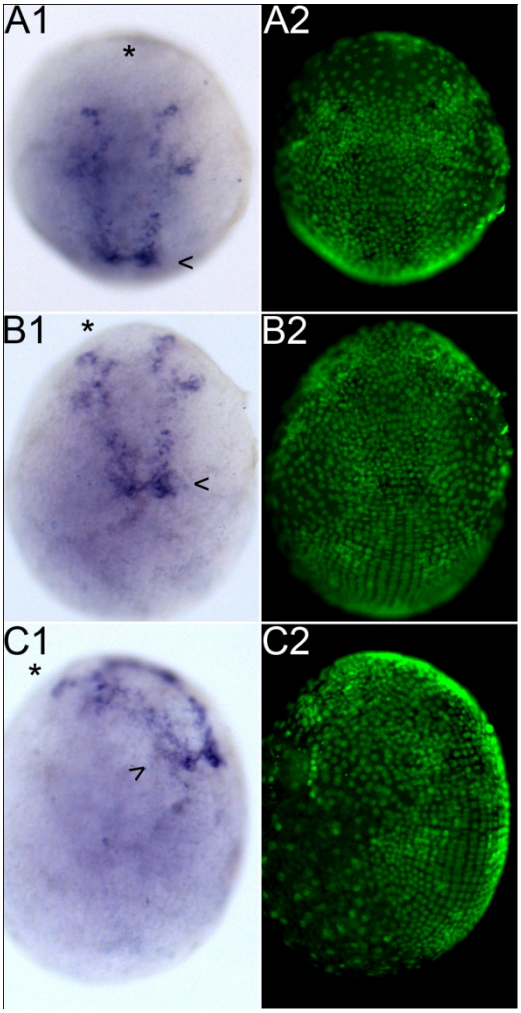



**Figure 57: Expression of *Ph six4* at stage S14**. The embryo is mounted in anterior orientation (A1, 2), ventral orientation (B1, B2) and lateral, tilted orientation (C1, C2, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph six4* expression. In all images, anterior is up. Shown are bright field images in the left panel (A1-C1) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A2–C2; nuclei are green, expression of *Ph six4* is detectable as it quenches the nuclear labelling). In all bright field images (A1-C1), the anterior end of the embryo is indicated by a black asterisk and the position of the future Mn segment is marked by a black arrowhead.

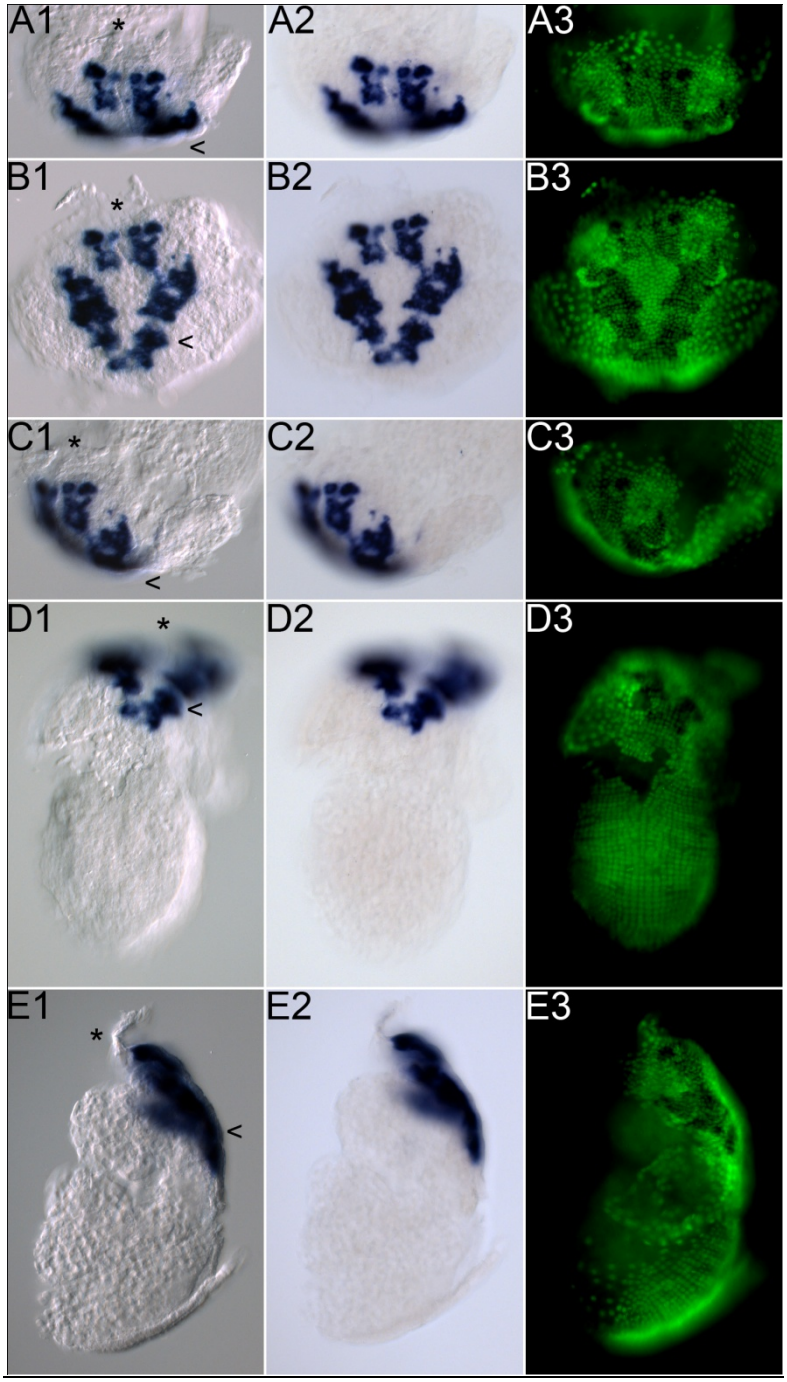



**Figure 58: Expression of *Ph six4* at stage S15**. The embryo is mounted in frontal orientation (A1-A3), ventral orientation, centred up the An2 segment (B1-B3), tilted anterior orientation (C1-C3), ventral orientation, centred up the trunk (D1-D3) and lateral orientation (E1-E3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph six4* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – E3; nuclei are green, expression of *Ph six4* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** In both pre-antennal hemispheres, the *Ph six4* expression profile encompasses a medially and apically located single cell, a lateral, apical group of cells and, posterior to these, an almost rectangular sheet of cells. **B** *Ph six4* expression is found in broad, bilateral stripes within the segments Mx1, Mn, An2 and An1. **D** On the level of the future Mx1 and Mx2 segments, the embryo has been torn on the left side during mounting.

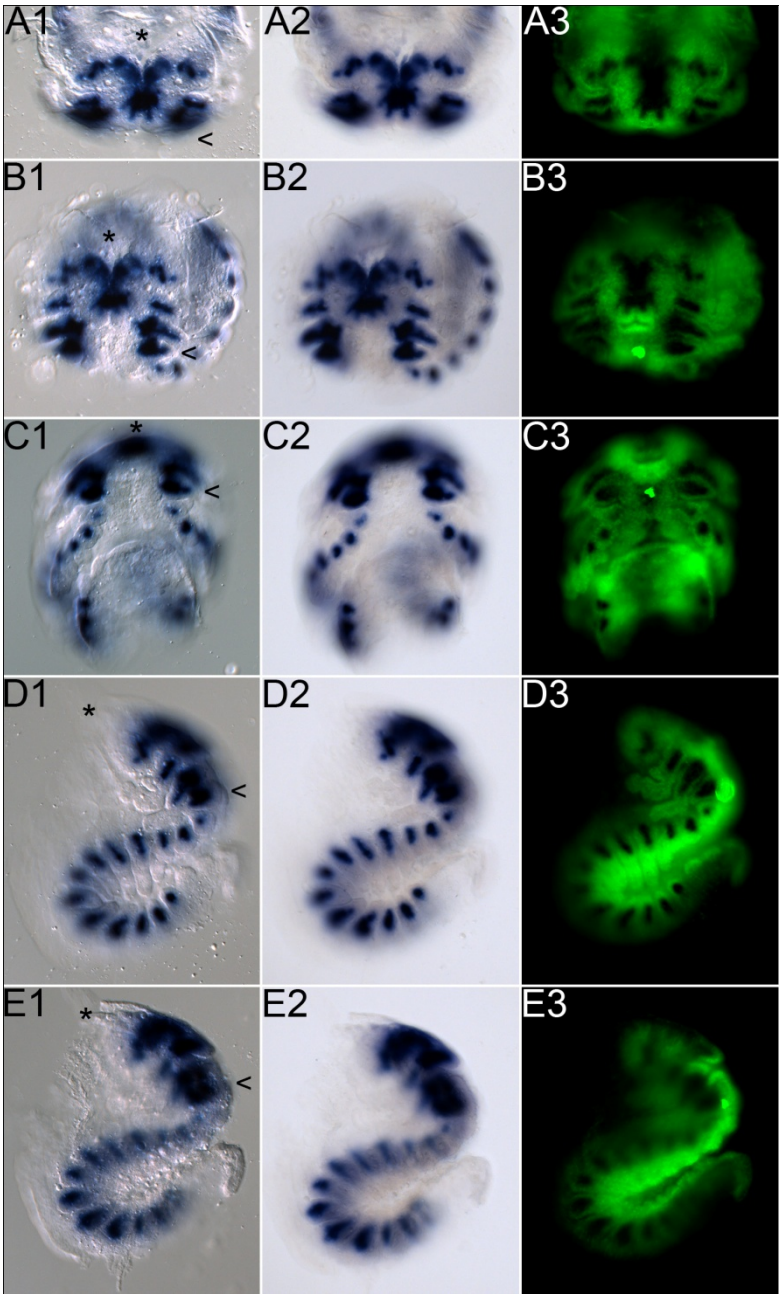



**Figure 59: Expression of *Ph six4* at stage S20**. The embryo is mounted in frontal orientation (A1-A3), ventral orientation, centred up the An1 segment (B1-B3), ventral orientation, native coiled posture, centred up the Mx1 segment (C1-C3), lateral orientation, focus on the An1 appendage (D1-D3, ventral is to the right) and lateral orientation, sagittal focus (E1-E3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph six4* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – E3; nuclei are green, expression of *Ph six4* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph six4* is expressed bilaterally in the central part of the pre-antennal hemispheres. **B** *Ph six4* is expressed in a medial domain that covers the apical pre-antennal brink and extends posterior into the base of the labrum. **C** The embryo is mounted in its native coiled posture. **D** *Ph six4* expression is found within the basal articles of each segmental appendage.

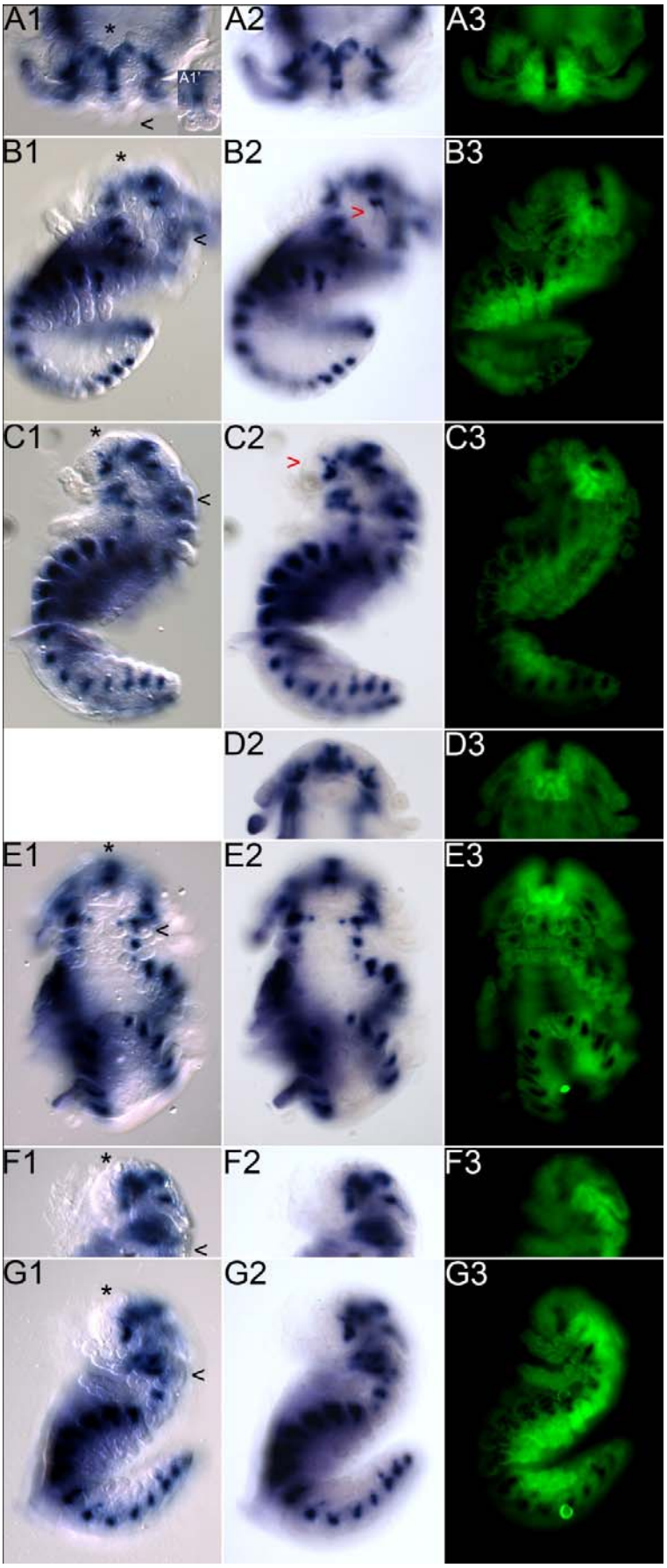



**Figure 60: Expression of *Ph six4* at stage S22**. The embryo is mounted in frontal orientation (A1-A3), tilted anterior orientation (B1-B3), tilted lateral orientation (C1-C3, ventral is to the right) ,ventral orientation, median focal plane (D2-D3), ventral orientation, focus on the paragnaths (E1-E3), lateral orientation, sagittal focal plane (F1-F3, ventral is to the right) and lateral orientation, focal plane on the An2 appendage (G1-G3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph six4* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-G1), bright field images in the middle panel (A2-G2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – G3; nuclei are green, expression of *Ph six4* is detectable as it quenches the nuclear labelling). In all DIC images (A1-G1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph six4* is expressed in a distinct stripe in the medial pre-antennal region of the embryonic head, with two oval domains extending into the anterior brink of the head. **B** *Ph six4* expression is found at the root of the labrum, with two fine strings of *Ph six4*–positive cells extending into the labral lobes (B2, red arrowhead). **C** Similar to stage 20 embryos, *Ph six4* expression in the centres of the pre-antennal hemispheres is composed of several clusters of *Ph six4* expressing cells (C2, red arrowhead).

### 3.3.2 *Parhyale hawaiensis gbx*

Because representatives of the *unplugged/gbx* family of genes share important and putatively ancestral functions in the organisation and patterning of brain regions (Hirth et al., 2003; Raible and Brand, 2004). I cloned and analysed the *Parhyale hawaiensis gbx* (*Ph gbx)* ortholog in this work.

#### 3.3.2.1 Isolation of *Ph gbx*

Because sufficient genomic sequence data is not currently available for *Parhyale hawaiensis*, the isolation of *gbx* sequences was approached by performing varied-stringency degenerate PCR screens (5.2.3.1). To this end, degenerate primers based on comparative amino acid alignments of HD of described UNPG and GBX proteins were used (5.2.4). *Ph gbx* was recovered from repeated varied-stringency degenerate PCR screens on a collection of *Parhyale* cDNA covering all stages of embryogenesis (5.2.2). The isolated fragments contained 100 bp of specific *Ph gbx* sequence and were sufficient for unambiguous homology identification (5.3.3.1).

In detail, seven sequences were recovered from repeated varied-stringency degenerate PCR screens. BLAST evaluation (5.3.3.1) revealed that they represent a novel *unpg/gbx* sequence. All obtained sequences are considered identical because the nucleotide variation between any two of these is not exceeding 1%. Since each sequence derives from a different priming event, they are considered independent.

Based on this primary sequence information, specific oligonucleotides were designed and used to prime *Ph gbx* RACE reactions (5.2.3.3, A4.2, table A3). Consequently, the remaining mRNA transcript sequences both 5' and 3' of the initial fragments covering the homeobox were obtained.

In detail, *Ph gbx* 5' cDNA sequence was fully recovered by two-step 5' RACE (5.2.3.3). Three independent *Ph gbx* 5' cDNA fragment clones were obtained. Their sequences show the same transcription start and are considered identical, since the nucleotide variation between any two of these is not exceeding 1% (see also A3, table A1).

*Ph gbx* 3' cDNA sequence was fully recovered by two-step 3' RACE. Three independent *Ph gbx* 3' cDNA fragment clones were obtained. The sequences of two of them,

Ph_gbx_3R11 and Ph_gbx_3R12, cover the same length and extend 350 bp farther 3' than the significantly shorter sequence of Ph_gbx_3R03. The fraction of sequence they share, however, is considered identical, since the nucleotide variation between any two of their sequences is not exceeding 1% (see also A3, table A1). All three of them show minor differences in poly(A) tailing.

Sequence information provided by all recovered 5' and 3' *Ph gbx* RACE clones was sufficient for assembling the complete *Ph gbx* cDNA sequence *in silico*. In order to verify the consistency of the established *Ph gbx* cDNA sequence, coherent *Ph gbx* ORF fragments were isolated via long-distance PCR reactions (5.2.3.4), performed on *Parhyale* cDNA collections (5.2.2). Twelve independent *Ph gbx* ORF sequences were obtained (A2.10.1). They are consistent with the findings from 5' and 3' RACE (A2.10.2). All sequences except that of clone Ph_gbx_ORF03 are considered identical, since the nucleotide variation between any two of these is not exceeding 1% (A3, table A1). The sequence of clone Ph_gbx_ORF03 differs in at least 2.4% of nucleotide positions from any other *Ph gbx* ORF sequences. It also shows an insertion of 3bp which encodes a glutamine residue that adds to a C-terminal stretch of glutamines (A2.10.2). These findings suggest that this clone represents an allelic variation of *Ph gbx*.

#### 3.3.2.2 Characterisation of *Ph gbx*

The comparison of all recovered *Ph gbx* sequences led to the identification of 91 sites of nucleotide exchange within the complete *Ph gbx* cDNA. 90 of those occur uniquely and are therefore considered sporadic (5.3.3.2). They are found predominantly in the ORF (82) where they account for 36 amino acid changes (missense alterations) and two premature stops (nonsense alterations). The remaining 44 nucleotide exchanges found within the ORF do not alter the translated amino acid sequence (silent exchanges). These findings suggest that these nucleotide exchanges are artificial. One nucleotide exchange (1591G/a) is present in more than one sequence. It is therefore considered polymorphic. Since the amino acid residue encoded at this position (362G) is not altered, it is likely that this nucleotide exchange represents a naturally occurring polymorphism of wild type *Ph gbx* alleles (5.3.3.2).

Based on these findings, the sequence of the clone Ph_gbx_ORF58 was chosen as the source of the *Ph gbx* reference sequence (*Ph gbx*_ref). It was extended 5' by adding 491 bp

derived from clone Ph_gbx_5R05 and 3' by adding 1243 bp derived from clone Ph_gbx_3R12 (A2.10.1). An Alignment of *Ph gbx*_ref with all recovered *Ph gbx* sequences confirms its statistical consensus relevance (A2.10.2). With the exception of Ph_gbx_ORF03, all recovered *Ph gbx* sequences show less than 1% difference from *Ph gbx2*_ref, confirming that they in fact derive from transcripts of *Ph gbx*. *Ph gbx*_ref has been used for phylogenetic studies.

The *Ph gbx* transcript is 3.5 kb in length and encodes a protein of 554 amino acids. The *Parhyale* GBX protein has a HD (60 amino acids) that his highly similar to HD of described UNPG and GBX proteins (3.3.2.3). *Ph* GBX has a C-terminal VPIPV protein motif (23 amino acids) that is conserved within described UNPG and GBX proteins (Figure 61, Figure 62, 3.3.2.3).

**Figure 61: Schematic view of *Ph gbx* transcript.** The length of the *Ph gbx* transcript is 3456 bases. Shown are: 5' UTR in grey (length 505 bases, nucleotide positions 1-505), ORF in black (length 1665 bases, encoding 554 amino acids; nucleotide positions 506-2170) and the 3' UTR in grey (length 1286 bases, nucleotide positions 2171-3456). The ORF region encoding the HD is depicted in blue (length 180 bases, encoding 60 amino acids; nucleotide positions 1346-1525). The ORF region encoding a conserved VPIPV motif is depicted in pink (length 69 bases, encoding 23 amino acids; nucleotide positions 1598-1666).

```
0001                                                GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGTGAAAGCTTCGTAAGTTTAG
0056          TTGAGCTTGTGAATGTAATGAAATATTTAGATATAAAAGACAATTGTACTGAACTGATACGTAATTTAGCCAGAAGTGATGATTTACATG
0146          CAGGCATCCAATTTACTCCCAGTGAACTTCTATGAGTCAAAACTGTGCTATCAGAATGAAGTAAGTGTTATTGTGAGGCTACTATTTTCA
0236          TATTTACGTAACGTACTATATTGCATATCCCAGACGTACATCTACGTTTAGACAGAAGTACGATCGTGTGACTATACGAACTAATTATTT
0326          AACCAATAACTCTGCAATTTCTAGTTCTTAGTGAATTATTTAAAAAGTTGGTGTACGTCATGCGAGATGAAACGTTTTGTGGTTATAATT
0416          GCTTGAACTTTCGCTATTTTTTACTGGACAACACAAAGTAAAAAAAAGTGCGCATTCGATCATTGTTCATCCCAGGGAATTATTTGAACG

0001          M  L  A  S  S  E  M  T  T  C  D  S  P  P  K  S  P  T  T  K  M  G  P  N  S  K  S  F  A  M
0506          ATGTTAGCGTCTTCAGAAATGACAACTTGCGATTCACCACCTAAATCTCCCACCACAAAAATGGGCCCGAACAGCAAGAGTTTTGCCATG

0031           E  Y  L  L  T  I  S  G  Q  D  K  N  D  C  N  R  S  K  S  M  K  T  S  A  A  S  T  L  L  S
0596          GAATATCTCCTCACCATCTCCGGTCAAGACAAAAATGACTGCAATAGATCAAAATCCATGAAAACTTCAGCAGCTTCGACGCTCCTGTCC

0061           R  E  T  D  T  H  P  N  H  L  P  A  S  H  M  L  D  I  S  R  Q  L  S  Q  I  T  Q  I  S  Q
0686          CGTGAAACAGACACGCACCCGAACCATCTGCCCGCTAGTCATATGCTCGACATCTCGCGCCAGTTGTCGCAAATAACCCAAATTTCTCAG

0091           N  L  A  L  P  H  L  Y  N  S  C  L  L  T  G  G  M  A  S  V  G  L  P  P  F  Y  G  H  P  P
0776          AATCTCGCACTTCCCCATCTGTACAACTCGTGCCTGCTGACGGGCGGCATGGCCTCGGTTGGGTTACCTCCGTTCTACGGGCACCCACCG

0121           L  A  P  H  Y  P  Y  P  A  P  S  T  V  P  F  P  N  P  I  T  G  S  L  P  N  G  F  S  A  G
0866          CTGGCCCCGCACTACCCGTACCCTGCCCCCTCCACCGTACCGTTCCCGAACCCGATCACGGGTTCGTTGCCCAACGGGTTCTCCGCGGGT

0151           G  L  Q  P  A  L  N  P  P  F  L  S  L  K  D  F  G  P  G  A  G  S  R  Y  F  G  G  S  P  T
0956          GGACTTCAGCCAGCTCTGAACCCACCGTTTTTGTCCTTGAAGGATTTTGGTCCTGGGGCTGGATCGAGATACTTCGGGGGATCTCCCACC

0181           A  G  G  E  N  P  R  N  A  I  L  R  S  T  L  S  S  E  T  P  A  D  L  P  S  P  F  S  G  A
1046          GCCGGGGGCGAAAACCCCCGTAACGCCATCTTGAGATCGACTCTGAGCAGCGAGACGCCTGCAGATTTACCGTCGCCCTTCTCGGGGGCG

0211           T  N  G  D  E  S  S  H  I  S  D  L  D  F  K  E  S  T  K  S  N  E  L  V  D  D  L  Q  I  D
1136          ACCAACGGAGATGAAAGCAGTCATATTTCGGACTTGGACTTCAAGGAGTCTACAAAATCAAATGAGCTAGTGGATGACCTGCAAATAGAC

0241           P  V  D  S  D  D  E  S  H  V  N  G  N  S  A  G  L  G  G  G  G  A  P  L  D  S  S  T  T  T
1226          CCCGTGGACAGCGATGACGAGAGTCACGTCAACGGCAACTCCGCGGGACTGGGGGGCGGTGGGGCGCCGCTGGACAGCTCCACCACTACC

0271           G  A  E  H  R  G  T  G  S  K  N  R  R  R  R  T  A  F  T  S  E  Q  L  M  E  L  E  R  E  F
1316          GGCGCAGAGCACCGCGGCACCGGCAGTAAAAACAGAAGACGCCGGACGGCCTTCACTTCGGAGCAGCTCATGGAACTCGAGAGGGAATTC

0301           Q  T  K  K  Y  L  T  L  S  E  R  S  H  I  A  Q  T  L  H  L  S  E  V  Q  V  K  I  W  F  Q
1406          CAAACGAAGAAATATCTGACGCTGAGCGAGCGGTCGCACATTGCTCAGACACTGCACCTCAGTGAGGTCCAAGTGAAAATTTGGTTCCAG

0331           N  R  R  A  K  W  K  R  V  K  A  G  L  V  G  S  S  G  G  G  L  V  G  G  T  N  K  G  G  A
1496          AACCGGCGGGCGAAATGGAAGCGCGTGAAGGCTGGCCTTGTCGGCTCGAGCGGCGGGGGCCTCGTGGGAGGTACAAACAAGGGGGGCGCC

0361           G  G  G  H  K  I  I  V  P  I  P  V  H  V  N  R  L  S  N  L  S  Q  A  H  Q  I  E  K  C  R
1586          GGAGGGGGGCACAAGATTATCGTCCCAATCCCAGTCCATGTGAACAGGCTAAGCAATCTGTCGCAGGCGCATCAGATAGAGAAGTGTCGA

0391           G  S  A  A  G  S  N  S  D  G  S  L  S  R  P  T  E  S  S  S  A  A  D  D  D  G  K  L  A  N
1676          GGGTCGGCGGCTGGAAGCAACAGCGACGGCAGCTTGAGCAGACCGACCGAGTCTTCTTCGGCTGCTGATGATGACGGAAAATTGGCCAAC

0421           Q  R  M  A  T  E  A  C  T  S  T  K  H  R  N  K  S  A  N  L  S  S  S  F  L  N  S  E  T  S
1766          CAGAGGATGGCCACAGAAGCGTGTACCTCCACCAAACACAGAAACAAATCAGCGAATTTGTCGTCGAGCTTCTTGAATTCAGAGACGTCG

0451           E  S  S  I  V  G  V  N  Q  H  S  Q  D  Q  Q  Q  Q  H  R  A  L  Q  S  R  H  S  A  F  Q  G
1856          GAAAGTTCGATCGTAGGTGTAAATCAACATAGCCAAGATCAACAGCAGCAACATCGCGCCCTACAGTCACGGCATTCTGCGTTCCAGGGG

0481           L  S  S  S  T  S  S  C  S  S  G  H  S  E  L  S  N  L  G  L  S  S  G  R  N  G  S  A  L  E
1946          CTGTCGTCCTCTACCTCATCGTGTTCCAGCGGCCACTCAGAACTGTCAAATTTAGGGCTGTCGTCGGGTCGAAATGGCTCTGCCCTGGAA

0511           A  H  L  L  G  H  T  K  E  V  T  E  Q  L  K  S  M  R  N  G  G  R  T  S  T  T  S  S  S  T
2036          GCCCACTTACTCGGCCATACCAAGGAGGTCACGGAACAGTTGAAGTCAATGAGAAATGGTGGTAGAACATCGACAACGTCGTCCTCGACC

0541           S  L  Q  C  Y  F  Q  N  H  E  R  I  S  P  #                                                0554
2126          TCGTTGCAATGCTATTTTCAAAATCATGAGAGGATTTCCCCGTAACCTTGATGTTTATAACTAAACAAACATATATCTATTATTTTTGGA
2216          ACAGATTGGTGATAAAGAATATCATCGTGCACTATTACTTTGAAAATCAAGATGAGATGACGACCTAACCATAATTTACAAATAAACTTT
2306          CCTCTCAGGGAAACTTTACGGTATCTGATCTAGTGCAGATCGTACAACAATTTCATCGTCACAGCAATGCACTTTTCAAAATCACGGGGA
2396          ATTTCTCTAGATAACCTTATCATTTATCATTAAAACATTGCTCACAAATAGCTTTATCATAATGGTTGCAGGCGGTCATCATCATCATTC
2486          CTGATTTAAGAATCAGGAAAGGATTTCTTCTTCGTTTGGAAACTCAGTAATAAAATCTCATTCACCAATATACCAAGATTTTCGTTGTTT
2576          TGTTGTATCTTTAATAAGTCATCATCCTCGTCATTGCATTATTATTTCAAAATCGTGAAGGAAATCTTACTTAGCTTGAAATTCGCCCAT
2666          AAAATTTCATTCACTAACAGCTTTGTGGCTTCCTAAGCTACGACAACCATATTATTAGTTTTTGTTTCCCCACTGCCATATTATTGGTAT
2756          AAATATATCAGTATCTTACACCGGAAACATCTTTTCACAATTTTTAATGACACCAACTAACTGAATCGCAATAATTAAACTACACTACTG
2846          TAACGAAAGTTAAAATCCTCTATGGGAAACCTCAGGAAGTAGAAGCTAAACTCTAGAATATACGGATCGTCACCATAATTTAGGCATTCT
2936          CGTTGGAATTATCTTAGTTGAAAGAAAAGAAAGCAACTAATTTTCTATGTTTTATTTCGCATTAAGTGGAAGATTCTTTCAGAAACTTTA
3026          CTATCAGTAGATCAAATTGTAACGCAAGTGGAAGGTACATAAATTTATCTTCAGTCAATGCTTCCCATATCACCACGTACGACGCAGCTA
3116          GTGCTTGAATATTAATAGGAAAAGGGAAAGGGCCTGCAGTAATTTAGGAAAAATGCATAATAGTTCTCAAGCACTGTAAATCACTTATAG
3206          GAATAGTAGTTCGGGAAAGAAAACAAAAGAATTACAAAAATCATTAGTAAAAGTAAGAAAATAGTTAAGGAAACAAAAGCACAGAAAATA
3296          AAGATGTCCCGCTGTTCAGCTAGGACTACTGTTCTGTATGTACAGCTTGTAAATTATATGGACCACATTGCGTAGCTTGGGCGAACTTAA
3386          CTAATCTGGCGGTACTTCCAAGCGATAGTTGTGTGTTTACTAACTGAAAAAAAAAAAAAAAAAAAAAAAAGT                3456
```

**Figure 62:** ***Ph gbx*** **cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph gbx* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The HD is shown in blue and a conserved protein motif covering the amino acids VPIPV in pink. The nucleotide sequences that encode these domains and motifs are shown in the respective colours. Numbers to the left give

the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

#### 3.3.2.3 Phylogeny of *Ph gbx*

BLAST searches (Altschul et al., 1997) using the complete amino acid sequence of *Parhyale* GBX as query against the protein databases of *Drosophila melanogaster* (blastp, taxid 7227) and mouse (blastp, taxid 10090) revealed highest sequence similarity to *Drosophila* UNPG and mouse GBX1 proteins, respectively. Murine GBX2 is found as the second-most similar protein within the mouse proteome. Other homeodomain-containing proteins were found with significantly lower BLAST score. Independently, a BLAST search was performed using the complete amino acid sequence of *Parhyale* GBX as query against the complete translated nucleotide database (tblastn). From this search, the putative *Pediculus humanus corporis* (human body louse) homeobox protein GBX-1 showed highest sequence similarity to *Parhyale* GBX.

For comparison, the protein sequence of *Parhyale* GBX was aligned with described and annotated UNPG and GBX proteins of various animal species using the CLUSTALW2 algorithm (Larkin et al., 2007). It reveals extreme amino acid conservation of the UNPG/GBX HD and high similarity of the VPIPV motifs (A6.3). Interestingly, in all examined species, the VPIPV was specific to UNPG/GBX proteins and was not found in other proteins. BLAST searches performed with the *Parhyale* GBX VPIPV motif as query against different protein and nucleotide databases corroborated this finding.

Based on a coherent sequence alignment fraction of the HD and VPIPV motifs of all examined UNPG/GBX proteins gathered earlier, a phylogenetic calculation was done (Mr Bayes, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The resulting tree was insufficient to resolve the phylogeny of the corresponding animal species (Figure 63, see also A6.1, table A5). This further confirms the fact that the HD sequences of the examined UNPG/GBX proteins are too conserved and invariant between species to allow robust phylogenetic reconcilements.

In order to further investigate the putative homology of *Parhyale* GBX and the described UNPG/GBX proteins of other species, a phylogenetic tree was calculated based on the alignment of HD of *Parhyale* GBX, *Drosophila* UNPG and murine GBX1 and GBX2 proteins, along with additional HD-containing *Drosophila* and mouse proteins that were

found being similar to *Parhyale* GBX during previous BLAST searches (using the CLUSTALW2 algorithm, Larkin et al., 2007, A6.2, table A6). Here, *Parhyale* GBX is aligned along with *Drosophila* UNPG and mouse GBX1/GBX2 in one clade (Figure 64). This substantiates the initial assumption that *Parhyale* GBX is in fact the ortholog of *Drosophila* UNPG. Interestingly, neither mouse GBX1 nor GBX2 were revealed as unambiguous orthologs to *Parhyale* GBX, suggesting that both are equally paralogous to *Parhyale* GBX. Within the proteomes of each examined arthropod species, only a single UNPG/GBX protein was found, suggesting that the murine GBX1 and GBX2 derive from a duplication event that is only present in the lineage leading to vertebrates and mammalians.

**Figure 63 (next page): Phylogenetic analysis of *Parhyale* GBX and UNPG/GBX proteins from other species.** Shown are: *Parhyale hawaiensis* GBX (Phaw_GBX), *Drosophila melanogaster* UNPG (Dmel_UNPG), *Drosophila virilis* UNPG (Dvir_UNPG), *Drosophila pseudoobscura* UNPG (Dpse_UNPG), *Anopheles gambiae* UNPG (Agam_UNPG), *Tribolium castaneum* UNPG (Tcas_UNPG), *Pediculus humanus* GBX1 (Phum_GBX1), *Nasonia vitripennis* UNPG (Nvit_UNPG), *Ixodes scapularis* Xvent-2B (Isca_Xvent-2B), *Platynereis dumerilii* GBX (Pdum_GBX), *Danio rerio* Gbx1 (Drer_Gbx ), *Danio rerio* Gbx2 (Drer_Gbx2), *Mus musculus* gastrulation brain homeobox 1 protein (Mmus_Gbx1), *Mus musculus* gastrulation brain homeobox 2 protein (Mmus_Gbx2) and *Nematostella vectensis* GBX (Nvec_GBX). Pdum_GBX was set as outgroup. For reference regarding the non-*Parhyale* protein sequences, see A6, tables A5 and A6. **A** Phylogram showing Clade credibility values. **B** Phylogram based on average branch lengths. Phylogenetic Analysis was done using Mr Bayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; 5.3.2; 100,000 generations).

**A**

```
/------------------------------------------------------------ Pdum_GBX
|
|                                                  /---------- Mmus_Gbx2
|                                        /---100---+
|                                        |         \---------- Drer_Gbx2
|-------------------82------------------+
|                                        |         /---------- Mmus_Gbx1
|                                        \----95---+
|                                                  \---------- Drer_Gbx1
|
|                                                  /---------- Nvec_Gbx
|                                                  |
|                                                  |---------- Dmel_UNPG
+                                                  |
|                                        /----92---+---------- Dpse_UNPG
|                                        |         |
|                                        |         |---------- Dvir_UNPG
|                              /----86---+         |
|                              |         |         \---------- Agam_UNPG
|                              |         |
|                   /----53---+         \-------------------- Phum_GBX1
|                   |         |
|                   |         |                   /---------- Phaw_GBX
|         /----87---+         \---------55--------+
|         |         |                             \---------- Nvit_UNPG
|         |         |
\----80---+         \---------------------------------------- Tcas_UNPG
          |
          \------------------------------------------------- Isca_Xvent-2B
```

**B**

```
/-------------- Pdum_GBX
|
|        /- Mmus_Gbx2
|   /----+
|   |    \ Drer_Gbx2
|---+
|   |  /-- Mmus_Gbx1
|   \--+
|      \-- Drer_Gbx1
|
|                              /------------------------------------ Nvec_Gbx
|                              |
|                              |-- Dmel_UNPG
+                              |
|                      /------+- Dpse_UNPG
|                      |      |
|                      |      |- Dvir_UNPG
|                /-----+      |
|                |     |      \------- Agam_UNPG)
|                |     |
|           /----+     \----- Phum_GBX1
|           |    |
|           |    |     /--------------------- Phaw_GBX
|     /-----+    \-----+
|     |     |          \------ Nvit_UNPG
|     |     |
\-----+     \----------- Tcas_UNPG
      |
      \----------- Isca_Xvent-2B

|-------| 0.100 expected changes per site
```

**Figure 63:** **Phylogenetic analysis of *Parhyale* GBX and UNPG/GBX proteins from other species.**

**A**

```
/------------------------------------------------------------ Dmel_DR
|
|                                                 /------------ Dmel_CG11085
|                                        /-----93-----+
|                                        |            \------------ Dmel_B-H1
|-----------------100----------------+
|                                        |            /------------ Mmus_Nkx1-2
|                                        \-----100----+
|                                                     \------------ Dmel_SLOU
+
|                                                     /------------ Mmus_Gbx1
|                                        /-----96-----+
|                                        |            \------------ Mmus_Gbx2
|                          /-----57----+
|                          |            \------------------------- Phaw_GBX
|            /-----94-----+
|            |             \------------------------------------- Dmel_UNPG
|            |
\-----78----+                                         /------------ Mmus_Meox1
             |                         /-----100----+
             |                         |            \------------ Mmus_Meox2
             \-----------100----------+
                                       |            /------------ Mmus_Mnx1
                                       \-----100----+
                                                    \------------ Dmel_EXEX
```

**B**

```
/--------------------------------- Dmel_DR
|
|                            /------------- Dmel_CG11085
|                /----------+
|                |           \------------------------------------ Dmel_B-H1
|----------------+
|                |                  /----- Mmus_Nkx1-2
|                \------------------+
|                                   \------ Dmel_SLOU
+
|                              /---- Mmus_Gbx1
|                        /-----+
|                        |     \--- Mmus_Gbx2
|                   /----+
|                   |    \--------------- Phaw_GBX
|        /----------+
|        |          \--- Dmel_UNPG
|        |
\--------+                                                /--- Mmus_Meox1
         |              /--------------------------------+
         |              |                                \----- Mmus_Meox2
         \--------------+
                        |              /-------- Mmus_Mnx1
                        \--------------+
                                       \----- Dmel_EXEX

|--------| 0.100 expected changes per site
```

**Figure 64: Phylogenetic analysis of *Parhyale* GBX and similar proteins from *Drosophila* and mouse.** Shown are: *Parhyale hawaiensis* GBX (Phaw_GBX), *Drosophila melanogaster* UNPG (Dmel_UNPG), *Drosophila melanogaster* EXTRA-EXTRA (Dmel_EXEX), *Drosophila melanogaster* gene product of CG11085 (Dmel_CG11085), *Drosophila melanogaster* SLOUCH (Dmel_SLOU), *Drosophila melanogaster* BARH1 (Dmel_B-H1), *Drosophila melanogaster* DROP (Dmel_DR), *Mus musculus* gastrulation brain homeobox 1 protein (Mmus_Gbx1), *Mus musculus* gastrulation brain homeobox 2 protein (Mmus_Gbx2), *Mus musculus* mesenchyme homeobox 1 protein (Mmus_Meox1), *Mus musculus* motor neuron and pancreas homeobox 1 protein (Mmus_Mnx1), *Mus musculus* mesenchyme homeobox 2 protein (Mmus_Meox2) and *Mus musculus* NK1 transcription factor related, locus 2 protein (Mmus_Nkx1-2). Dmel_DR was set as outgroup. For reference regarding the non-*Parhyale* protein sequences, see A6, tables A5 and A6. **A** Phylogram showing Clade credibility values. **B** Phylogram based on average branch lengths. Phylogenetic Analysis was done using Mr Bayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; 5.3.2; 100,000 generations).

#### 3.3.2.4 Expression of *Ph gbx*

Expression of *Ph gbx* is first detectable in stage 19 embryos. During this stage, the stomodeum has begun to invaginate medially at the level between the An1 and An2 segments (Figure 65 A3; 3.1.3; Browne et al., 2005). *Ph gbx* is expressed bilaterally at the base of each An1 appendage in two distinct domains. These expression domains are arranged in anterior-posterior direction and border the developing labrum and the invaginating stomodeum laterally (Figure 65 A1-A3). In addition, *Ph gbx* expression is found within the An1 appendage as a small, dot-like domain in the penultimate distal appendage article (Figure 65 B1'). In the Mn segment, weak *Ph gbx* expression is detectable in cells that lie at the base of the developing mandibular appendage (Figure 65 C1, C2). No expression of *Ph gbx* was found in embryos of earlier stages.

In stage 22 embryos, the stomodeum is entirely covered by the labrum that has completed its elongation and directly abuts the paragnaths (Figure 66 C1, C3; 3.1.4; Browne et al., 2005). Expression of *Ph gbx* at the base of the An1 appendages is maintained. In the stage 22 embryonic head, these medial *Ph gbx* expression domains are located anterior to the stomodeum (Figure 66 A1-3, B1-3; 3.1.4). *Ph gbx* is also expressed in a segmental pattern, encompassing bilateral expression domains that lie within the ventral and medial compartments of each segment. The *Ph gbx*-positive cells in this area are putatively part of the ventral nervous system (Figure 66 D1-3, E1-3; see also Browne et al., 2006; Schaeper et al., 2010). In addition, *Ph gbx* is expressed in small, dot-like domains within all segmental appendages. In the An1 and An2 appendages, *Ph gbx* is expressed within all appendage articles except the most distal ones (Figure 66 B1, B2). In thoracic appendages, *Ph gbx* expression is found in the penultimate distal articles (Figure 66 E1, E2; F1, F2). During Parhyale embryogenesis, no *Ph gbx* expression was found within the labrum (Figure 65 A1, A2; Figure 66 A1', B1'), in the stomodeal field (Figure 65 B1, B1', B2; Figure 66 B1, B2) or in the proctodeal field (Figure 65 C1-3; Figure 66 D1-3).

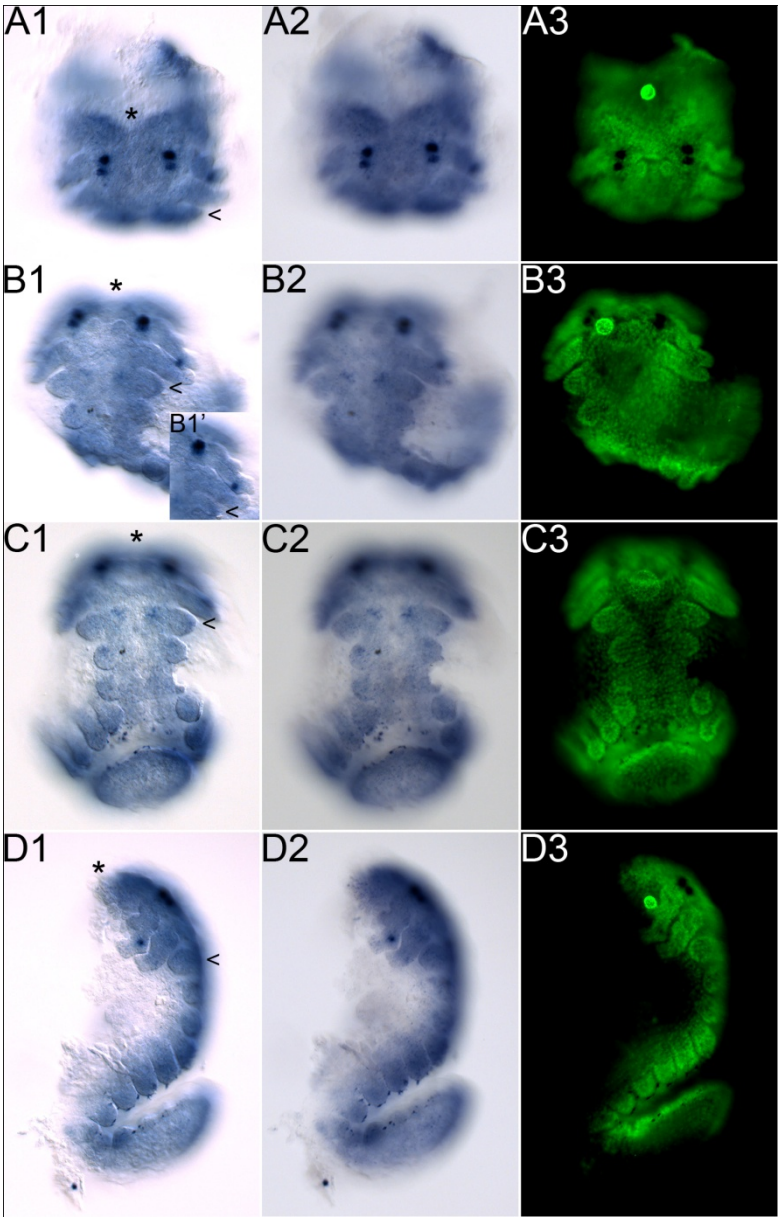



**Figure 65: Expression of *Ph gbx* at stage S19**. The embryo is mounted in frontal (A1-A3), tilted anterior (B1-B3), ventral (C1-C3) and lateral (D1-D3, ventral is to the right) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph gbx* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2-D2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – D3; nuclei are green, expression of *Ph gbx* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph gbx* is expressed in two distinct domains at the base of each An1 appendage that border labrum and stomodeum (A3, yellow asterisk) laterally. **B** *Ph gbx* expression is found within the An1 appendage as a small, dot-like domain in the penultimate distal appendage article (B1'). In the Mn segment, weak *Ph gbx* expression is detectable in cells that lie at the base of the developing mandibular appendage. **C** The embryo is centred up the Mn, Mx1 and Mx2 segments. **D** The focal plane is on the An1 and An2 appendages.

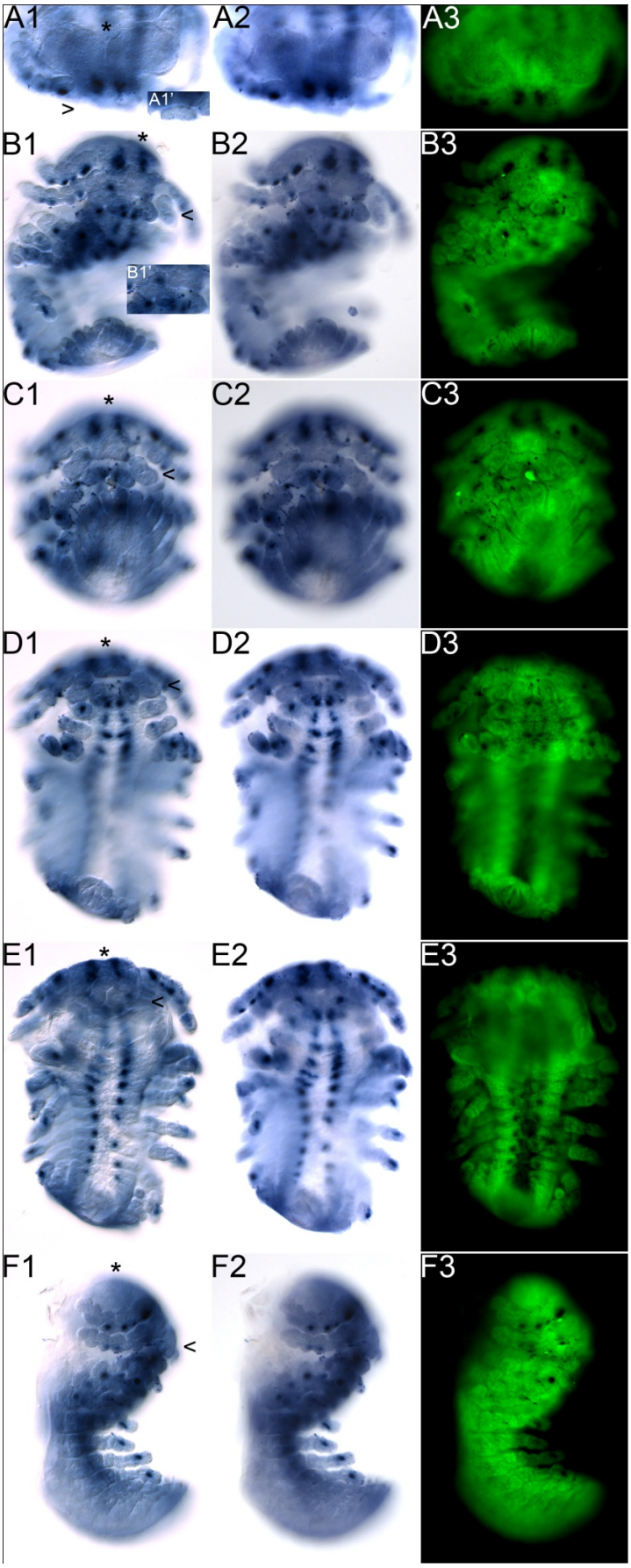



**Figure 66: Expression of *Ph gbx* at stage S22**. The embryo is mounted in frontal (A1-A3), in tilted anterior (B1-B3), in ventral, native coiled posture (C1-C3), as well as in ventral, dissected flat, focal plane on Mx1 (D1-D3), ventral, dissected flat, focal plane on T5 (E1-E3) and lateral (F1-F3, ventral is to the right) orientation in order to facilitate better understanding of the three-dimensional pattern of *Ph gbx* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-F1), bright field images in the middle panel (A2-F2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – F3; nuclei are green, expression of *Ph gbx* is detectable as it quenches the nuclear labelling). In all DIC images (A1-F1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** Expression of *Ph gbx* at the base of the An1 appendages is maintained. The labrum is free of *Ph gbx* expression (A1'). **B** In the An1 and An2 appendages, *Ph gbx* is expressed within all appendage articles except the most distal ones. No *Ph gbx* expression is detectable in the developing stomodeal field (B1'). **C** The embryo is mounted in its native, coiled posture. The labrum abutting the mandibles and the paragnaths is in the centre. The proctodeal developing field is free of *Ph gbx* expression. **D** Focal plane is on Mx1. **E** Focal plane is on T5. *Ph gbx* is expressed in each segment in two restricted, medial expression domains. In thoracic appendages, *Ph gbx* expression is found in the penultimate distal articles. **F** The focal plane is on the An1 and An2 appendages.

## 3.4 *Parhyale hawaiensis knirps* genes

In insects, members of the *knirps* family of genes have been shown to be involved in various aspects of head development (Ben-David and Chipman, 2010; Cerny et al., 2008; Gonzalez-Gaitan et al., 1994; Rothe et al., 1994). However, a direct comparison of *knirps* functions in the fruit fly *Drosophila melanogaster* and the red flour beetle *Tribolium castaneum* reveals significant differences even within insect species(Cerny et al., 2008). In order to address conserved and divergent roles of *knirps* genes in arthropod head development, I cloned and analysed *Parhyale hawaiensis knirps* genes in this work.

### 3.4.1 Isolation and characterisation of *Parhyale hawaiensis knirps* genes

Because sufficient genomic sequence data is not currently available for *Parhyale hawaiensis*, the isolation of putative *kni* sequences was approached by performing varied-stringency degenerate PCR screens (5.2.3.1). To this end, degenerate primers based on comparative amino acid alignments of both Zinc finger domains (ZNFD) and *knirps box* domains (KNID) of described Knirps proteins were used. Two *Parhyale hawaiensis knirps* sequences, *Parhyale hawaiensis knirps1* (*Ph kni1*) and *knirps2* (*Ph kni2*) were recovered from repeated varied-stringency degenerate PCR screens on a collection of *Parhyale* cDNA covering all stages of embryogenesis (5.2.2). The isolated fragments contained 211 bp of specific *Ph kni1* and *Ph kni2* sequence and were sufficient for unambiguous homology identification (5.3.3.1, see also A2.11 and A2.12). Several independent cycles of a depletive lowered-stringency PCR screen (5.2.3.2) were performed in order to isolate additional *Parhyale knirps* sequences. However, no novel sequences indicating additional *Parhyale knirps* paralogs were recovered.

In detail, 11 sequences were recovered from repeated varied-stringency degenerate PCR screens. They fall into two groups of sequences, *Ph kni1* and *Ph kni2*, with seven sequences representing *Ph kni1* and four sequences representing *Ph kni2*. Within the 211 bp fragment, any individual *Ph kni1* and *Ph kni2* sequences differ from each other in at least 19.9% nucleotide positions, strongly supporting that they actually derive from different genes. Since each sequence of the *Ph kni1*-group derives from a different priming event, they are considered independent. All independent sequences of the *Ph kni1*-group are considered

identical because the nucleotide variation between any two of these is not exceeding 1%. An exception is Ph_kni1_vsdP07. Its sequence differs from any other sequence of the *Ph kni1*-group in at least 2.4% of nucleotide positions. For *Ph kni2*, all recovered sequences are considered independent and identical since they all derive from different priming events and the nucleotide variation between any two of these is not exceeding 1% (A3, table A1). BLAST evaluation (tblastx; taxid: 7227, A6.1 table A5), revealed highest sequence similarity with *Drosophila melanogaster knirps-like* (*Dm knrl*) for both *Ph kni1* and *Ph kni2*. This finding indicates that *Ph kni1* and *Ph kni2* do not represent direct orthologs of single members of the *Drosophila knirps* gene family (specifically, of *Dm kni*, *Dm knrl* or *Dm eg*, see also 3.4.6).

Based on this primary sequence information, specific oligonucleotides were designed and used to prime *Ph kni1* and *Ph kni2* RACE reactions (5.2.3.3, A4.2, table A3). Consequently, the remaining mRNA transcript sequences both 5' and 3' of the initial fragments covering ZNFD and KNID were obtained.

### 3.4.2 Isolation of *Ph kni1*

From two-step 5' RACE, seven independent *Ph kni1* 5' cDNA fragment clones were obtained. Their nucleotide sequence varies considerably (A3, table A1), with differences between any two individual sequences going up to 7.3% of nucleotides. Notably, the sequence of Ph_kni1_5R30 differs from all other recovered *Ph kni1* 5' cDNA sequences to the greatest extend. A comparison of all obtained *Ph kni1* sequences reveals a high frequency of polymorphic nucleotide positions within the fraction of sequence covered by the *Ph kni1* 5' cDNA fragment clones (3.4.3). If neither the sequence of Ph_kni1_5R30 nor the polymorphic nucleotide positions are taken into account, any two sequences of the remaining *Ph kni1* 5' cDNA fragment clones vary less than 1%. All *Ph kni1* 5' cDNA fragment clones show identical transcription starts except Ph_kni1_5R30 whose sequence extends slightly shorter 5' than the other sequences. These findings indicate that the recovered *Ph kni1* 5' cDNA fragment clones derive from transcripts of different alleles of the same gene.

*Ph kni1* 3' cDNA sequence was fully recovered by two-step 3' RACE, supplemented by consecutive two-step 3' extensive RACE. In detail, four independent *Ph kni1* 3' cDNA fragment clones were obtained from the initial two-step 3' RACE. They all cover the same

fraction of sequence, with minor variations in poly(A) tailing. The nucleotide variation of any two of the sequences is around 1-2%, which is slightly above average as compared to other 3' RACE sequences examined in this work. Similar to the findings from *Ph kni1* 5' RACE, several polymorphic sites of nucleotide exchange were determined (3.4.3), indicating that these RACE sequences actually derive from transcripts of different alleles of the same gene. Because all four 3' RACE sequences translate into a consistent amino acid sequence that does not show a translation stop before the poly(A) tails, they were recognised as prematurely aborted ORF fragments (5.3.3.2). In order to recover the full *Ph kni1* 3' cDNA sequence, additional two-step 3' extensive RACE was performed. From this approach, two independent *Ph kni1* extended 3' cDNA sequences were obtained. They vary in about 2% of nucleotide positions which goes well with all other *Ph kni1* RACE sequences that were recovered in this work. Both cover an identical fraction of *Ph kni1* ORF and show the same translation stop. However, they vary in length and, consequently, in the size of their 3' UTRs, by 142bp.

Sequence information provided by all recovered 5' and 3' *Ph kni1* RACE clones was sufficient for assembling the complete *Ph kni1* cDNA sequence *in silico*. In order to verify the consistency of the established *Ph kni1* cDNA sequence, coherent *Ph kni1* cDNA and ORF fragments were isolated via long-distance PCR reactions (5.2.3.4), performed on *Parhyale* cDNA collections (5.2.2). Three independent *Ph kni1* cDNA sequences and seven independent *Ph kni1* ORF sequences were obtained (A2.11.1). They are consistent with the findings from 5' and 3' RACE (A2.11.2). Nucleotide variations between any two of these are around 1-2%, which goes well with previous findings from *Ph kni1* RACE. Accordingly, this suggests that these sequences derive from transcripts of different alleles of the same gene. (5.3.3.3, see also A3, table A1).

### 3.4.3 Characterisation of *Ph kni1*

The comparison of all recovered *Ph kni1* sequences led to the identification of 124 sites of nucleotide exchange within the complete *Ph kni1* cDNA, corresponding to 5.6% nucleotide variation. 72 of those occur uniquely and are therefore considered sporadic (5.3.3.2). They are found predominantly in the ORF (51) where they account for 28 amino acid changes (missense alterations) and one premature stop (nonsense alteration). 22 nucleotide exchanges found within the ORF do not alter the translated amino acid sequence (silent exchanges).

These findings go well with the expected randomness of artificial nucleotide exchanges, suggesting that they are artefacts. The remaining 52 nucleotide exchanges are found in more than one independent sequence and are therefore considered polymorphic. They are predominant in the 5' UTR (42). Of the six polymorphic nucleotide exchanges present in the ORF, five do not alter the translated amino acid sequence. The remaining nucleotide exchange leads to an alternation of valine and phenylalanine for the amino acid residue at position 212, which is not within one of the described KNI protein domains (ZNFD, KNID, PIDLS motif). These findings strongly suggest that they represent naturally occurring polymorphisms of wild type *Ph kni1* alleles (5.3.3.2).

There is no indication for the existence of differentially spliced *Ph kni1* transcripts or other types of *Ph kni1* transcript isoforms since the comparison of all recovered *Ph kni1* sequences supports a consistent *Ph kni1* transcript with only minor variations in 5' and 3' extension.

Based on these findings, the complete sequence of the clone Ph_kni1_cDNAJZB5 was chosen as the *Ph kni1* reference sequence (*Ph kni1_ref*). It covers all fractions of cDNA sequence that are shared by at least two independent *Ph kni1* cDNA fragment clones. Except two sporadic nucleotide exchanges within the 3' UTR, it only carries putatively allelic polymorphisms and therefore translates into the consensus *Parhyale* Kni1 protein sequence. As an existing cDNA clone sequence, it is preferred over an in silico modified consensus sequence and, for this reason, not altered further. *Ph kni1_ref* has been used for phylogenetic studies.

The *Ph kni1* transcript is 2.2 kb in length and encodes a protein of 397 amino acids. The *Parhyale* Kni1 protein has a typical C4-type N-terminal Zinc finger domain (ZNFD, 66 amino acids) with closest resemblance to Knrl ZNFD. Directly adjacent to the ZNFD, *Parhyale* Kni1 has a knirps box domain (KNID, 19 amino acids) with high sequence similarity to the KNID of described kni family proteins of other species. C-terminally, a conserved PIDLS motif is found within the hydrophobic region of the *Parhyale* Kni1 protein. This is typical for Kni proteins of arthropods (Figures 67, 68; 3.4.6).

**Figure 67: Schematic view of *Ph kni1* transcript.** The length of the *Ph kni1* transcript is 2228 bases. Shown are: 5' UTR in grey (length 698 bases, nucleotide positions 1-698), ORF in black (length 1194 bases, encoding 397 amino acids; nucleotide positions 699-1892)

and the 3' UTR in grey (length 336 bases, nucleotide positions 1893-2228). The ORF region encoding the ZNFD is depicted in blue (length 198 bases, encoding 66 amino acids; nucleotide positions 729-926). The ORF region encoding the KNID is depicted in orange (length 57 bases, encoding 19 amino acids; nucleotide positions 945-1001). The ORF region encoding a conserved PIDLS motif is depicted in pink (length 36 bases, encoding 12 amino acids; nucleotide positions 1716-1751).

```
0001                                AGTTGTGCTCGTGCTGGCCGAAAGAGTGCATACTCACGCGTCCGAGCTGTGGTTTGCACTTCCTGGCA
0069          GGATATCTACTGGATGCAATAGCATTGCAGTTACACATTGATATCATTTTGAGGAAATTCACCGTTACCCTAATTTCATAATCTGGGAGC
0159          AGGAATTTATCGCCCGCCACAGAGTTGGCGTGTATGTCTTCCCTTGATTGCCACTAAATGTAATTGAACATATTTTTGACACGGATGGTA
0249          ACCGCAAAAATCGCCTAACTGTACTGAAGCCGCCATGAAGAGCAAATGTCCAAGTGCCCCAAGAAGCTTGCATCTCTTTGCTGGATCGTC
0339          ATTCGACTAGAATATCACTCTGTGGAAAGTCAAGTTGTGCTGAGTGTCTGTTGAACTATTAACCAAGGCATCGCTAATATAAAGTGCCAG
0429          TGTAGGAAAAAAAGTTACTGTTGAACCAAACTTACAGTCTACAAAACTTTTATAGAAAGATAATTGTTTATAATTTTGTGTTGAATTGAA
0519          AACAATAGGACAACATCCTTGAAATTCAGCAATTATTTTTTCACACTGTTAAAGACTCAATACTGACAGTAATTCAAACTTTGAAATGCT
0609          GTCGTCATATCTTGGGCAGCTGTAGTTATTTGAAGCACAGTTTTTTGAAAATTAAGTCAAGTGTCACTAAGCAGAAATCCTCGAGTCTGC

0001          M  T  V  A  N  Y  L  R  M  N  Q  L  C  K  V  C  S  E  P  A  A  G  F  H  F  G  A  F  T  C
0699          ATGACAGTCGCAAATTATCTCAGGATGAATCAGTTGTGCAAGGTGTGCTCTGAGCCGGCAGCAGGATTCCACTTTGGAGCTTTTACGTGT

0031           E  G  C  K  S  F  F  G  R  T  Y  N  N  L  S  Q  I  H  E  C  K  N  G  G  Q  C  V  I  N  K
0789          GAAGGCTGCAAGTCCTTCTTCGGGAGAACATACAACAACTTATCCCAAATTCATGAATGCAAGAATGGAGGCCAGTGTGTGATCAACAAG

0061           Q  N  R  T  S  C  K  A  C  R  L  R  K  C  L  F  V  G  M  S  K  S  G  S  R  Y  G  R  R  S
0879          CAAAATCGAACTTCGTGTAAGGCTTGTCGCCTTCGGAAATGCCTTTTCGTTGGAATGTCCAAAAGTGGATCAAGATATGGGAGAAGATCA

0091           N  W  F  K  I  H  C  A  L  Q  E  Q  A  A  N  L  N  R  S  T  E  G  A  G  N  I  N  L  Q  L
0969          AACTGGTTTAAAATTCACTGCGCACTCCAAGAGCAAGCGGCCAACTTGAACAGAAGCACGGAGGGTGCTGGGAATATCAATTTGCAGCTG

0121           S  A  A  R  E  N  L  I  R  L  S  P  S  R  C  V  D  D  E  G  K  L  G  S  P  A  L  S  S  P
1059          AGCGCAGCACGGGAAAACCTCATTAGATTATCCCCAAGCAGGTGTGTCGACGACGAAGGCAAGCTTGGTTCCCCAGCACTTTCATCTCCT

0151           D  S  N  T  S  D  T  S  V  D  V  E  P  R  R  N  I  S  G  T  S  I  Q  T  S  H  S  Q  L  T
1149          GATAGCAACACGTCGGACACAAGTGTAGATGTCGAACCTAGAAGAAATATATCTGGTACTTCTATTCAGACTTCGCATTCGCAACTAACA

0181           L  A  Q  H  E  S  L  V  A  R  E  L  S  K  R  T  G  N  A  L  A  T  V  G  G  R  S  L  E  P
1239          CTCGCCCAACACGAATCTTTAGTAGCTCGGGAACTTTCCAAAAGAACAGGAAATGCACTAGCTACTGTAGGAGGGAGAAGTCTGGAGCCA

0211           P  V  P  G  T  D  F  V  A  G  L  S  L  Y  G  I  H  S  G  F  S  L  L  Q  A  S  Q  L  Y  Q
1329          CCAGTTCCCGGTACTGATTTCGTTGCTGGGCTTTCGCTTTATGGAATTCACTCCGGCTTTTCGTTACTACAAGCATCACAATTATACCAG

0241           R  L  Y  P  P  F  L  Y  P  G  I  R  P  V  M  D  P  I  A  P  P  Q  Q  S  P  S  P  T  Q  I
1419          CGTTTATACCCACCGTTTTTGTACCCAGGTATCCGGCCTGTAATGGACCCAATTGCACCACCTCAACAATCGCCATCACCTACTCAAATA

0271           K  S  L  P  I  K  S  P  S  N  P  P  S  P  A  H  S  L  P  P  P  S  T  K  R  E  Q  V  S  N
1509          AAATCTCTACCTATTAAGTCCCCCTCCAATCCACCCTCTCCTGCCCATTCTTTGCCTCCCCCGTCGACAAAAAGGGAGCAAGTATCGAAC

0301           P  S  S  P  Q  S  L  P  A  R  H  F  S  D  R  T  A  H  S  P  S  S  S  S  P  S  P  P  L  I
1599          CCATCTTCACCTCAATCTTTACCTGCCCGTCATTTCTCGGACCGAACGGCACACTCTCCGTCCTCCTCATCTCCATCGCCGCCCTTAATT

0331           S  R  K  R  A  A  D  S  P  P  E  Q  D  A  P  I  D  L  S  V  K  R  P  R  F  S  S  T  T  S
1689          TCCCGTAAACGAGCAGCTGACTCTCCTCCTGAGCAAGATGCTCCCATTGATTTAAGCGTGAAGCGACCGCGCTTTTCGTCAACCACTTCA

0361           V  P  L  P  D  L  T  P  I  P  L  P  F  N  F  L  H  L  A  A  P  S  T  S  S  T  P  S  V  P
1779          GTTCCATTGCCTGATCTCACCCCTATTCCGCTGCCTTTTAATTTTTTACATTTAGCAGCGCCATCAACATCTTCAACACCTTCGGTACCT

0391           V  D  L  T  S  N  A  &                                                                  0397
1869          GTAGACTTGACTTCCAATGCTTGATAGTTTCGGAATCGGTGATGCAAGAGCGTATGTATGTCTGCTTGACGCTTTTTCATTGCATTGGAT
1959          GCATATCGCTAGCAAAAGTCAAGGTCTCTTGTCACCAAGATTAGGTGACTGAAATTACAGCACTGCTCACATCAAAGACTGAAACAACGT
2049          CGAGAAATTCGTGCTGTAAATAGATGCTGTACTGTTACTGAGTAGCACAACCCGATAGCAAACTATTAGCCAGTTAAAAATTCTAAAATT
2139          GAGAATTCTACTAGTCCTAAGCTTGCACCCTACTGTTTTCAGAAGCCTCAGGGTGCAAGAAGTTACCTTGTATACTTCTGGTCGTGCCAG   2228
```

**Figure 68:** ***Ph kni1*** **cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph kni1* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The ZNFD is shown in blue, the KNID in orange and a conserved motif covering the amino acids PIDLS in pink. The cysteine residues within the ZNFD are highlighted in blue. The nucleotide sequences that encode these domains are shown in the respective colours. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

### 3.4.4 Isolation of *Ph kni2*

From two-step 5' RACE, five independent *Ph kni2* 5' cDNA fragment clones were obtained. They all show identical transcription starts. Any two of their sequences may vary up to 2%, which is slightly above average as compared to other RACE sequences recovered in this work. Still, this finding suggests that all sequences derive from transcripts of the same gene.

*Ph kni2* 3' cDNA sequence was fully recovered by two-step 3' RACE. Five independent *Ph kni2* 3' cDNA fragment clones were obtained. However, only from clone Ph_kni2_3R21 the complete sequence was retrieved. Minor sequence sections of the other clones remained of low quality despite repeated sequencing efforts. Since all parts of the entire 3' cDNA sequence are covered by at least two independent *Ph kni2* 3' cDNA fragment clones, the sequence information was considered sufficient for assembling the complete *Ph kni2* cDNA sequence *in silico*. Within the fraction of sequence they share, any two 3' RACE sequences vary in less than 1% of nucleotide positions and are therefore considered identical. There are minor variations in poly(A) tailing.

Sequence information provided by all recovered 5' and 3' *Ph kni2* RACE clones was sufficient for assembling the complete *Ph kni2* cDNA sequence *in silico*. In order to verify the consistency of the established *Ph kni2* cDNA sequence, coherent *Ph kni2* cDNA and ORF fragments were isolated via long-distance PCR reactions (5.2.3.4), performed on *Parhyale* cDNA collections (5.2.2). Two independent *Ph kni2* cDNA sequences and three independent *Ph kni1* ORF sequences were obtained (A2.12.1). They are consistent with the findings from 5' and 3' RACE (A2.12.2). Because the nucleotide variation between any two ORF or cDNA sequences, respectively, does not exceed 1%, they are considered identical (5.3.3.3, A3, table A1).

### 3.4.5 Characterisation of *Ph kni2*

The comparison of all recovered *Ph kni2* sequences led to the identification of 64 sites of nucleotide exchange within the complete *Ph kni2* cDNA, corresponding to 2.9% nucleotide variation. 60 of those occur uniquely and are therefore considered sporadic (5.3.3.2). They are found predominantly in the ORF (46) where they account for 28 amino acid changes

(missense alterations). The remaining 18 nucleotide exchanges found within the ORF do not alter the translated amino acid sequence (silent exchanges). These findings go well with the expected randomness of artificial nucleotide exchanges, suggesting that they are artefacts. The remaining four nucleotide exchanges are found in more than one independent sequence and are therefore considered polymorphic. They are predominant in the 5' UTR (3). The single polymorphic nucleotide exchange (A/G) present in the ORF does not alter the translated amino acid sequence. These findings strongly suggest that it represents a naturally occurring polymorphism of wild type *Ph kni2* alleles (5.3.3.2). There is no indication for the existence of differentially spliced *Ph kni2* transcripts or other types of *Ph kni2* transcript isoforms since the comparison of all recovered *Ph kni2* sequences supports a consistent *Ph kni2* transcript sequence.

Based on these findings, the sequence of the clone Ph_kni2_cJZ1209 was chosen as the source of the *Ph kni2* reference sequence (*Ph kni2_ref*). Sporadic and putatively artificial nucleotide exchanges were substituted with the 'consensus' nucleotides that are present in all other obtained sequences (546T>C, 837C>T, 1062T>C).It was extended 5' by adding 50 bp derived from clone Ph_kni2_5R01 and 3' by adding 32 bp derived from clone Ph_kni2_3R21 (A2.12.1). An alignment of *Ph kni2_ref* with all recovered *Ph kni2* sequences confirms its statistical consensus relevance (A.12.2). All recovered *Ph kni2* sequences show less than 1% difference from *Ph kni2_ref*, confirming that they in fact derive from transcripts of *Ph kni2*, except the sequence of clone Ph_kni2_5R03 (1.5%). *Ph kni2_ref* has been used for phylogenetic studies.

The *Ph kni2* transcript is 2.7 kb in length and encodes a protein of 692 amino acids. The *Parhyale* Kni2 protein has a typical C4-type N-terminal Zinc finger domain (ZNFD, 66 amino acids) with closest resemblance to Knrl ZNFD. Directly adjacent to the ZNFD, *Parhyale* Kni2 has a *knirps* box domain (KNID, 19 amino acids) with high sequence similarity to the KNID of described *kni* family proteins of other species. C-terminally, a conserved PIDLS motif is found within the hydrophobic region of the *Parhyale* Kni2 protein. This is typical for Kni proteins of arthropods (Figure 69, 70; 3.4.6).

**Figure 69: Schematic view of *Ph kni2* transcript.** The length of the *Ph kni2* transcript is 2656 bases. Shown are: 5' UTR in grey (length 244 bases, nucleotide positions 1-244), ORF in black (length 2079 bases, encoding 692 amino acids; nucleotide positions 245-2323) and the 3' UTR in grey (length 333 bases, nucleotide positions 2324-2656). The ORF region encoding the ZNFD is depicted in blue (length 198 bases, encoding 66 amino acids; nucleotide positions 251-448). The ORF region encoding the KNID is depicted in orange (length 57 bases, encoding 19 amino acids; nucleotide positions 467-523). The ORF region encoding a conserved PIDL motif is depicted in pink (length 36 bases, encoding 12 amino acids; nucleotide positions 1967-2002).

**Figure 70 (next page): *Ph kni2* cDNA and derived amino acid sequence.** The sequence is in FASTA format and represents the *Ph kni2* cDNA, derived from the mRNA transcript. 5' and 3' UTR are shown in grey, ORF in black. The translated amino acid sequence is printed bold and above the corresponding nucleotide sequence. Individual amino acids are above the central nucleotide of the respective codon. The putative start and stop codons are shown in green and red, respectively. The ZNFD is shown in blue, the KNID in orange and a conserved motif covering the amino acids PIDL in pink. The cysteine residues within the ZNFD are highlighted in blue. The nucleotide sequences that encode these domains are shown in the respective colours. Numbers to the left give the relative nucleotide and amino acid sequence positions and share the font parameters of the corresponding sequence. The ends of the amino acid and the nucleotide sequences are indicated by numbers to the right of the corresponding line.

```
0001                                GGAATTTTCACTGTTCCCTTTACCACGTTTTACTGAACAAATTCGGTTAAAAATTTACCCGTAC
0065        AAATAAGGGATGAAAATTGACAGATAATGAAGTTCGATGCTCCAGAATTAAGTCACTTTATTTAGACAGACGAAAGCTTTGTATTCGGCC
0155        ATCATTTTTCAAAATCAGGTTTTGAAGTTCCGGTTCAGAGGGCTTCATAGTCAACCTTTGGATGTGAAGAGAAACAAGATAGCAAGCGAG

0001        M  N  Q  L  C  K  V  C  A  E  P  A  A  G  F  H  F  G  A  F  T  C  E  G  C  K  S  F  F  G
0245        ATGAACCAACTGTGCAAGGTCTGTGCTGAACCTGCCGCTGGTTTCCATTTTGGAGCATTCACTTGCGAAGGATGCAAGAGTTTCTTCGGG

0031        R  T  Y  N  N  L  S  Q  I  H  E  C  K  N  G  G  Q  C  V  I  N  K  Q  N  R  T  S  C  K  A
0335        AGAACATACAACAACTTATCGCAAATTCACGAATGCAAGAACGGAGGCCAGTGTGTGATCAACAAACAAAACAGGACTTCGTGCAAGGCC

0061        C  R  L  R  K  C  L  V  V  G  M  S  K  T  G  S  R  Y  G  R  R  S  N  W  F  K  I  H  C  L
0425        TGTCGTCTCAGGAAATGTCTAGTTGTGGGCATGTCTAAAACGGGGTCACGGTACGGCCGCAGGTCTAATTGGTTCAAGATCCATTGTTTA

0091        L  Q  E  Q  C  G  Y  S  G  T  A  G  Q  L  P  T  F  R  S  P  V  T  P  T  L  Q  D  G  S  L
0515        CTACAAGAACAATGCGGGTACAGCGGCACTGCCGGGCAGCTACCAACGTTCCGCAGCCCCGTCACTCCCACACTGCAAGACGGCTCCTTA

0121        P  F  V  N  T  Q  L  R  D  A  L  R  M  D  V  P  L  T  P  V  E  H  S  A  S  P  S  P  R  I
0605        CCTTTCGTCAACACTCAGTTGAGGGACGCCCTAAGGATGGATGTGCCCCTAACGCCAGTCGAACACAGTGCTTCTCCCAGTCCTCGAATA

0151        P  I  T  S  A  A  P  G  G  V  C  I  R  P  P  T  S  V  E  M  G  P  A  G  V  T  V  S  E  P
0695        CCTATCACGTCTGCTGCCCCTGGGGGTGTGTGCATCCGACCCCCCACAAGTGTGGAGATGGGCCCAGCTGGAGTAACTGTCTCAGAACCG

0181        H  L  S  P  N  F  P  M  G  Q  I  G  A  V  E  R  R  I  Q  E  E  T  L  N  A  L  R  R  G  V
0785        CACTTGTCGCCAAATTTCCCCATGGGACAAATAGGAGCGGTTGAGCGTCGAATTCAGGAAGAAACTCTCAATGCACTACGGAGAGGTGTG

0211        G  M  P  G  S  E  L  E  A  T  Y  R  E  Q  I  L  E  A  L  R  R  S  H  N  M  Q  P  P  R  L
0875        GGAATGCCCGGAAGTGAACTCGAGGCGACGTATCGAGAACAAATTTTGGAAGCTCTCAGGCGGTCCCACAATATGCAGCCACCTCGGCTC

0241        L  D  P  S  A  V  Q  T  L  R  R  S  P  P  I  A  E  P  V  R  R  I  G  A  E  Y  E  R  V  R
0965        CTGGACCCCTCTGCGGTCCAAACGCTGAGAAGGTCACCGCCTATAGCTGAACCTGTAAGAAGAATAGGGGCTGAATACGAAAGAGTCCGA

0271        G  V  S  A  E  A  E  I  R  M  N  S  N  P  P  Y  D  W  D  P  R  L  Q  E  A  P  V  F  S  E
1055        GGAGTATCTGCTGAAGCTGAAATTAGAATGAATTCAAACCCTCCATACGACTGGGATCCCAGATTGCAAGAAGCTCCGGTGTTCAGTGAA

0301        A  S  V  V  K  S  E  S  P  H  S  S  K  V  A  L  S  V  S  V  S  T  S  Q  E  V  D  S  S  S
1145        GCGTCGGTAGTGAAATCTGAATCTCCTCATAGTTCAAAAGTTGCTCTCAGTGTAAGTGTTAGCACTTCTCAAGAAGTAGACAGTAGCAGC

0331        S  N  Q  I  K  I  E  C  S  R  S  E  T  E  T  G  S  K  S  T  P  E  V  K  S  E  R  R  S  V
1235        TCAAATCAAATTAAAATAGAGTGCAGCCGGAGTGAAACTGAAACAGGTAGTAAAAGTACTCCAGAGGTTAAATCAGAAAGAAGGTCAGTA

0361        G  N  R  S  P  V  S  L  T  P  K  S  Q  A  S  A  R  L  Q  D  E  T  G  A  P  M  T  E  E  E
1325        GGAAATCGTTCTCCTGTAAGTCTCACTCCAAAGTCTCAGGCCAGTGCTAGATTACAAGATGAGACAGGTGCTCCCATGACCGAAGAAGAA

0391        R  L  S  F  L  K  S  A  S  K  I  S  A  N  N  T  A  T  R  D  G  S  A  V  N  S  F  H  G  I
1415        CGATTATCTTTCCTGAAGTCAGCCAGTAAAATATCTGCAAACAATACAGCTACACGCGATGGGTCAGCTGTTAACAGTTTTCATGGCATC

0421        P  G  L  F  P  H  R  S  S  S  R  D  D  I  L  Y  G  P  L  N  N  N  I  F  P  F  A  S  Y  P
1505        CCTGGCTTATTTCCTCATCGCTCCTCTTCCAGAGACGATATTCTTTATGGTCCTTTGAACAACAACATTTTCCCATTTGCATCGTATCCT

0451        P  V  W  P  S  I  L  P  F  P  S  L  A  K  I  Y  P  P  P  P  F  S  S  P  A  V  S  G  A  P
1595        CCAGTATGGCCTTCAATTCTTCCATTCCCTTCCCTGGCAAAAATCTATCCTCCACCTCCATTTAGTTCACCTGCGGTTTCTGGAGCACCA

0481        I  S  L  D  G  F  K  S  P  L  Y  L  P  T  E  L  R  N  T  V  L  P  N  T  L  P  D  V  T  K
1685        ATCTCACTTGATGGTTTCAAGAGTCCCCTGTATTTGCCAACTGAGCTCAGAAACACAGTGTTGCCAAACACATTGCCGGATGTTACAAAA

0511        E  D  S  S  N  K  K  R  I  L  D  A  I  L  Q  V  Q  R  E  S  A  C  G  P  R  S  P  L  G  G
1775        GAAGATTCTAGCAACAAGAAAAGAATTTTAGATGCTATCCTTCAGGTTCAGAGAGAAAGTGCATGTGGACCTCGATCTCCTCTGGGAGGA

0541        L  S  P  P  G  V  T  L  A  L  S  L  P  Q  H  V  T  T  S  T  S  T  L  P  M  V  T  A  T  T
1865        TTATCTCCTCCTGGAGTGACGCTCGCGCTAAGCCTTCCCCAGCACGTGACGACATCGACATCTACGCTGCCGATGGTGACAGCCACAACG

0571        T  T  S  T  A  P  G  E  Q  P  I  D  L  T  V  R  R  K  R  K  M  A  K  R  D  I  M  Y  Q  P
1955        ACGACATCCACAGCACCGGGTGAGCAGCCCATCGATCTCACGGTTCGGAGAAAGAGAAAAATGGCCAAGCGTGACATTATGTACCAGCCA

0601        N  E  S  D  P  Q  E  G  M  E  D  E  Y  E  Q  E  Q  K  D  D  A  D  P  K  E  E  S  I  L  K
2045        AACGAAAGTGATCCTCAGGAAGGCATGGAAGATGAGTATGAACAAGAACAGAAAGACGACGCAGATCCGAAGGAAGAATCTATCTTGAAA

0631        D  E  E  V  L  E  E  E  P  V  D  E  R  V  S  K  E  I  L  V  K  G  T  S  E  I  V  D  L  K
2135        GACGAAGAGGTTTTGGAAGAGGAGCCTGTGGACGAAAGGGTAAGCAAAGAAATCCTAGTGAAAGGAACTTCAGAGATTGTCGATCTTAAA

0661        T  S  L  E  C  A  Q  S  L  P  E  P  R  F  E  V  P  M  K  I  M  K  L  E  T  A  G  D  G  L
2225        ACGAGTTTGGAGTGTGCCCAGTCCCTGCCCGAGCCGAGGTTCGAGGTGCCAATGAAGATAATGAAATTAGAAACTGCTGGTGATGGCCTC

0691        I  T  &                                                                                  0692
2315        ATCACCTGACGCCTGCTAAGTTAATTCACAGAGTCTCATGGGAGACTGTAATGTGCCATATTATGTAAGGTTTTGTTAAGTTACATTCTC
2405        AGATGAGTGTAGGGAGCGGCGGGTAAAGTATTACAGATAATGCAATGAGAGCCATATGAACTGCATGAGATGTGTACAGTGTCTTGCGAT
2495        CAATACACTGGCTCCTTACTGAGAATATAATTATGTTACAAATGCCATTATTCCAGTAAACTGCTCAGTAACTGTAAGGACTGAATTGAG
2585        TAGTTAAGTCTAACTGTTGTTTATTGTCTAAATTTCTATTGCTTAGCAGCTATTTCTTATAAATTCATTTCC                2656
```

**Figure 70:** ***Ph kni2*** **cDNA and derived amino acid sequence.**

### 3.4.6 Phylogeny of *Parhyale hawaiensis knirps* genes

Importantly, *knirps* genes have so far only been found in arthropods. Even within insect species with well-studied genomic data available, individual repertoires of *knirps* genes vary considerably (Cerny et al., 2008; Rothe et al., 1989). BLAST searches (Altschul et al., 1997) using the complete amino acid sequence of *Parhyale* KNI1 as query against the protein databases *of Drosophila melanogaster* (blastp, taxid 7227) and mouse (blastp, taxid 10090) revealed highest sequence similarity to *Drosophila* EG and murine estrogen receptor 1 (Esr1), respectively. For *Parhyale* KNI2, an analogous BLAST search yielded *Drosophila* EG, as well, and murine estrogen receptor 2 (Esr2). Independently, a BLAST search was performed using the complete amino acid sequences of *Parhyale* KNI1 and KNI2 as query against the complete translated nucleotide database (tblastn). From this search, the putative *Apis mellifera* KNRL (PREDICTED: *Apis mellifera* similar to Knirps-related protein (LOC412475), mRNA; NCBI Reference Sequence: XM_395932.2) showed highest sequence similarity to *Parhyale* KNI1. *Drosophila mojavensis* GI11566 (Dmoj\GI11566, mRNA; NCBI Reference Sequence: XM_002008848.1; highest sequence similarity to Drosophila melanogaster EG) showed highest sequence similarity to *Parhyale* KNI2.

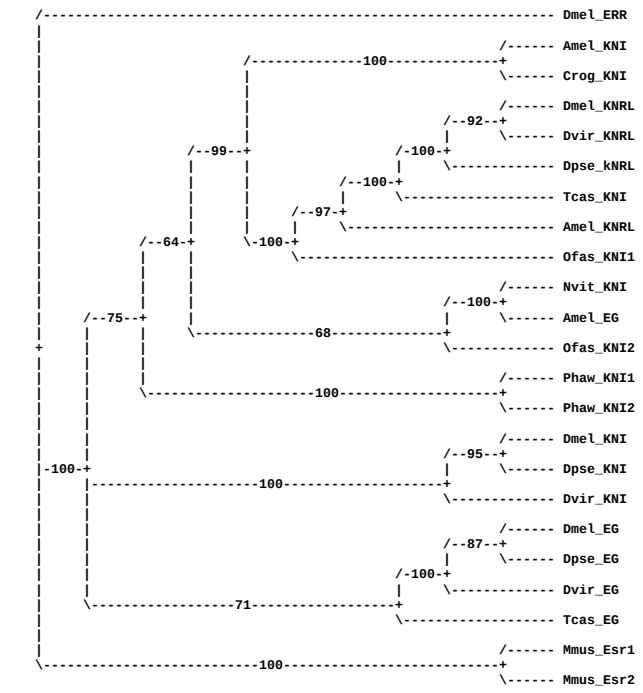
Arthropod *knirps* genes encode proteins that carry a Zinc-finger domain and an adjacent *kni* box that has not been found in any other transcription factors or within deuterostomian species. Accordingly, a BLAST search (tblastn) using the *Parhyale* KNID as query against all available deuterostomian databases did not reveal anything of significant similarity. In contrast, an individual arthropod species may have a repertoire of up to three different *kni* genes (Rothe et al., 1989). In order to find out if *Ph* KNI1 and *Ph* KNI2 resemble orthologs of individual insect KNI proteins, the phylogenetic relation of *Ph* KNI1 and *Ph* KNI2 to different KNI-family proteins of a wide range of arthropod species was calculated (Mr Bayes, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The phylogenetic calculation was based on an alignment of *Ph* KNI1 and *Ph* KNI2 with other arthropod KNI proteins using the CLUSTALW2 algorithm (Larkin et al., 2007). Specifically, a coherent fragment of the alignment covering the translation starts and reaching to the end of the PIDL motives was used for calculating the phylogeny (A6.3). Both murine estrogen receptors that represent the closest relatives of the *Parhyale* KNI proteins in mouse were included along with their *Drosophila* ortholog (*Drosophila melanogaster* estrogen-related receptor, ERR). In the resulting phylogram, murine Esr1, murine Esr2 and *Drosophila* ERR form an outgroup

(Figure 71). The arthropod KNI proteins constitute two well-defined groups, the dipteran KNIRPS (in contrast to KNIRPS RELATED) proteins and the dipteran and coleopteran EG proteins. The remaining dipteran KNRL proteins are grouped together in a defined sub-branch within the large and divergent group of all other arthropod KNI representatives (Figure 71). The phylogenetic structure of this divergent group indicates several independent *knirps* gene duplications. Interestingly, the phylogram shows with absolute confidence that *Ph kni1* and *Ph kni2* derive from a genetic event specific for the amphipod lineage leading to *Parhyale* (Figure 71 A). In this regard, both proteins are co-orthologous to Drosophila KNRL.

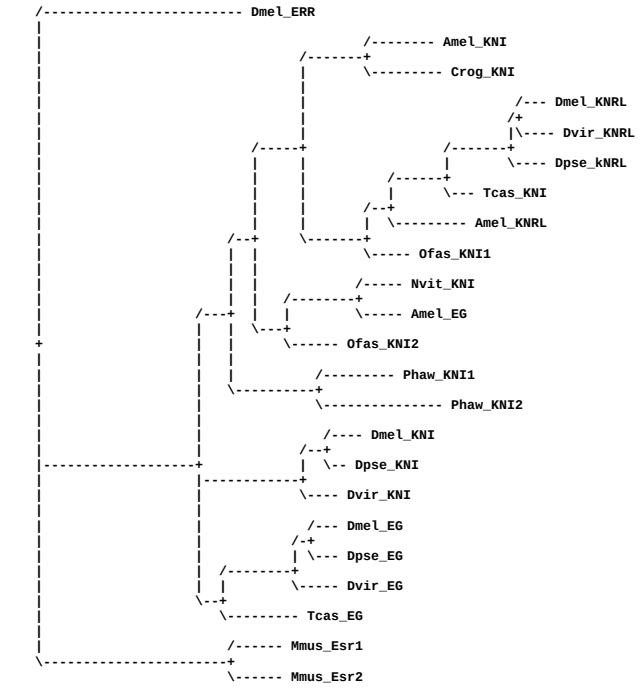
**Figure 71 (next page): Phylogenetic analysis of *Parhyale* KNI1 and KNI2 together with estrogen receptors and KNI proteins from other species.** Shown are: *Parhyale hawaiensis* KNI1 (Phaw_KNI1), *Parhyale hawaiensis* KNI2 (Phaw_KNI2), *Drosophila melanogaster* KNI (Dmel_KNI), *Drosophila melanogaster* KNRL (Dmel_KNRL), *Drosophila melanogaster* EG (Dmel_EG), *Drosophila melanogaster* ERR (Dmel_ERR), *Drosophila virilis* KNI (Dvir_KNI), *Drosophila virilis* KNRL (Dvir_KNRL), *Drosophila virilis* EG (Dvir_EG), *Drosophila pseudoobscura* KNI (Dpse_KNI), *Drosophila pseudoobscura* KNRL (Dpse_KNRL), *Drosophila pseudoobscura* EG (Dpse_EG), *Tribolium castaneum* KNI (Tcas_ KNI), *Tribolium castaneum* EG (Tcas_EG), A*pis mellifera* KNI (Amel_KNI), A*pis mellifera* KNRL (Amel_KNRL), A*pis mellifera* EG (Amel_EG), *Nasonia vitripennis* KNI (Nvit_KNI), *Oncopeltus fasciatus* KNI1 (Ofas_KNI1), *Oncopeltus fasciatus* KNI2 (Ofas_KNI2), *Caligus rogercresseyi* KNI (Crog_KNI), *Mus musculus* estrogen receptor 1 (Mmus_Esr1) and *Mus musculus* estrogen receptor 1 (Mmus_Esr1). Dmel_ERR was set as outgroup. For reference regarding the non-*Parhyale* protein sequences, see A6, tables A5 and A6. **A** Phylogram showing Clade credibility values. **B** Phylogram based on average branch lengths. Phylogenetic Analysis was done using Mr Bayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; 5.3.2; 300,000 generations).

**Figure 71:** **Phylogenetic analysis of *Parhyale* KNI1 and KNI2 together with estrogen receptors and KNI proteins from other species.**

**A**




**B**



### 3.4.7 Expression of *Ph kni1*

Expression of *Ph kni1* is first detectable in stage 14 embryos. During this stage, the head lobes begin to condensate. The medial cleft that separates the head lobes starts to close as new cells arise at the level of the An2 segment (Figure 72 A3, B3; 3.1.2, Browne et al., 2005). *Ph kni1* is expressed bilaterally in clusters of cells within the An1 and An2 segments (Figure 72 A1-A3, B1-B3). These cell clusters represent the respective segmental appendage anlagen at the onset of their distal projection. Similar expression of *Ph kni1* is found in the appendage fields of the Mn segment, albeit at comparatively low level (Figure 72 C1-C3). A group of cells that is situated medial to the An2 and Mn appendage fields and that expands anterior to the closing medial cleft also shows *Ph kni1* expression (Figure 72 A1-A3, B1-B3).

Embryos of stage 15 are characterised by the eventual closure of the medial cleft which results in the fusion of the head lobes at the level of the An1 and An2 segments and the lateral projection of the pre-antennal hemispheres. The appendage buds of the An1, An2 and Mn segments are clearly visible as distal outgrowths from the embryonic head (Figure 73 A3, D3; R1.3; Browne et al., 2005). Novel *Ph kni1* expression domains have arisen. Individual cells in the medial, medial anterior and lateral parts of the pre-antennal region of the embryonic head show expression of *Ph kni1*. The level of *Ph kni1*-expression in the lateral cells appears highest (Figure 73 B1-B3, D3). Expression of *Ph kni1* in the developing appendage fields of An1, An2 and Mn is maintained (Figure 73 A1-A3, B1-B3). Within the appendage fields of the mandibular segment, expression of *Ph kni1* appears to be split into two rows that consist of three to four cells aligned in an anterior to posterior fashion (Figure 73 A2). The *Ph kni1* expressing cells that are located medially to the appendage fields of the An2 and Mn segments have clustered more tightly (Figure 73 A1-A3). At the posterior end of the embryo, expression of *Ph kni1* becomes first visible in a medially located cell during this stage (Figure 73 C1, C3). This area flags the proctodeal development field.

At stage 17, the elongating germ band has begun to fold inwards toward the egg interior at segment T5. This posterior flexure is visible as a clear furrow in the germ band (Figure 74 D1-D3; E1-E3; R1.4; Browne et al., 2005). The medial pre-antennal expression of *Ph kni1* has expanded. It now resembles a coherent domain encompassing a large central composition of cells and small anterior extensions (Figure 74 A1-A3; C1-C3). The lateral pre-antennal expression domain of *Ph kni1* is constituted by an increased number of cells. Here, *Ph kni1* continues to be expressed at high levels (Figure 74 A1-A3; C1-C3; E1, E2). The appendage

fields of the An1, An2 and Mn segments show extensive *Ph kni1* expression (Figure 74 B1-B3; C1-C3). In the An2 appendage field, a smaller group of *Ph kni1*-positive cells is posterior and medially adjacent to the broader expression domain covering the complete appendage field (Figure 74 B1-B3; D1-D3). Several cells that lie between the appendage fields of the An2 and Mn segments remain positive for *Ph kni1* at a low level of expression (Figure 74 B1-B3; D2). Novel *Ph kni1* expression is detectable in the appendage fields of Mx2, T1, T2, T3 and T4. Its intensity is comparatively low in Mx2 and T4 (Figure 74 D1-D3; E1-E3). At the posterior end of the embryo, *Ph kni1* expression has expanded. It resembles a coherent, crescent-shaped domain at the site of the future proctodeum (Figure 74 F1-F3).

At stage 18, the flexure of the embryo is fully established (Figure 75 E3, G1-G3; 3.1.3; Browne et al., 2005). The stomodeum begins to invaginate. Two stomodeal projections become visible on lateral positions within the deepening stomodeal opening (Figure 75 A3, B3; 3.1.3, Browne et al., 2005). At the onset of stage 18, the expression profile of *Ph kni1* in the head of the embryo exhibits a complex topology. *Ph kni1* is expressed bilaterally in small, isolated groups of cells at the lateral periphery of the pre-antennal part of the head (Figure 75 B1', F1, F2) as well as in larger groups of cells that lie at the anterior, dorsal end of the head hemispheres (Figure 75 A1-A3, B1', B3', F1-F3, G2). The medial part of the pre-antennal region that borders the An1 segment shows an extensive, distinct pattern of *Ph kni1* expression that is composed of several bilateral clusters of *Ph kni1*-expressing cells and reaches into the An1 segment (Figure 75 A1-A3, B1-B3). A medial group of cells that lies on the level of the An1/An2 segmental border flags the site of the invaginating stomodeum. This group of cells itself is free of *Ph kni1* expression. It is surrounded by four clusters of *Ph kni1*-positive cells. Laterally on both sides of the developing stomodeal field, two cell clusters exhibit strong expression of *Ph kni1*. They represent the lateral stomodeal projections that arise during this stage. Directly anterior to the stomodeal field, a unilateral group of cells expresses *Ph kni1*. They directly abut the more anterior area that gives rise to the labrum. A fourth group of *Ph kni1*-expressing cells abuts the stomodeal field posterior (Figure 75 A2, A3, B1-B3). Within the An1 appendages, the proximal articles show *Ph kni1* expression (Figure 75 A1-A3, B1-B3). The An2 appendages show strong *Ph kni1* expression in all cells except the most distal ones. Additionally, at the base of each An2 appendage, one group of cells that is located anterior, one that is located posterior and a third, smaller group of cells that lies medial within the An2 segment also exhibit *Ph kni1* expression (Figure 75 B1-B3). In the Mn appendages, *Ph kni1* expression is found in all peripheral cells, with the central part of

the appendage being free of *Ph kni1* expression. Similar to the An2 segment, three groups of cells that show *Ph kni1* expression are found at the base of each Mn appendage, one anterior, one posterior and a third group of cells that lies medially to the others (Figure 75 C1-C3). In the segments that lie posterior to Mn, *Ph kni1* expression is found exclusively in the appendage anlagen (Figure 75 C1-C3). As these appendages elongate, the *Ph kni1* expression domains gain the shape of distally situated crescents, covering two to three cell rows of peripheral appendage cells while excluding the most distal tips of the appendages (Figure 75 D1-D3, E1-E3, F1-F3). In the proctodeal development field at the posterior end of the embryo, *Ph kni1* is strongly expressed in two large bilateral domains and a smaller medial domain that lies anterior to those. Cells that are located between these domains also show *Ph kni1* expression, albeit comparatively weaker (Figure 75 E1-E3).

At the onset of stage 19, the lateral parts of the pre-antennal lobes have expanded. As a consequence, they now appear wedge-shaped and extend in part to the dorsal face of the egg. The labrum has protruded distally and begins to elongate towards the stomodeal opening. The paragnaths become visible just medial to the mandibles (Figure 76 B3; C2, C3; D2, D3; 3.1.3; Browne et al., 2005). *Ph kni1* expression within two groups of cells at the anterior end of each head hemisphere is maintained. However, following the dorsal expansion of the head lobes, the small, isolated groups of *Ph kni1*-positive cells now lie anterior to the larger groups of cells as compared to stage 18 (Figure 76 A1, A2). The labrum is free of *Ph kni1* expression (Figure 76 B1-B3). The expression pattern of *Ph kni1* in the medial part of the pre-antennal region remains comparable to the previous stage. Due to the dorsal expansion of the pre-antennal region, it has taken shape of two parallel anterior-posterior stretches of cells that show strong *Ph kni1* expression. These stretches include the bases of the An1 appendages (Figure 76 C1-C3) and are separated by a central region of cells that does not express *Ph kni1* (Figure 76 B1-B3). Similar to stage 18 embryos, *Ph kni1* expression is found in the lateral stomodeal projections as well as in unilateral groups of cells that lie at the anterior and posterior of the invaginating stomodeum (Figure 76 B2, B3; C2, C3). As in embryos of stage 18, cells within the proximal article of the An1 appendages show *Ph kni1* expression (Figure 76 C1-C3). The An2 appendages are now composed of three elements. *Ph kni1* expression is found in cells that connect the first and second appendage articles (Figure 76 C1-C3). In addition, strong *Ph kni1* expression is maintained in two groups of cells that are located at the base of the An2 appendages, one anterior and the other one posterior (Figure 76 D1-D3). In the Mn appendages, *Ph kni1* expression is found in all peripheral cells, with the central part of

the appendage being free of *Ph kni1* expression (Figure 76 E1-E3). Directly at the base of each Mn appendage, strong *Ph kni1* expression is found in one anterior and one posterior group of cells. Medial to the Mn appendage bases, two groups of *Ph kni1*-positive cells flag the emerging paragnaths (Figure 76 D1-D3, E1-E3). Like in stage 18 embryos, in segments that lie posterior to Mn *Ph kni1* expression is found exclusively in the appendage anlagen (Figure 76 E1-E3). As these appendages have elongated, *Ph kni1* is expressed in broad, sheath-like domains, excluding the most distal parts of the appendage (Figure 76 F1-F3). *Ph kni1* expression in the proctodeal development field has decreased considerably as compared to stage 18 (Figure 76 E1-E3, G1-G3).

At stage 20, the labrum is beginning to expand over the stomodeal opening. Posterior of the labrum, the lower edge of the stomodeum is visible as a raised medial ridge. The An2 appendage has undergone rapid elongation (Figure 77 B2, D2, F2, 3.1.4, Browne et al., 2005). As in the previous stage, *Ph kni1* expression is still found within small groups of cells at the anterior end of each head hemisphere (Figure 77 A1, A2). Within the distal tip of the labrum, weak *Ph kni1* expression has arisen (Figure 77 B1, B2; D1, D2). In the medial part of the pre-antennal head, *Ph kni1*-positive cells now surround the stomodeal opening (Figure 77 B1, B2; C1, C2; D1, D2). In contrast to stage 19 embryos, the bases of the An1 appendages are free of *Ph kni1* expression (Figure 77 A1, A2; F1, F2). At the bases of the An2 appendages as well as the appendages of the Mn, Mx1 and Mx2 segments, expression of *Ph kni1* is maintained (Figure 77 B1, B2; C1, C2; E1, E2). Accordingly, *Ph kni1* expression is visible in all segmental appendages posterior to the Mx2 segment. In contrast to the previous stage, it is restricted to the proximal parts of the thoracic appendages. However, in the more posterior, abdominal segments, *Ph kni1* expression appears sheath-like (Figure 77 E1, E2; F1, F2). At this stage, the gill anlagen of the T2-T6 appendages are visibly branching off the most proximal, i.e. coxal appendage articles. They exhibit slight *Ph kni1* expression (Figure 77 E1 and E2, red arrowheads pointing at the gill anlagen of T3). Strong *Ph kni1* expression is found in the proctodeal development field (Figure 77 C2; D1, D2; E2). Novel expression of *Ph kni1* can now be found in small groups of cells that are located dorsal of the maxilliped anlagen, in the lateral regions of the T1 segment (Figure 77 C1, red arrowhead; E1, E2).

In stage 22 embryos, the labrum has completed its posterior extension. Its lobes directly abut the paragnaths anterior, with the mandibles prominently protruding laterally on both sides of the labrum. The labrum covers the entire opening of the stomodeum. The foregut has elongated inwards the thickened frontal head region (Figure 78 B3, C3, D3; 3.1.4; Browne et

al., 2005). *Ph kni1* is expressed in manifold areas and tissues of the embryo. *Ph kni1* expression is found in two lateral stripes within the labral lobes (Figure 78 A1, A2, small images). Within the developing head, *Ph kni1* is expressed in nodular cell clusters that surround the stomodeal opening (Figure 78 G1-G3, Figure 79 A1-A3). In the T1 segment, *Ph kni1* is expressed in a posterior, lateral stripe that is composed of two rows of cells. This stripe extends from the base of the T1 appendage (maxilliped) laterally to the dorsal brink of the T1 segment (Figure 78 B1-B3, E1-E3). This area separates the segments of the embryo that later form the morphologically distinct head capsule from the pereon. In most appendages, including the An1 and the thoracic appendages, *Ph kni1* is expressed in cells that connect individual appendage articles (Figure 78 B1-B3, C1-C3, D1-D3). *Ph kni1* expression remains strong at the base of the An2 appendage and in the proximal An2 appendage article (Figure 78 B1-B3). *Ph kni1* expression domains are found within the Mn appendages, the paragnaths, the Mx1 and Mx2 appendages and the developing maxillipeds (T1) (Figure 78 C1, C2, D1, D2). The proctodeum shows strong *Ph kni1* expression (Figure 78 C1-C3, D1, D2, G1-G3). In those appendages that develop gills (T2-T8), *Ph kni1* is expressed at the branching point between coxa and gill anlage and extends to the proximal part of the gill. In this regard, *Ph kni1* expression appears most intensive in the gills of T3-T7 (Figure 78 B1-B3, G1, Figure 79 C1-C3, D1-D3, the read arrowheads point towards the proximal part of the T4 gill anlage). In broad sheets of medial, ventral cells of the thoracic and abdominal segments, strong *Ph kni1* signal is apparent (Figure 78 C1-C3, D1-D3, Figure 79 B1-B3, the red arrowhead shows *Ph kni1*-positive cells in the T1 segment). At this stage, however, these areas have begun to produce cuticle. Therefore, it cannot be excluded that this signal is, at least partially, unspecific and due to incomplete removal of cuticle during dissection.

At stage 23, the embryo's head has started to close dorsally (Figure 80 E1, E2; 3.1.4; Browne et al., 2005). In nuclear stainings of anterior views of stage 23 embryos, well-defined morphological structures have become visible in the proto- and deutocerebral derivates of the embryonic brain (3.1.4; Browne et al., 2005). The pattern of *Ph kni1* expression is highly similar to that of stage 22 embryos (Figure 80 A1, A2; B1, B2; C1, C2). Compared to the previous stage, *Ph kni1* expression in cells that are located at branching points of appendages and between individual appendage articles is more apparent (Figure 80 C1, C2; D1, D2). In addition, it is now found in all appendages (for *Ph kni1* expression in abdominal appendages, see Figure 80 B2). An exception is the first antennae, where *Ph kni1*-positive cells are found only within the two most proximal elements of the appendage (Figure 80 C1, C2).

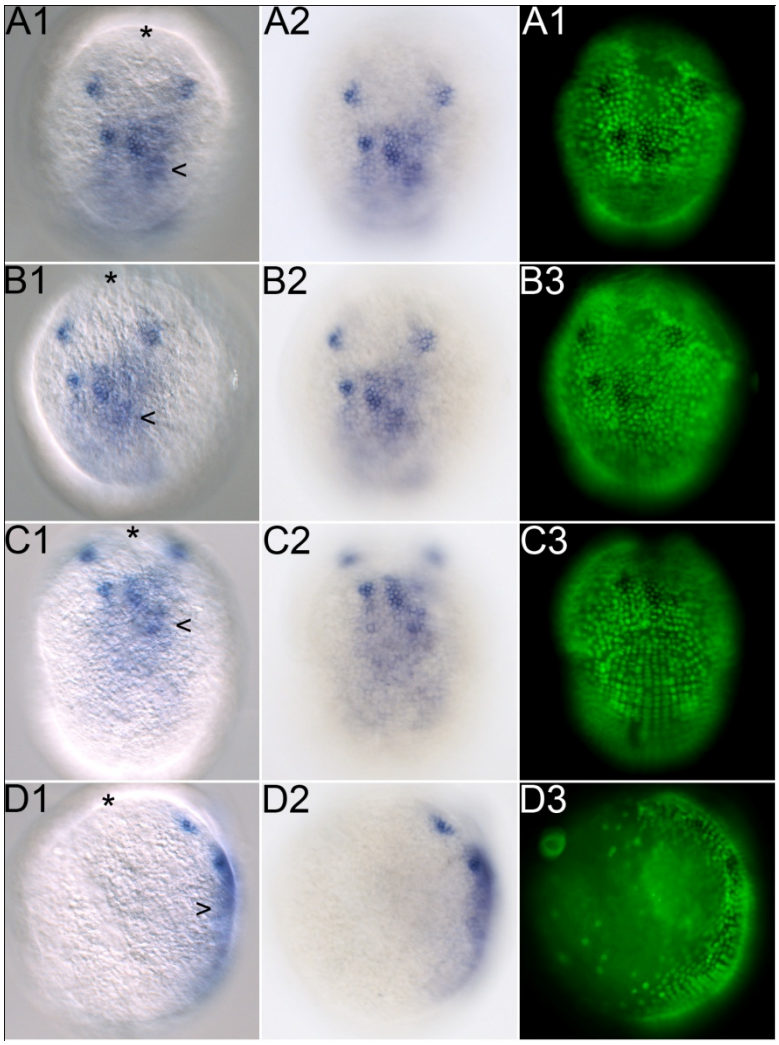



**Figure 72: Expression of *Ph kni1* at stage S14**. The embryo is mounted in anterior orientation (A1-A3), in anterior, tilted orientation (B1-B3), in ventral orientation (C1-C3) and in lateral orientation (D1-D3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni1* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2-D2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–D3; nuclei are green, expression of *Ph kni1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph kni1* is expressed in the appendage anlagen of An1 and An2 . **B** *Ph kni1* is expressed in a group of cells medial to the An2 and Mn appendage fields and expanding anterior to the closing medial cleft. **C, D** Low-level expression of *Ph kni1* is found in the appendage fields of the Mn segment.

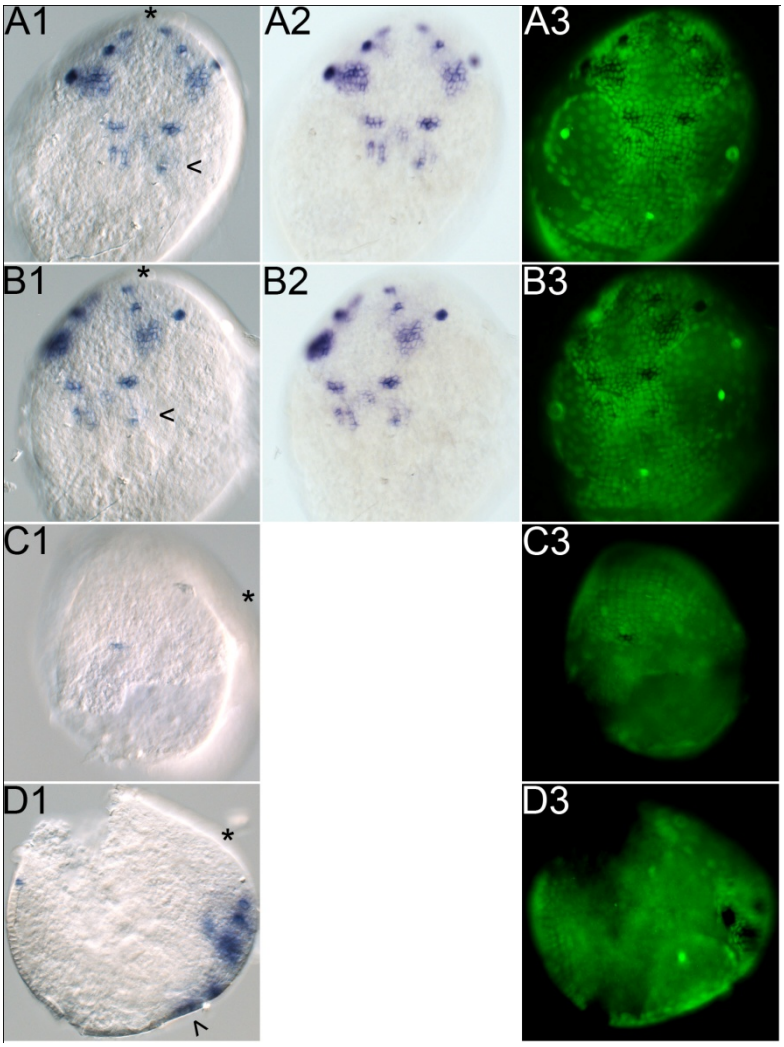



**Figure 73: Expression of *Ph kni1* at stage S15**. The embryo is mounted in ventral orientation, centred up the An2 and Mn segments (A1-A3), in ventral, tilted orientation (B1-B3), in posterior orientation (C1, C3) and in lateral orientation (D1, sagittal focal plane and D3, focal plane on the An1 appendage; in both images, anterior is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni1* expression. In images C1 and C3, ventral is up. In images D1 and D3, dorsal is up. In all other images, anterior is up. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2, B2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–D3; nuclei are green, expression of *Ph kni1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead.

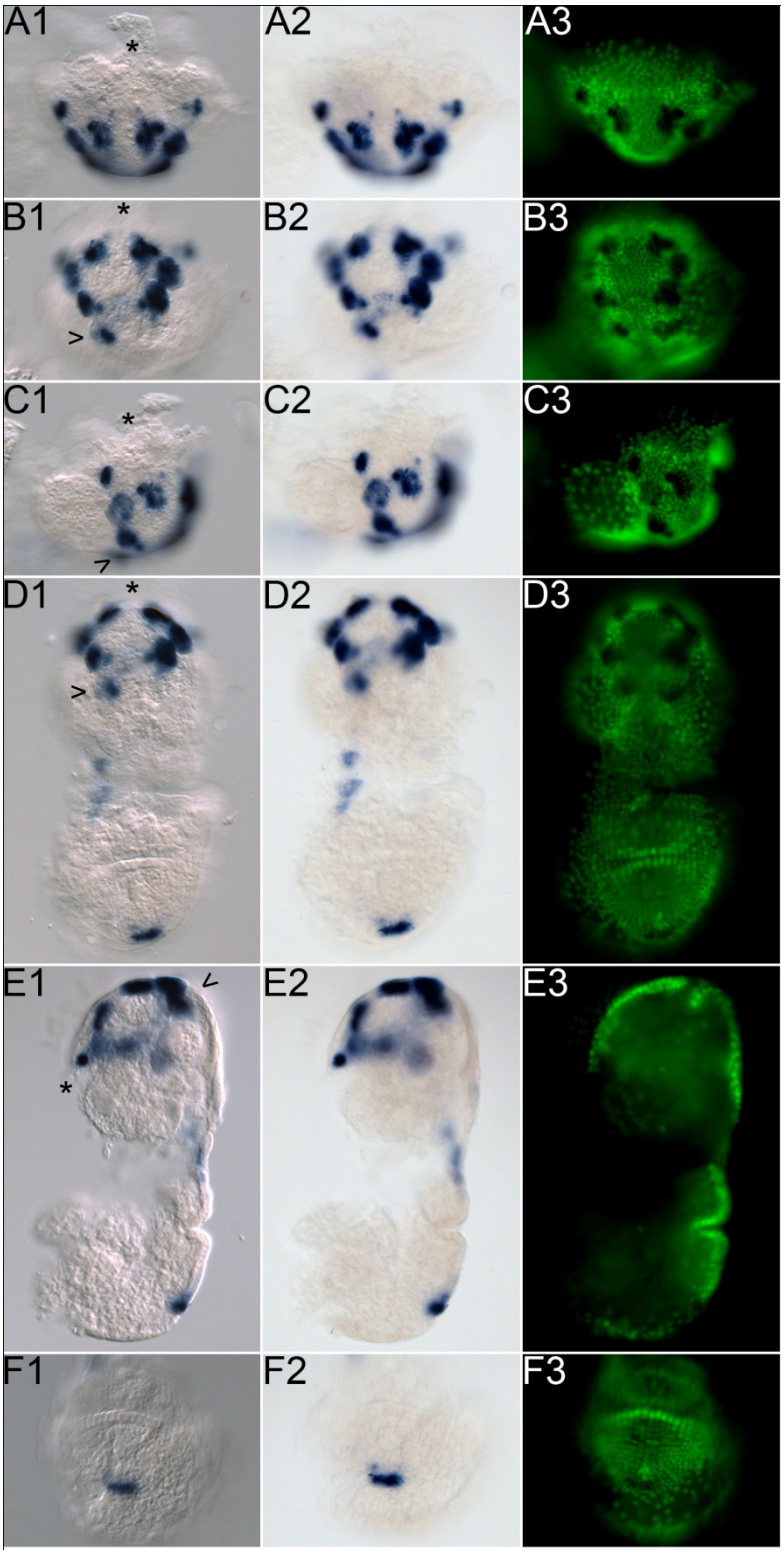



**Figure 74: Expression of *Ph kni1* at stage S17**. The embryo is mounted in anterior orientation (A1-A3), in anterior, ventral orientation, centred up the An2 and Mn segments (B1-B3), in anterior, tilted orientation (C1-C3), in ventral orientation (D1-D3), in lateral orientation, sagittal focal plane (E1-E3, ventral is to the right) and in posterior orientation (F1-F3) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni1* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-F1), bright field images in the middle panel (A2-F2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–F3; nuclei are green, expression of *Ph kni1* is detectable as it quenches the nuclear labelling). In the DIC images A1-E1, the anterior end of the embryo is indicated by a black asterisk. If visible, the position of the Mn segment is marked by a black arrowhead.

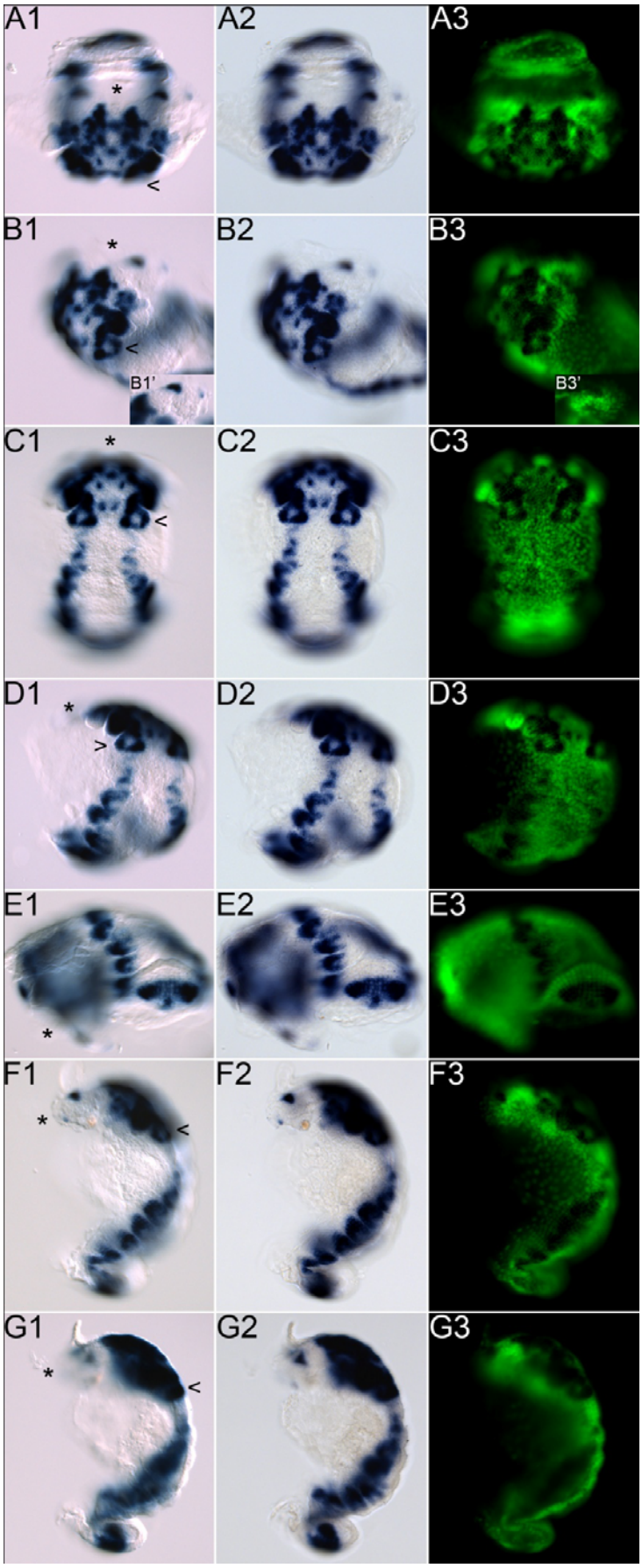



**Figure 75: Expression of *Ph kni1* at stage S18**. The embryo is mounted in anterior orientation (A1-A3), in anterior, tilted orientation (B1-B3), in ventral orientation (C1-C3), in ventral, tilted orientation (D1-D3), in posterior, slightly tilted orientation, focus on the ventral flexure (E1-E3), in lateral orientation (F1-F3, ventral is to the right) and in lateral orientation, sagittal focal plane (G1-G3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni1* expression. With exception of images E1-E3, where the ventral side is up, anterior is up in all images. Shown are DIC images in the left panel (A1-G1), bright field images in the middle panel (A2-G2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – G3; nuclei are green, expression of *Ph kni1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-G1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph kni1* is expressed in two lateral stripes within the labral lobes (see also small images). **B, C** *Ph kni1* is expressed in numerous parts and cell groups of the An1, An2 and Mn appendages. **D** In the segmental appendages posterior to Mn, crescent-shaped *Ph kni1* expression domains are found in peripheral rows of cells, excluding the distal-most regions. **E** In the proctodeal field, *Ph kni1* is strongly expressed in two large bilateral domains and a smaller medial domain that lies anterior to those. **F, G** *Ph kni1* is expressed in the lateral periphery and in an anterior cluster of cells within the pre-antennal head lobes.

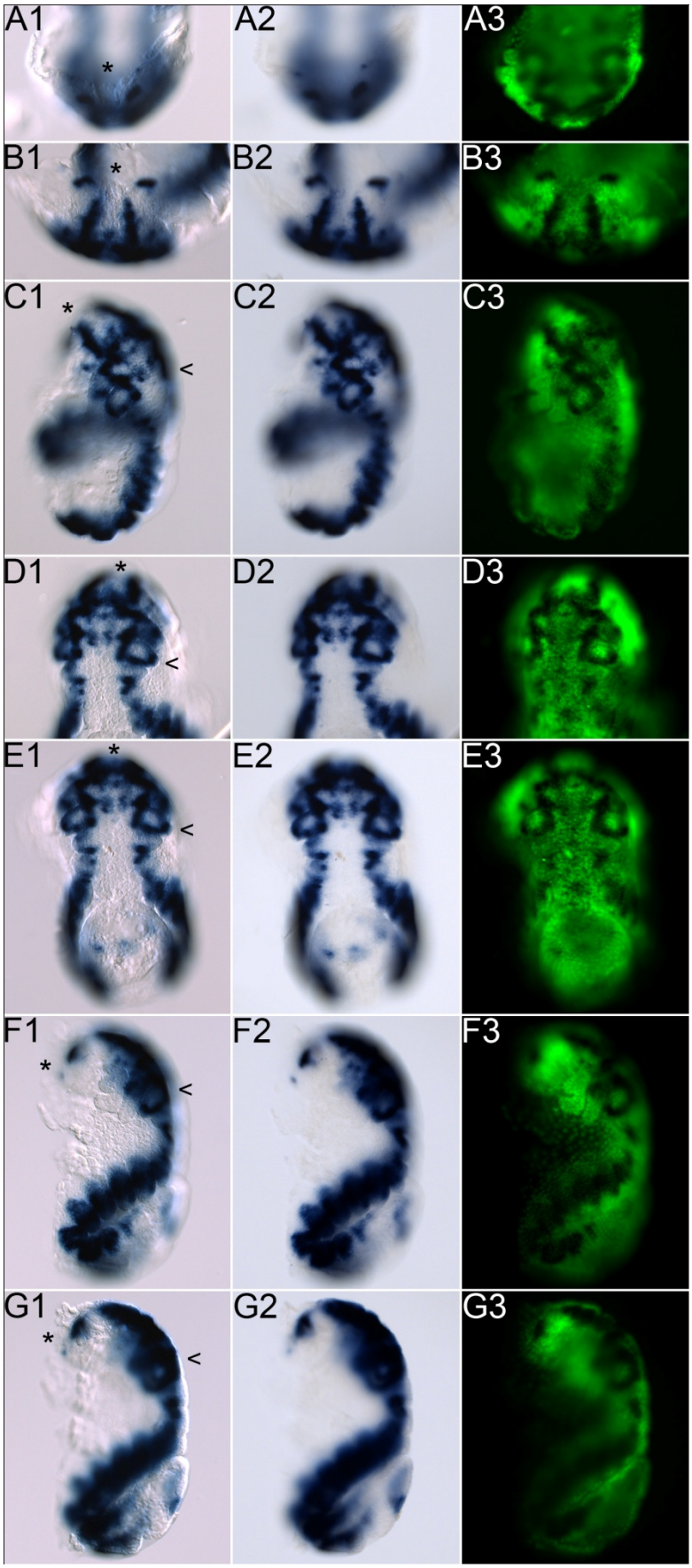



**Figure 76: Expression of *Ph kni1* at stage S19**. The embryo is mounted in dorsal orientation, showing the anterior head (A1-A3), in anterior orientation (B1-B3), in anterior, tilted orientation (C1-C3), in ventral, slightly tilted orientation (D1-D3), in ventral orientation (E1-E3), in lateral orientation (F1-F3, ventral is to the right) and in lateral orientation, sagittal focal plane (G1-G3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni1* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-G1), bright field images in the middle panel (A2-G2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – G3; nuclei are green, expression of *Ph kni1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-G1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** *Ph kni1* is expressed in one small isolated and in one larger group of cells within the anterior part of each head lobe. **B** *Ph kni1* expression in the medial part of the pre-antennal head has taken shape of two parallel anterior-posterior stretches that include the bases of the An1 appendages and are separated by a central region of cells that are free of *Ph kni1* expression. **C** *Ph kni1* expression is found in cells within the proximal article of the An1 appendages and in cells that connect the first and second An2 appendage articles. **D** *Ph kni1* expression is maintained in two groups of cells that are located at the base of the An2 appendages. **E** *Ph kni1* is expressed in peripheral cells of the Mn appendage as well as in two groups of cells at the base of the Mn appendage. The emerging paragnaths also show *Ph kni1 expression.* **F** In segments posterior to Mn *Ph kni1* is expressed exclusively in the appendage anlagen in broad, sheath-like domains, excluding the most distal parts of the appendage. **G** The proctodeal development field shows decreased *Ph kni1* expression as compared to stage 18.

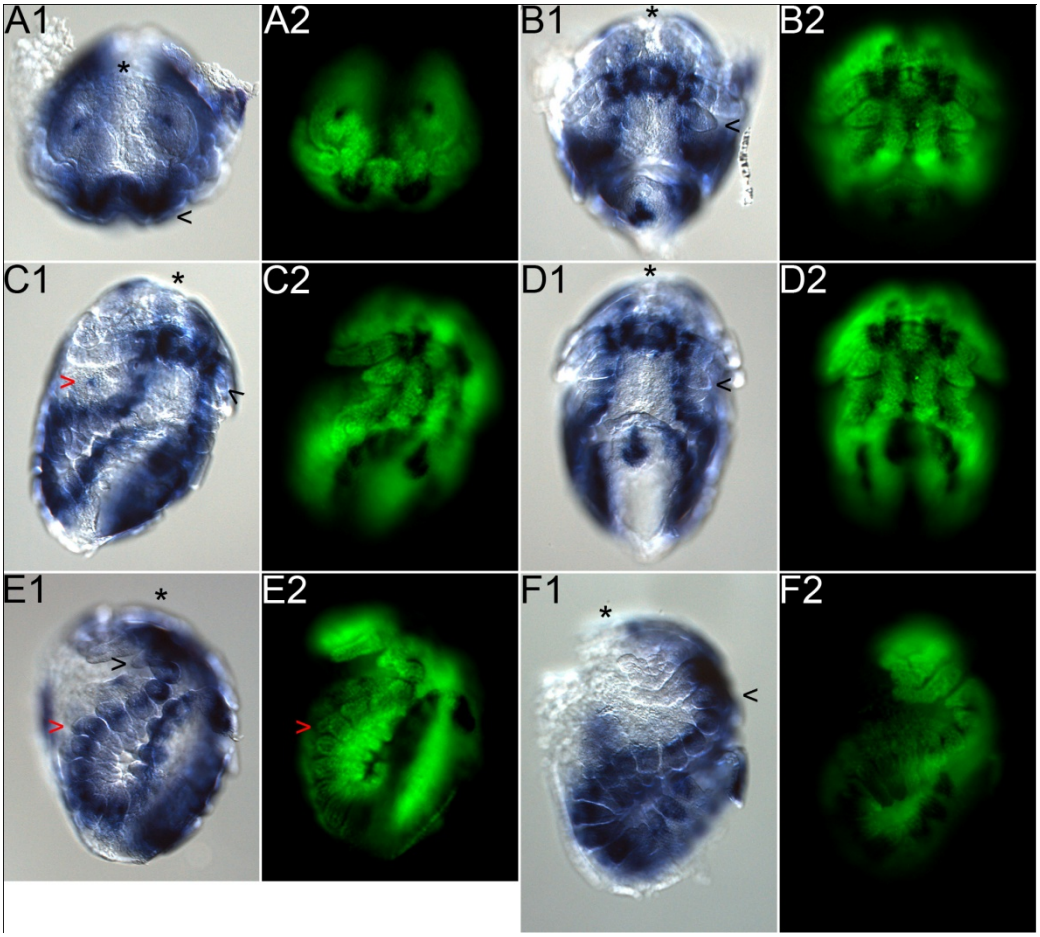



**Figure 77 : Expression of *Ph kni1* at stage S20**. The embryo is mounted in anterior orientation (A1, A2), in ventral orientation, focus on the stomodeal opening (B1, B2), in tilted lateral orientation (C1, C2, ventral is to the right), in an alternative ventral orientation, focus on the Mn segment (D1, D2), in an alternative tilted lateral orientation, focus on the T4 appendage (E1, E2, ventral is to the right) and in lateral orientation, focus on the An2 appendage (F1, F2, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni1* expression. In all images, anterior is up. Shown are DIC images (A1-F1) and, to the right, nuclear labelling (SYTOX®) of the same embryo (A2–F2; nuclei are green, expression of *Ph kni1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-F1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead. **A** The embryo is centred up the pre-antennal head hemispheres. **B** Focus is on the stomodeal opening. **C** Novel *Ph kni1* expression is found in small clusters of cells in the lateral region of the T1 segments (red arrowhead). **D** Focus is on the Mn appendages. **E** Focus is on the T3 appendage. *Ph kni1* is expressed weakly in the gill anlagen of the T3 appendage. **F** Focus is on the An2 appendage.

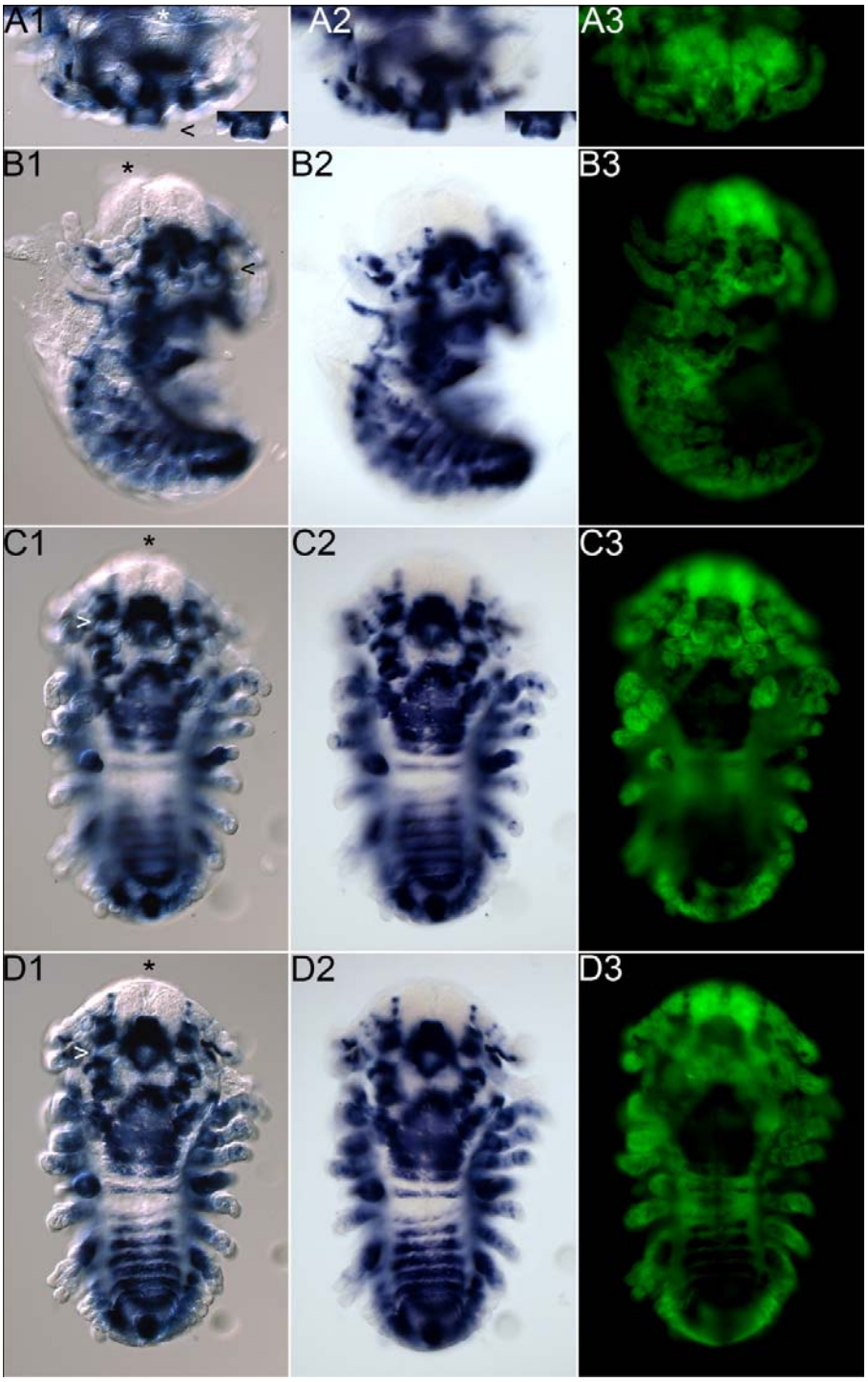



**Figure 78: Expression of *Ph kni1* at stage S22**. The embryo is mounted in anterior orientation (A1-A3), in anterior, tilted orientation (B1-B3), in ventral orientation, focus on labrum and paragnaths (C1-C3), in ventral orientation, focus on T3 (D1-D3), in lateral orientation, focus on the An1 and An2 appendages (E1-E3, ventral is to the right), in lateral orientation, sagittal focal plane (F1-F3, ventral is to the right) and in dorsal orientation, focus on the root of the stomodeal tube (G1-G3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni1* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-G1), bright field images in the middle panel (A2-G2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3 – G3; nuclei are green, expression of *Ph kni1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-G1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead (continued on next page).

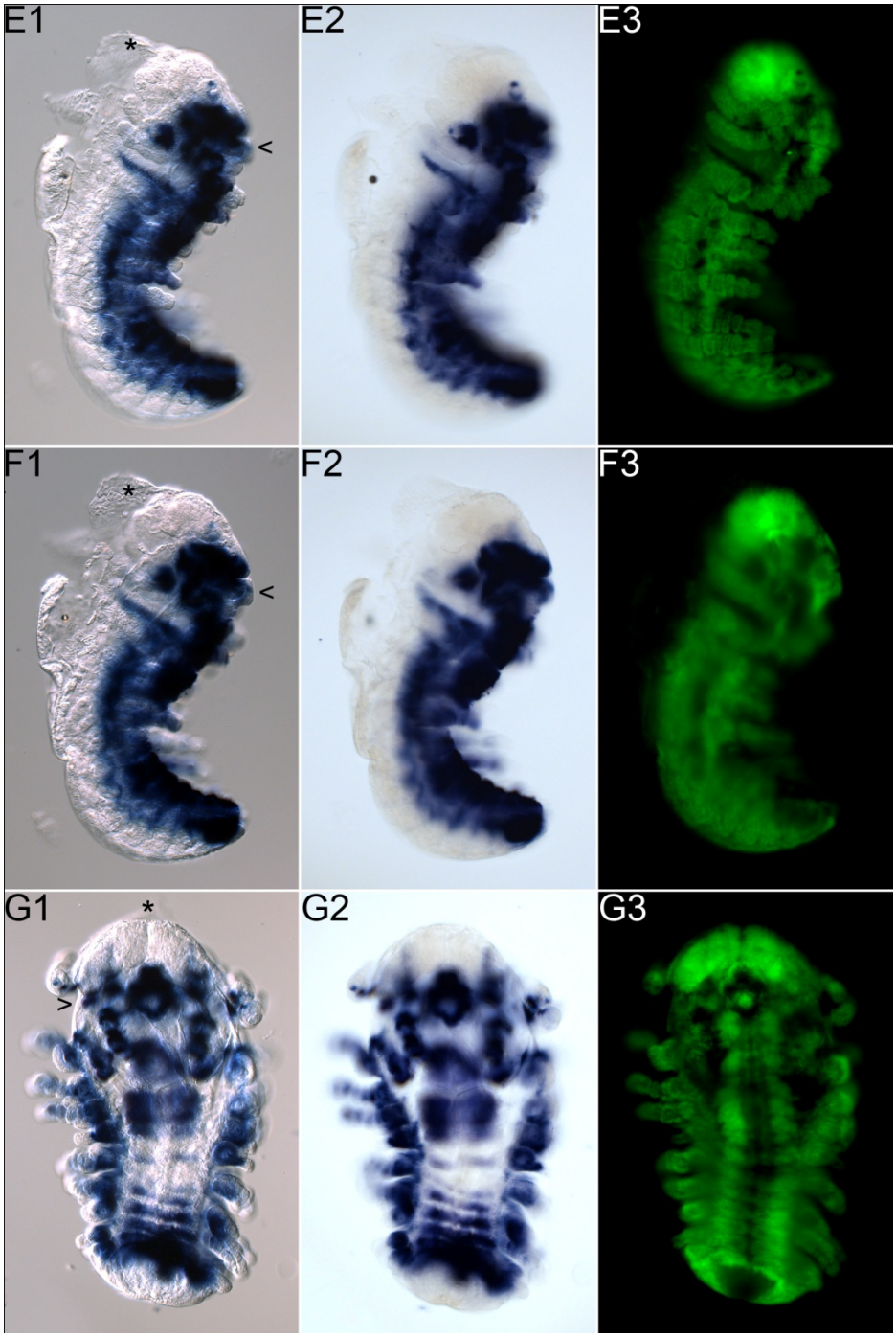



**Figure 78 (continued from previous page):** **A** *Ph kni1* expression in the head lobes and around the invaginating stomodeal opening has a complex topology. **B** In the T2-T8 appendages *Ph kni1* is expressed at the branching point between coxa and gill anlage. **C** Focus is on the labrum and paragnaths. *Ph kni1* is expressed within the Mn appendages, the paragnaths, the Mx1 and Mx2 appendages and the maxillipeds. **D** *Ph kni1* is expressed in cells that connect individual appendage articles. The proctodeum shows strong *Ph kni1* expression. **E, F** *Ph kni1* is expressed in an anterior, lateral stripe that extends from the base of the first gnathopod laterally to the dorsal brink of the T2 segment. **G** *Ph kni1* is expressed in cell clusters that surround the stomodeal opening.

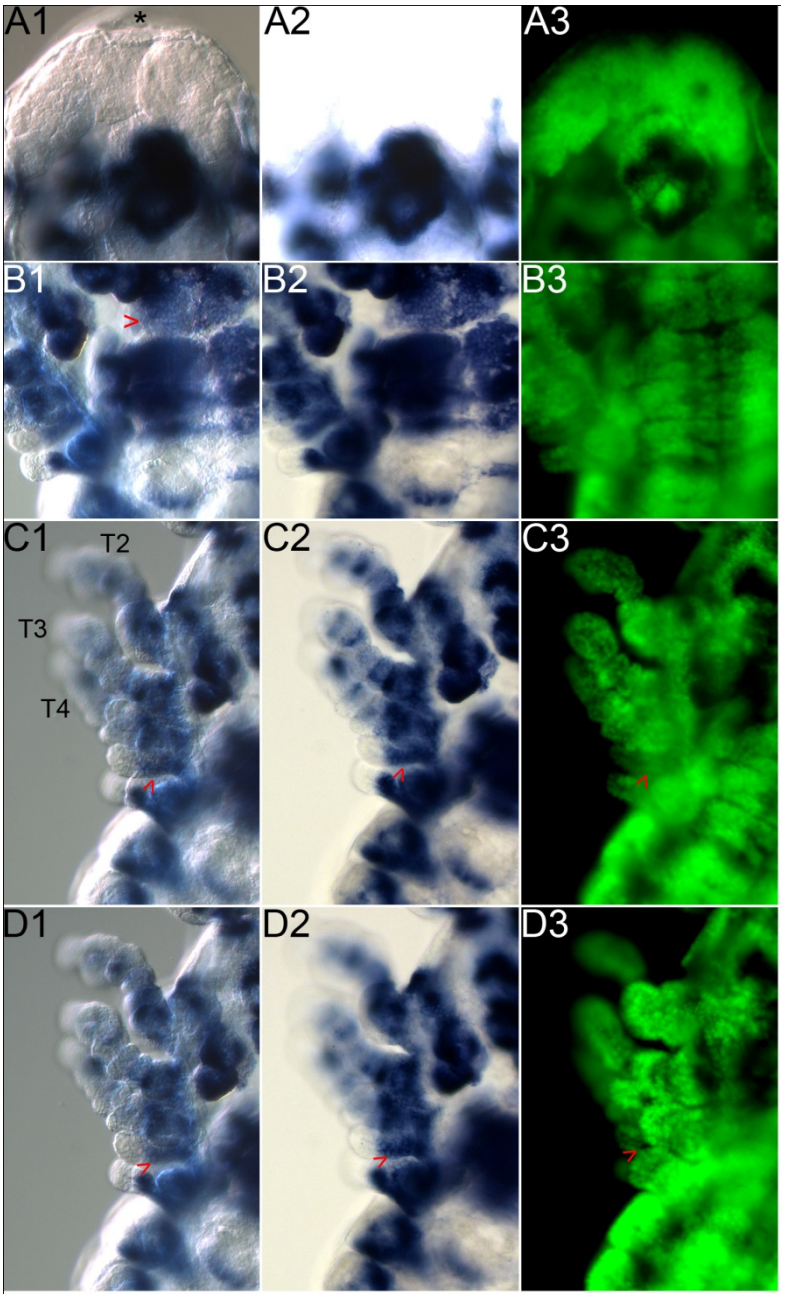



**Figure 79: Expression of *Ph kni1* at stage S22, detailed images**. All images have been taken with a x20 magnification. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2-D2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–D3; nuclei are green, expression of *Ph kni1* is detectable as it quenches the nuclear labelling). In all images, anterior is up. **A** Dorsal view, detail of the stomodeal opening. **B** Ventral view, detail of the central segmental area of T1 (red arrowhead) as well as T2-T5. **C, D** Ventral view, details of the gill anlagen of the T2, T3 and T4 appendages in different focal planes. The read arrowheads point to the proximal part of the T4 gill anlage.

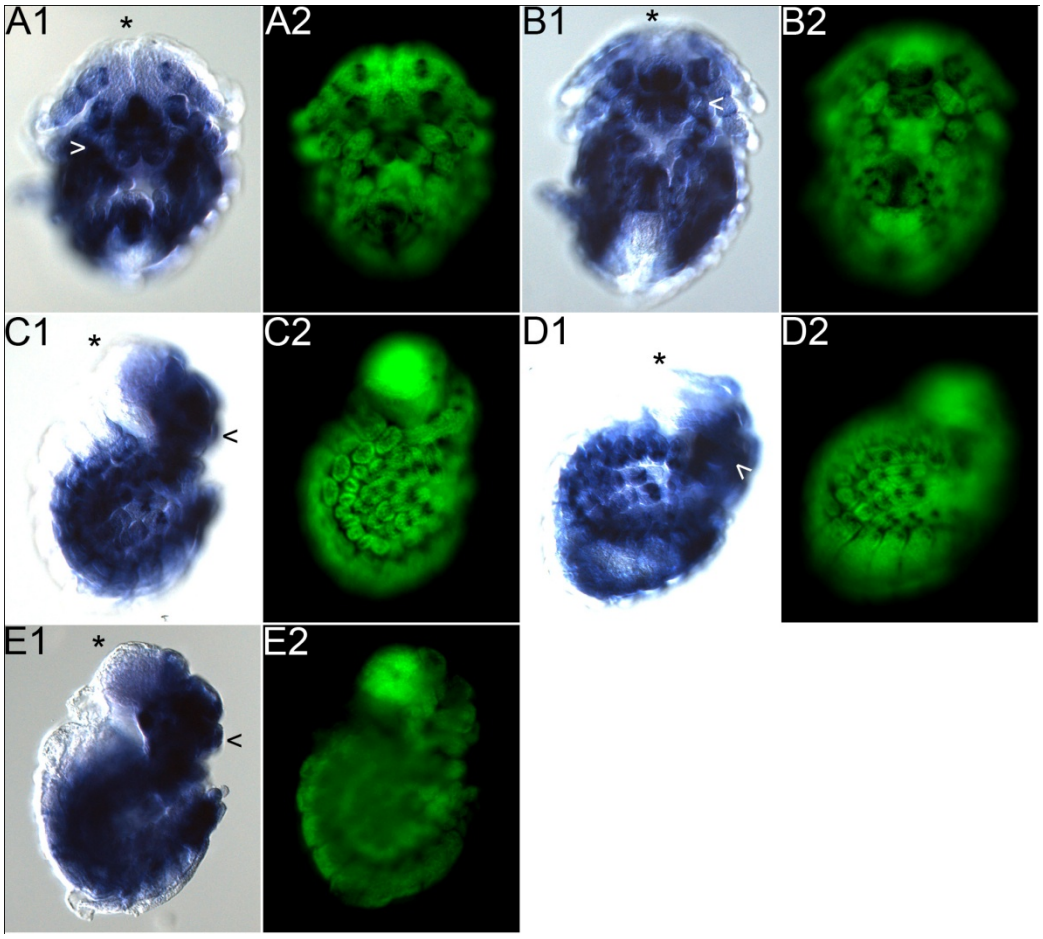



**Figure 80: Expression of *Ph kni1* at stage S23.** The embryo is mounted in ventral orientation, focus on the bases of the second antennae (A1, A2), in ventral orientation, focus on the labrum (B1, B2), in lateral orientation, focus on the proximal thoracic appendage articles (C1, C2), in a tilted, lateral orientation (D1, D2) and in lateral orientation, sagittal focal plane (E1, E2) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni1* expression. In all images, anterior is up. In images C1, C2, D1, D2, E1 and E2, ventral is to the right. Shown are DIC images (A1-E1) and, to the right, nuclear labelling (SYTOX®) of the same embryo (A2–E2; nuclei are green, expression of *Ph kni1* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by an arrowhead.

### 3.4.8 Expression of *Ph kni2*

Expression of *Ph kni2* is first detectable in stage 21 embryos. The pre-antennal head lobes have expanded dorsally and appear oval-shaped. The labrum covers the stomodeal opening and has nearly completed its elongation (Figure 81 A3, E3, 3.1.4, Browne et al., 2005). *Ph kni2* is expressed in the proctodeal field (Figure 81 B1-B3, D1-D3). Faint *Ph kni2* signal is found in between the bases of the An1 and An2 appendages (Figure 81 A1-A3, C1-C3). There is no detectable *Ph kni2* expression prior to stage 21.

In stage 22 embryos, the mandibles protrude laterally on both sides of the labrum. The labrum has completed its posterior extension, its lobes directly abutting the paragnaths anterior (Figure 82 C3, 3.1.4, Browne et al., 2005). *Ph kni2* is expressed in the proctodeum (Figure 82 C1-C3) and at the bases of all segmental appendages, typically extending into the proximal appendage articles (Figure 82 B1-B3, D1-D3, E1-E3).

At stage 23, tergites and coxal plates become visible in thoracic and abdominal segments (Figure 83 E1-E3, Browne et al., 2005). *Ph kni2* expression is found in many areas and groups of cells within the embryo. In general, groups of *Ph kni2* expressing cells are found at the distal parts of appendage articles, in places where the next distal appendage element is connected or branches off as well as at the distal tips of thoracic appendages (Figure 83 E1-E3, F1-F3). At the bases of the An2 appendages and in the Mn, Mx1 and Mx2 appendages, strong *Ph kni2* signal is detectable (Figure 83 B1-B3, C1-C3). *Ph kni2* is expressed in cells surrounding the foregut, putatively in cells that contribute to the developing stomatogastric nervous system (Figure 83 D1-D3). *Ph kni2* expressing cells are found in the coxal plates (Figure 83 G1-G3) and at the bases of tergites (Figure 83 H1-H3). The proctodeum exhibits strong *Ph kni2* signal (Figure 83 D1-D3, E1-E3). Importantly, a patch of strong signal that is found asymmetrically in the right An1 appendage is most likely a remnant of cuticle that has not been removed during dissection and has stained unspecifically (Figure 83 D1-D3).

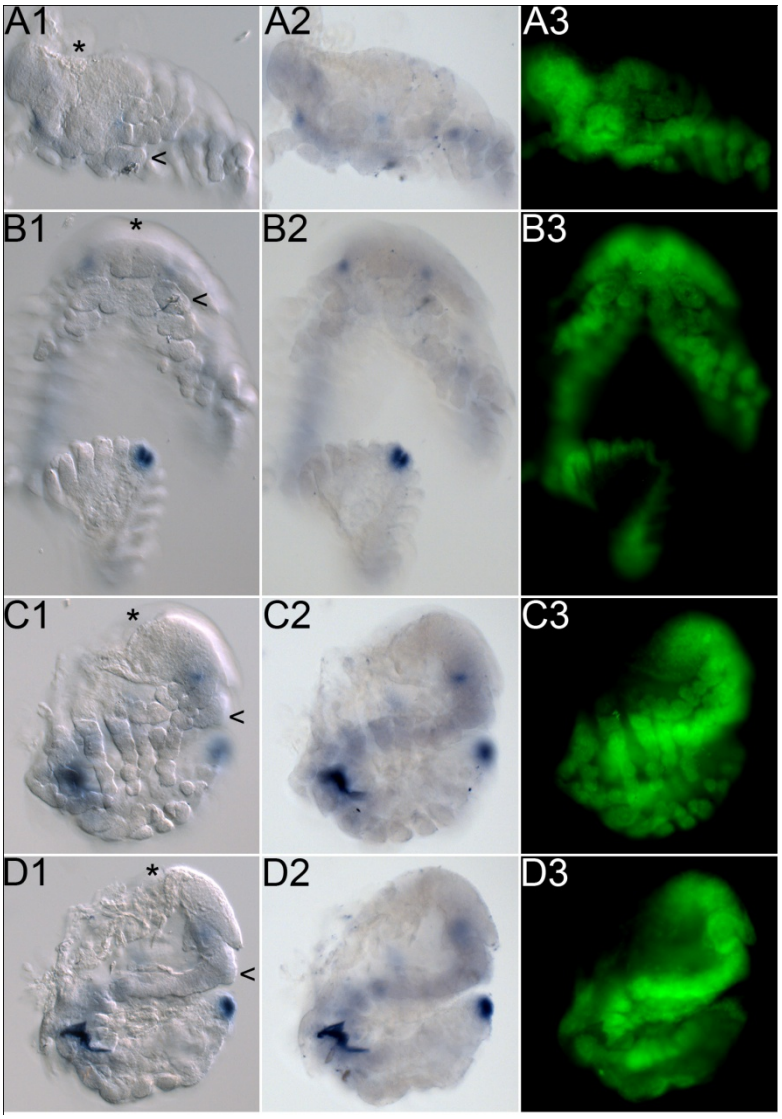



**Figure 81: Expression of *Ph kni2* at stage S21**. The embryo is mounted in anterior orientation (A1-A3), in ventral orientation (B1-B3), in lateral orientation, focal plane on the proximal articles of the An1 and An2 appendages (C1-C3, ventral is to the right) and in lateral orientation, sagittal focal plane (D1-D3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni2* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-D1), bright field images in the middle panel (A2-D2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–D3; nuclei are green, expression of *Ph kni2* is detectable as it quenches the nuclear labelling). In all DIC images (A1-D1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead.

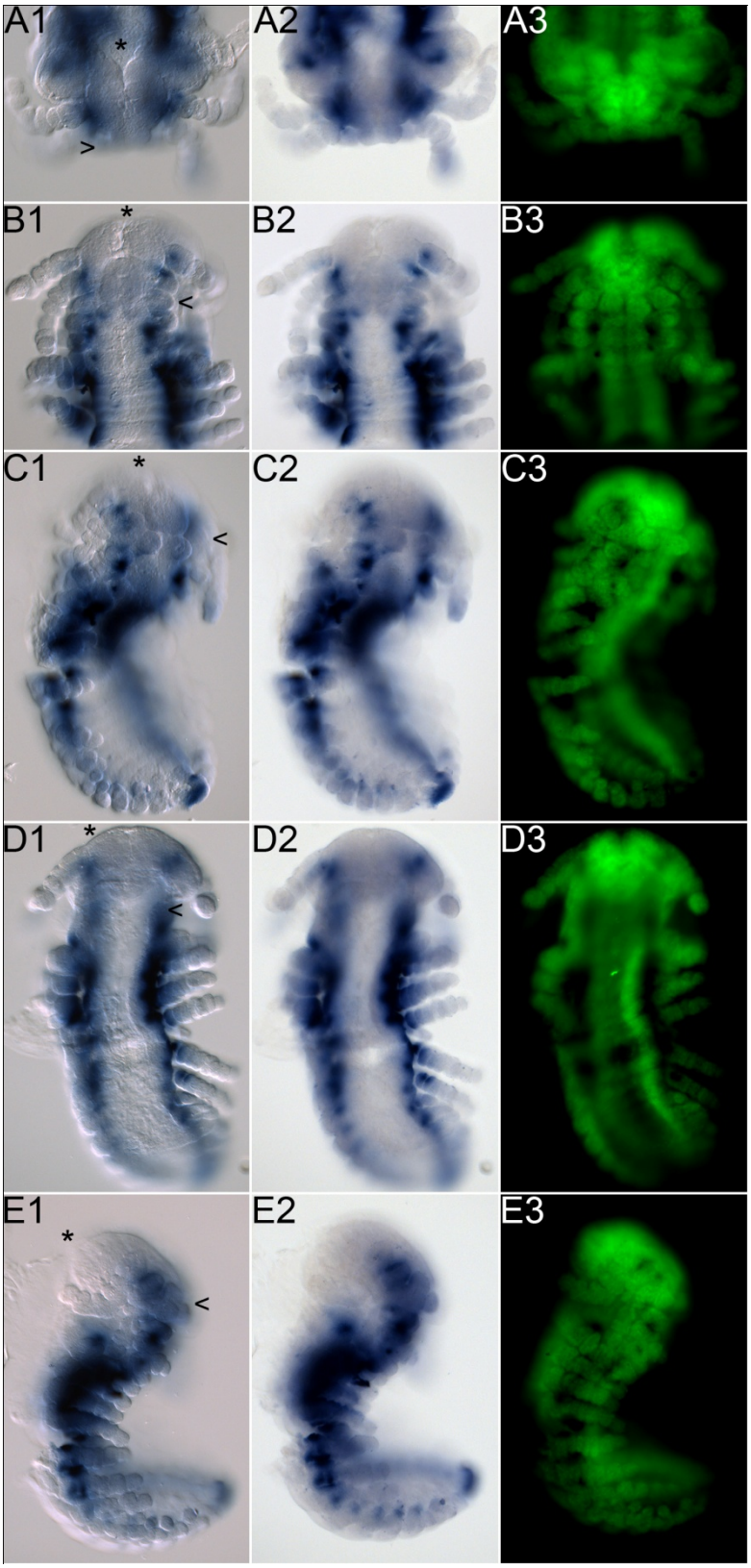



**Figure 82: Expression of *Ph kni2* at stage S22**. The embryo is mounted in anterior orientation (A1-A3), in ventral orientation, focus on the proximal An1 and An2 articles (B1-B3), in ventral, tilted orientation (C1-C3), in ventral orientation, focal plane on the bases of the T3-T5 appendages (D1-D3) and in lateral orientation, focal plane on the An1 appendage (E1-E3, ventral is to the right) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni2* expression. In all images, anterior is up. Shown are DIC images in the left panel (A1-E1), bright field images in the middle panel (A2-E2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–E3; nuclei are green, expression of *Ph kni2* is detectable as it quenches the nuclear labelling). In all DIC images (A1-E1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead.

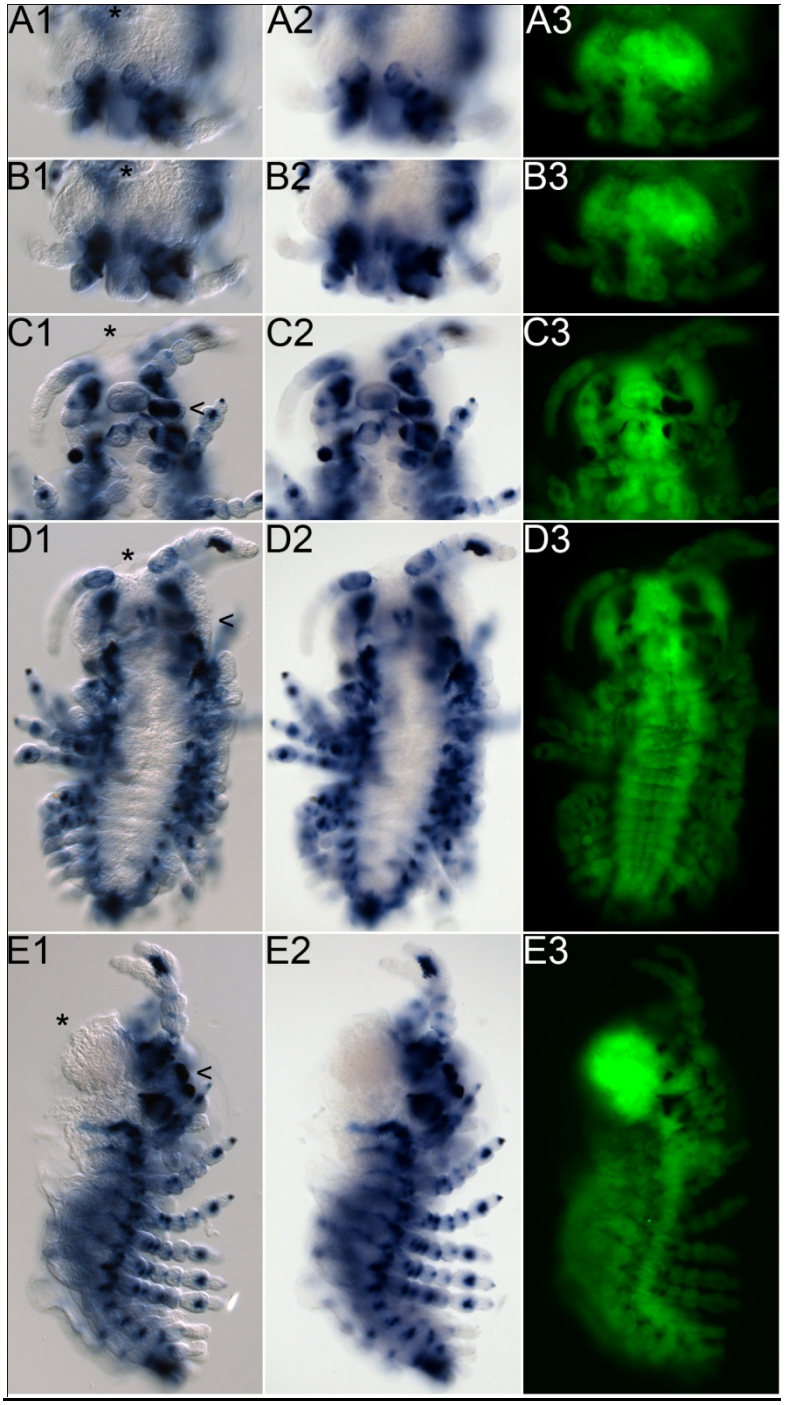



**Figure 83: Expression of *Ph kni2* at stage S23**. The embryo is mounted in anterior orientation, focal plane on the proximal An1 appendage articles (A1-A3), anterior orientation, focal plane on the labrum (B1-B3), in ventral orientation, focal plane on the tip of the labrum and the paragnaths (C1-C3), in ventral orientation, focal plane on the proximal An1 appendage articles (D1-D3), in lateral orientation, focal plane on the thoracic appendages (E1-E3, ventral is to the right) and in dorsal, tilted orientation, focal plane on the thoracic appendages (F1-F3) in order to facilitate better understanding of the three-dimensional pattern of *Ph kni2* expression. Magnifications (x20) of the dorsal, tilted orientation of the embryo shown in F1-F3 (x10 magnification) are shown in G1-G3 (focal plane on the coxal plates of T4-T8) and H1-H3 (focal plane on the bases of the tergites of segments T4-T8). In all images, anterior is up. (Continued on next page).

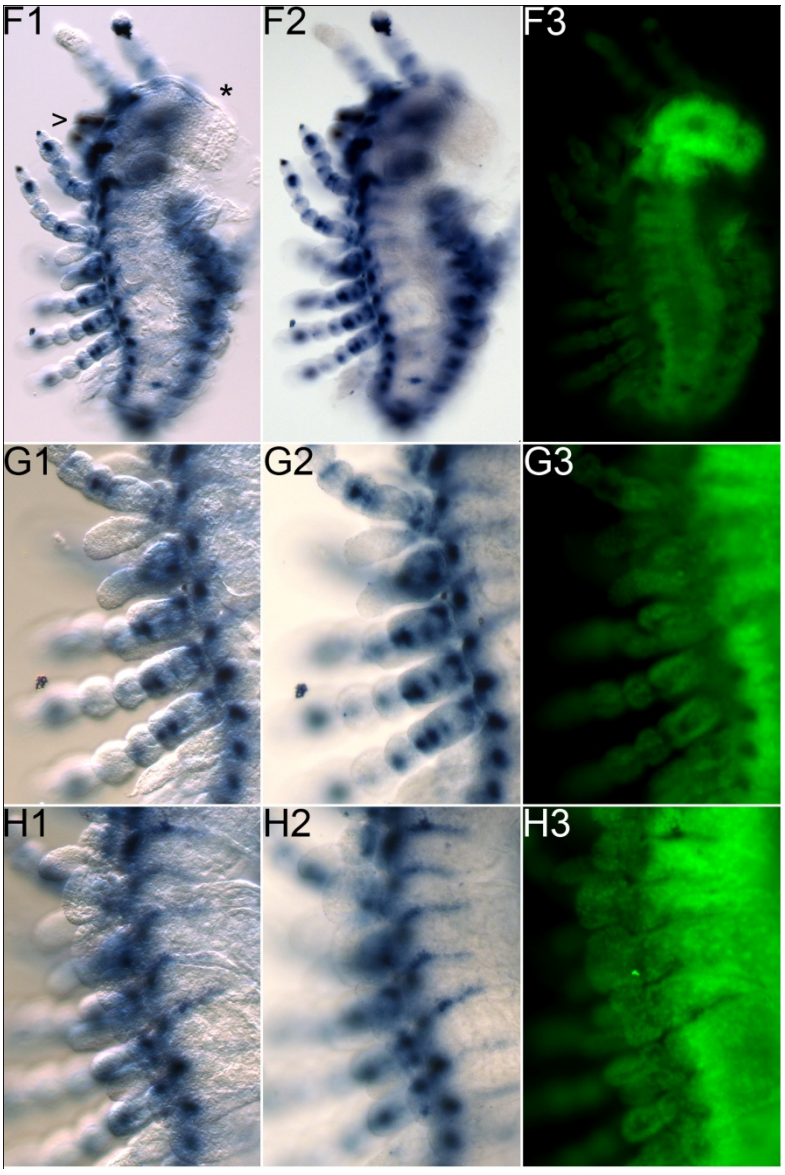



**Figure 83 (continued from previous page):** Shown are DIC images in the left panel (A1-H1), bright field images in the middle panel (A2-H2) and nuclear labelling of the same embryo (SYTOX®) in the right panel (A3–H3; nuclei are green, expression of *Ph kni2* is detectable as it quenches the nuclear labelling). In DIC images A1-F1), the anterior end of the embryo is indicated by a black asterisk and the position of the Mn segment is marked by a black arrowhead.

## 3.5 Techniques to address gene loss-of-function via RNA interference in *Parhyale hawaiensis*

Many of the classical approaches to address gene function are not yet established for *Parhyale hawaiensis*. A method to study loss of function of genes that has proven highly efficient in a number of arthropod species is RNA interference (RNAi, Belles, 2010; Bucher et al., 2002; Tomoyasu et al., 2008). One aim of this work, therefore, has been the establishment of RNAi-based techniques in *Parhyale hawaiensis*. The different approaches taken to this end are described below.

### 3.5.1 RNAi mediated loss-of-function by injection of dsRNA

Application of double-strand RNA (dsRNA) into cells, tissues or whole animals has been shown to trigger RNAi in a wide range of species (Belles, 2010; Brown et al., 2003; Bucher et al., 2002; Catteruccia and Levashina, 2009; Tomoyasu and Denell, 2004; Tomoyasu et al., 2008). However, not all species are equally sensitive to RNAi (Belles, 2010). In this work, the feasibility of an RNAi approach based on injection of double-strand RNA was tested for *Parhyale hawaiensis*.

#### 3.5.1.1 RNAi mediated by injection of dsRNA in embryos

One-cell and two-cell embryos were injected with dsRNA derived from a 3' fraction of the *Ph otd1* transcript and monitored individually for their development post-injection (5.5.3). At varying time points after injection, single embryos were collected, fixed, dissected and subjected to *Ph otd1* WMISH (5.4.3) in order to assess degradation of *Ph otd1* transcripts. The embryos were also examined for developmental and morphological phenotypes resulting from loss of *Ph otd1* function. As a control, embryos were injected with dsRNA derived from *EGFP* (Chalfie et al., 1994) and treated analogously. Both groups of embryos manifested a comparable amount and range of developmental aberrations and showed similarly high levels of lethality. None of both showed a measurable decrease of *Ph otd1* expression as compared to wild-type embryos (not shown). Likewise, in none of both, specific morphological

phenotypes resulting from a loss of *Ph otd1* function could be found (not shown). Therefore, I conclude that, at least for *Ph otd1*, injection of dsRNA does not cause specific RNAi in *Parhyale* and that the observed effects resemble injection artefacts. This approach, therefore, has been dismissed.

#### 3.5.1.2 RNAi mediated by injection of dsRNA in adult females

In *Tribolium castaneum*, parental RNAi, i.e. RNAi effects that manifest in the offspring of female pupae or adults injected with dsRNA, is a highly efficient tool to target transcript degradation. Advantages are the specific loss-of-function from the earliest stages of embryogenesis onwards and the evasion of mechanical stress caused by injecting embryos. In addition, parental RNAi has been shown to provide an abating continuum of phenotypic severity from the first batch of an injected female's offspring to the last (Bucher et al., 2002). In previous experiments, *Parhyale hawaiensis* embryos have reacted very sensitive to mechanical manipulation, revealing a high level of unspecific negative impacts after dsRNA injection that could have masked specific RNAi effects. For this reason, it was tested here whether an experimental approach based on parental RNAi could provide an efficient alternative to dsRNA injection in *Parhyale* embryos.

In a first approach, wild-type females and an equal amount of females carrying a stable transgene expressing *DsRed* (Horn et al., 2002) under control of a muscle-specific driver (*pMiII-mse_DsRed*, courtesy of Anastassios Pavlopoulos) were isolated carrying embryos in their brood pouches. They were then kept individually until they released their offspring. Then, just before injection, they were sedated and set up with little moisture in a ventral posture (5.4.7). Wild-type females were injected with dsRNA targeting *Ph otd1*, while the transgenic *pMiII-mse_DsRed* females were injected with dsRNA targeting *DsRed*. For setting up injection parameters and as a control, wild-type females were analogously injected with Phenol Red (phenolsulfonphthalein, Merck, Germany). Injected females were kept and monitored individually until they moulted (2.5.4). Most females (92.5%, n=40) survived the first three days. More than half of the remaining, however, died subsequently during their attempt to moult. The survivors were put together with male *Parhyale*. The majority of these females was attacked and killed by the males. The remaining ones (n=3) left the mating procedure without producing fertilised eggs. All of them died during the following days.

Because females were impaired in mating after injection, the experimental set-up was altered to overcome this problem. Both wild-type and transgenic *pMiII-mse_DsRed* females were isolated and treated as above. They were put together with male *Parhyale* prior to the injection of dsRNA. Females that left successful matings and had not yet finished the transfer of embryos into their brood pouch (n=12 wild type females, n=8 transgenic females; within less than one hour after mating) were injected with dsRNA targeting *Ph otd1* and *DsRed*, respectively, and subsequently monitored as described above. They showed a similar survival rate after three days as compared to the experiment described above (90%). Most of the females produced small batches of fertilised eggs (typically <5), but died within a week after injection. Those that survived displayed gradual degradation of their offspring within the brood pouch. Only one embryo that was the offspring of a transgenic *pMiII-mse_DsRed* female hatched from a laid batch of 11 embryos. It did not show any decrease in muscle-specific DsRed expression (not shown). Many transgenic females lost local, muscle-specific *DsRed* expression after injection. In all examined cases, however, this coincided with paralysis and, in a few cases, necrosis and degradation of the affected muscular tissue (not shown). From these findings, there is no indication that parental injection of dsRNA in *Parhyale* did trigger RNAi in the offspring. Further, the effects observed in the injected female *Parhyale* and their ovaries were artificial. Therefore, this method has been dismissed.

### 3.5.2 RNAi mediated by heat shock inducible transgene-based expression of hairpin RNAs

As shown above, the attempts to induce RNAi by directly injecting dsRNA into *Parhyale* embryos and into ovaries of *Parhyale* females were unsuccessful. In addition, *Parhyale* demonstrated a problematic level of sensitivity to mechanical strain. For these reasons, an alternative, transgene-based approach was pursued. In the mosquito *Anopheles gambiae*, RNAi was effectively triggered by stable transformation of the germ line with transgenes that allowed induction of the expression of RNA hairpins (Catteruccia and Levashina, 2009). In a previous study, various forms of hairpin constructs had been tested and evaluated (Brown et al., 2003). The main focus lay on the efficiency of different spacer elements that separate the two arms of the hairpins. Both small spacer fragments (8-12bp) and endogenous functional introns used as spacers proved to be highly effective. Based on these

results, constructs carrying an endogenous *Parhyale* heat shock inducible regulator element (Pavlopoulos et al., 2009) driving hairpins targeted at *Ph otd1* (*pMi[3xP3_DsRed; hsp70_hp-otd1],* 5.2.6.3*)* and *DsRed* (*pMi[3xP3_EGFP; hsp70_hp-DsRed],* 5.2.6.4*)* were created. 3xP3 driven *DsRed* and *EGFP* expression served as a transformation marker in these constructs. Both stable germ line transformation and transient expression (systemic and clonal) of these constructs were approached.

For stable germ line transformation (5.5.1), one-cell embryos were injected with the respective *pMinos*-based construct and mRNA of *Minos* transposase (Pavlopoulos and Averof, 2005). After hatching, injected embryos were examined for expression of the transformation marker. In this work, very few (n=4) of the hatched embryos displayed EGFP expression. However, none of those survived the first moult after hatching. Hatched embryos that were negative for the transformation marker were pooled and kept separately. They were put together for mating as soon as they reached sexual maturity. Their F1 generation was then screened for transgenic marker expression. However, no transgenic offspring was identified. Therefore, no population of stable transgenic *Parhyale* carrying an RNA hairpin construct was established.

Alternatively, one-cell and two-cell embryos were injected as above and directly subjected to three consecutive heat shocks (1h, 37°C), starting 18h post-injection with 24h time lags between them. 18h after the last heat shock, a fraction of these embryos was fixed, dissected and subjected to WMISH in order to address hairpin expression and endogenous transcript degradation. To this end, one probe was used that targets both endogenous and hairpin *Ph otd1* sequence, and another one that recognises only endogenous *Ph otd1* transcript. Embryos subjected to WMISH visualising both endogenous *Ph otd1* transcript and ectopically expressed *Ph otd1* RNA hairpin molecules showed a combination of the wild-type *Ph otd1* expression pattern and additional mosaic-like expression of the *Ph otd1* hairpin (Figure 84 A, B). In individual embryos, the number and size of these additional areas of *Ph otd1* signal varied as well as the intensity of the signal. Importantly, all embryos showed at least one ectopic *Ph otd1* signal (not shown). Embryos subjected to WMISH specific for endogenous *Ph otd1* transcripts showed expression domains highly similar to wild-type, with regard to spatial distribution and intensity. Both groups of embryos subjected to WMISH, i.e. those showing both endogenous and ectopic *Ph otd1* signal at the same time as well as those exhibiting only endogenous *Ph otd1* signal, revealed disturbances in the head morphology, as

displayed by nuclear labelling. For this reason, the expression pattern of *Ph otd1* appeared slightly distorted in the majority of analysed embryos (Figure 84).

The remaining embryos were allowed to develop further. However, all of them revealed severe defects in development during the following days and eventually died and degraded. As a control, wild-type embryos were injected with *pMiII-mse_DsRed* (courtesy of Anastassios Pavlopoulos) and treated analogously. The fraction of these control embryos that were subjected to *Ph otd1* WMISH showed wild-type *Ph otd1* signal, but also a similar level of morphological aberrations compared to the non-control embryos. Of those control embryos that were allowed to develop further, 31.6% hatched (n=38). They showed morphological defects at an occurrence comparable to embryos injected with water.

From these findings, I conclude that clonal RNA expression was induced after applying heat shocks to embryos that had been injected with the respective constructs. However, effective RNAi could not be observed. Importantly, it remains unclear if RNA hairpins were successfully formed. Therefore, and because no population of *Parhyale* transgenic for the hairpin constructs could be established, this approach was not followed up further.

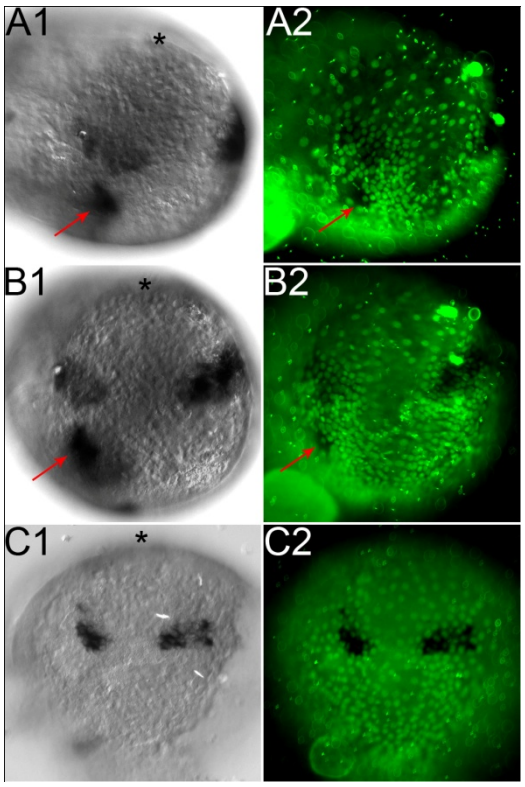



**Figure 84: WMISH of *Parhyale* embryos injected with a *pMinos*-based RNA hairpin expression construct targeting *Ph otd1*.** Bright field images are on the left side of the panel (A1, B1, C1) and images showing the nuclear labelling (SYTOX®) of the same embryos are on the right (A2, B2, C2). **A**, **B** Injected *Parhyale* embryo that has been subjected to WMISH targeting both endogenous and ectopically expressed *Ph otd1* transcripts. Posterior of the endogenous wild-type expression pattern of *Ph otd1*, ectopic *Ph otd1* expression signals were detected (red arrows). **C** Injected Parhyale embryo that has been subjected to WMISH targeting only the endogenous *Ph otd1* transcripts. Aside from minor disturbances of the head morphology (C2), the WMISH signal resembles the wild type *Ph otd1* expression pattern.

### 3.5.3 Splicing of a *Ph kni2* intron in the context of an artificial EGFP expression construct

When considering the cloning strategy for inheritable inducible hairpin expression constructs, the choice of the spacer sequence that separates both hairpin arms has been shown to be crucial (Brown et al., 2003). The prerequisites for the sequence spacer include practicability with regard to cloning success and stable propagation of the hairpin construct within bacteria on the one hand, and a high efficacy of RNA hairpin formation within a living cell after induction of the transgene on the other hand. In previous approaches, a sequence containing 8 non-palindromic nucleotides was used as a spacer. While high efficacy of hairpin formation when using this strategy has been shown in insects (Catteruccia and Levashina, 2009; Kennerdell and Carthew, 2000), stable cloning and propagation of the construct in bacteria have been impaired and required high effort as well as the use of specifically designed bacteria strains (SURE 2 Supercompetent Cells, Stratagene, Agilent Technologies, Santa Clara, USA; 5.2.6). A strategy that has been described and tested as at least evenly efficient implies the use of endogenous, functionally spliced introns as spacers. Supposedly, the advantage of an intron used as spacer between the two arms of a hairpin construct is that it separates palindromic nucleotide sequences wide enough to allow for stable propagation in bacteria, and that the process of removing that intron during splicing promotes the formation of the RNA hairpin (Brown et al., 2003). This approach, however, is only feasible for species where introns have been isolated and characterised, or for those where genomic sequence data is available.

In the course of verifying BAC plasmids carrying gene loci of interest via a PCR based approach, a putative intron-containing sequence fragment could be isolated for *Ph kni2* (Jonas Schwirz, personal communication). The fragment was cloned and sequenced, revealing that the *Ph kni2* ORF sequence (3.4.5) was disrupted by an insertion of a non-coding sequence element of 1kb size. This element carried putative splicing donor and acceptor sites (Figure 85). In order to test if this element was a functionally spliced intron in the context of *Parhyale hawaiensis*, an artificial EGFP-intron composite gene was established and put under the control of a muscle specific *Parhyale* enhancer (*pimpMiII*[*mse_EG-kni2intron-FP*], 5.2.6.6). Following a *Minos*-based transgenesis approach (5.5.1), this test construct was injected in one-cell and two-cell *Parhyale* embryos.

30% of hatching, injected *Parhyale* expressed a functional EGFP protein in their muscles (Figure 86), confirming that the isolated element is an endogenous, functionally spliced *Ph kni2* intron. Control injections were done using the original Minos-based EGFP expression construct that lacks the intron element (*pimpMiII*[*mse_EGFP*], 5.2.6.5). 18% of these embryos showed muscular EGFP expression after hatching.

In some cases, only a subset of all muscles expressed EGFP. A likely explanation for this observation is that the actual event of stable integration of the transgene into the genome can occur at varying time points post-injection (e.g., Pavlopoulos and Averof, 2005). Depending on the exact time point of integration, only subsets of embryonic cells may have integrated the transgene, resulting in mosaic transformation of the embryo.

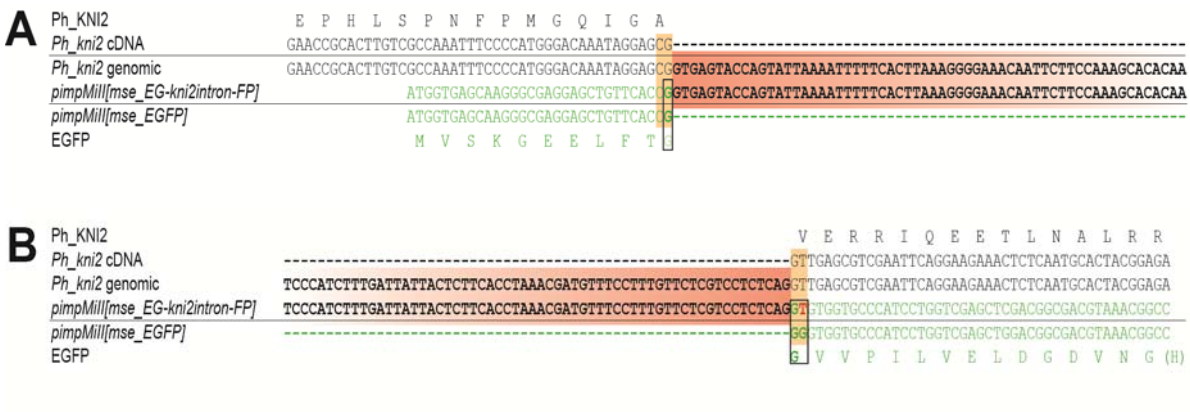



**Figure 85: Design of *pimpMiII*[*mse_EG-kni2intron-FP*].** This construct was generated to test the functionality of a putative *Ph kni2* intron, recovered from a PCR fragment (see text). Shown are the **A** 5' and **B** 3' parts of an alignment of (from top to bottom): the *Ph kni2* cDNA sequence, the corresponding *Ph kni2* genomic environment, the EGFP sequence of *pimpMiII*[*mse_EG-kni2intron-FP*], containing the putative *Ph kni2* intron and, lastly, the corresponding sequence of the original *pimpMiII*[*mse_EGFP*] construct, lacking the intron sequence. Above the nucleic acid sequence alignment, the amino acid sequence of *Ph kni2* is shown for illustration. One-letter codes of individual amino acids are directly above the centre of the corresponding codon. Analogously, the amino acid sequence of EGFP is shown below the alignment. Again, one-letter codes of individual amino acid residues are beneath the centre of the corresponding codon. The intron sequences are highlighted red. *Ph kni2* cDNA and protein sequences are grey, EGFP DNA and protein sequences are green. Hyphens (-) indicate sequence gaps. The putative *Ph kni2* intron was cloned into a specific position within the EGFP ORF in order to retain the original genomic environment adjacent to the *Ph kni2* intron. In the alignment, these positions are highlighted orange. In detail, it was seen to it that two bases 5' (CG) and two bases 3' (GT) of the intron in *pimpMiII*[*mse_EG-kni2intron-FP*] are identical to the *Ph kni2* genomic situation. In order to achieve this with regard to the EGFP ORF, the third nucleotide of a glycine codon of the original EGFP ORF (black rectangle) was changed from G to T (printed red). This does not alter the amino acid sequence of EGFP in *pimpMiII*[*mse_EG-kni2intron-FP*] in case the putative *Ph kni2* intron is removed after successful splicing.

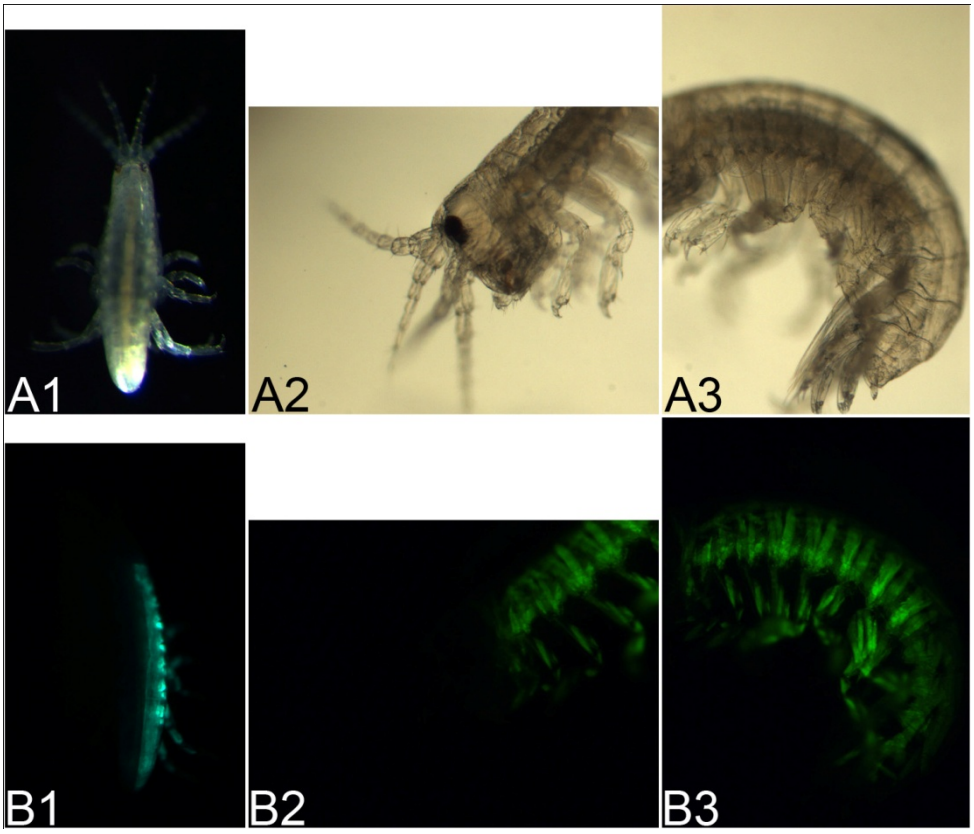



**Figure 86: Muscle-specific expression of functional EGFP in *Parhyale* hawaiensis injected at S1 with *pimpMiII*[*mse_EG-kni2intron-FP*].** Shown are DIC images (A1-A3) and fluorescent images (B1-B3) of an animal three days after hatching. Functional EGFP protein is expressed in a subset of the animal's muscles. Specifically, trunk and leg muscles of the right body half exhibit EGFP expression. **A** Shown are: dorsal view of the complete animal (A1), a lateral view of the head (A2) and a lateral view of the trunk (A3). A1 was taken with the epifluorescence stereomicroscope MZ16FA (Leica Microsystems, Germany, 5.6). Both A2 and A3 images were taken with the inverse microscope Observer Z1 (Carl Zeiss MicroImaging, Germany, 5.6), under x5 magnification. **B** Shown are the same views and orientations as A1-A3.

### 3.5.4 RNAi mediated loss-of-function by injection of siRNA

Previous efforts to induce RNAi in *Parhyale*, either by injection of dsRNA (3.5.1) or induction of RNA hairpin expression from transgenic constructs (3.5.2), were unsuccessful. For this reason, an alternative RNAi approach based on small interfering RNA molecules (siRNA) was designed. In diverse animal species, siRNA-mediated RNAi is used as an effective tool for degrading specific mRNA, therefore reducing respective protein levels and causing loss-of-function phenotypes. Recently, this strategy has been effectively applied to elucidate the role of the *Parhyale Hox* gene *Ultrabithorax* in segmental appendage morphology (Liubicich et al., 2009). Additionally, siRNA-mediated RNAi was used to show conserved functions of the *Parhyale distalless* paralogue *Ph dll3* in distal appendage development (Liubicich et al., 2009). Here, siRNA-mediated RNAi using STEALTH siRNA duplexes (Invitrogen; 5.5.2; A4.3, table A4) was approached for *Ph six3* (3.3.1.2) and *Ph kni1* (3.4.2) in order to address their functional role during *Parhyale* head development.
Following the provider's recommendations, three individual siRNA duplex molecules were designed and ordered for each gene, targeting different regions of the respective transcripts. In order to circumvent off-target effects that could arise from affecting transcripts with high sequence similarities, only non-conserved parts of the *Ph six3* and *Ph kni1* transcripts were used as basis for the generation of siRNA duplex molecules (see also A4.3, table A4).

For each gene, *Parhyale* embryos were injected with any single siRNA molecule species and with combinations of all three. Global degradation of the targeted mRNA by applying siRNA to one-cell as well as clonal degradation by injecting two-cell embryos was addressed via WMISH (5.4.3). The WMISH signal of siRNA-injected embryos was compared to that of wild-type embryos (see 3.3.1.7 and 3.4.7 for the dynamics of *Ph six3* and *Ph kni1* expression profiles). The following criteria were evaluated: (i) reduced signal intensity, (ii) prolonged duration until injected embryos reached a signal intensity comparable to wild-type and (iii) occurrence of mosaic-patterned loss of specific signal within siRNA-injected embryos. Embryos that showed one of these criteria were considered as weakly affected, those that showed more than one as strongly affected and those that did not show any expression of the gene of interest after injection with siRNA were considered as being affected to the highest degree. Damaged embryos that could not be used for the aforementioned expression pattern comparison were assessed separately. The possibility of

unspecific mRNA degradation was addressed by subjecting a fraction of embryos injected with siRNA of one gene to WMISH specific for the other.

In order to investigate developmental defects caused by loss of protein function due to RNAi, the morphology of embryos that displayed reduced WMISH signal was examined and compared to wild-type embryos. To this end, nuclear labelling of the embryos was used (5.4.5). In addition, siRNA-treated embryos that had been allowed to hatch were analysed for their morphology and compared to wild-type animals of the same age. A feeding assay (3.5.4.5) was conducted in order to address the question whether a loss of function of the genes of interest was able to cause impediment of behavioural properties of hatched animals.

#### 3.5.4.1 Assessment of unspecific injection phenotypes

Control injections of *Parhyale* embryos with water revealed unspecific effects that resulted from puncturing and inflating *Parhyale* embryos. As a consequence, around 10% of embryos injected with water usually died shortly after injection. In addition, another 20% of embryos injected with water accumulated developmental defects during subsequent days which eventually led to their degradation before hatching or to their death in the attempt to hatch (Figure 88 A). Of those control animals that hatched successfully, usually 2-3% exhibited morphologic phenotypes, the most frequent being gut malformations, body axis distortions or allometric effects caused by premature hatching (Figure 87). Aside from these, *Parhyale* embryos demonstrated a considerable sensitivity to manipulation in previous experiments in general (3.5.1, 3.5.2). For these reasons, embryos were monitored for developmental defects or lethality (5.5.3) at several specific time points after siRNA injection: Lethality of embryos that occurred during the first 24h after injection was considered a direct consequence of mechanical stress. Likewise, defects during development or lethality that occurred before the initial expression of the targeted gene of interest were considered unspecific. In contrast, embryos that showed affected development or lethality after this second checkpoint post-injection were most likely exhibiting specific effects based on transcript degradation and loss of function. Since the process of hatching requires a concerted, fine-tuned action of many different developmental processes, lethality shortly before or during hatching was assessed additionally.

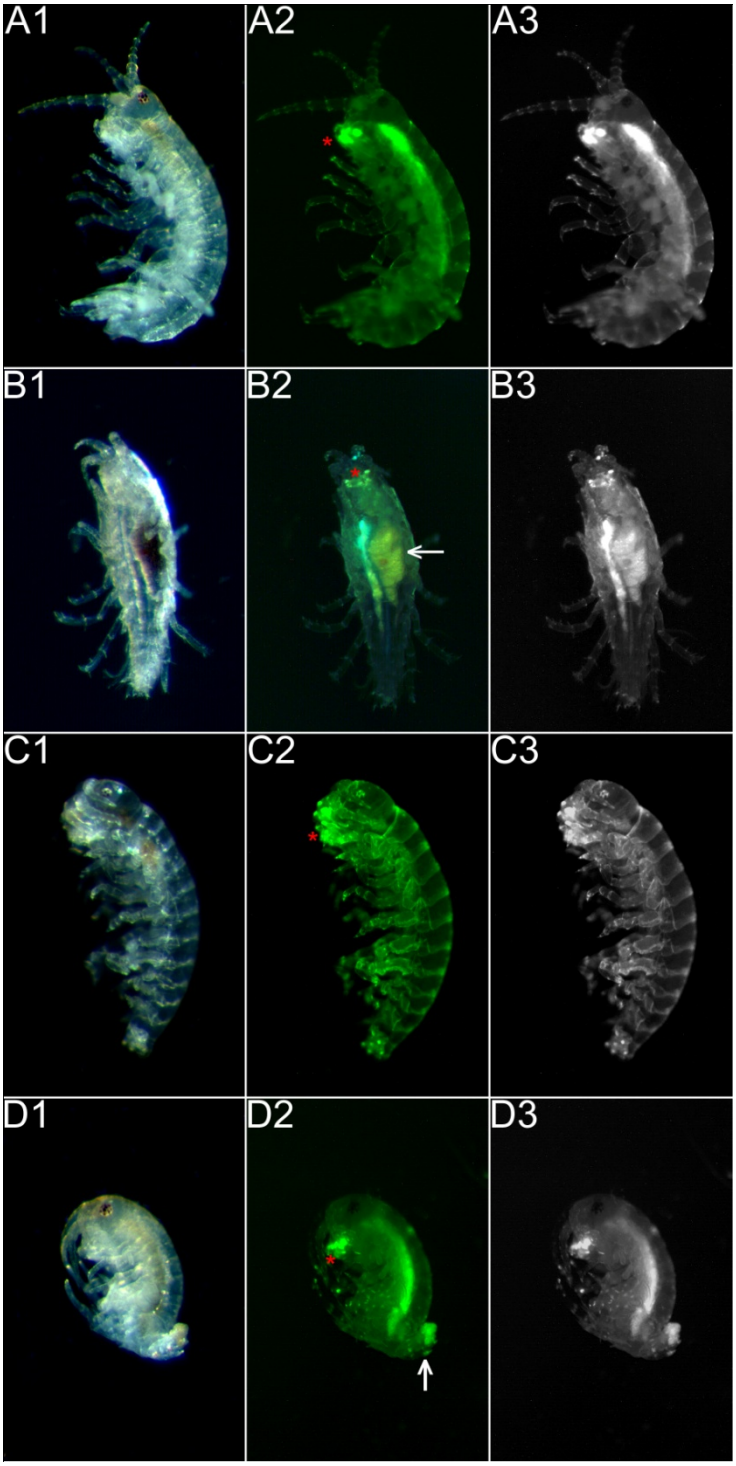



**Figure 87: Unspecific phenotypes occurring after injection of S1 *Parhyale* embryos with water (control injections).** All animals shown in this figure had been injected with water at one-cell stage (S1) and exhibited a range of unspecific morphological and developmental phenotypes after hatching. The animals shown here had hatched 24h prior to photographing. The images were taken using the epifluorescence stereomicroscope MZ16FA (Leica Microsystems, Germany, 5.6). To the left, images of hatched animals using indirect light (without filter) from a cold light source are shown (A1, B1, C1, D1, 5.6). In the centre of the panel, images of the same animals using a mercury-vapour lamp and EGFP-LP (5.6) to capture the auto fluorescence of the cuticle are shown (A2, B2, C2, D2). A red asterisk marks the buccal mass that resembles the functional unit of tightly clustered feeding appendages in adult *Parhyale*. To the right, the same auto fluorescence images are shown in greyscale (A3, B3, C3, D3). In all images, anterior is up. **A** Control-injected *Parhyale* showing wild-type morphology after hatching. **B** Hatched control-injected *Parhyale* exhibiting gut malformations. In this animal, the gut phenotype is visible as a bulbous constriction that contains a large amount of yolk remnants (white arrow, B2). **C** Prematurely hatched control-injected *Parhyale*. This animal had hatched before completing morphogenesis. It has short appendages that are not fully oriented as in wild-type, antennae that are distinctly bent backwards, a bubble-shaped head capsule, erroneous allometric ratios of head and trunk segments and an underdeveloped abdomen (compare to A). **D** Hatched control-injected *Parhyale* displaying a severe axis deformation, among other unspecific phenotypes. The abdominal segments are bent dorsally (white arrow).

### 3.5.4.2 Effects of siRNA injection targeting *Ph six3*

*Ph six3* expression arises shortly after the onset of stage 11 within single cells that are located medially in the pre-antennal part of each head lobe. The expression of *Ph six3* remains largely restricted to this area of the embryonic head and parallels the development and establishment of medial head and brain structures as well as the foregut (3.3.1.7). In order to address the functional role of *Ph six3* in these developmental processes, a siRNA-based RNAi approach was followed.

Derived from non-conserved regions of the *Ph six3* mRNA, three specific siRNA molecules were designed and ordered using BLOCK-iT™ RNAi Designer (Invitrogen, 5.5.2), termed *Ph six3* si4, *Ph six3* si6 and *Ph six3* si9. Their sequence was cross-checked with available sequence information of other Parhyale genes, especially the only other known representative of six-class genes in Parhyale, Ph six4. Single-cell embryos were injected with each individual *Ph six3* siRNA molecule species (100 µM concentration) and with combinations of all three (concentrations 33 µM each and 100 µM each). A small number of two-cell embryos (stage S2) were injected with *Ph six3* si4 in order to address clonal RNAi effects. Additionally, few embryos that had just exited gastrulation were injected with combinations of all three siRNA molecule species (concentrations 33 µM each) in order to find out if siRNA molecules could transgress cell membranes and, therefore, whether they could be successfully applied to embryos of later stages (see also 3.5.4.4). Injected embryos were incubated at 28°C (5.4.6). Their development was monitored and documented at 24h and 96h post-injection, and from there, on consecutive days until the completion of embryogenesis. The process of hatching was documented separately.

In summary, of all embryos evaluated via WMISH (n=207), 51.2 % displayed a reduction in *Ph six3* expression. This percentage is composed of 28.5% of embryos that showed weak reduction, 14% that showed strong reduction of *Ph six3* expression and 8.7% of embryos that did not show any specific *Ph six3* signal at all (Figure 89). All of those embryos who did not show any specific *Ph six3* signal at all had been fixed and dissected when they had reached stages 11 to 13. At these stages, *Ph six3* is expressed bilaterally in single or very small groups of cells (3.3.1.7). This implies that these embryos expressed the relatively lowest levels of *Ph six3* in comparison to embryos of later stages, that they still had a relatively low amount of cells in general and had that they had progressed in development for the shortest period of time after injection of the siRNA. 36.2 % of all examined embryos did not show any

decrease or reduction in levels and pattern of *Ph six3* expression as compared to wild-type. The remaining embryos could not be evaluated (8.7 %) or showed a dramatically altered morphology (3.9 %) that did not allow for direct comparison to wild-type.

In embryos of stages 20 to 23, the wild-type *Ph six3*-expression pattern encompasses a bilateral medial domain in the pre-antennal part of the head as well as expression in the anterior stomodeal field. Later *Ph six3* expression is found in the oesophageal roof and the foregut as well as in bilateral groups of cells in the pre-antennal hemispheres. Prospectively, these areas of the pre-antennal hemispheres give rise to the medullar regions of the protocerebrum. Typically, siRNA-mediated loss of *Ph six3* signal abolished the medullar expression completely and reduced the size of the medial and stomodeal expression domains in affected embryos (Figure 91, Figure 32, Figure 93).

Notably, each single siRNA molecule species had a different efficacy when applied to *Parhyale* embryos (Figure 88 B, D, F; Figure 89). Injection of *Ph six3* si4 (100 µM) caused reduction of *Ph six3* expression in 75.7% (n=33) of examined embryos, whereas 62.4% of embryos injected with *Ph six3* si9 (100 µM; n=56) and 45% of all embryos injected with *Ph six3* si6 (100 µM; n=60) showed a decrease in *Ph six3* expression. In two-cell embryos injected with *Ph six3* si4 a similar decrease of *Ph six3* expression was restricted to one side of the embryo (Figure 90, see also below). Application of a combination of all three *Ph six3* siRNA molecule species led to a reduction of *Ph six3* expression in 36.4% of embryos (concentration of 33µM each) and 21.6% of embryos (concentration of 100µM each), respectively. It is important to note that for each approach of siRNA injections, a relatively smaller percentage of embryos that show reduction in *Ph six3* signal correlated to a higher degree of lethality post-injection (Figure 88, Figure 89). This is most apparent from the comparison of embryos injected with different concentrations of the combination of all three siRNA molecule species: 7.7% of embryos injected with a combination of all three siRNA molecule species, 33 µM each, died between 96h (which corresponds to the observed complete establishment of the initial *Ph six3* expression) and hatching (Figure 88 C), whereas 17.2% of those injected with the triple amount of siRNA died in the same period post-injection (Figure 88 E). Control embryos injected with water showed a significantly lower tendency of lethality during their development. In contrast to embryos injected with siRNA, virtually all control embryos that exhibited wild-type development and phenotypes at 24h post-injection remained wild-type and hatched (Figure 88 A). Embryos that had been injected with siRNA targeting *Ph kni1* all showed wild-type expression of *Ph six3* (Figure 90, Figure

91, n=20). Similarly, a small fraction of embryos injected with siRNA targeting *Ph six3* transcripts were subjected to *Ph kni1* WMISH. They showed wild-type *Ph kni1* expression (not shown, n=15).

Several of those embryos that had been injected one-sided at stage 2 exhibited asymmetric *Ph six3* signal at early stages (Figure 90). This indicates that in these embryos, siRNA-mediated degradation of *Ph six3* transcripts was, in fact, restricted to the descendants of the injected cell. However, the majority of examined embryos of this approach exhibited wild-type *Ph six3* expression at early stages. At later stages, neither asymmetric nor global decrease of *Ph six3* signal was observed in any of the embryos injected at S2. An analogous approach of larger scale was carried out for siRNA targeting *Ph kni1* (3.5.4.3).

The small number of embryos that had been injected with siRNA after they had completed gastrulation did not unambiguously exhibit a reduction of *Ph six3* signal. A more comprehensive approach addressing the feasibility of systemic siRNA-mediated RNAi in embryos injected after gastrulation was taken for *Ph kni1* (3.5.4.4).

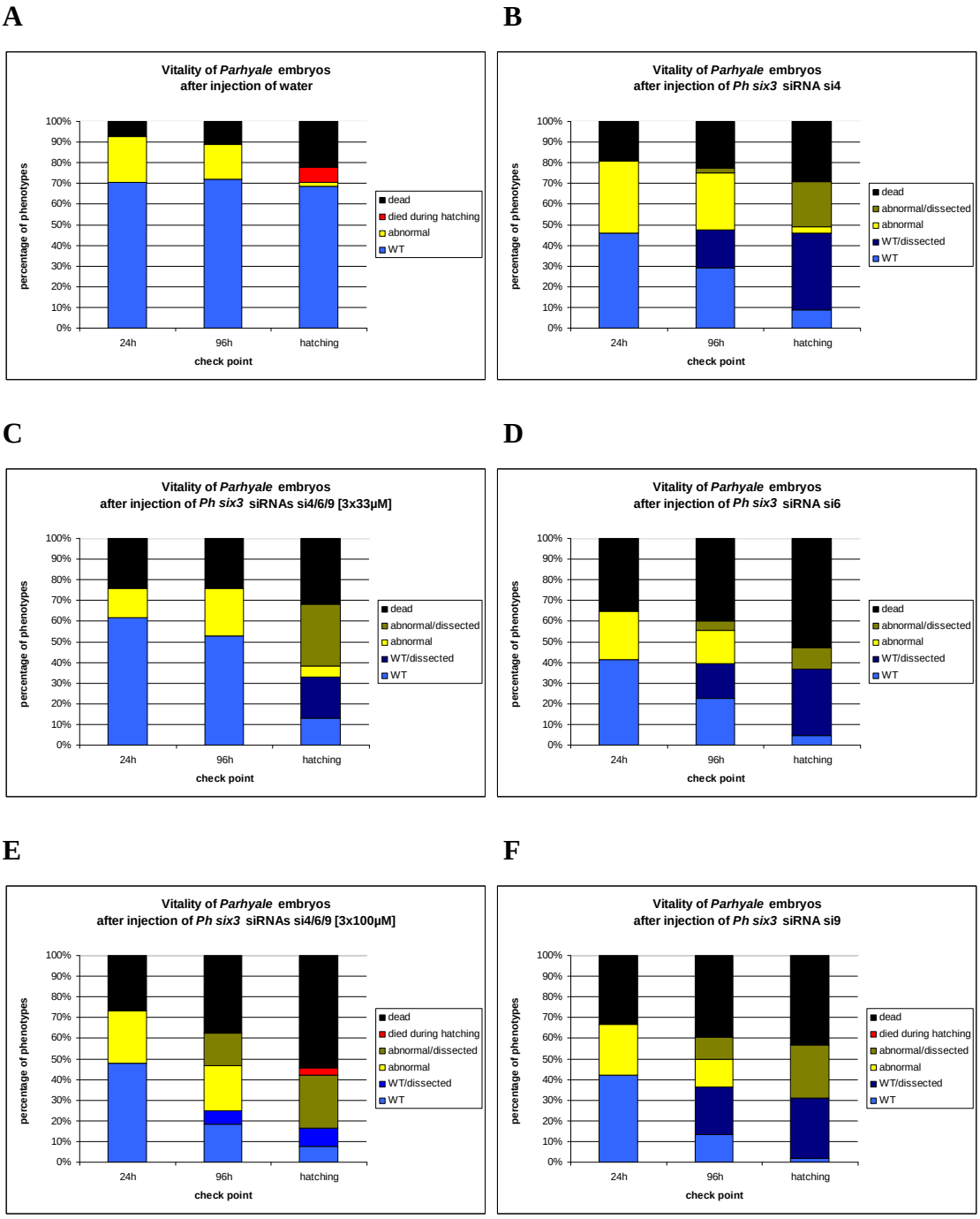



**Figure 88:** **Vitality of *Parhyale hawaiensis* embryos after injection of different species and combinations of siRNA targeting *Ph six3* (B-F) compared to control injections with water (A).** At several points of time after injection (24h, 96h and immediately after hatching), the status of injected embryos was evaluated according to the following criteria: wild-type phenotype (blue), embryos exhibiting abnormal or defective development (yellow), dead embryos (black) and embryos that died during the attempt to hatch (red). Embryos that had been removed from the incubation plates for dissection and WMISH at the time of the evaluation are indicated as dark blue (removed as wild-type) and dark yellow (removed exhibiting developmental defects). All embryos shown here were injected at one-cell stage. **A** Vitality of *Parhyale hawaiensis* embryos after control injection of water (n=54), **B** after injection of siRNA si4 (concentration 100µM; n=168), **C** after injection of a combination of siRNAs si4, si6 and si9 (concentration 100µM; n=91), **D** after injection of siRNA si6 (concentration 100µM; n=168), **E** after injection of a combination of siRNAs si4, si6 and si9 (concentration 300µM; n=238) and **F** after injection of siRNA si9 (concentration 100µM; n=157).

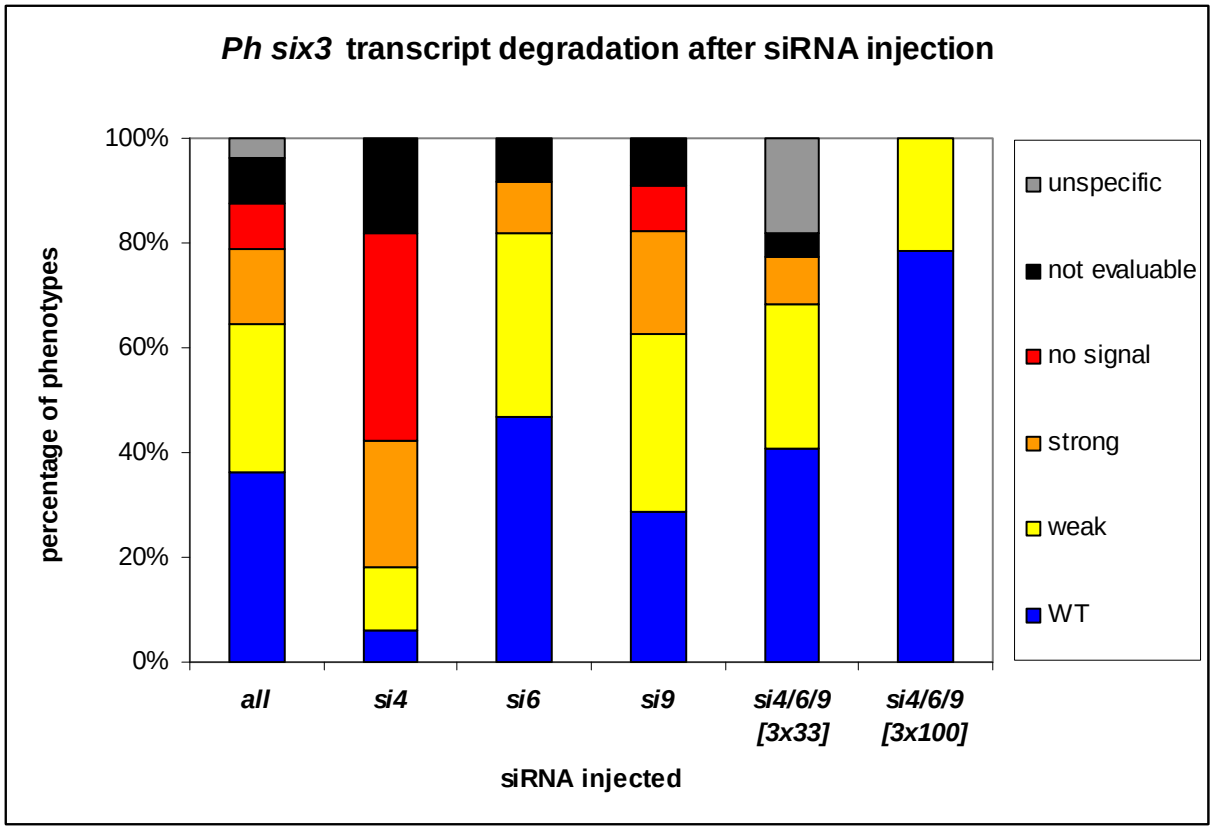



**Figure 89: Evaluation of WMISH signal reduction of siRNA-injected embryos compared to WT, according to the parameters described above (3.5.4).** The first column to the left ('all') shows the summarised results of all examined embryos (n=207). The columns that follow to the right each show the results for one set of injections. From left to right, shown are: results for embryos injected with si4 (100 µM, n=33), si6 (100 µM, n=60), si9 (100 µM, n=56), the combination of si4, si6 and si9 (33 µM each, n=44) and the combination of si4, si6 and si9 (100 µM each, n=14). For each column, the fraction of injected embryos that showed WT WMISH signal is depicted in blue, those showing a slightly reduced WMISH signal are represented in yellow, a strong reduction in WMISH signal is represented in orange, and the fraction of injected embryos that did not show any WMISH signal is depicted in red. The fraction of embryos that could not be evaluated due to damage is shown in black. Several embryos that had been injected with a combination of all three siRNA molecule species (33 µM each) and had been fixed and dissected at stage 18 showed strong but unspecific WMISH signal. They are represented in grey.

Impacts on the morphology that correlated to siRNA-mediated RNAi were examined and assessed on embryos that had been subjected to WMISH and showed a decrease in *Ph six3* expression and on those embryos injected with siRNA that had been allowed to hatch.

Prior to stage S15, the ectodermal germ band of the *Parhyale* embryo resembles a two-dimensional layer of cells. Embryos of stages earlier than S15 that showed a reduction in *Ph six3* signal exhibited wild-type morphology. Specifically, no deletion of single cells or of tissues could be observed (Figure 90). Accordingly, the medial area of the head was established normally in all examined embryos (see also 3.3.1.7). However, at these stages, *Ph six3* is expressed in only a small number of cells. For this reason, the deletion or degradation of only few cells might not be easy to detect by examining images of nuclear labelling.

In contrast, in embryos of subsequent stages, during which neural and stomodeal characters had begun to develop, a decrease in *Ph six3* expression was typically associated with a reduction of the size and cell number of associated tissues. In embryos of stages 18 to 20, a decrease of *Ph six3* signal typically correlated with a considerable reduction of the size of the medial, *Ph six3*-positive head regions (Figure 91). In several embryos, a slight distortion and deformation of surrounding tissues could be observed. For instance, the stomodeal opening appeared to be slightly enlarged as compared to wild-type, with the lateral stomodeal projections protruding outward more prominent (Figure 91). In addition, the labrum appeared to be set back and its protrusion over the stomodeum delayed. As a consequence, the morphology of the medial head of affected embryos was more similar to that of wild-type embryos of earlier stages (Figure 92 A).

In later stages of embryogenesis (S22), the medial area of the pre-antennal part of the head that normally expresses *Ph six3* appeared similarly reduced in size. As a consequence, the elongating labrum projected from this area of the embryonic head at an altered angle as compared to wild-type embryos (Figure 92 B).

In embryos of stage 23, the foregut resembles an arched tube that reaches into the central part of the emerging head capsule (3.1.4). Several embryos that revealed a strong reduction of *Ph six3* expression within the foregut roof also showed malformations of the foregut which appears collapsed or wrinkled (Figure 93). None of the examined embryos of stages 16 or later showed a complete reduction of *Ph six3* expression. Accordingly, more severe morphological phenotypes resulting from complete *Ph six3* loss-of-function could not be observed.

From each set of injections, a small percentage of siRNA-treated embryos that had been allowed to hatch showed unspecific morphological phenotypes, comparable to control embryos injected with water (1.9-5.5%). Aside from these effects, no phenotypes related to parts of the embryo that express *Ph six3* were observed. Specifically, the labrum was present and appeared wild-type in all analysed embryos.

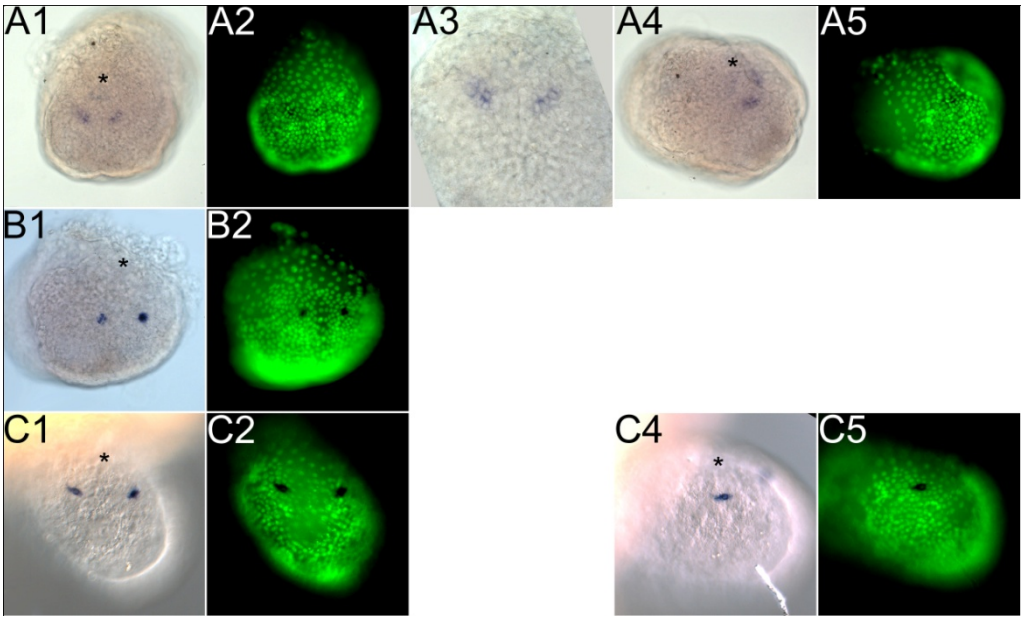



**Figure 90: Decrease of *Ph six3* WMISH signal in S12 embryos after injection of siRNA targeting *Ph six3*.** In all images, anterior is up. In the images A1, A4, B1, C1 and C4 the anterior end of the embryo is indicated by a black asterisk. Shown are: **A** Embryo injected at stage S1 with si6 (100µM). **A1,2** Frontal orientation of the embryo, showing both head lobes. **A1** Bright field image. **A2** Nuclear labelling (SYTOX®). **A3** Close-up of the embryo showing the remnant *Ph six3* signal in detail (bright field image, x20 magnification). **A4,5** lateral orientation of the same embryo showing the right head lobe. **A4** Bright field image. **A5** Nuclear labelling (SYTOX®). **B** Embryo injected one-sided at stage S2 with si6 (100µM). The embryo is oriented frontal, slightly tilted to the left. The *Ph six3* WMISH signal is asymmetric, indicating a reduction of *Ph six3* transcripts within the right hemisphere of the embryo (clonal effect). **B1** Bright field image. **B2** Nuclear labelling (SYTOX®). **C** Control embryo injected with siRNA targeting *Ph kni1* (si1, 100µM, S1) and subjected to *Ph six3* WMISH. The embryo shows the wild-type *Ph six3* expression pattern. **C1, 2** Frontal orientation of the embryo, showing both head lobes. **C1** DIC image. **C2** Nuclear labelling (SYTOX®). **C4, 5** Lateral orientation of the same embryo showing the right head lobe. C4 DIC image. C5 Nuclear labelling (SYTOX®). All embryos shown exhibit wild-type morphology.

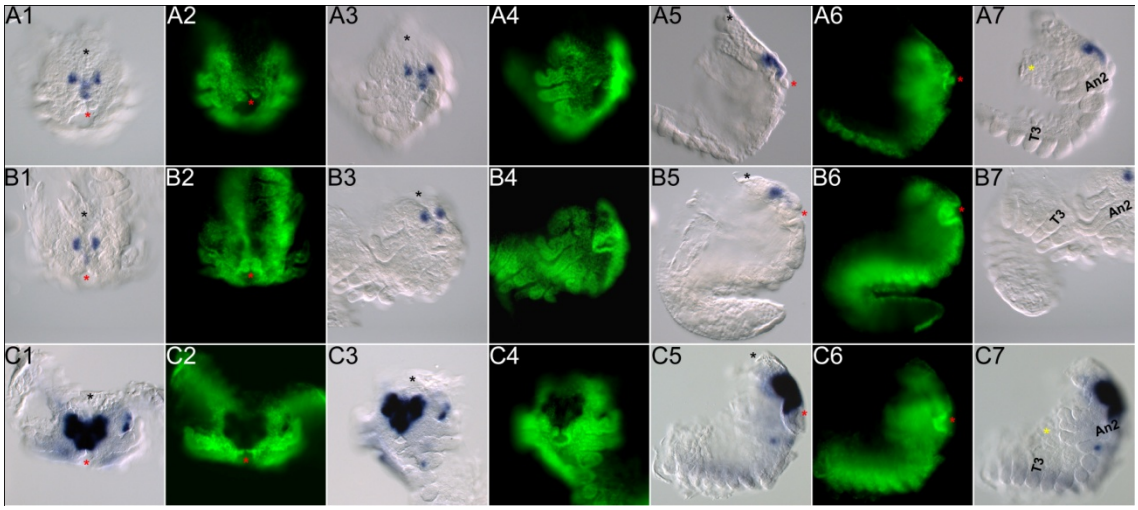



**Figure 91: Decrease of *Ph six3* WMISH signal in S20 and S21 embryos after injection of siRNA targeting *Ph six3* and observed morphological impacts.** In all images, anterior is up. All images having odd numbers are DIC images; all images having even numbers show the respective nuclear labelling (SYTOX®). A7, B7 and C7 DIC images are used to define the specific stage of the embryo. In all other DIC images, the anterior end of the embryo is indicated by a black asterisk. In images having the numbers 1, 2, 5 and 6, the centre of the stomodeal opening is indicated with a red asterisk. **A** Stage 20 **e**mbryo injected at stage 1 with si9 (100µM) exhibiting strong reduction of *Ph six3* WMISH signal. The embryo is ruptured directly posterior of the stomodeal opening. **A1, 2** Frontal orientation of the embryo. **A3, 4** Frontal orientation, tilted to the left. **A5, 6** Lateral orientation, sagittal focal plane. **A7** Lateral orientation, focal plane on the An2 appendage. In order to confirm the embryo's stage on the basis of appendage and midgut anlagen morphology, the An2 and the T5 appendages are labelled and the lateral midgut anlagen are marked with a yellow asterisk (3.1.3; 3.1.4, Browne et al., 2005). **B** Stage 21 **e**mbryo injected at stage 1 with si9 (100µM) exhibiting strong reduction of *Ph six3* WMISH signal. The appearance of the stomodeal region resembles that of a younger embryo. This is likely a morphological consequence of the size reduction of *Ph six3*-positive tissue observed in this strongly affected embryo. **B1, 2** Frontal orientation of the embryo. **B3, 4** Frontal orientation, tilted to the left. **B5, 6** Lateral orientation, sagittal focal plane. **B7** Lateral orientation, focal plane on the An2 appendage. In order to confirm the embryo's stage on the basis of appendage morphology, the An2 and the T5 appendages are labelled (see also Introduction, Browne et al., 2005). **C** Untreated embryo of stage 20 with wild-type *Ph six3* WMISH signal (see also 3.3.xy). **C1, 2** Frontal orientation of the embryo. **C3, 4** Frontal orientation, tilted to the right. **C5, 6** Lateral orientation, sagittal focal plane. **C7** Lateral orientation, focal plane on the An2 appendage. In order to confirm the embryo's stage on the basis of appendage and midgut anlagen morphology, the An2 and the T5 appendages are labelled and the (ruptured) lateral midgut anlagen are marked with a yellow asterisk (3.1.3; 3.1.4; Browne et al., 2005).

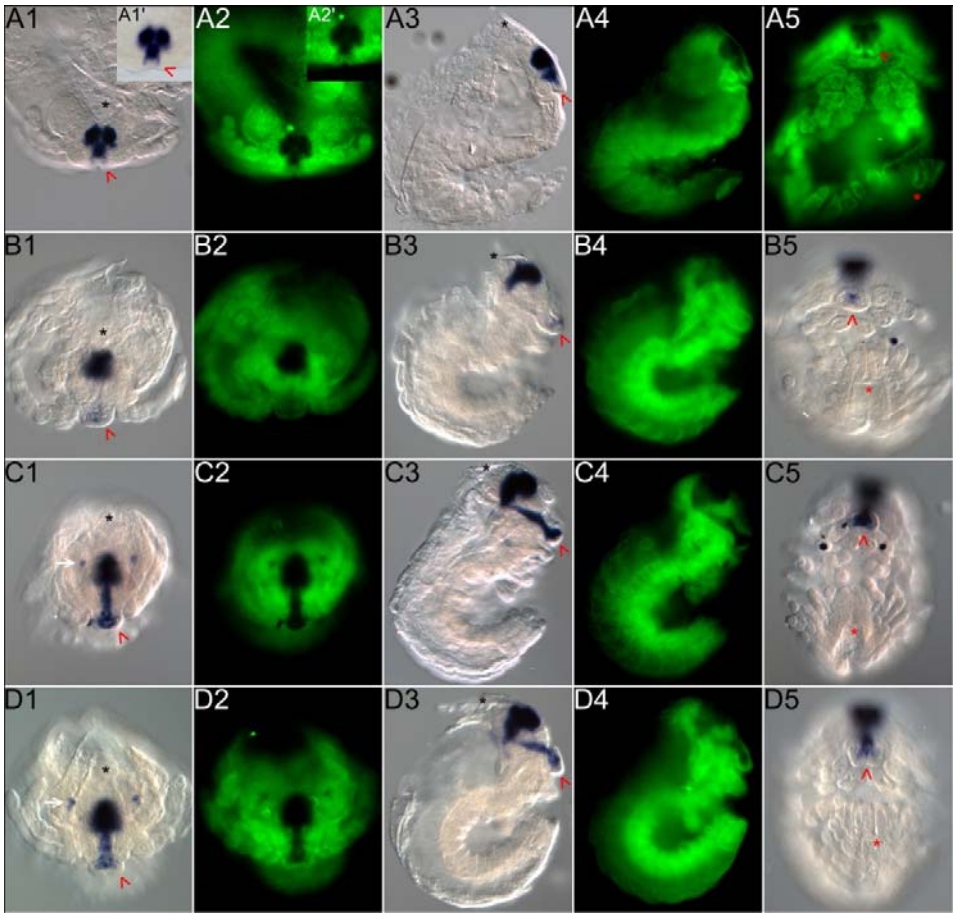



**Figure 92: Decrease of *Ph six3* WMISH signal in S22 embryos after injection of siRNA targeting *Ph six3* and observed morphological impacts.** In all images, anterior is up. All images having numbers 1 and 3 are DIC images. Here, the anterior end of the embryo is marked by a black asterisk and the distal tip of the labrum is indicated by a red arrowhead. All images having numbers 2 and 4 show the respective nuclear labelling (SYTOX®). A5 (nuclear labelling), B5, C5 and D5 (DIC images) are used to define the specific stage of the embryo. In these, the hindgut proctodeum is marked by a red asterisk and the distal tip of the labrum is indicated by a red arrowhead. **A** Early stage 22 embryo injected at stage 1 with si6 (100µM) exhibiting weak reduction of *Ph six3* WMISH signal. Notably, as a putative consequence of loss of *Ph six3* function, the morphology of the medial head of the embryo appears more similar to that of wild-type embryos of earlier stages (see also text, 3.3.1.7 and figures 6 and 7 A). **A1, 2** Anterior orientation of the embryo. **A1', 2'** Detail of the medial head region expressing *Ph six3*, focal plane is on the labrum. **A3, 4** Lateral orientation, sagittal focal plane. **A5** Ventral orientation, focal plane on the labrum. The hindgut proctodeum that becomes visible during S21 is marked with a red asterisk (3.1.4, Browne et al., 2005). **B** Stage 22 embryo injected at stage 1 with si6 (100µM) exhibiting strong reduction of *Ph six3* WMISH signal. Notably, as a putative consequence of loss of *Ph six3* function, the medial area of the embryonic head appears reduced in size and the labrum protrudes at a different angle as compared to wild-type. (see also text and D). **B1, 2** Anterior orientation of the embryo. **B3, 4** Lateral orientation, sagittal focal plane. **B5** Ventral orientation, focal plane on the labrum. The hindgut proctodeum that becomes visible during S21 is marked with a red asterisk (3.1.4; Browne et al., 2005). **C** Control embryo injected with siRNA targeting *Ph kni1* (si1, 100µM, S1) and subjected to *Ph six3* WMISH. The embryo shows the wild-type *Ph six3* expression pattern. A white arrow marks the right medullar expression domain of *Ph six3* that was only present in embryos exhibiting wild-type *Ph six3* expression. **C1, 2** Anterior orientation of the embryo. **C3, 4** Lateral orientation, sagittal focal plane. Note that although strong *Ph six3* WMISH signal is present in the labrum, its morphology appears slightly altered as compared to wild-type. **C5** Ventral orientation, focal plane on the labrum. The hindgut proctodeum that becomes visible during S21 is marked with a red asterisk (3.1.4; Browne et al., 2005). **D** Untreated S22 embryo exhibiting the wild-type *Ph six3* expression pattern. A white arrow marks the right medullar expression domain of *Ph six3* that was not found in embryos injected with siRNA targeting *Ph six3*. **D1, 2** Anterior orientation of the embryo. **D3, 4** Lateral orientation, sagittal focal plane. **D5** Ventral orientation, focal plane on the labrum. The hindgut proctodeum that becomes visible during S21 is marked with a red asterisk (3.1.4; Browne et al., 2005).

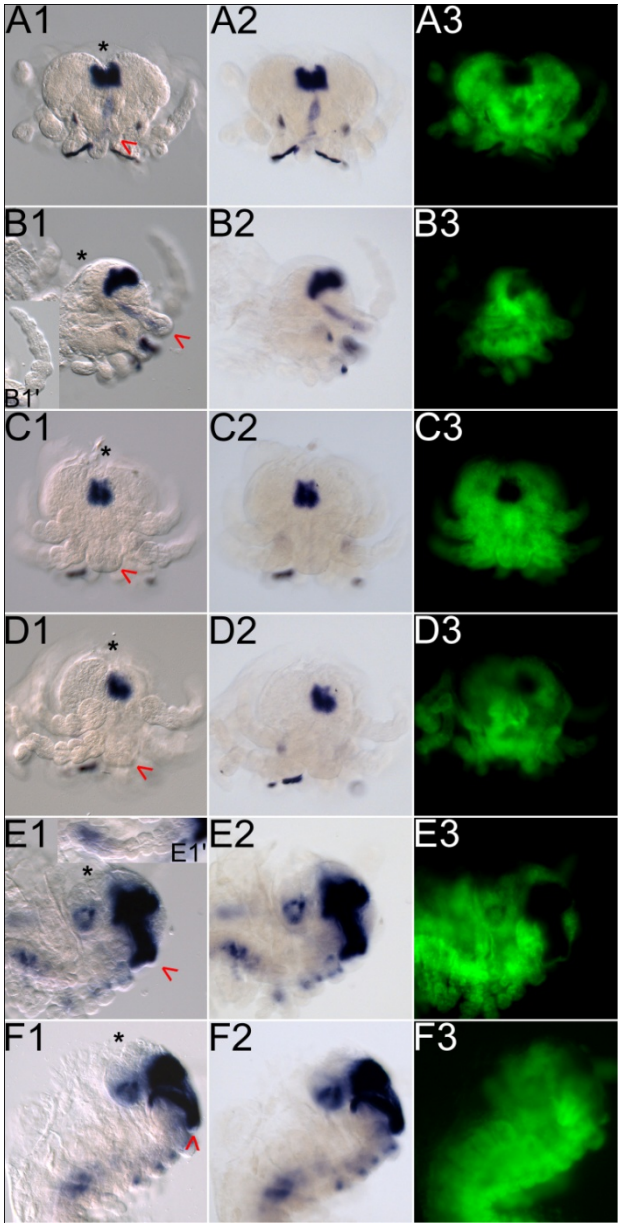



**Figure 93: Decrease of *Ph six3* WMISH signal in S23 embryos after injection of siRNA targeting *Ph six3* and observed morphological impacts.** In all images, anterior is up. All images having number 1 are DIC images. Here, the anterior end of the embryo is indicated by a black asterisk and the distal tip of the labrum is marked with a red arrowhead. All images having number 2 are bright field images and all images having number 3 show the respective nuclear labelling (SYTOX®). The stage of the affected embryos was confirmed on the base of the An2 appendage morphology (compare B1', C1 to E1'; see Browne et al., 2005, Fig. 11). **A** Stage 23 embryo injected at stage 1 with si9 (100µM) exhibiting strong reduction of *Ph six3* WMISH signal. The embryo is mounted in anterior orientation. **B** Same embryo as in A, mounted in lateral orientation (sagittal focal plane). **B1'** Detail of the An2 appendage (in focus). **C** Stage S23 embryo injected at stage S1 with a combination of si4, si6 and si9 (100µM) exhibiting strong reduction of *Ph six3* WMISH signal. The embryo is mounted in anterior orientation. **D** Same embryo as in C, mounted in anterior orientation, tilted to the left. **E** Untreated stage S23 embryo exhibiting the wild-type *Ph six3* expression pattern. The embryo is mounted in anterior orientation, tilted to the left. **E1'** Detail of the An2 appendage (in focus). **F** Same embryo as in E, mounted in lateral orientation (sagittal focal plane).

### 3.5.4.3 Effects of siRNA injection targeting *Ph kni1*

During *Parhyale* development, *Ph kni1* is expressed dynamically in multiple regions, fields and tissues of the embryo, including the pre-antennal part of the head, the stomodeum during its formation, tissues related to the stomatogastric nervous system and throughout appendage development (3.4.7). In order to address the functional role of *Ph kni1* in these developmental processes, in particular in head development, a siRNA-based RNAi approach was followed.

Derived from non-conserved regions of the *Ph kni1* mRNA, three specific siRNA molecules were designed and ordered using BLOCK-iT™ RNAi Designer (Invitrogen, 5.5.2), termed *Ph kni1* si1, *Ph kni1* si2 and *Ph kni1* si8. Their sequence was cross-checked with available sequence information of other *Parhyale* genes, especially the only other known *knirps* representative in *Parhyale*, *Ph kni2*. Single-cell embryos were injected with each individual *Ph kni1* siRNA molecule species (100µM concentration) and with combinations of all three (100 µM each). Two-cell embryos were treated analogously to address clonal RNAi effects. In order to address systemic RNAi effects and the feasibility of triggering RNAi after injection of multicellular embryos of later stages, an additional set of experiments was conducted: embryos were injected with siRNA at various time points after they had completed gastrulation (3.5.4.4). Injected embryos were incubated at 28°C. Their development was monitored and documented at 24h and 96h post-injection, and from there, on consecutive days until the completion of embryogenesis (5.5.3). The process of hatching was documented separately.

In summary, of all embryos that had been injected at one- and two-cell stage and that were evaluated via WMISH (n=124), 37.9 % exhibited reduced *Ph kni1* expression. This percentage is composed of 26.6% of embryos that showed weak reduction and 11.3% that showed strong reduction. None of the embryos exhibited a complete loss of specific *Ph kni1* signal (Figure 94). 39.5 % of the examined embryos did not show any decrease or reduction in levels and pattern of *Ph kni1* expression as compared to wild-type. The remaining embryos could not be evaluated (16.9 %) or showed a dramatically altered morphology (5.6 %) that did not allow for direct comparison to wild-type. Similar to previous observations for Ph six3, each single siRNA molecule species lead to variable results when applied to Parhyale embryos (Figure 94). Also, two-cell embryos treated with siRNA displayed less reduction of *Ph kni1* signal than one-cell embryos. In detail, injection of *Ph kni1* si1 (100 µM) caused

reduction of *Ph kni1* expression in 81.8% (n=11, injection at one-cell stage) and 33.3% (n=12, injection at two-cell stage) of examined embryos. In contrast, injection of *Ph kni1* si2 (100 µM) caused reduction of *Ph kni1* expression in 53.8% (n=13, injection at one-cell stage) and 24.6% (n=57, injection at two-cell stage) of examined embryos. Injection of *Ph kni1* si8 (100 µM) caused reduction of *Ph kni1* expression in 47.4% (n=19, injection at one-cell stage) and 33.3% (n=12, injection at two-cell stage) of examined embryos. Embryos that had been treated with combinations of all three *Ph kni1* siRNA molecule species were not used for this evaluation, but had all been allowed to hatch in order to be examined for morphological phenotypes (see below).

An siRNA-mediated decrease in *Ph kni1* expression was typically represented by a general reduction in signal intensity. In contrast to the situation for siRNA-mediated *Ph six3* decrease, no change in expression domain size was observed. In addition, affected embryos frequently showed a complete loss of small *Ph kni1* expression domains or a mosaic-like absence of *Ph kni1* signal in larger expression domains (Figures 95 and 96).

Similar to the results of siRNA injections targeting *Ph six3*, the lethality of injected embryos varied depending on the siRNA molecule in use, but was significantly higher than the lethality of control embryos injected with water (not shown). Embryos that had been injected with siRNA targeted at *Ph kni1* all showed wild-type expression of *Ph six3* (3.5.4.2) and vice versa (not shown).

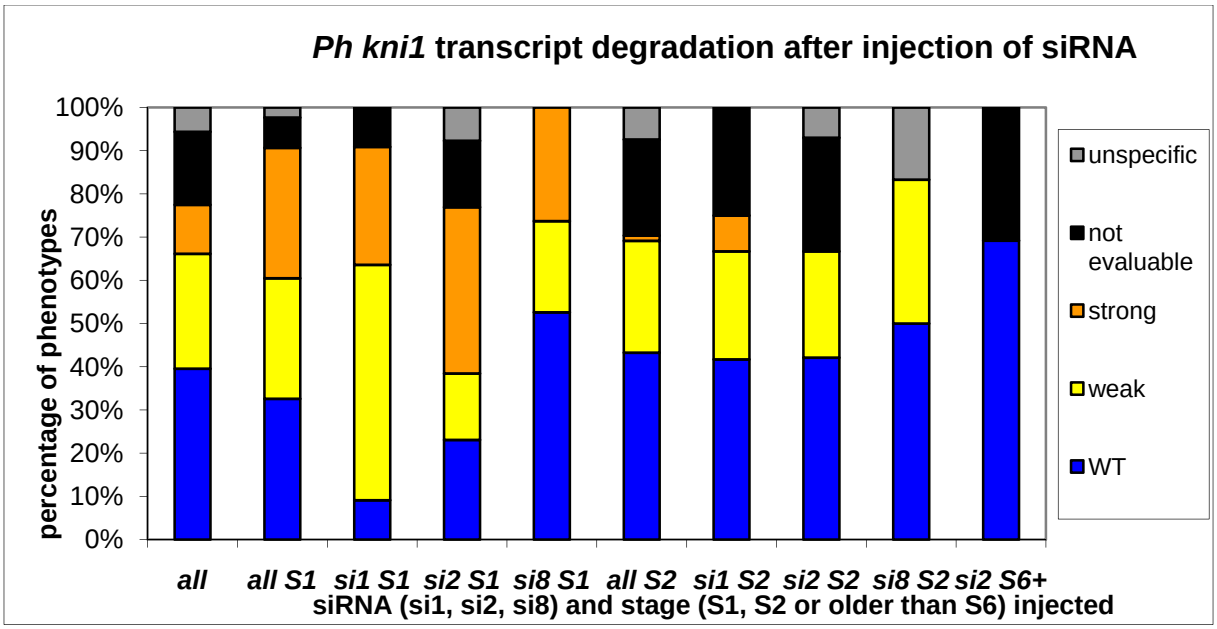



**Figure 94: Evaluation of WMISH signal reduction of *Ph kni1*-specific siRNA-injected embryos compared to WT, according to the parameters described above.** The first column to the left shows the summarised results of all examined embryos (n=124). The next column to the right shows the summarised results of all embryos that had been injected at S1 (n=43). The columns that follow to the right show the results for the following sets of injections: embryos injected at S1 with si1 (100 µM, n=11), si2 (100 µM, n=13) and si8 (100 µM, n=19). The next column to the right shows the summarised results of all embryos that had been injected at S2 (n=81). The columns that follow to the right show the results for the following sets of injections: embryos injected at S2 with si1 (100 µM, n=12), si2 (100 µM, n=57) and si8 (100 µM, n=12). The last column to the right shows the results for the embryos injected at S6 and later stages with si2 (100 µM, n=26). For each column, the fraction of injected embryos that showed WT WMISH signal is depicted in blue, those showing a slightly reduced WMISH signal are represented in yellow and a strong reduction in WMISH signal is represented in orange. None of the examined embryos exhibited a complete loss of WMISH signal. The fraction of embryos that could not be evaluated due to damage is shown in black. Several embryos of various sets of injections showed strong but unspecific WMISH signal. They are represented in grey.

**Figure 95:** **Decrease of *Ph kni1* WMISH signal in *Parhyale* embryos after injection of siRNA targeting *Ph kni1*.**

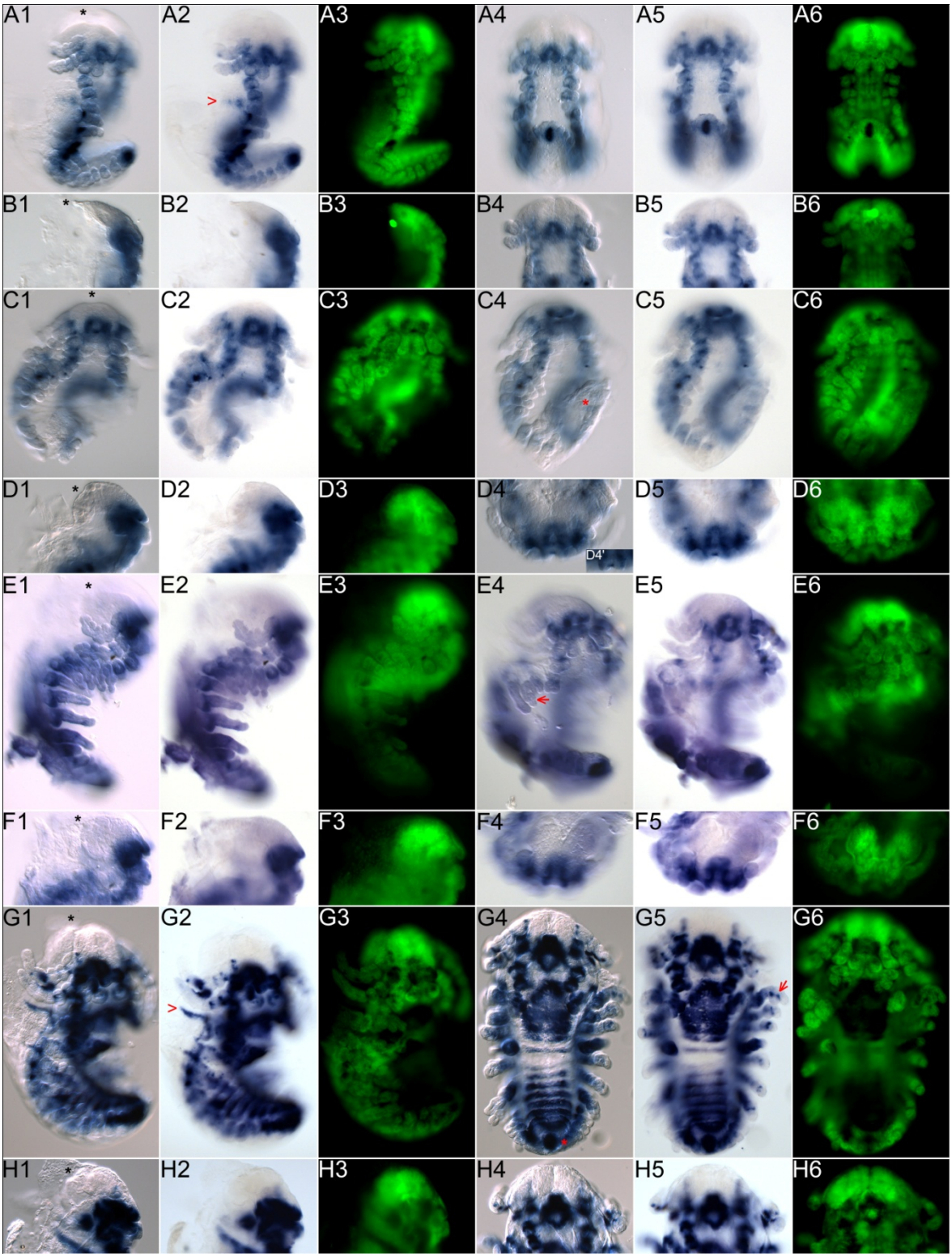


**Figure 95 (previous page): Decrease of *Ph kni1* WMISH signal in *Parhyale* embryos after injection of siRNA targeting *Ph kni1*.** Shown are DIC images (#1, #4), bright field images (#2, #5) and nuclear labelling (#3, #6) of three different embryos of stage 22, all of them exhibiting strong reduction of endogenous *Ph kni1* WMISH signal after injection of *Ph kni1*-specific siRNA (A-F) as well as an untreated S22 embryo exhibiting wild-type *Ph kni1* expression. For illustration, each embryo is shown in 4 different orientations. In all images, anterior is up. In all DIC images to the left (#1), the anterior end of the embryo is indicated by a black asterisk. **A, B** S22 embryo injected at stage 1 with si1 (100µM). **A1, 2, 3** Tilted lateral orientation of the embryo. A1 DIC image. A2 Bright field image. As compared to wild-type, the lateral *Ph kni1* expression domain of segment T1 appears strongly reduced and fragmented (see also G2, red arrowhead). A3 Nuclear labelling (SYTOX®). **A4, 5, 6** Ventral orientation. A4 DIC image. A5 Bright field image. A6 Nuclear labelling (SYTOX®). **B1, 2, 3** Lateral orientation, sagittal focal plane. B1 DIC image. B2 Bright field image. B3 Nuclear labelling (SYTOX®). **B4, 5, 6** Dorsal orientation, focal plane on the stomodeal opening. B4 DIC image. B5 Bright field image. B6 Nuclear labelling (SYTOX®). **C, D** S22 embryo injected at stage 1 with si2 (100µM). **C1, 2, 3** Tilted ventral orientation of the embryo, focal plane on the An2 appendage. C1 DIC image. C2 Bright field image. C3 Nuclear labelling (SYTOX®). **C4, 5, 6** Alternative tilted ventral orientation, focal plane on the proctodeum. C4 DIC image. This embryo does not show *Ph kni1* expression in the hindgut proctodeum (red asterisk). In contrast, wild-type *Ph kni1* expression is strong in this area of the embryo (see also G4, red asterisk). C5 Bright field image. C6 Nuclear labelling (SYTOX®). **D1, 2, 3** Lateral orientation, sagittal focal plane. B1 DIC image. B2 Bright field image. B3 Nuclear labelling (SYTOX®). **D4, 5, 6** Anterior orientation, focal plane on the medullar hemispheres. D4 DIC image. D4' Detail of the labrum, taken from a DIC image of the same embryo and orientation, but different focal plane. D5 Bright field image. D6 Nuclear labelling (SYTOX®). **E, F** Late S22 embryo injected at stage 1 with si8 (100µM). **E1, 2, 3** Lateral orientation of the embryo, focal plane on the An1 appendage. E1 DIC image. E2 Bright field image. E3 Nuclear labelling (SYTOX®). **E4, 5, 6** Tilted ventral orientation, focal plane on the mandibles. E4 DIC image. This embryo does not show *Ph kni1* expression in the developing gnathopod of T3 (red arrow). In contrast, *Ph kni1* expression is found at the branching point of the two distal-most appendage articles of T3 gnathopods in wild-type embryos (see also G5, red arrow). E5 Bright field image. E6 Nuclear labelling (SYTOX®). **F1, 2, 3** Lateral orientation, sagittal focal plane. F1 DIC image. F2 Bright field image. F3 Nuclear labelling (SYTOX®). **F4, 5, 6** Anterior orientation, focal plane on the medullar hemispheres. F4 DIC image. F5 Bright field image. F6 Nuclear labelling (SYTOX®). **G, H** Untreated S22 embryo exhibiting the wild-type *Ph kni1* expression pattern. **G1, 2, 3** Tilted ventral orientation of the embryo, focal plane on the An1 appendage. G1 DIC image. G2 Bright field image. The wild-type lateral *Ph kni1* expression domain of segment T1 is indicated by a red arrowhead. G3 Nuclear labelling (SYTOX®). **G4, 5, 6** Ventral orientation. G4 DIC image. The wild-type *Ph kni1* expression domain of the hindgut proctodeum is indicated by a red asterisk. G5 Bright field image. The wild-type *Ph kni1* expression that is found at the branching point of the two distal-most appendage articles of T3 gnathopods is marked by a red arrow. G6 Nuclear labelling (SYTOX®). **H1, 2, 3** Lateral orientation, sagittal focal plane. H1 DIC image. H2 Bright field image. H3 Nuclear labelling (SYTOX®). **H4, 5, 6** Dorsal orientation, focal plane on the stomodeal opening. H4 DIC image. H5 Bright field image. H6 Nuclear labelling (SYTOX®).

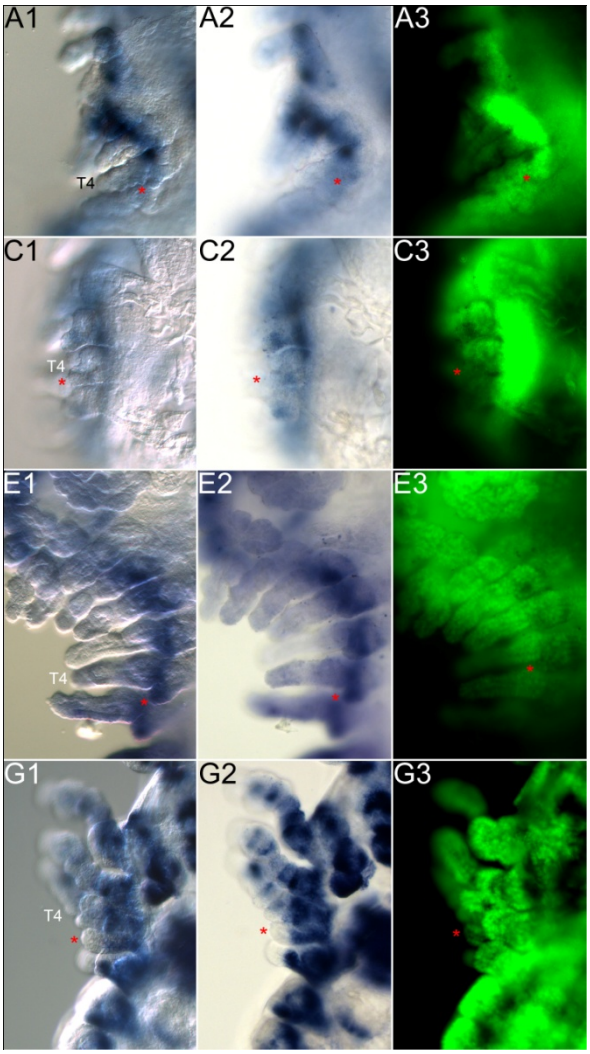



**Figure 96: Details of *Parhyale* embryos exhibiting a decrease of *Ph kni1* WMISH signal after injection of siRNA targeting *Ph kni1*.** Shown are DIC images to the left, bright field images in the central panel and, to the right, nuclear labelling (SYTOX®) of the embryos depicted in figure 95: three different stage 22 embryos, all of which exhibit strong reduction of endogenous *Ph kni1* WMISH signal after injection of *Ph kni1*-specific siRNA (A, C, E) as well as an untreated S22 embryo exhibiting wild-type *Ph kni1* expression (G). The images show details of the left thorax in dorsal view. In all images, anterior is up. In all DIC images, the T4 appendage is labelled. In all images, the gill of the T4 appendage is marked by a red asterisk. **A** Detail of embryo shown in figure 95 A, B: S22 embryo injected at stage 1 with si1 (100µM). **C** Detail of embryo shown in figure 95 C, D: S22 embryo injected at stage S1 with si2 (100µM). **E** Detail of embryo shown in figure 95 E, F: late S22 embryo injected at stage S1 with si8 (100µM). **G** Detail of embryo shown figure 95 G, H: untreated S22 embryo exhibiting the wild-type *Ph kni1* expression pattern.

Of those siRNA-treated embryos that had been allowed to hatch, many showed no or only unspecific morphological phenotypes. Few animals displayed severe phenotypes that were unique, i.e. that were not observed in other animals. These were probably also unspecific. The percentage of unspecific phenotypes was comparable to control embryos injected with water (not shown). Several animals, however, displayed pereopods that appeared bent or arranged abnormally. *Ph kni1* expression is found in all appendages throughout development (3.4.7). Based on these considerations and preliminary findings, the role of *Ph kni1* in appendage morphogenesis was addressed via a more directed approach. One- and, to a lesser extent, two-cell embryos were injected with a combination of all three siRNA molecule species targeting *Ph kni1* (100 µM each). Those that survived and hatched successfully (n=32) were examined specifically for pereopod phenotypes (5.5.4, Figure 97, Figure 98). 28.1% showed wild-type pereopods (n=9). 56.3% displayed single pereopods that were bent or arranged abnormally (weak pereopod phenotype, n=18). 12.5% displayed several affected pereopods but still at least one wild-type pereopod (n=4, medium pereopod phenotype). In a single animal (3.1%), all pereopods were affected to varying degrees (strong pereopod phenotype). In embryos that had been subjected to clonal injections of Ph *kni1*-specific siRNA, medium pereopod phenotypes manifested one-sided (Figure 98 C). Aside from pereopod phenotypes, a small fraction of animals showed unspecific phenotypes unrelated to pereopod morphology, comparable to control embryos injected with water (not shown).

Control embryos were injected with water or, alternatively, with siRNA targeting *Ph six3* (combination of all three siRNA species, 100 µM each). They were treated and examined as above. Of those animals that had been injected with water and that had successfully hatched (n=47), 59.6% showed wild-type pereopods (n=28) and 34% showed weak pereopod phenotypes (n=16). None of the control animals showed medium or strong pereopod phenotypes. The remaining 6.4% (n=3) displayed pereopod phenotypes that resulted from premature hatching. In these cases, the ‘amphipodal’ bidirectional arrangement of anterior and posterior pereopod pairs had not been established; instead, all pereopods pointed to the same direction. In addition, the allometric relations of head, thorax and abdomen were largely different from those of wild-type animals. For these reasons, this phenotype was considered an injection artefact (see also 3.5.4.1). As an additional control, a small number of embryos that had been injected with siRNA targeting *Ph six3* and that had hatched successfully (see above, n=6) was examined analogously. Similar to control animals injected

with water, 66.7% (n=4) showed wild-type pereopods while the remaining 33.3% (n=2) showed weak pereopod phenotypes.

From these findings, I conclude that a decrease of *Ph kni1* activity interferes with the correct morphogenesis and relative arrangement of pereopods in *Parhyale*. However, several other effects, putatively caused by injection alone, appear to interfere with pereopod development to a lesser degree as well.

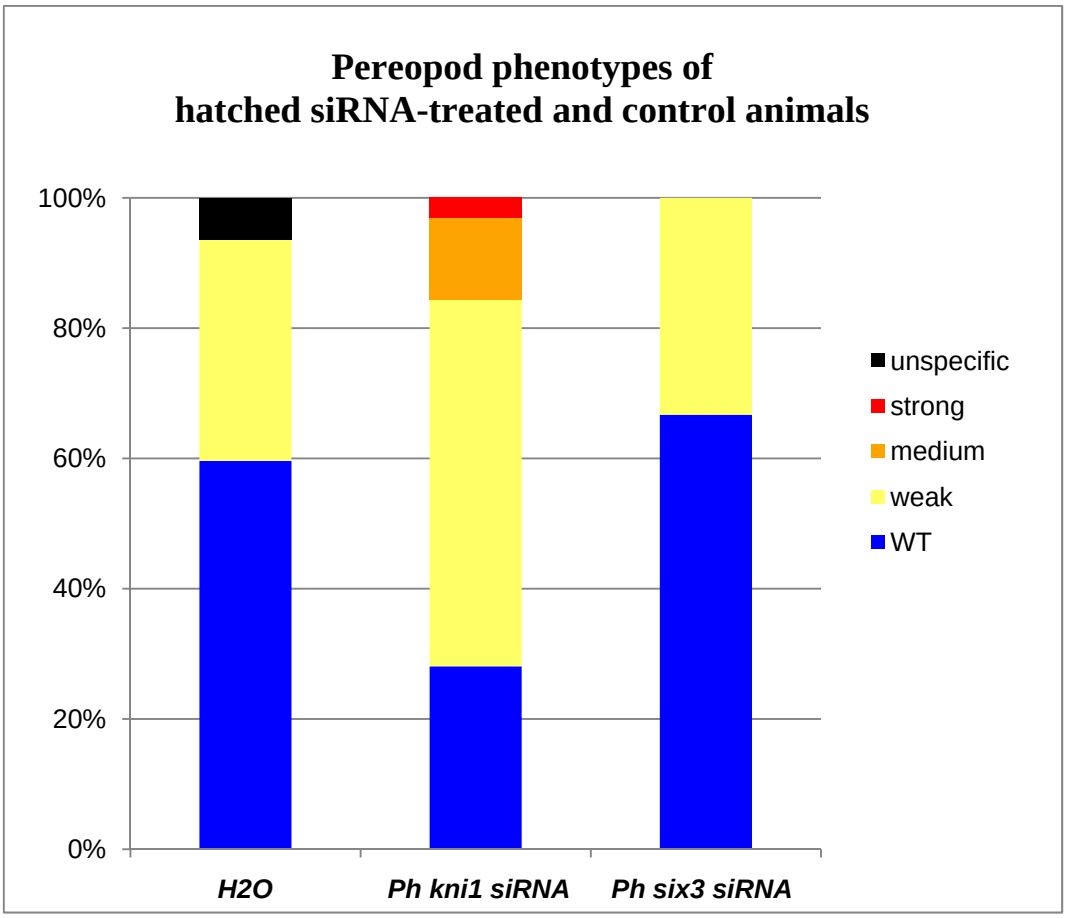



**Figure 97: Occurrence of pereopod phenotypes in animals that had been injected with siRNA targeting *Ph kni1* (middle column) as compared to control animals injected with water (left column) or siRNA targeting *Ph six3* (right column).** From all three sets of injections, animals exhibiting single malformed or misarranged pereopods hatched (weak phenotype). However, this phenotype occurred almost twice as frequent in animals that had been injected with *Ph kni1*-specific siRNA (56.3%, n=32) as compared to control animals injected with water (34%, n=47) or *Ph six3*-specific siRNA (33.3%, n=6). Only in animals that had been injected with siRNA targeting *Ph kni1*, medium (12.5%) and strong (3.1%) pereopod phenotypes were identified. In all columns, the fraction of examined animals exhibiting wild-type morphology is depicted in blue, weak phenotype in yellow, medium phenotype in orange and the fraction exhibiting the strong phenotype in red. The fraction of animals that showed unspecific morphological phenotypes is depicted in black.

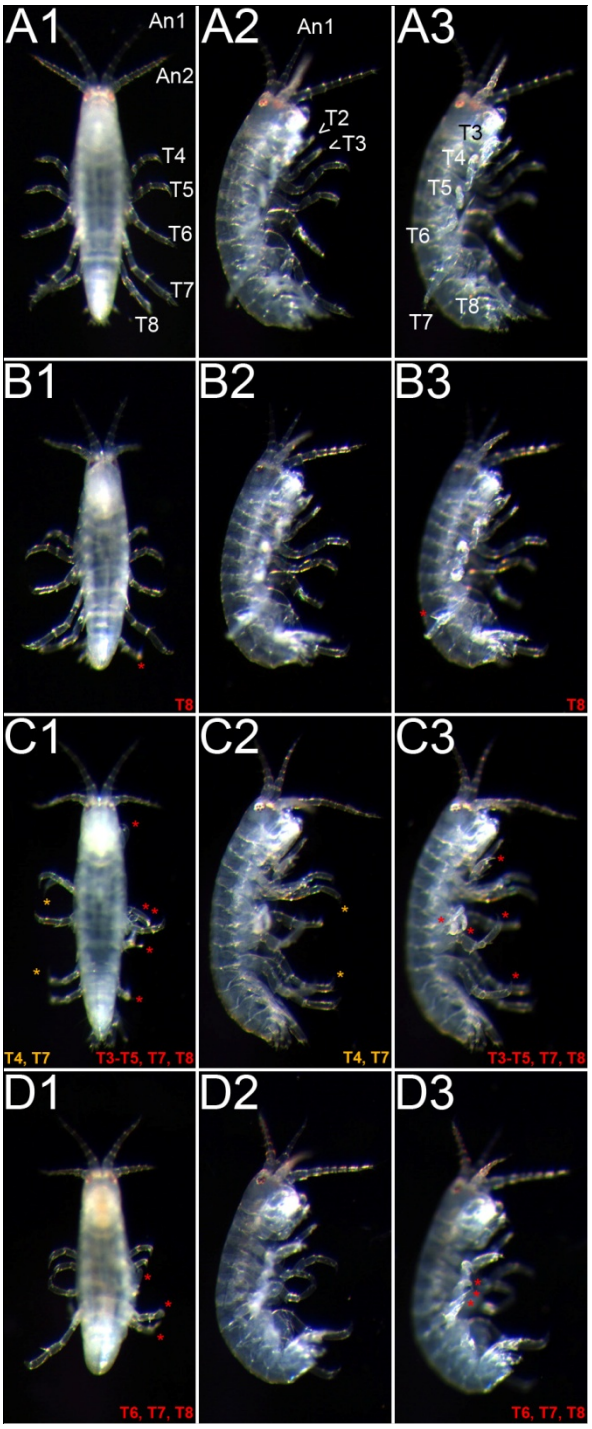



**Figure 98: Pereopod phenotypes of hatched *Parhyale* after injection of siRNA targeting *Ph kni1*.** All animals shown in this figure had hatched 3 days prior to photographing. At the moment the images were taken, all animals were alive, reacted to mechanical stimulation and exhibited motility in all appendages (5.5.4). The images were taken using the epifluorescence stereomicroscope MZ16FA (Leica Microsystems, Germany, 5.6) with indirect illumination by a cold light source and without filter. Images to the left show dorsal views of the examined animals (A1, B1, C1, D1), images in the centre show lateral views, focal plane on the appendages of the left body side (A2, B2, C2, D2) and images to the right show lateral views, focal plane on the appendages of the right body side (A3, B3, C3, D3). Orange and red asterisks mark morphologically affected appendages of the left and right body sides, respectively. In all images, anterior is up. **A** Untreated *Parhyale* exhibiting wild-type appendage arrangement and morphology. **A1** For illustration, the first and second antennae are labelled (An1, An2) as well as the pereopods of the thoracic segments T4 to T8. **A2** For illustration, the left first antenna is labelled (An1) as well as the left gnathopods of the thoracic segments T2 und T3. **A3** For illustration, the appendages of the thoracic segments T3 to T8 (right body side) are labelled. **B** *Parhyale* exhibiting the weak pereopod phenotype. The animal had been injected at stage 1 with siRNA targeting *Ph kni1*. The right pereopod of segment T8 is bent anterior (**B3**). In wild-type animals, the pereopods of T7 and T8 of one body side conclude an angle of about 60° in resting position (see also A3). **C** *Parhyale* exhibiting the medium pereopod phenotype in pereopods of T4 and T7 (left body side; C1, C2) as well as the pereopods of to T4, T5, T7 and T8 (right body side; C1, C3). In addition, the left gnathopod of T3 is abnormally bent outwards and, therefore, visible in dorsal view (C1). The animal had been injected at stage 1 with siRNA targeting *Ph kni1*. **D** *Parhyale* exhibiting the medium pereopod phenotype in pereopods of T6 to T8, right side only. The animal had been injected at stage 2 with siRNA targeting *Ph kni1*.

#### 3.5.4.4 Addressing systemic siRNA-mediated RNAi

In the experiments above, injection of siRNA into one-cell and two-cell embryos was shown to cause specific degradation of the target mRNA. However, it appeared that the longer the period of time was between injection of the embryos and their dissection and analysis, the weaker the observed RNAi effect was. A plausible reason for this effect might be the gradual degradation of siRNA due to their half-life (for more information, see invitrogen's supplement on STEALTH siRNA) as well as the titration of the initial amount of siRNA over developmental progress, as an increasing number of cells of an individual share an initial, absolute amount of siRNA molecules. This was reflected by the observation that only those injected embryos that were dissected and investigated at early points of development after injection showed a complete degradation of the targeted transcript (3.5.4.2), whereas embryos that had been allowed to develop further until dissection and WMISH in general tended to display milder decreases in endogenous gene transcripts. In order to circumvent this and achieve stronger RNAi effects in embryos of later stages, embryos that had passed gastrulation and reached stages S8, S16 and S22 were injected with *Ph six3* and *Ph kni1*-specific siRNA. After a minimum of 24h post-injection to allow for effective RNAi response, a fraction of the injected embryos was dissected and subjected to WMISH as described earlier (3.5.4). None of these embryos showed a decrease in *Ph six3* or *Ph kni1* expression (not shown), respectively. The remaining embryos were allowed to hatch. Typically, 25 – 30% of these showed various morphological phenotypes, ranging from defects in gut formation to deletions of whole segments, appendages or smaller cuticular regions and twisted body axes, abnormally bent and arranged appendages and allometric effects. These phenotypes were found with comparable percentages in embryos injected post-gastrulation with *Ph six3* and *Ph kni1*-specific siRNA as well as control embryos injected with water. I conclude that these effects are unspecific and represent artefacts. Accordingly, there is no indication that siRNA molecules are transported over cell boundaries in *Parhyale* embryos that have completed gastrulation. Therefore, I conclude that systemic RNAi is not triggered after injection of siRNA into multi-cellular, late-staged *Parhyale* embryos. Also, multicellular embryos appear more sensitive to injection than S1 and S2 embryos.

#### 3.5.4.5 Conception of a feeding assay in order to examine behavioural phenotypes after siRNA-mediated RNAi

*Ph six3* is expressed during the development of medial structures of the pre-antennal head and the protocerebrum and during the invagination of the foregut (3.3.1.7). *Ph kni1* expression parallels the establishment of the stomodeal opening and is found during the development of neural characters related to the stomatogastric nervous system (3.4.7). However, hatched animals that had been subjected to siRNA injections targeting either gene did not show specific morphological phenotypes in those regions of the head (3.5.4.2, 3.5.4.3). In order to address the question, whether the feeding behaviour of these animals was affected due to a reduction in *Ph six3* or *Ph kni1* function during development, one-cell embryos were injected with combinations of all three siRNA molecule species (100 µM each) targeting *Ph six3* or *Ph kni1*, respectively and allowed to develop and hatch. Single animals were then fed with small pieces of pigmented food three days after hatching (Tetra Rubin, Tetra GmbH, Melle, Germany) and compared to wild-type and control animals. Before, animals of interest had been deprived of food, but checked for vitality.

Wild-type animals (n=10) recognized the offered food within seconds, grabbed it, probed it with their antennae and started ingesting immediately. Usually, food pigments were visible in the gut 1 min after the animal had started feeding. The following day, all of them had produced excrements.

Of those animals that had been injected with siRNA targeting *Ph six3* (n=4), 2 ignored the food at first. After repeated offers, both eventually started to probe the food, grab and ingest it. The remaining animals reacted like wild-type. Several minutes after ingestion, food pigments were visible in the guts of all of them. The following day, all of them had produced excrements. However, compared to wild-type animals, their movements and reactions were slowed.

Of those animals that had been injected with siRNA targeting *Ph kni1* (n=4), one reacted to food only after repeated offers, but eventually started to feed like wild-type animals. Similar to above, its actions and movements were visibly slowed. Another animal persistently ignored the offered food. A third animal showed interest in the offered food, but did not ingest it. Similarly, the fourth animal recognised the food offer, probed the food and started processing it with its mouthparts, but was unable to ingest it. No food pigments appeared in its gut. Over the next days, the animal did not produce excrements.

Embryos that had been injected with both species of siRNA, therefore targeting both *Ph six3* and *Ph kni1*, showed a much lower rate of survival and hatching. Still, 6 of those were examined as described above. 2 behaved like wild-type animals. 3 animals showed weak and slow reaction to offered food and did not ingest it. The sixth animal ignored the food completely.

As a control, embryos that had been injected with water and showed varying degrees of unspecific morphologic phenotypes were tested (n=3). One showed wild-type morphology and reacted to offered food in a wild-type fashion. The second showed unspecific defects in the gut. Still, this animal identified, probed, grabbed and ingested the offered food like wild-type animals. Shortly after, food pigments were visible in the gut. However, due to malformations of the midgut, it did not produce excrements. The third one exhibited a phenotype of premature hatching (R). It reacted to food like wild-type animals. After several seconds, however, its behaviour became uncoordinated, and it eventually stopped feeding. Still, food pigments were visible in its gut after several minutes and it produced excrements. As an additional control, non-injected animals that had been incubated during embryonic development like injected embryos (n=3; 5.4.6; 5.5.3) were subjected to the same procedure. All of those reacted to the food offer with wild-type behaviour.

In summary, the majority of the animals that had been injected with siRNA targeting *Ph six3* and/or *Ph kni1* were slowed in their behaviour (57%, n=14) as compared to control animals, which all showed wild-type reaction time and immediate sensory and muscular coordination. Apart from being slowed, animals subjected to *Ph six3* siRNA had successfully fed and produced excrements. In contrast, animals subjected to *Ph kni1* siRNA showed more specific phenotypes related to feeding behaviour, namely the incapability of ingestion or complete ignorance of the presence of the food. Half of the animals that had been injected with siRNA targeting both genes exhibited combinations of both slowness and impediment of ingestion.

From these findings, I conclude that the feeding assay presented here represents an easy and cheap method to address neuronal phenotypes after siRNA-mediated RNAi. Even though no morphologic phenotypes were observed for any of the animals injected with siRNA, many of them exhibited behavioural defects that might well correlate with neural organs which express *Ph six3* (protocerebrum) or *Ph kni1* (stomatogastric nervous system) during development. In order to confirm the preliminary findings of this work, a more comprehensive approach with a larger number of animals will be necessary.

# 4 Discussion

## 4.1 Molecular aspects of candidate gene isolation in *Parhyale hawaiensis* and its implications for comparative developmental analyses

### 4.1.1 In need of an isogenic *Parhyale hawaiensis* wild-type population

*Parhyale hawaiensis* was isolated as a pest species from the John G. Shedd Aquarium, Chicago, IL (Browne et al., 2005). Today's laboratory populations were typically established from several hundred wild-type adults. Currently, BAC-based fractions of the *Parhyale* genome are sequenced and assembled (Parchem et al., 2010). To this end, an isogenic subpopulation of *Parhyale* had been established by continuous inbreeding of siblings over several generations (Ron Parchem, personal communication; Parchem et al., 2010). However, because the process of BAC sequencing is not completed, no genomic sequence data is currently available for *Parhyale*. For this reason, I cloned genes of interest via PCR using degenerate primers (5.2.1). The sequence information gathered from the genes analysed here show a high occurrence of polymorphisms, most of which are found in the non-coding regions of the cDNA. The majority of those polymorphisms found within the ORF of the genes of interest do not result in amino acid exchanges. Therefore, I conclude that they represent allelic variations of the same genes as they occur naturally within a non-isogenic wild type population. A higher number of nucleotide exchanges was found only sporadically, i.e. in single sequence clones of a given gene.

An example is the 5' RACE clone Ph_kni1_5R31, which shows more than 4% nucleotide differences to the consensus *Ph kni1* cDNA sequence (*Ph kni1_ref*; 3.4.3), while all other 5' and 3' RACE as well as ORF and cDNA sequences recovered for *Ph kni1* show less than 1% differences (A3, table A1). For that reason, this sequence is more likely to derive from imprecise transcription events that happened during reverse transcription and, to a lesser extent, PCR reactions - therefore resembling an artefact - than to represent a rare existing allelic variation. Reverse transcriptase is known to be error-prone (Boyer et al., 1996).

As a result of high SNP frequency, gene transcripts within a single animal might not be targeted by specific siRNA. Therefore, RNAi and transcript degradation in this animal could be impaired. For these reasons, it is of great importance to provide at least one isogenic *Parhyale* population.

### 4.1.2 Splicing diversity and transcript isoforms of *Parhyale* candidate genes for head development

Another striking observation was that the transcripts of several of the analysed transcription factors exist in isoforms. This phenomenon was most obvious for *Ph otd2* (3.2.1.4; Browne et al., 2006) and *Ph awh* (3.2.4.2, A2.7), but smaller variations were also observed for *Ph otd1* (3.2.1.3, A2.1). In many species, similar phenomena have been shown to result from alternative splicing of primary gene transcripts (Graveley et al., 2011; McManus and Graveley, 2011; Pan et al., 2008; Ramani et al., 2011). For some factors involved in neuronal development, alternative splicing has been linked to distinct functional adjustments of the same original transcription factor to different developmental contexts (e.g., Sunmonu et al., 2011). The isoform variations found here might also represent different alleles of an individual gene. However, based on the expression and putative roles of *Ph otd1*, *Ph otd2* and *Ph awh* in head and brain development, my data favours the model of alternative splicing to suit differential functions.

### 4.1.3 Consequences of varying transcription factor family repertoires

Comprehensive phylogenetic analyses of homeodomain transcription factor families in cnidarian and bilaterian species have allowed to devise models that describe how the number of paralogs of a specific gene family found in one species may have expanded from a basic repertoire of this family in a common ancestor (e.g., Ryan et al., 2006). Although mechanisms of genome multiplication and locus duplication receive increasing evidence and enable to explain an existing repertoire in an individual species, based on its phylogenetic position (e.g., Otto and Yong, 2002), reliable genomic sequence data is required in order to confirm such a prediction eventually. For *Parhyale hawaiensis*, differences in the number of paralogs of a

given candidate gene have been found even in comparison to closely related arthropod species, more precisely *Drosophila melanogaster* and *Tribolium castaneum*, but genomic data is not available in order to clearly verify the exact number of paralogs. Here, I want to discuss general findings of paralog proliferation and its specific implications for comparative analyses including *Parhyale*.

In many cases where additional paralogs of a gene family have appeared in a species, sub-functionalisation can be observed (e.g., Force et al., 1999). This means that proposed conserved functional and corresponding expression profile properties of a single, ancestral "founder" gene is distributed asymmetrically to the paralog derivatives in the examined species. Impressive examples include the *Hox3/zerknüllt (zen)* genes in insects (for an overview, see Hughes and Kaufman, 2002). Here, it is an item of current discussion that from a single ancestral *zen*-gene in a basal insect species, locus duplication led to the dipteran gene pair *bicoid* (*bcd*) and *zen*, both of which exhibit only a fraction of the original expression pattern which encompassed dorsal and anterior expression during blastoderm stage: *Drosophila zen* is expressed dorsally restricted in the extra-embryonic anlagen, whereas *Drosophila bcd* is exclusively expressed in the anterior of the embryo (Casillas et al., 2006; McGregor, 2005; Rushlow et al., 1987; Stauber et al., 1999). According to the current model of *bcd* evolution, *bcd* was then able to acquire de novo anterior determinant function in dipterans (putatively replacing a more ancestral role of *otd*), while the role of *zen* remained crucial for extra-embryonic-tissue development (Casillas et al., 2006; McGregor, 2005). In *Tribolium*, no *bcd* homolog exists (Bucher et al., 2005; Schröder, 2003); instead, *Tc'zen1* is expressed both anterior and dorsal during early development. Although it is potentially active during many developmental processes, it is necessary only for patterning extra-embryonic tissue, more precisely for the specification of the serosa (van der Zee et al., 2005). This finding is closely correlated with the differences in embryogenesis modes: intermediate development (*Tribolium*) vs. long-germ-development (Drosophila; for an overview over both modes, see Schröder et al., 2008). Interestingly, additional *zen* paralogs exists in both *Drosophila* (*zerknüllt-related*/*zen2*) and *Tribolium* (*Tc'zen2*). *Tc'zen2* acts later during development than *Tc'zen1*. It initiates the fusion of the extra-embryonic membranes, i.e. the serosa and the amnion (van der Zee et al., 2005). *Drosophila zen2* is expressed in a similar fashion as *Drosophila zen*. However, its function is still unknown (Casillas et al., 2006). In higher dipterans, the *zen* duplication leading to *zen2* has apparently occurred later with regard to evolutionary radiation than the event leading to *bcd* (Negre et al., 2005). Therefore, the

events leading to the *Tc'zen1*/*Tc'zen2* duplication in *Tribolium* and to the *zen/zen2* duplication *Drosophila* are, in all likelihood, independent.

For the candidate gene approach performed in this work, this means that in case a recovered *Parhyale* homolog displays derived expression patterns and putative functions as compared to other pancrustaceans, it cannot be excluded that additional paralogs of this gene might exist and that those still exhibit the conserved pancrustacean mode. Although differing numbers of paralogs between *Parhyale* and individual insect species have been found for several candidates, such as *aristaless* and *orthodenticle*, this problem is most apparent for *Ph kni1* and *Ph kni2*. Neither of these exhibit early regional head expression, as insect homologs do. To address the existence of an additional *Parhyale kni* paralog that might still confer a conserved function in head segmentation, several rounds of depletive varied-stringency PCR (MM) were performed. From these, no additional *kni* gene was recovered. Further, the third most common sequences to be recovered were in all cases entirely unrelated to *kni* genes and likely the result of unspecific binding of the degenerate primers (not shown). Still, additional and weakly expressed *kni* family members might be present in *Parhyale* and not be represented in the transcriptome at a quantity that allows isolation via PCR-based techniques. However, the existence of an early and regionally expressed *Ph kni* paralog is unlikely, as its transcript number would be expected to supersede that of *Ph kni2* which is expressed only during late periods of development and in a small number of cells, effectively reflecting a subset of *Ph kni1*-positive cells. In addition, phylogenetic analyses of *kni* genes (3.4.6) suggest that individual *kni* repertoires vary strongly, even within insects, and that dipteran *kni* genes (as opposed to *knrl* and *eg*) represent a novel feature of this lineage that can, e.g., not be found in the coleopteran *Tribolium* (see also Cerny et al., 2008; Rothe et al., 1992; Rothe et al., 1994).

Conclusively, the transcriptome-based approaches used for the isolation of *Parhyale* homologs of candidate head developmental genes in this work are not sufficient to unambiguously define and prove the exact number of paralogs of a given gene family, such as the *kni* family, although several attempts have been undertaken to minimize the chance to miss additional paralogs. Eventual clarification can only be achieved by detailed analysis of complete genomic sequence data. However, this is currently not available for *Parhyale*.

## 4.2 Evaluation of functional approaches for developmental studies in *Parhyale hawaiensis*

The successful establishment of molecular, genetic and histological techniques, such as WMISH (MM, Rehm et al., 2009ca) and transgenesis (MM, Pavlopoulos and Averof, 2005) as well as the provision of a comprehensive and detailed staging system (Introduction, Browne et al., 2005) has given the amphipod *Parhyale hawaiensis* a promising perspective as a newly emerging convenient, highly practical and unique crustacean model for comparative developmental studies. A reliable set of techniques to facilitate robust functional analysis had not been established at the beginning of this work. However, such a set is of course crucial for rendering any model system useful for biological studies in the long term. For this reason, I pursued various RNAi-related approaches to contribute to addressing functional studies in *Parhyale* in this work.

### 4.2.1 Sensitivity of *Parhyale* to injection-based techniques

Addressing *Parhyale* gene function via techniques that involve injection of embryos has revealed high sensitivity of *Parhyale* to physical manipulation and mechanical stress (3.5.1; 3.5.2). Injection resulted in a considerable range of unspecific developmental defects and morphological phenotypes that occurred with comparable frequency after control injections. These artefacts are likely to mask specific effects (see also 3.5.4.1), therefore impeding functional analyses. In order to relieve those animals that were examined for specific phenotypes from the stress of physical manipulation, two alternatives to embryonic injection were tested here: (i) parental injections and (ii) a *Minos*-transposon-based approach for stable transgenesis with inducible ectopic expression constructs (4.2.3, see below). In *Tribolium*, injection of dsRNA into female pupae and female adults has been highly effective (Bucher et al., 2002). This procedure triggers RNAi within the ovaries, providing batches of progeny embryos that display inheritance of the effects of RNAi. Depending on the time between injection and egg deposition, these embryos typically exhibit loss-of-function effects within a phenotypic range. Strong phenotypes were found for embryos of the first egg lay and got increasingly milder within subsequent batches. In *Parhyale*, female adults were treated analogously. They were punctured laterally into the soft tissue that connects the tergites of

segments T4 and T5. dsRNA was released close to their easily visible ovaries. Although the initial survival of injected females was high, all of them gradually exhibited degradation of close-by tissues, e.g. ovaries and, to a lesser extent, muscles as well as infection and necrosis of the area of the puncture. Embryos that developed from these females did not show RNAi effects. In contrast, female *Tribolium* pupae and adults reveal robust long-term vitality and retain functional ovaries after injection, although they suffer from comparable mechanical stress (e.g., Tomoyasu et al., 2008). *Tribolium* exhibits high sensitivity to systemic RNAi (Bucher et al., 2002; Tomoyasu and Denell, 2004). Accordingly, female sterility or lethality after injection of dsRNA is believed to specifically result from RNAi in pupal and adult tissues (Gregor Bucher, personal communication; Bucher et al., 2002). In *Parhyale*, this technique is also impeded by the fact that injection of dsRNA has so far not been found to trigger degradation of endogenous transcripts via RNAi (3.5.1).

In addition, injected *Parhyale* females were frequently attacked by male mating partners. A general observation is that weakened, infected or physically wounded individuals are attacked and killed by other *Parhyale*. On several occasions, this was also observed as a reaction of wild-type animals to transgenic ones, usually injected animals that exhibited expression of the transformation marker. Behavioural traits that provide social barriers to pathogen transmission are diverse and well-described throughout animal species (Loehle, 1995). However, this phenomenon has not been studied in *Parhyale* populations so far.

### 4.2.2 Feasibility of siRNA-mediated RNAi in *Parhyale*

Initial attempts to cause RNAi in *Parhyale* embryos via injection of dsRNA were unsuccessful. Similar problems have been observed for other species as well (Tomoyasu et al., 2008). Efficiency and repertoire of cellular components required for causing RNAi by the presence of dsRNA molecules are considered the prerequisite for successful application of this technique. In fact, they may vary considerably among species (Tomoyasu et al., 2008). In detail, Dicer activity is required for the cleavage of long dsRNA molecules into small interfering RNA molecules (siRNA). These are used by the RISC complex to target and degrade endogenous RNA molecules that show sequence homology to the siRNA (for an overview, see Hammond, 2005). Lack of functional components of this cellular machinery may be the cause for the impediment of dsRNA-mediated RNAi in *Parhyale hawaiensis*.

In mammalian cells, ectopic dsRNA triggers the antiviral response. As a consequence, the physiology of these cells is drastically changed, leading to a wide range of phenotypes unrelated to the original dsRNA-mediated RNAi target and, therefore, potentially masking specific phenotypes. This characteristic trait impedes dsRNA-mediated RNAi approaches in mammalian cells (Gantier and Williams, 2007; Karpala et al., 2005). Accordingly, a similar response might be triggered in *Parhyale* as well, providing an alternative explanation for the impediment of specific RNAi via injection of dsRNA as well as the high level of sensitivity that was observed for *Parhyale* embryos after injection of dsRNA.

A feasible way to overcome the lack of susceptibility to dsRNA-triggered RNAi in a given species is the direct use of siRNA molecules targeting the transcripts of the genes of interest and therefore virtually shortcutting the cellular processing of dsRNA (e.g., Tuschl, 2001). In *Parhyale*, application of siRNA successfully reduced the transcript levels of the targeted genes (3.5.4). However, although a wide range of transcript reduction strength as well as correlating morphological changes of head and brain tissues, specifically, were observed in embryos subjected to WMISH (3.5.4.2), none of the injected animals exhibited visible cuticular defects and alterations in head morphology after hatching. At the same time, the majority of siRNA-injected embryos died before hatching (3.5.4.2; 3.5.4.3).

In contrast, siRNA-mediated loss-of-function of *Parhyale Ubx* (*Ph Ubx*) and *Parhyale distalless-early* (Ph dll-e; Liubicich et al., 2009) as well as *Ph kni1* (this work) resulted in visible appendage phenotypes. These could be easily assessed from examining cuticles of newly hatched adult *Parhyale*.

In *Parhyale*, the level of *Ph Ubx* expression in anterior thoracic segments corresponds to the morphology of the respective segmental appendages: segments posterior to T3 show the highest levels of *Ph Ubx* expression and develop pereopods, while the T3 and T2 segments exhibit lower *Ph Ubx* expression levels and develop gnathopods (Liubicich et al., 2009). During *Parhyale* development, *Ph Ubx* expression remains absent in the appendages of the T1 segment. These appendages become maxillipeds, acquiring a specialised feeding morphology. In many crustacean lineages, the anterior boundary of *Ubx* expression correlates to the number of maxilliped-bearing anterior thoracic segments (Averof and Patel, 1997). Reduction of *Ph Ubx* transcript levels via siRNA-mediated RNAi has been shown to lead to homeotic transformations of thoracic appendages toward more anterior fates (Liubicich et al., 2009). Conversely, ectopic expression of *Ph Ubx* is sufficient to cause homeotic transformations of anterior appendages toward more posterior fates, including maxilliped-to-leg transformations

(Pavlopoulos et al., 2009). Importantly, maxillipeds are typically specified in the absence of *Ph Ubx*, but can also develop under low or transient *Ph Ubx* expression (Pavlopoulos et al., 2009). These findings strongly suggest that *Ph Ubx* determines appendage morphogenesis in a dose-dependent manner. Accordingly, even gradually lowering the transcript levels of *Ph Ubx* via siRNA-mediated RNAi is sufficient to cause homeotic transformations without the need of completely disrupting *Ph Ubx* function in the respective segments and appendages.

*Ph dll-e* is a *Parhyale hawaiensis distalless* paralog that is expressed early during development in the distal parts of the developing appendages, highly similar to the situation in insects (e.g., Panganiban and Rubenstein, 2002). Likewise, *Ph dll-e* loss-of-function via siRNA-mediated RNAi leads to truncated appendages in *Parhyale* hatchlings (Liubicich et al., 2009). This goes well with identified *distalless* functions in various other arthropods (Beermann et al., 2001; Cohen et al., 1989).

Injection of siRNA targeting *Ph kni1* resulted in the reduction or the complete loss of *Ph kni* WMISH signal in cells located at joint positions between appendage articles (3.5.4.3). In accordance, it caused malformations and misarrangement of appendages in hatched animals.

From these findings, I conclude that siRNA-mediated loss-of-function of developmental genes does have the capacity to cause visible morphological and cuticular phenotypes in hatched *Parhyale hawaiensis*. Strikingly, such phenotypes were only observed in segmental appendages, but not in the head or head-related structures. This could be for several reasons. In the following, I want to describe three models that summarise the possible reasons: First, *Ph kni1* and *Ph six3* functions in head development do not affect structures that correlate to cuticular areas in an adult animal. Specifically, although *Ph six3* expression was reduced or absent in the elongating labrum of S20-S23 embryos that had been injected with *Ph six3*-specific siRNA, the labrum was present in all of those embryos, albeit projected at a different angle compared to wild-type (3.5.4.2). In contrast, protocerebral characters, more precisely the hemi-ellipsoid bodies, were drastically reduced in size in accordance with a strong decrease of *Ph six3* expression after siRNA-mediated RNAi. This suggests that *Ph six3* function may not convey significantly to establishing and maintaining head regions that correspond to cuticular structures, such as the labrum, and, further, that a size reduction of brain characters, caused by *Ph six3* loss-of-function, might not correlate to visible cuticular defects.

Second, disturbing appendage development by loss-of-function experiments may not jeopardize the viability of the embryo itself, thus more frequently leading to observable phenotypes of appendage morphology in hatched animals. In contrast, severe regional, segmental or neuronal phenotypes might in fact manifest in developing Parhyale embryos as a result of siRNA-mediated loss-of-function of developmental genes; however, since they affect more complex and interactive developmental processes, they may rather cause premature death of *Parhyale* embryos, i.e. before they reach S30 and hatch.

Third, in cases where different genetic and developmental responses to varying dosis levels of a specific genetic regulator exist in *Parhyale*, siRNA-mediated reduction of the activity of such a gene might simply trigger an alternative "wild-type-like", although ectopic, developmental process in the affected area of the embryo. This model may well apply to the examples of homeotic transformations caused by Ph Ubx loss-of-function (as well as gain-of-function) and to the malformed and misarranged appendages found in embryos after siRNA-mediated *Ph kni1* transcript reduction. Here, affected proximal appendage article joints might mimic the morphogenetic situation found in more distal article joints that express lower levels of *Ph kni1* (Figure 78).

Apart from the models described above, the fact that injection of siRNA caused death in large fractions of *Parhyale* embryos before hatching suggests that *Parhyale* is in principle more sensitive to mechanical stress by injection and to RNAi-mediated manipulation of gene function during development than other species, such as *Tribolium* or *Drosophila*. The fact that *Parhyale* develops directly and, accordingly, does not develop an intermediate, cuticularized larva that could be used for the analysis of morphological phenotypes, adds to this notion.

In addition to the problems described above, the half life of siRNA molecules and the fact that systemic RNAi did not occur after injection of post-gastrulation embryos (3.5.4.4) both hamper the analysis of gene function during late embryogenesis. A plausible, however - to date - speculative explanation for the deficiency of systemic RNAi might be that molecular core components for its transmission, e.g. by channelling siRNA molecules across cellular borders, are not functional in *Parhyale* cells. This has also been suggested for the difference in RNAi response between *Tribolium* (systemic) and *Drosophila* (restricted). However, although *Tribolium* displays a high efficacy in systemic RNAi response and several putative dsRNA transporters have been identified, the underlying mechanisms of this process have still not been understood (Tomoyasu et al., 2008).

### 4.2.3 Overcoming RNAi- and injection-related disadvantages of *Parhyale hawaiensis*: a technical perspective

Compared to *Tribolium*, other arthropod species such as *Drosophila* show a significantly lower to no measurable sensitivity to systemic RNAi, rendering injection of dsRNA into embryos or adults inefficient (Tomoyasu et al., 2008). In *Drosophila* and *Anopheles,* alternative transgene-based approaches that allow the induced ectopic expression of RNA hairpins have been successful in causing RNAi (Catteruccia and Levashina, 2009; Kennerdell and Carthew, 2000). For this reason, an analogous strategy has been approached in *Parhyale* in this work. However, the establishment of populations of transgenic *Parhyale* carrying inducible RNA hairpin expression constructs has not been successful (3.5.2). Induction of transient mosaic expression of *Ph otd1* RNA hairpins in injected embryos via heat shocks has been accomplished, but no specific degradation of endogenous transcripts could be observed. Importantly, it was not possible to assess correct formation of RNA hairpin loops by WMISH. In *Anopheles gambiae*, use of endogenous introns as spacers has been discussed to enhance RNA hairpin formation (Brown et al., 2003). Here, a *Ph kni2* sequence fragment has been characterised to be an endogenous intron that is effectively spliced *in vivo* (3.5.3). This might provide an effective tool to increase the feasibility of transgene-based RNA hairpin strategies in future approaches.

## 4.3 Anterior-posterior axial polarity: the amphipod way

In bilaterian species, the establishment of anterior-posterior, of dorso-ventral and, as a result, of left-right polarity constitutes the prerequisite for subsequent axes formation as well as regional and segmental patterning processes. Depending on the species, the underlying molecular processes take place during the initial phases of embryogenesis or even before, specifically during oogenesis, early cleavage or gastrulation (e.g., seeGilbert, 2006; Wolpert et al., 2002). To elucidate conservation and plasticity of anterior and head development in bilaterians and, in particular, among pancrustaceans, I want to begin with discussing the molecular aspects of anterior-posterior axis formation in *Parhyale*. Here, I will compare the

data that already exists from *Parhyale hawaiensis* together with my findings to the respective processes in the insects *Drosophila melanogaster* and *Tribolium castaneum*.

In the *Drosophila* embryo, anterior-posterior polarity is created by maternal contributions to the oozyte. These provide axial information by acting as pole determinants. On the molecular and genetic level, these determinants resemble localized mRNA encoding transcriptional and translational regulators. Both anterior (*bicoid, bcd*) and posterior (*nanos, nos*) determinants have been identified and studied intensively (Driever and Nüsslein-Volhard, 1988; Irish et al., 1989; Klingler, 1990; Porcher and Dostatni, 2010). Through gradients of these high-level hierarchy factors that form within the syncytial blastoderm during early embryogenesis, relative anterior and posterior positional information is established in *Drosophila*. This information is used for subsequent regionalisation and segmentation of the germ band along the anterior-posterior axis (e.g., reviewed in Jaeger, 2011). Interestingly, *bcd,* which acts as a morphogen, appears to be a novel feature limited to higher dipterans (McGregor, 2005). Based on its biochemical properties, it has been reasoned that *bcd* may have substituted *otd* for its function as an early anterior determinant (Lynch and Desplan, 2003; Treisman et al., 1989).

Like *Drosophila*, *Tribolium* embryos initially undergo superficial cleavage resulting in a syncytial blastoderm. In contrast to *Drosophila*, the germ rudiment emerges at the posterior, ventral side of the egg. Initially, it covers only the future head (procephalic and gnathocephalic) as well as the prospective thoracic segments. More posterior segments are established and specified during subsequent development. The complementary anterior and dorsal parts of the egg give rise to extra-embryonic tissue (reviewed in Schröder et al., 2008). Notably, in both *Drosophila* and *Tribolium*, *zerknüllt* genes, which represent *Hox3* homologs, determine future extra-embryonic tissue already at an early phase of development (Doyle et al., 1986; Rushlow et al., 1987; van der Zee et al., 2005). In *Tribolium*, a posterior-anterior protein gradient of *CAUDAL* is established during blastoderm. *Tribolium caudal* is required for the establishment of all segments posterior to the procephalon (Copf et al., 2004), whereas *zerknüllt* and *mex-3* repress *caudal* function anterior and are discussed as candidates for an anterior determination system as in *Drosophila* (Schoppmeier et al., 2009). However, there is ongoing debate whether *orthodenticle* genes might fill the role of an ancestral anterior determinant (Kotkamp et al., 2010; Lynch and Desplan, 2003; Schröder, 2003).

In *Parhyale*, early cleavage in the embryo is in stark contrast to the insect modes mentioned above: three total and unequal cleavages lead to four macromeres and four

micromeres, each of which exhibit individual and specific morphologies. Importantly, at this early stage of embryogenesis, all bilaterian polarities (anterior-posterior, dorso-ventral and left-right) are realised already due to the specific determination and relative position of the eight blastomeres. Each of these gives rise to a cell progeny of distinct, invariant fate (Gerberding et al., 2002; 2.5.5.1). Interestingly, the Ep macromere, which gives rise to posterior-most ectodermal structures, is the only blastomere unambiguously determined to relevant anterior-posterior fate. This fact suggests the existence of a posterior determinant in *Parhyale* that is asymmetrically distributed during the initial embryonic cleavages. In contrast, all future anterior and head structures originate from progeny of the Er and El macromeres (Gerberding et al., 2002; 2.5.5.1). Currently, the *Parhyale* homolog of *nanos* is examined as a candidate determinant that provides posterior positional information by being asymmetrically distributed during the first embryonic cleavages (Matthias Gerberding, personal communication). As mentioned above, *otd* has been discussed as a possible ancestral anterior determinant. In *Parhyale*, neither of both *otd* paralogues is maternally expressed; accordingly, they are not present in the embryo during early cleavages, before zygotic expression is initiated. This indicates that neither *Ph otd1* nor *Ph otd2* have functions in the initial establishment of anterior-posterior polarity. It is important to note that the anterior macro- and micromeres in Parhyale embryos of 8-cell-stage (S4) give rise to the visceral mesoderm and the endoderm. This might suggest that the later anterior-posterior polarization of the embryo is preceded by a more basal, vegetal-animal distinction within the zygote, as is the case for the majority of vertebrate species (Gilbert, 2006; Wolpert et al., 2002). In addition, *zen* homologues appear to have retained their original *Hox3* homeotic function in crustaceans (Hughes and Kaufman, 2002; Papillon and Telford, 2007). In order to address the existence of early anterior determinants in *Parhyale*, a candidate gene approach might not be feasible. Instead, differential transcriptome and proteome screens of *Parhyale* blastomeres promise an unbiased, comprehensive approach to this issue. As mentioned above, respective studies are currently in progress (Matthias Gerberding, personal communication).

## 4.4 Early anterior regionalization in *Parhyale hawaiensis*

### 4.4.1 Involvement of *Ph otd1* in early anterior regionalisation

At the onset of gastrulation, descendants of all blastomeres undergo complex relative position changes and asymmetric divisions. Consequently, this leads to evenly distributed cells that are located on the surface of the egg and contain only a small cytoplasmic fraction, enclosing an appearingly homogenous and anuclear central yolk mass (Browne et al., 2005). As a result of gastrulation, the ectodermal part of the *Parhyale* germ forms at stage 8 from descendants of the three ectodermal blastomeres, resembling a circular, relatively dense expanse of clustered cells (Browne et al., 2005; Gerberding et al., 2002). Within the condensing germ disc, head regions and segments have to be defined. Accordingly, this process is accompanied by the initiation of zygotic gene expression (e.g., Browne et al., 2006).

At the onset of stage S8, zygotic *Ph pby1* expression arises in all cells of the newly established germ disc. Shortly after, expression of *Ph otd1* can be observed in descendants of Er and El at the anterior end of the germ disc. In this work, this was the earliest observed zygotic expression of a *Parhyale* gene within the anterior part of the embryo, hence first reflecting anterior polarity and patterning within the embryo's ectoderm. Because *otd* genes have been shown to play important roles in early anterior patterning in a great number of species (Acampora et al., 2000a; Cohen and Jürgens, 1990; Finkelstein and Boncinelli, 1994; Schinko et al., 2008; Simeone et al., 2002; Steinmetz et al., 2011), these observations raise the question if *Ph otd1* has the function of a zygotic top-hierarchy factor for regional anterior development and patterning in *Parhyale*.

Based on expression data of *Ph pby1*, *Ph otd1* and several additional early zygotically expressed genes, I want to propose two alternative models to help increase the understanding of the molecular processes underlying the developmental dynamics of early anterior patterning in *Parhyale*. I will then compare these models and discuss their implications for future studies.

The first model comprises two phases. Uniform expression of *Ph pby1* precedes the initiation of patterning processes in the germ disc. During the first phase of early anterior patterning, *Ph pby1* expression retracts from cells of the anterior-most part of the germ disc to the same extend that novel *Ph otd1* expression arises there. At this point of *Parhyale*

development, the germ rudiment consists of cells with anterior identity and of a large remaining population that represents a yet unspecified 'reservoir' of cells. For a short time, both expression domains abut each other directly. This moment reflects the earliest observed realisation of anterior-posterior axial polarity by zygotically expressed genes. Subsequently, new cells arise between the anterior-most *Ph otd1*-positive cells and the *Ph pby1* domain, indicating the onset of the second phase.

During the second phase, expression of *Ph pby1* retracts further from lateral parts of the germ disc. Eventually, *Ph pby1* expression is restricted to posterior cells of the germ disc. *Ph otd1* expression is found in an increasing number of cells at the anterior pole of the germ disc. At the same time, the central region of the germ disc that expresses neither of both genes expands. Expression of *Parhyale labial* (*Ph lab*) was shown to arise in cells that are located between the *Ph otd1* and *Ph pby1* expression domains. During the entire embryogenesis, *Ph lab* remains active. It specifies the future An2 segment (William E. Browne, personal communication). This suggests that during this phase, the identities of the procephalic segments, in particular the pre-antennal and/or ocular, the An1 and the An2 segments are established. No gene that is specifically expressed in the region of the An1 segment has been described so far. Based on studies in *Tribolium*, a *Parhyale* homologue of *empty spiracles* may be an interesting candidate. In *Tribolium*, a situation similar to *Ph otd1* and *Ph pby1* in this model has been described regarding early head development. Here, *Tc'otd1* and *Tribolium buttonhead* (*Tc'btd*) are expressed in mutually exclusive domains that directly abut each other. As embryonic development progresses, both expression domains become separated as *Tc'btd* expression moves posterior. *Tc'btd* expression becomes restricted to cells of the mandibular segment (Schinko et al., 2008).

According to the second, alternative model, the initial, apparently uniform expression of *Ph pby1* in the condensing germ disc does not reflect all of the ectodermal part of embryo; instead, it only represents the population of posterior cells already at the beginning, while the remaining, more anterior cells are still distributed over the surface of the egg at a low cellular density. Two separate early processes take place in the *Parhyale* embryo at the same time: the morphological constitution of the densely clustered, recognizable germ rudiment on the one hand, and the patterning of the procephalic head segments on the other. In this regard, expression of head segment-specific genes including *Ph otd1* (pre-antennal and ocular area), a yet unknown factor for the An1 segment (a putative candidate being an *ems* homolog, see above) and *Ph lab* determining the An2 segment arise synchronously in cell groups that are

still in the process of condensing to the germ disc. The expression of *Ph hbn* in bilateral, anterior-lateral clusters of cells with low density (4.4.2) may be an indication supporting this model. However, this model requires an underlying, yet unknown molecular mechanism of high-level hierarchy which provides anterior-posterior positional information and leads to the establishment of a polarised naupliar region within *Parhyale* embryos that have just exited gastrulation. Consequently, *Ph otd1* would rather act as a segmental head gene than a major upstream anterior determinant.

To corroborate one of these two models, additional information on post-gastrulation gene expression is crucial. Based on data of early blastodermal gene expression in insects, a specific set of candidate genes could be determined and their homologs analysed during these early stages of *Parhyale* embryogenesis. With regard to the findings of a comprehensive study on anterior head patterning in *Tribolium*, this set should include *Parhyale* homologs of *empty spiracles*/*emx*, *eyeless*/*pax6*, *retinal homeobox*, *sine oculis/six1*, *eyes absent*, *tailless*/*tlx*, *cubitus interruptus*/*gli3*, *lim1/5*, *goosecoid*, *sloppy paired*, *scarecrow/nkx2.1* and *mirror/irx* (Posnien, 2009). For higher resolution of *Ph otd1* and *Ph pby1* interaction and dynamics, a more comprehensive set of WMISH data, including minimum age intervals of used embryos should prove useful. Live imaging of *Parhyale* embryo during gastrulation and early germ disc patterning (from stages 5 through 9) would be extremely helpful in reconciling cell movements. This technique has been successfully applied to zebrafish embryos (e.g., Rieger et al., 2011) and has provided impressive insights in cellular details during development. At this point, my data - especially the findings on *Ph hbn* expression – favours the second model.

### 4.4.2 *Ph hbn* represents a conserved factor involved in anterior head and brain patterning

In *Drosophila*, *homeobrain* expression arises very early during syncytial blastoderm stage within the dorsal head region (Walldorf et al., 2000). Equally, expression of *Tribolium hbn* (*Tc'hbn*) is detected very early at the blastoderm stage in anterior bilateral domains (Sebastian Kittelmann, personal communication). *Ph hbn* expression follows the insect pattern as it also starts very early in development, during patterning of the condensing germ disc (stage 8). Along *Ph pby1* and *Ph otd1*, it is one of the earliest zygotically expressed genes examined in this work. Similar to the situation in insects, bilateral clusters of *Ph hbn*-positive

cells are initially found at lateral, anterior positions of the germ disc. Interestingly, the *Ph hbn* expression domains are found within the periphery of the germ disc that is characterised by a low density of cells as compared to the more densely clustered cells of the centre of the condensing germ disc. In subsequent stages, *Ph hbn*-positive cells have arranged into arched clusters that are still located on lateral-anterior positions, but now well within the newly emerging head lobes of the densely clustered germ rudiment. Under the premise that *Ph hbn* expression remains constant and unaltered within these cells during stages 8 through 12, this allows us to reconcile cellular movements that take place during germ disc condensation. It also substantiates the model, according to which cells are recruited from outside the germ rudiment and dynamically add to the embryo during early stages of *Parhyale* development (Browne et al., 2005; 4.4.1).

Due to its early zygotic expression in relatively large bilateral regions of the developing head, *Ph hbn* might be an early regionalising factor, specifying a head region anterior and lateral to that of *Ph otd1*. *Ph hbn*-expressing cells appear, however, to lie nested within the larger *Ph otd1* domain during *Parhyale* development. WMISH visualising both expression domains in single embryos (5.4.4) could help elucidate this issue. Further, robust functional analyses regarding *Ph hbn* are required for addressing *Ph hbn* function and the relation of *Ph hbn* and *Ph otd1*. In this work, however, neither of these approaches has been successful.

In *Drosophila*, *Tribolium* and *Parhyale*, *hbn* expression becomes more dynamic during later stages of embryogenesis (Walldorf et al., 2000; Sebastian Kittelmann, personal communication). With the exception of *Drosophila hbn*, which in addition becomes temporarily expressed in the ventral nerve chord (Walldorf et al., 2000), *hbn* expression remains restricted to the anterior head and brain in these species. Because ventral nerve chord expression has not been detected in other arthropods so far, this may be a derived, novel *hbn* function that has evolved in the lineage leading to *Drosophila*. In contrast, putatively conserved expression of *hbn* in distinct areas and neuromeres of the anterior brain has been reported for lophotrochozoans (Fröbius and Seaver, 2006) and echinoderms (Burke et al., 2006) as well.

To summarise these findings and observations, I conclude that *hbn* is a highly conserved factor within bilaterians and possibly even metazoans (Mazza et al., 2010) that acts as an ancestral player in anterior head and brain development and has probably been lost secondarily within chordates (Mazza et al., 2010).

## 4.5 *Ph six3* and *Ph gbx* do not specify procephalic head regions

In several arthropod species, expression of *six3* arises during early embryogenesis in a broad regional domain adjacent anterior to the domain of otd genes at roughly the same time (Seo et al., 1999; Steinmetz et al., 2010). In *Tribolium*, loss of *six3* function during blastoderm stage leads to the deletion of the entire anterior-median part of the larval cuticle, encompassing dorsal head cuticle, the labrum and the anterior-most part of the protocerebrum (Posnien, 2009). Similar functions of *six3* genes as early regional requisites for anterior head and brain patterning have also been suggested for various other species (Steinmetz et al., 2010). From these observations, it has been reasoned that *six3* might play the role of a major ancestral player in anterior patterning of bilaterians. Accordingly, it is currently discussed as a conserved member of an ancestral anterior segmentation mechanism. Similar, *gbx* genes have been found to be expressed in a broad regional domain directly posterior to the *otx*-positive region during early *Platynereis dumerilii* development (Steinmetz et al., 2011). This leads to the interesting question whether the genes *six3*, *otd* and *gbx* indeed represent basic elements of a conserved anterior regionalisation and segmentation mechanism that might have evolved only once in the bilaterian ancestor.

In *Parhyale*, *otd1* is expressed in a fashion that fits the model mentioned above well, i.e. in a broad anterior domain during early embryogenesis. The domain initially covers the entire, newly emerged head lobes (stage 9) and becomes subsequently restricted to the pre-antennal area of the *Parhyale* embryo excluding the anterior-most cells, encompassing also the anterior first antennal segment and abutting the stripe-shaped engrailed domain of An1 (Browne et al., 2006).

Based on their expression, however, *Ph six3* and *Ph gbx* cannot be considered as early regionalisation factors like *Ph otd1*. Although expressed early at the onset of trunk segmentation, *Ph six3* remains restricted to single cells and later very small groups of cells that lie medially within the *Ph otd1* domain. Only later during embryogenesis, as soon as the former medial cleft between the two head lobes has been replaced with ectodermal tissue (from stage 15), does the *Parhyale* embryo express *Ph six3* in a large, median domain that eventually reaches more anterior than the *Ph otd1* domain. *Ph gbx* is expressed much later during *Parhyale* development and is found exclusively in small, distinct groups of cells.

Based on these findings, *Parhyale* might resemble a derived state regarding the expression of conserved early regionalisation factors in pancrustacean or bilaterian species. The temporal shift of establishing the anterior median area later than the head lobes might have resulted in an apparently delayed expression of the chief anterior *six3* domain in *Parhyale*. Therefore, I propose the model that *Ph six3* expression represents a derived state typical for the lineage leading to *Parhyale*. Further evidence from other malacostracans is needed to corroborate or invalidate this. Importantly, though restricted, *Ph six3* expression starts shortly after *Ph otd1* expression during *Parhyale* development. *Ph gbx*, in contrast, shows no early expression. This fact argues against *gbx* as a conserved early regionalisation gene like *otd* and *six3*.

## 4.6 Development of the anterior median area and the stomodeal field in *Paryhale hawaiensis*

### 4.6.1 Two populations of cells comprise the anterior median region in *Parhyale hawaiensis*

In *Parhyale*, morphologically distinct head structures become first visible in the shape of bilateral head lobes. They emerge early during embryogenesis at stage 9 and cover the areas of the future first and second antennal segments as well as the pre-antennal part of the embryonic head including the ocular segment. Until stage 14, the expanding head lobes remain separated by a medial cleft. At the onset of stage 14, mitoses can be observed within the medial cleft at the level of the second antennal segment. In slightly older embryos, the medial cleft at this position has disappeared, as *Ph kni1*-positive cells have emerged. At the same time during embryogenesis, a small population of cells has appeared within the medial cleft at the anterior end of the head lobes. Before the fusion of both head lobes has been completed at stage 15, a small circular area between both head lobes at the level of the first antennal segment remains of low cell density. From these observations, I propose the following model to reconcile the establishment of anterior median tissue: At the onset of stage 14, two developmental fields begin to fill the medial cleft with embryonic cells. At the

anterior end of the embryo, both head lobes become connected by a population of cells that – from its relative position – is likely to promote the establishment and development of the labral appendage during subsequent stages. Posterior at the level of the second antennal segment, an alternative developmental organizer accounts for the establishment of medial tissue. In contrast to the anterior developmental field, these cells are characterized by expression of *Ph kni1*. Both cell populations expand, until at stage 15, the former separation of the embryonic head lobes is no longer visible. The nature of these developmental fields remains unclear, since genes that are expressed specifically in the cell population located at the anterior end have not been characterised in more detail so far, and, in addition, siRNA mediated RNAi targeted at *Ph kni1* did not reveal specific phenotypes in the posterior median area of the embryonic head.

### 4.6.2 Reconciling stomodeal development based on *Ph six3*, *Ph six4* and *Ph kni1* expression

In *Parhyale* embryos, the median anterior tissue forms only after the initial phase of head patterning and the establishment of the head lobes has been completed (4.6.1). It gives rise to at least two different head structures: anterior, to the labrum and dorsal median epidermis, and posterior, to elements derived from the stomodeal development field (SDF). Here, I want to combine expression data from *Ph six3*, *Ph six4* and *Ph kni1* to describe a model that helps reconcile the steps from the initial definition of the SDF until the invagination of the foregut, its elongation and projection as well as the establishment of neuronal structures associated to the stomatogastric nervous system (SNS).

As early as stage 12, scattered *Ph six4* expression marks those cells of the presumptive mandibular, second and first antennal and pre-antennal head segments that directly border the prospective median anterior region, at this stage still preceded by a morphological cleft separating both head lobes. At stage 15, the anterior median region of the embryonic head is fully established. It is sharply abutted by bilateral *Ph six4* expression domains, giving it a triangular shape. Subsequently, *Ph six3* expression arises in cells of the anterior median region, resulting in mutually exclusive *Ph six3* and *Ph six4* expression domains at the level of the An1 and An2 segments. Importantly, this exclusive expression pattern does not apply for the anterior-most portion of the pre-antennal head, where cells co-express both genes. At

stage 18, *Ph kni1* expression marks lateral groups of cells that will subsequently form the morphologically distinguishable lateral oesophageal projections. According to the model described here, this process marks the initiation of the patterning and definition of the SDF itself. Under the premise that the circularly located *Ph kni1*-positive groups of cells resemble neurogenic populations that will give rise to neurons of the stomatogastric nervous system and that will eventually surround the stomodeal opening, it is possible to delineate the periphery of the SDF itself at this point of *Parhyale* embryogenesis. *Ph kni1* expression in a small cluster of cells that is located directly in the centre of the putative SDF may indicate and precede the morphological process of subsequent foregut invagination. At this stage, *Ph six3* expression covers the anterior half of the SDF and reaches to the anterior end of the embryonic head. *Ph six4* expression still overlaps with the *Ph six3* domain in the anterior-most median part of the pre-antennal head, but is directly adjacent and mutually exclusive to it on the level of the SDF. During stage 19, the invagination of the foregut can be followed from morphological observations (Figure 6; 3.1.3). *Ph kni1* remains expressed in the lateral oesophageal projections and in the putative neurogenic cell populations of the prospective SNS. At the onset of stage 20, the invagination of the foregut is evident and the stomodeum is visible as a gradually contracting ring. The *Ph kni1*-positive oesophageal projections now appear as distal outgrowths that are embedded on lateral positions within the foregut. Until the end of stage 20, the labrum has elongated to a point that its bi-lobed distal tip covers the stomodeum entirely. The posterior rim of the stomodeum is still visible as an elevated, horizontal ridge of cells. The process of foregut invagination and stomodeal contraction is also reflected by the relative positions of the *Ph six3* and *Ph six4* domains: at the end of stage 20, the bilateral *Ph six4* domains are no longer separated by *Ph six3*-positive cells, but appear fused in the anterior median region, while *Ph six3*-positive median tissue has submerged into the head. From this population of cells, the hemi-ellipsoid bodies form as bilateral median protocerebral derivates; posterior, *Ph six3*-positive cells now constitute the roof of the foregut that projects as an arched tube into the emerging head capsule. *Ph kni1* expression continues in tissue derived from the lateral oesophageal projections and in the emerging stomatogastric neuromeres.

To verify the identity of *Ph kni1* expressing cells of the SDF as factually neurogenic, suitable neuronal and neuroblast markers have to be found for *Parhyale* and their activity analysed and traced in detail. Preliminary *Ph kni1* loss-of-function data did not reveal morphologically visible defects in stomodeum formation and SNS neuromere establishment.

The fact that hatched animals reveal impediment in food ingestion after siRNA-mediated *Ph kni1* knock-down does imply a crucial role of *Ph kni1* in SNS development, however. Alternatively, *Ph kni1* expression might mark mesodermal cells of this head area. The fact that *Parhyale myocyte enhancing factor 2* (*Ph mef2*), a downstream factor of *Parhyale twist* (*Ph twi*), is expressed in a similar pattern in the head, may add to this hypothesis. *Ph mef2* expression starts in mesodermal cells after they have begun to proliferate. *Ph mef2* remains active in the developing muscles of *Parhyale* (Price and Patel, 2008).

To summarize, the model described here helps reconcile and visualize the processes that occur during stomodeum formation, development of median anterior brain structures and foregut invagination. In order to verify and corroborate this model, additional expression data and functional analyses of genes specifically involved in the processes mentioned above are necessary.

### 4.6.3 Preliminary *Ph six3* loss-of-function data suggests a crucial role in the development of median anterior protocerebral derivates and the foregut

The early expression profile of *Ph six3* indicates only partial functional conservation with *six3* genes of insect species, namely *Drosophila* and *Tribolium*. The situation in *Parhyale* argues against an involvement of *Ph six3* in early regionalization of the head in an area anterior to *Ph otd1* expression. Not only does *Ph six3* expression arise later than *Ph otd1* and apparently within the *Ph otd1* expression domain, but also is the anterior median region – whose patterning is governed by *six3* in *Tribolium* (Posnien, 2009) - not present until the completion of stage 15. This is in stark contrast to the blastodermal stages of *Drosophila* and *Tribolium*, during which the embryo resembles a regular superficial layer of cells, on which the expression of early developmental genes, such as *six3*, is quasi projected in broad, coherent regions (Posnien, 2009; Seo et al., 1999). Because *Ph six3* expression after stage 15 shows much more similarity to the corresponding situation in insects, this contrast during early development might in large parts be due to the drastically different modes of early embryogenesis of *Parhyale* compared to the insect species mentioned above. However, *Ph six3* expression is not found in newly emerging cell groups as they gradually contribute to the anterior median region (stage 14), but only considerably later. In addition, the posterior

part of the anterior median region is marked by *Ph kni1* expression during its establishment and remains free of *Ph six3* activity. In *Tribolium*, loss of *Tc'six3* function during early embryogenesis leads to the deletion of a large, coherent region comprising the labrum, the anterior brain and anterior and dorsal median head epidermis (Posnien, 2009). Disruption of *Ph six3* function via siRNA-mediated RNAi did not lead to similar phenotypes, although the amount of *Ph six3* transcripts was shown to be dramatically decreased. Instead, the anterior median and dorsal epidermis as well as the labrum were unaffected in all examined embryos that had been injected with siRNA targeting *Ph six3*. For the *Drosophila six3* homolog *optix*, comprehensive loss-of-function data is not available. Manipulation of *Drosophila* OPTIX DNA-binding specificity has been shown to affect eye development, but not anterior head regionalization (Kenyon et al., 2005).

Even though siRNA-mediated RNAi targeting *Ph six3* did in no observed embryo lead to deletions of the anterior half of the anterior median region, it caused drastic decrease in the size of the hemi-ellipsoid bodies. Along with the finding that many embryos subjected to *Ph six3* RNAi also showed malformed foreguts during later stages of development, this strongly suggests that *Ph six3* is required late in *Parhyale* development during the establishment and development of the median protocerebral derivatives, i.e. the hemi-ellipsoid bodies, as well as the foregut. Hatched animals that had been subjected to injection of siRNA targeting *Ph six3* at S1 showed impediment in food recognition. Because the hemi-ellipsoid bodies are believed to receive sensory input from olfactory neurons of the antenna in crustaceans (Strausfeld, 2009), this observation goes well with the suggested role of *Ph six3* in the development of these brain characters.

Conclusively, these data and findings from comparative analysis of *six3* expression in *Parhyale*, *Tribolium* and *Drosophila* suggest that *Tc'six3* represents a derived state, according to which *Tc'six3* has taken over upstream functions in early anterior pattering, thus governing the establishment and persistence of a coherent region and acting as a head gap gene. The situation of *Ph six3*, accordingly, might represent a more conserved state: the role of *six3* genes in anterior brain development (Steinmetz et al., 2010).

### 4.6.4 *Ph six4* is expressed during mesoderm and muscle development and might contribute to an ancestral bilaterian anterior neuronal region

Together with *sine oculis/Six1* (*so/Six1*) and *optix/Six3*, *six4* constitutes a conserved repertoire of SIX-domain and homeodomain encoding genes that are believed to play important roles in head development (Kawakami et al., 2000; Seo et al., 1999; Stierwald et al., 2004). Varying numbers of representatives of each of these three different SIX-HD factor types are present in diverse species, such as mammals (Kawakami et al., 2000), the cnidarian *Cladonema radiatum* (Stierwald et al., 2004) and *Drosophila* (Seo et al., 1999). Therefore, they have supposedly been present in the last common cnidarian-bilaterian ancestor (see also Ryan et al., 2006). Based on expression and functional data that has been gathered so far, all three SIX-HD families appear to be involved in head development. However, the exact functions may vary between different species. Homologs of *six4*, in particular, serve two major categories of developmental function: First, involvement in mesoderm differentiation and myogenesis (Clark et al., 2006; Grifone et al., 2005; Kirby et al., 2001; Maire et al., 2010; Seo et al., 1998), and second, a role in anterior head and brain patterning (Ghanbari et al., 2001; Posnien et al., 2011; Schlosser and Ahrens, 2004; Seo et al., 1998).

*Ph six4* expression during development goes well with the first putatively conserved role of *six4* genes: *Ph six4*-positive cells arise within all segmental appendages shortly after the distal projection of these is morphologically visible. In addition, its expression in segmental appendages is highly similar to that of *Parhyale twist* (*Ph twi*); more precisely, *Ph twi* is expressed in a subset of *Ph six4*-positive cells. *Ph twi* has been found to be active in a subset of segmental mesoderm (Price and Patel, 2008). This strengthens the idea that *Ph six4* might confer a conserved function in mesodermal and muscle development, probably upstream of *Ph twi*.

*Ph six4* head expression arises earlier during development than the expression in segmental mesodermal tissue. Here, three different *Ph six4*-positive cell populations can be distinguished: the first group consists of cells that are located within the distally projecting appendage anlagen of the An1, An2 and Mn segments. These populations most likely also represent mesodermal derivates. The fact that post-mandibular appendage development is delayed in relation to head appendage development explains why strong *Ph six4* expression arises earlier here. Importantly, mesodermal cells expressing *Ph six4* might well exist outside this obvious first group of cells as well.

The second group encompasses cells that are directly adjacent to *Ph six3*-positive cells of the anterior median region and that are located medially of the bases of the head segment appendages. The sharp, appearingly mutually exclusive expression domains of *Ph six4* and *Ph six3* in this area might indicate an underlying genetic mechanism conveying cross-regulation of these genes. Because siRNA-mediated RNAi targeting *Ph six3* has been shown to efficiently reduce the expansion and intensity of the anterior median *Ph six3* expression domains in injected embryos, at least during later stages, this question could be addressed by WMISH targeting *Ph six4* in stage 15-18 embryos that had been injected with siRNA specific for *Ph six3*.

The third group of *Ph six4*-positive cells in the developing *Parhyale* head comprises several median and, during later stages, medio-lateral clusters of cells that lie at the anterior end of the pre-antennal region of the head. Recent studies on *Tribolium six4* (*Tc'six4*) expressing cells within a comparable region of the anterior head suggest the existence of a specific developmental region that is conserved between vertebrates and insects, termed panplacodal region (Posnien et al., 2011). This region is characterised by a genetic network constituted by the same set of homologous genes in insects and vertebrates. Aside from *six4*, this region is marked by the activity of *eyes absent* (*eya*) and *so/Six1* (Posnien et al., 2011). In vertebrates, it gives rise to cranial placodes: specialized ectodermal regions that give rise to elements of the neuroendocrine system as well to the visual, olfactory and acoustic sense organs (reviewed in Streit, 2007). Although the developmental relevance of this region in *Tribolium* is still the subject of discussion, a detailed examination of the corresponding region in *Parhyale*, in particular of gene expression and function of the other "panplacodal" markes aside from *Ph six4*, would prove extremely informative of a possible conservation of such a distinct developmental region within the anterior head of bilaterians.

Taken together, I conclude that *Ph six4* confers conserved functions in mesodermal and head development, even though the exact relative temporal and spatial expression pattern may be different as compared to insects. In addition, *Ph six4* expression goes well with the model of an ancestral anterior placode field involved in the development of various neuronal characters.

## 4.7 *Ph awh* is a highly conserved representative of the family of LIM-HD encoding transcription factors

*Parhyale arrowhead* belongs to the divergent *awh*/*Lhx6/8* subgroup of LIM-HD encoding genes (Hobert and Westphal, 2000). Representatives of this family and, specifically, of the *awh/Lhx6/8* group are present throughout the bilaterian phylogeny (Hobert and Westphal, 2000). Expression analyses of these factors during mammalian development have suggested crucial roles in head development. In detail, *Lhx6* and *Lhx8* are expressed in the first branchial arch and the basal forebrain, where they are believed to be key components of signalling cascades that determine aspects of morphogenesis and differentiation. They have also been shown to be crucial for maxillary and mandibulary processes and are expressed in discrete regions of the telencephalon (Grigoriou et al., 1998). More recent studies have shown that *Lhx8* is required for the development of cholinergic neurons in the mouse forebrain (Zhao et al., 2003) as it represents a pivotal factor for cholinergic differentiation (Manabe et al., 2007; Manabe et al., 2008). In *Caenorhabditis elegans*, specific olfactory neuron fate is determined by the *awh/Lhx6/8* homolog *lim-4*. These data go well with *Ph awh* expression in the developing hemi-ellipsoid bodies during *Parhyale* embryogenesis. They represent median protocerebral elements of the crustacean brain that are discussed to be involved in olfactory sensation (Strausfeld, 2009).

Although *Drosophila Awh* is expressed in the nervous system as well, the only reported function to date concerns the establishment and proliferation of subsets of imaginal cells: abdominal histoblasts and salivary gland imaginal rings (Curtiss and Heilig, 1995, 1997). Notably, ectopic activity of *Awh* in other subsets of imaginal cells leads to apoptosis of these and to the deletion of corresponding adult structures. In detail, a neomorphic *Awh* allele leads to the specific loss of the retinal precursors before their differentiation (Curtiss and Heilig, 1995). Still, despite *Awh* expression in large bilateral procephalic domains during an intermediate phase of Drosophila embryogenesis – a situation that is highly reminiscent of *Parhyale awh* expression - no correlating function has been determined so far. Aside from that, *Parhyale awh* and *Drosophila Awh* share segmental expression in dispersed, medio-laterally located cells during mid and late development. In *Drosophila*, *Awh* expression remains only in the abdominal histoblasts at a later point of embryogenesis. It would be interesting to address *Parhyale awh* function in the respective cells, especially since *Parhyale* is an arthropod that develops directly.

Conclusively, these findings and observations suggest that *awh/Lhx6/8* genes were already present in the last common ancestor of all bilaterians. Although a conserved function for differential neurogenesis processes appears plausible, these genes appear, at least in arthropods, to have been recruited for more divergent developmental processes during evolution as well.

## 4.8 *Ph gbx* expression suggests a conserved role in deutocerebral patterning and VNS development

Within the group of homeodomain transcription factors, proteins encoded by homologs of *unpg/gbx* genes have been characterised as members of the Antennapedia/Hox-related class (ANTP/Hox-related; Galliot et al., 1999; Holland and Takahashi, 2005; Introduction). Comprehensive phylogenetic analyses of homeobox sequences including recent data from the diploblastic sea anemone *Nematostella vectensis* suggest that *unpg/gbx* genes were already part of the genetic repertoire of the cnidarian-bilaterian ancestor (Ryan et al., 2006). Important functions of these genes in head patterning, brain regionalisation as well as neurogenesis have been identified in various bilaterian species, suggesting that *unpg/gbx* genes are highly conserved head development genes (e.g., Burroughs-Garcia et al., 2011; Hirth et al., 2003; Lichtneckert and Reichert, 2005; Steinmetz et al., 2011; Urbach, 2007; Urbach and Technau, 2003). Here, I want to give an overview of several of the *unpg/gbx* functions and the corresponding developmental processes that have been identified in different species, compare them with the situation in *Parhyale* and, from that, derive a model for a putative set of conserved functions of these genes within pancrustaceans.

In many diverse species, expression of *unpg/gbx* arises in a domain that is posterior and adjacent to *otd/otx* expression (Lichtneckert and Reichert, 2005; Steinmetz et al., 2011; Urbach, 2007). Expression studies in the lophotrochozoan *Playtenereis dumerilii* opened the discussion for a possible ancestral regionalization and segmentation mechanism of the anterior head via *six3*, *otd/otx* and *unpg/gbx* that is directly followed posterior by segment-specific *hox*-gene expression (Steinmetz et al., 2011; Steinmetz et al., 2010). Although *Ph gbx* expression in the head is found posterior to that of *Parhyale otd* genes, it is not expressed during early embryogenesis. Also, it arises in restricted groups of cells which argues against a

role as a regional or segmental determinant. Therefore, *Ph gbx* does not substantiate the *Platynereis* model (4.5). Rather, its situation is highly similar to that of *Drosophila unpg* (see below).

In vertebrates, *otx* and *gbx* genes interact at the midbrain-hindbrain boundary during brain regionalization. As a consequence, a major developmental organizer is established, the MHB or isthmic organizer (for an overview, see Raible and Brand, 2004). Assuming the bilaterian ancestor already had a condensed anterior brain, the corresponding area in insects has been proposed to be the deuto-/ tritocerebral border (Hirth et al., 2003; Lichtneckert and Reichert, 2005). Although the mode of interaction and the genetic network comprising *otd* and *unpg* in *Drosophila* appears to be largely identical to the vertebrate situation, and the genetic interface in which this interaction takes place can be found at a similar relative position, no developmental organizer comparable to the MHB appears to be present in insects (Urbach, 2007). In addition, *Drosophila unpg* arises later in development than *otd*. Also, the *otd/unpg* interface of gene regulation is slightly shifted anterior as compared to vertebrates (Urbach, 2007). Based on the expression data retrieved for *Ph gbx* in this work, the *Parhyale* situation is highly similar to the one found in *Drosophila*: within the deutocerebral neuromere that derives from the An1 segment, anterior *otd1* expression borders and possibly overlaps with posterior *gbx* expression (3.3.2.4, Browne et al., 2006). The existence of regulatory interactions of *Ph otd1* and *Ph gbx* as well as the functional significance of it can only be addressed via efficient functional analyses, though. Still, the expression of *otd1* and *gbx* in *Parhyale* during the establishment and development of deutocerebral characters might indicate a conserved network of genetic interaction that has already been present in the pancrustacean ancestor.

*Ph gbx* also shows segmental expression in small bilateral, median ventral groups of cells that arises later than its expression in the head. Comparing it with *Parhyale Sp8* expression (Schaeper et al., 2010) suggests that these *Ph gbx*-positive cells are part of the developing ventral nervous system (VNS). In *Drosophila*, *unpg* expression in cells of the VNS is found as well. However, it arises earlier than the expression in the head (Chiang et al., 1995; Cui and Doe, 1995). From these findings, I conclude that *unpg/gbx* function in VNS patterning is likely to be conserved between *Parhyale* and *Drosophila*, and that the differences in timing reflect the different modes of embryogenesis of these two species.

Expression of *Ph gbx* during late appendage development appears to be specific for *Parhyale*. Considering the putative role of *Ph gbx* in brain development and neurogenesis in

the VNS, this expression might also reflect involvement of *Ph gbx* in neurogenic processes. Of course, this remains highly speculative. In *Drosophila*, *unpg* functions during late development have been found as well. In detail, *unpg* activity is required for tracheal branching to supply segmental neuromeres (Chiang et al., 1995). Both situations might represent additional *unpg/gbx* functions that have been acquired to fit specific developmental needs in the lineages leading to *Drosophila* and *Parhyale*, respectively. Although a common evolutionary origin for these functions appears implausible, given the fundamental differences in respiration between both species, a more comprehensive comparison of *unpg/gbx* genes of additional arthropod species might provide unexpected answers to this question.

Conclusively, the *Ph gbx* expression data argues for a conserved patterning mechanism involving *Ph otd1* in the An1 neuromere, i.e. the deutocerebral anlage, within insects and crustaceans. In addition, *unpg/gbx* function within the VNS is putatively shared among pancrustaceans as well.

## 4.9 *Ph al1* is not involved in head development

Based on phylogenetic analyses, two *Parhyale* representatives of the *aristaless*-related family of homeodomain transcription factors that were recovered in this work, *Ph al1* and *Ph al2*, were identified as orthologs of *Drosophila aristaless* (*Dm al*, Schneitz et al., 1993). *Dm al* is expressed in the imaginal discs of antennae and legs, in areas that represent the distal parts of the future appendages. Here, the expression of *Dm al* is regulated by *distalless*. Its function is crucial for the development of the distal-most parts of these appendages. *Dm al* expression is also found in a proximal area of the appendage imaginal discs (Campbell et al., 1993; Schneitz et al., 1993). Comparable situations are also found in other insects, such as *Tribolium* and *Gryllus* (Angelini et al., 2009; Miyawaki et al., 2002). For this reason, the role of *aristaless* in the formation of the most distal parts of appendages is considered conserved within insects (Beermann and Schröder, 2004).

In contrast, vertebrate *arx* is expressed in various parts of the developing brain (El-Hodiri et al., 2003; Miura et al., 1997). In humans, several hereditary diseases exhibiting phenotypes of intellectual disability have been linked to mutations in the *arx* gene (e.g., Shoubridge et al., 2010). Importantly, neuronal expression of an *aristaless* homologue has

also been described in an ecdysozoan species, the nematode *C. elegans* (Melkman and Sengupta, 2005).

Similar to *aristaless* expression in insects, expression of *Ph al1* is restricted to appendages and, during late embryogenesis, to small groups of cells located in the dorsal part of the segments. During all examined stages, no expression has been found in head or brain structures. In contrast to the situation in insects, *Ph al1* expression in the distal-most regions of the elongating appendages is found only temporarily in a small fraction of cells. Different to *Dm al*, *Ph al1* is expressed in all segmental appendages. No expression has been found for *Ph al2* so far.

Notably, murine *arx* expression in the developing forebrain is regulated by Dlx genes (Cobos et al., 2005). These represent the vertebrate homologues of insect *distalless* genes (Panganiban and Rubenstein, 2002). This observation suggests that the regulation of *aristaless* may, at least in part, be conserved within bilaterians (Cobos et al., 2005).

From these findings, I propose the following model: *aristaless* function in the development of appendages is a conserved feature between insects and malacostracans, possibly even within arthropods. The function of vertebrate *arx* genes in the developing brain might be derived. Interestingly, *Drosophila munster*, a close relative of *Dm al* (e.g., Walldorf et al., 2000; Results), plays an important role in larval eye development (Goriely et al., 1999). According to phylogenetic analyses, it is co-orthologous to murine *arx*. This may indicate that both *aristaless* and *munster* derive from a single gene that shared both brain and appendage function. However, it remains unknown if the *Drosophila* situation of *aristaless* and *munster* exists in other arthropods as well. So far, no *Parhyale* homologue of *munster* has been identified. Also, additional *aristaless* paralogs with functions in head and brain development might exist in *Parhyale*.

## 4.10 Findings from *Parhyale knirps* genes suggest conserved roles in SNS development and appendage branching but not in head patterning and segmentation

From *Parhyale hawaiensis*, two *kni* paralogs were isolated. *Ph kni1* is expressed dynamically in many different tissues and developmental fields in the pre-antennal head, within the anterior median area, in all appendage anlagen and later during appendage morphogenesis in cells that are located at joints or connections between appendage articles. *Ph kni2* expression is found in a subset of *Ph kni1*-positive cells and arises late during development (3.4.8).

*Tribolium castaneum knirps* (*Tc'kni*) is expressed during blastoderm stage in a stripe-shaped domain within the prospective head region of the embryo. Loss of *Tc'kni* function via RNAi results in the deletions of the antennal and mandibular segments (Cerny et al., 2008). Similar expression domains of *kni* paralogs of other insects (Ben-David and Chipman, 2010; Gonzalez-Gaitan et al., 1994; Rothe et al., 1994) have raised the question whether head gap-gene functions may represent an ancestral role of the arthropod-only *kni* genes.

However, even though *kni* is expressed in gap-like anterior domains during early embryogenesis and blastoderm stages, respectively, in several insect species (Oncopeltus, Ben-David and Chipman, 2010; Tribolium, Cerny et al., 2008; Drosophila, Rothe et al., 1994), head gap function of *kni* has been found only for *Tc'kni* so far (Cerny et al., 2008). In *Parhyale*, neither *kni1* nor *kni2* show similar expression domains. Preliminary results from siRNA-mediated RNAi targeting *Ph kni1* did in no case reveal gap-like phenotypes in the *Parhyale* head. Instead, *kni1* – which is the only paralog active during early *Parhyale* embryogenesis - is expressed only in small groups of cells at a high level of dynamics. Although it cannot be excluded that further *kni* paralogs exist in *Parhyale* that might show an early expression pattern similar to insects, repeated and independent attempts to isolate additional *kni* sequences via depletive varied-stringency PCR did not provide any indication of additional members of the *kni* family in *Parhyale*.

In contrast to these differences in early anterior expression, *Ph kni1* shares expression in tissues associated with the stomatogastric nervous system (SNS) with *Drosophila kni* and *knrl* (Gonzalez-Gaitan et al., 1994). Notably, hatched *Parhyale* animals that had been injected with siRNA targeting *Ph kni1* exhibited difficulty in feeding and ingestion. This might point to a *Ph kni1* function in proper SNS development as well. Involvement of *kni* in SNS development

has not been addressed in other species so far. However, late expression of *kni* in *Tribolium* in the VNS suggests that it is associated with neuronal development.

After siRNA-mediated *kni1* RNAi, *Parhyale* animals showed defects in appendage morphogenesis and alignment. This goes well with the expression of *kni1* in cells located at joints and branching points between appendage articles. Interestingly, *Drosophila kni* and *knrl* are essential for mediating DPP signalling during tracheal placode development which is required for cell migration and tracheal branch formation. In detail, they are active in the dorsal and ventral branches of developing tracheal placodes. EGF signalling is, in turn, mediated via *spalt* (*sal*) and is required for anterior-posterior cell migration and morphogenesis. Several of these opposing signalling processes are coordinated by direct interaction of *Drosophila kni/knrl* and *sal*; accordingly, *kni* has been shown to repress *sal* activity in dorsal tracheal cells (Chen et al., 1998). On a very abstract point of view, a similarity could be seen in *Parhyale*, where *kni1* is expressed at appendage branching points and in the proximal gill anlagen. In order to find out if *Ph kni1* acts in a conserved way as coordinator between different signals during appendage development, maybe even representing a conserved genetic network including *sal*, additional homologs of known factors involved in *Drosophila* tracheal development have to be isolated and examined for expression and function during *Parhyale* appendage and gill development. Strikingly, several *Parhyale* orthologs of *Drosophila* genes involved in tracheal development have been shown to be expressed in gill anlagen in *Parhyale* (Franch-Marro et al., 2006). Moreover, there is ongoing discussion about a close evolutionary connection between arthropod appendages, arthropod respiratory organs and insect wings, and, more precisely of an evolutionary origin of trachea and gills (Averof and Cohen, 1997; Damen et al., 2002; Franch-Marro et al., 2006).

During head development, *Ph kni1* expression follows a dynamic and complex topology. A similar expression pattern has been shown for *Ph mef2*, in addition to expression of *Ph mef2* in mesodermal cells (Price and Patel, 2008). In many species, *mef2* genes play important roles in mesoderm development (Brand-Saberi and Christ, 1999; Elgar et al., 2008; Olson et al., 1995). However, *mef2* appears to be involved in neurogenesis as well (Heidenreich and Linseman, 2004). Therefore, the question remains open whether *Ph kni1* is also expressed during a phase of mesodermal development, or whether *mef2* shares functions in neurogenesis with *kni1* in Parhyale.

From these findings and observations, it seems unlikely that *knirps* genes play an ancestral role in early anterior patterning and head regionalisation. Instead, expression and

functional data suggest a conserved role in neurogenesis, more precisely in the subset of SNS neurons. The possible homology of crustacean arthropods and insect wings and/or tracheal placodes remains an intriguing question that might be addressed by comparative, more directed approaches including *kni* homologs of insects and crustaceans. The high number and dynamics of *kni*-expression domains in *Parhyale* supports the idea of *kni* genes as evolutionary young, pleiotropic genes.

## 4.11 Conclusions and Outlook

Although the principal mechanisms underlying bilaterian head development, more precisely anterior determination, head regionalization, head segmentation and the specification and substructuring of head and brain characters, remain largely unknown, the findings from head gene expression in *Parhyale* and, in a limited way, indications as to the function of two of them (*Ph six3* and *Ph kni1*), helps retracing the evolutionary steps of head development as well as conservation and plasticity of single elements. In general, expression of the *Parhyale* homologs of highly conserved head development genes, such as *Ph otd1* and *Ph otd2* (Browne et al., 2006), *Ph six3* and *Ph six4*, *Ph gbx* and *Ph awh* supports the idea that these genes and gene families represent ancestral players in comparable processes of head development and the patterning of developmental neuronal fields (4.6.4) that may have been already present in the last common ancestor of bilaterians. In general, the expression data of the *Parhyale* genes reflect a situation that is similar to that in insects rather than in vertebrates or other non-arthropod species. In this regard, the findings from *Parhyale* support and strengthen models, according to which these genes had acquired specific functions in head development in the last common ancestor of insects and crustaceans that are different from those found outside pancrustacean or arthropod species. Especially the expression profile of *Ph gbx* gives important insights in the divergent evolutionary path of this gene family between pancrustaceans (deutocerebral patterning), lophotrochozoans (early patterning of the post-stomodeal region) and chordates (establishment of the MHB organizer). To determine the extent to which these conserved elements of head development also apply to other non-pancrustacean species, additional data from chelicerates and myriapods is required. In contrast, several findings from *Parhyale* point towards additional and putatively more

ancestral roles of these genes outside head development. E.g., this is the case for the expression pattern of *Ph six4* which further supports a conserved role of these genes in mesoderm development of bilaterians.

Even for *Parhyale* genes that appear conserved compared to insects, the data presented here give examples of functional divergence, plasticity and fragmentation that can occur during the evolutionary adaptation within even closely related species. An example is *Ph six3*, whose expression profile differs in several aspects from that of insect *optix/six3* genes, reflecting in part the differences in early embryonic development. This goes well with preliminary functional data that argues against an important role of *Ph six3* in early anterior head patterning.

Lastly, members of the *kni* family of genes which have only been found in arthropods so far, convey pleiotropic and evolutionary divergent functions during development (e.g., Ben-David and Chipman, 2010; Cerny et al., 2008; Chen et al., 1998; Fuss et al., 2001; Gonzalez-Gaitan et al., 1994; Lunde et al., 1998; Nauber et al., 1988). In particular, involvement in head patterning and segmentation appears to be specific for insects, although the loss-of-function phenotypes of *kni* genes there show considerable divergence as well. Expression data from *Ph kni1* and *Ph kni2* as well as preliminary functional results suggest alternative conserved roles of *kni* genes in SNS development and, although more speculatively, as possible conserved downstream coordinators of appendage branching.

However, to verify and confirm the implications suggested from *Parhyale* expression data, efficient and comprehensive functional analyses have to follow up. Compared to other emerging non-insect arthropod models (e.g., the common house spider Achaearanea tepidariorum, McGregor et al., 2008), *Parhyale* is less suited for these approaches. Even though *Parhyale* offers several advantages compared to other crustacean systems, the lack of genomic sequence data, the small number of offspring per female and the low efficacy of RNAi-based techniques renders comprehensive functional and comparative approaches ineffective and tedious, if not impracticable. Because, in addition, *Parhyale hawaiensis* represents a derived pancrustacean species with regard to development, it is – in my opinion - unlikely to remain an arthropod model for future studies. However, along with data from additional non-model species, findings from *Parhyale* will in any case be valuable contributions to controversial discussions and models that are focused on specific comparative developmental or evolutionary issues.

# 5 Materials and Methods

## 5.1 *Parhyale hawaiensis* culture

Rearing of *Parhyale hawaiensis* was done as described in Browne et al. (2005), with the following modifications. Wild-type populations of *Parhyale hawaiensis* were reared at 26°C-28°C in rectangular plastic containers with lids (inner dimensions 0.6 m x 0.4 m x 0.2 m) containing a 2 cm layer of crushed coral substrate (grain size average 4 mm in diameter) and filled with 10 l 3% sea salt water. The 3% sea salt water was prepared from 300 g artificial sea salt ("Aqua Medic Reef Salt", Aqua Medic, Bissendorf, Germany), and 10 l desalinated water. Continuous water circulation and aeration was provided by circulation pumps (specialized trade for fishkeeping). Every two weeks, the water of the *Parhyale* stocks was completely exchanged by freshly prepared, pre-warmed 3% salt water.

Wild-type *Parhyale* were fed on average twice a week with a mixture of flake food (Tetra Rubin, Tetra GmbH, Melle, Germany), ground seaweed (*Fucus vesiculosus,* specialized trade), *Spirulina* (specialized trade for fishkeeping) and, optionally, liquid feeding supplements containing trace minerals and vitamins (specialized trade for fishkeeping), blended in a small amount of water. Usually, one 10 l *Parhyale* stock received 100 g of food each week. If needed, the amount of food given was increased in order to raise population density, frequency of mating and average number of produced embryos per female. In this case, water exchanges were done more often (as seen necessary).

## 5.2 Molecular biology

Unless mentioned otherwise, standard protocols were followed (Sambrook et al., 2000). Likewise, standard buffers and solutions were prepared and used as described in Sambrook et al. (2000).

### 5.2.1 Isolation of *Parhyale hawaiensis* candidate gene homologs

Complete cDNA sequences of *Parhyale* homologs of candidate genes involved in head development were recovered through three subsequent steps:

(i) To obtain primary *Parhyale*-specific gene sequences, varied-stringency PCR (5.2.3.1) using degenerate primers (5.2.4, A4.1, table A2) and double-strand cDNA collections as template (5.2.2) were performed. The degenerate primers were based on comparative protein sequence alignments of known insect and vertebrate homologs of the genes of interest. In particular, alignments of conserved protein motifs were preferred for the generation of degenerate primers.

(ii) Based on primary *Parhyale* gene sequence data, specific oligonucleotides were generated and used for 5' and 3' RACE reactions (5.2.3.3).

(iii) From sequence clones obtained through RACE, complete and coherent cDNA sequences were composed *in silico*. Then, these were verified via long-distance PCR using primers specific for the 5' and 3' ends of the hypothetical cDNA and double-strand cDNA collections as template.

From comparisons of all recovered cDNA sequences of a given gene, an artificial cDNA reference sequence was derived that resembled the statistically most relevant (likely) DNA sequence, both with regard to transcript isoforms and allelic variation (5.3.3.2). The reference sequences were used for phylogenetic analyses (5.3.2). For molecular applications (e.g. WMISH), a cDNA clone or, if unavailable, an ORF clone of the gene of interest with the highest sequence similarity to the respective reference sequence was used.

### 5.2.2 Preparation of *Parhyale hawaiensis* cDNA collections

From collections of Parhyale hawaiensis embryos covering all stages of development, mRNA was isolated and subsequently utilized for the generation of first-strand and double-strand cDNA collections. To this end, the "Micro Poly(A)Pure™ Small Scale mRNA Purification Kit" (Ambion; Austin, USA) was used. Typically, 2 µg highly purified mRNA was recovered from ca. 1000 Parhyale embryos (corresponding to a volume of 200 µl) that had been washed several times in PBS.

First-strand cDNA was generated via reverse transcription using Parhyale mRNA collections and the „SMART™ PCR cDNA Synthesis Kit" (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France), or the „SMART™ RACE cDNA Amplification Kit" (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France), if 5' and 3' specific first-strand cDNA collections were desired (e.g. for subsequent RACE PCR). In both cases, the manufacturer's instructions were followed accurately.

Double-strand cDNA was generated via PCR using Parhyale first-strand cDNA collections and the „SMART™ PCR cDNA Synthesis Kit" (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France), following the manufacturer's instructions.

### 5.2.3 Polymerase chain reactions (PCR)

PCR techniques using complex or heterogeneous templates, such as varied-stringency degenerate PCR (vsdPCR), depletive vsdPCR, RACE and long-distance PCR reactions targeting coherent ORF and complete cDNA sequences, were performed using the "Advantage® 2 PCR Kit" (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France) according to the instructions and recommendations of the manufacturer. PCR used for fast identification of transformant plasmids ("colony PCR") was done using the "Mango*Taq*™ DNA Polymerase" (Bioline, London, UK), following the manufacturer's instructions. For all other PCR applications, the "Taq DNA Polymerase (recombinant)" (FERMENTAS GMBH, European office, St. Leon-Rot) was used, following the manufacturer's instructions. The "Eppendorf Mastercycler gradient" (Eppendorf AG, Hamburg, Germany) was used for vsdPCR. For all other PCR applications, the "Eppendorf Mastercycler personal" (Eppendorf AG, Hamburg, Germany) was used as PCR cycler.

#### 5.2.3.1 Varied-stringency degenerate PCR (vsdPCR)

To obtain primary *Parhyale*-specific sequence fragments of gene-of-interest homologs, PCR using degenerate primers targeting conserved cDNA regions, double-strand cDNA collections covering all stages of *Parhyale* embryogenesis and the "Advantage® 2 PCR Kit" (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France) were performed. Triple volume (150 µl) standard reactions (manufacturer's instructions) were partitioned into 12 smaller reactions and distributed over an annealing temperature gradient ranging from 30°C to 50°C. This mediated a sufficient variation of stringency over the individual partitioned reactions to allow the isolation of homologous *Parhyale* gene sequences without exact knowledge of their similarity to known representatives of these genes in other species. For this PCR technique, the "Eppendorf Mastercycler gradient" (Eppendorf AG, Hamburg, Germany) was used.

#### 5.2.3.2 Depletive vsdPCR

Typically, sequence fragments of several genes were recovered from vsdPCR. The relative amount of an individual sequence species within the recovered vsdPCR amplificate depended on (i) the annealing temperature of the reaction it was recovered from and (ii) its sequence similarity to the degenerate primers. If desired to reduce the relative amount of abundant sequence species from a vsdPCR amplificate and, as a consequence, enrich the less frequent, more divergent sequence species, whole amplificates were purified (5.2.5) and subjected to specific DNA restriction reactions. Restriction enzymes targeting the most abundant sequence species were used. To minimize off-target restriction of rare sequence species, at least two different restriction enzymes specific for any single abundant sequence species were used in separate reactions. The remaining unrestricted sequence fragments were recovered and directly cloned in pCRII (5.2.6.1) or, if required, subjected to another round of PCR amplification (specifications as for vsdPCR, 5.2.3.1) prior to cloning in pCRII (5.2.6.1).

#### 5.2.3.3 Rapid elongation of cDNA ends (RACE)

To obtain 5' and 3' cDNA fragments of *Parhyale* gene of interests, RACE PCR were performed using the „SMART™ RACE cDNA Amplification Kit" (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France) and the "Advantage® 2 PCR Kit" (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France), following the manufacturer's instructions and recommendations. In detail, first-strand 5' and 3' cDNA collections covering all stages of Parhyale embryogenesis were used as templates along with specific reverse (orientation complementary and reverse to mRNA) and forward (orientation according to mRNA) primers, respectively. PCR programs were set up anew for each individual RACE PCR reaction. In general, the "touch-down" specification was used in RACE PCR programs: initially, few cycles (3-5) with maximum annealing temperature (72°C) were run; over the next 10 cycles, the annealing temperature was gradually lowered to $T_m$ (oligonucleotide in use) - 5°C; eventually, 20-30 cycles with the constant final annealing temperature were done. For more information on "touch-down" PCR programs, see the „SMART™ RACE cDNA Amplification Kit" manual.

In cases, when the intial 5' or 3' RACE PCR were insufficient to amplify the desired DNA fragment unambiguously, subsequent RACE PCR were performed, using the purified amplificate of the initial RACE PCR as template and "nested" primers that targeted template sequences 5' and 3' of the original primers, respectively. Again, "touch-down" PCR programs were used for these reactions.

#### 5.2.3.4 Long-distance PCR

To isolate complete and coherent cDNAs or cDNA fragments covering only the coherent ORF of *Parhyale hawaiensis* genes of interest, long-distance PCR were performed. To this end, the "Advantage® 2 PCR Kit" (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France) was used. The manufacturer's recommendations for long-distance PCR were followed. Usually, "touch-down" PCR programs were used (see also 5.2.3.3.

### 5.2.4 Oligonucleotides

Oligonucleotides were synthesized by Eurofins MWG Operon (Ebersberg, Germany). Oligonucleotides that were used as degenerate primers were derived from conserved protein motifs, based on alignments of homologous proteins from different species (see also 5.2.1). Precisely, degenerate primers represented a proportionate mix of different oligonucleotide species exhibiting all potential nucleic acid permutations that translated into the protein sequence of choice. As an exception, at positions that would have required all four types of dNTPs (N-positions according to the degenerate nucleic acid code), an inosine residue was included instead of creating four different dNTP variants. A complete list of degenerate oligonucleotides, their sequence, descriptions of their development from respective protein sequences and their *Parhyale* cDNA targets is in the appendix (A4.1, table A2). A complete list of specific oligonucleotides, their sequence, the reactions they were used for and the genes they were specific for is also in the appendix (A4.2, table A3).

### 5.2.5 DNA purification

DNA and DNA-containing reactions and solutions, such as cDNA collections prior to the use as PCR templates, primary RACE PCR amplificates to be used for a subsequent RACE PCR or plasmid DNA used for transgenesis, were purified using the „QIAquick PCR Purification Kit“ (Qiagen, Hilden, Germany), following the manufacturer’s instructions.

### 5.2.6 Cloning

For cloning, restriction enzymes (FERMENTAS GMBH, European office, St. Leon-Rot and New England Biolabs, Ipswich, USA), the “QIAEXII Gel Extraction Kit” (Qiagen, Hilden, Germany), T4 ligase (FERMENTAS GMBH, European office, St. Leon-Rot) and chemically competent DH5 alpha *E. coli* were used, following standard protocols (Sambrook et al., 2000) or the manufacturers’ instructions, respectively. Plasmid DNA was obtained using the “NucleoSpin® Plasmid” (Macherey-Nagel, Düren, Germany) and the “QIAGEN Plasmid Maxi Kit” (Qiagen, Hilden, Germany). For cloning hairpin constructs (MM), “SURE

2 Supercompetent Cells" (Stratagene, Agilent Technologies, Santa Clara, USA) were used, following the manufacturer's instructions and recommendations.

#### 5.2.6.1 Cloning of PCR amplificates

Primary gene sequences (5.2.3.1; 5.2.3.2), RACE fragments (5.2.3.3), coherent ORF fragments and complete cDNAs (5.2.3.4) of *Parhyale* homologs of genes of interest were cloned into *pCRII* using the "TA Cloning® Kit Dual Promoter (pCR®II)" (Invitrogen GmbH, Karlsruhe) according to the manufacturer's instructions. Blue-white selection was used as described in the manufacturer's protocol.

#### 5.2.6.2 Cloning of *pSL_shuttle*

The plasmid *pSL_shuttle* was generated to facilitate cloning of constructs for heat-shock inducible RNA hairpin expression. It was created as follows:

(i) ***pSLfa_SV40_fa***: A fragment containing an SV40 polyadenylation signal (*SV40*) was generated via PCR, using the oligonucleotides 6114 and 6115 (A4.2, table A3) and *pMi[3xP3_EGFP]* (constructed by Schinko, J.) as template. This fragment was cloned into *pSLfa1180fa* (constructed by Horn and Wimmer, 1999) via *Nde*I/*Bsi*WI, creating *pSLfa_SV40_fa*.

(ii) ***pSL_shuttle*** (*pSLfa_hsp70_SV40_fa*): A 2.5 kb fragment covering the *Parhyale hawaiensis* hsp70 upstream region (Pavlopoulos et al., 2009) which contains a *Parhyale* specific heat-shock-inducible regulatory element, was generated via PCR, using the oligonucleotides 6950 and 6951 (A4.2, table A3) and *pSL_Hsp70_En2* (*courtesy of Anastassios Pavlopoulos*) as template. It was cloned in *pSLfa_SV40_fa* via *Asc*I/*Nco*I, creating *pSLfa_hsp70_SV40_fa*. This plasmid was termed *pSL_shuttle*.

### 5.2.6.3 Cloning of *pMi[3xP3_DsRed; hsp70_hp-otd1]*

The construct *pMi[3xP3_DsRed; hsp70_hp-otd1]* was created through the following steps:

(i) ***pSL_shuttle_KanR***: A fragment covering a genetic element that mediates Kanamycin resistence (*KanR*) was generated via PCR, using the oligonucleotides 9600 and 9601 (A4.2, table A3) and the plasmid *pGBKT7* (BD Biosciences Clontech, Heidelberg, Deutschland) as template. This fragment was cloned in *pSL_shuttle* (5.2.6.2) via *Bam*HI/*Acc*65I, creating *pSL_shuttle_KanR.*

(ii) ***pSL_shuttle_hpotd1-KanR-1dtoph***: Two complementary *otd1* hairpin arm fragments were generated analogously via PCR, using the oligonucleotides 1030 and 1032 for one hairpin arm and 1031 and 1033 for the other. The first hairpin arm was cloned in *pSL_shuttle_KanR* between *hsp70* and *KanR* elements via *Nco*I/*Bam*HI. The second hairpin arm was cloned between the *KanR* and *SV40* elements via *Acc65*I/*Not*I. As a result, *pSL_shuttle_hpotd1-KanR-1dtoph* was created.
Importantly, the *KanR* element separated both hairpin arms, facilitating subsequent cloning steps as well as stable propagation of the respective plasmids. In addition, it was used for positive selection of all plasmid generations that contained the complete hairpin sequence.
Also, the oligonucleotides 1032 and 1033 introduced unique restriction sites for the "blunt cutters" *Pma*CI and *Sna*I, respectively. During the last cloning step resulting in the final *pMi[3xP3_EGFP; hsp70_hp-otd1]*, these restriction enzymes were used to remove the *KanR* element entirely and, via cis-ligation, constitute the 'mature' *otd1* hairpin, in which both hairpin arms were separated by 8bp of non-palindromic nucleic acid sequence.

(iii) ***pMi[3xP3_DsRed; hsp70_hpotd1-KanR-1dtoph]****:* The fragment containing *hsp70_hpotd1-KanR-1dtoph* was generated from *pSL_shuttle_hpotd1-KanR-1dtoph* via *Asc*I/*Not*I restriction. It was cloned in *pMi[3xP3_DsRed]* (constructed by Schinko, J.) via *AscI*/*Not*I, creating *pMi[3xP3_DsRed; hsp70_hpotd1-KanR-1dtoph]*.

(iv) ***pMi[3xP3_DsRed; hsp70_hp-otd1]:*** The *KanR* element was removed from *pMi[3xP3_DsRed; hsp70_hpotd1-KanR-1dtoph]* via *Sna*I/*Pma*CI restriction. By cis-ligation using "SURE 2 Supercompetent Cells" (Stratagene, Agilent Technologies, Santa Clara, USA; see also 5.2.6), *pMi[3xP3_DsRed; hsp70_hp-otd1]* was created. This construct was used for RNAi approaches based on stable transgenic as well as transient heat-shock induced expression of RNA hairpins.

#### 5.2.6.4 Cloning of *pMi[3xP3_EGFP; hsp70_hp-DsRed]*

The construct *pMi[3xP3_EGFP; hsp70_hp-DsRed]* was cloned analogously to *pMi[3xP3_DsRed; hsp70_hp-otd1]*, using *pMi[3xP3_DsRed]* (constructed by Schinko, J.) and the oligonucleotides 8100/8104 and 8101/8105 for generating *DsRed* hairpins and *pMi[3xP3_EGFP]* (constructed by Schinko, J.) as the final transposon plasmid carrying the hairpin.

#### 5.2.6.5 Cloning of *pimpMiII[mse_EGFP]*

To create *pimpMiII[mse_EGFP]* (A5), a fragment containing restriction sites for *Asc*I, *Eco*RI, *Pme*I, *Dra*I, *Bsi*WI, *Psp*OMI, *Apa*I, *Mlu*I, *Nco*I and *Fse*I was cloned into *pMiII[mse_EGFP]* (constructed by Schinko, J.) via *Asc*I/*Fse*I restriction. The fragment was obtained by annealing the complementary oligonucleotides 9700 and 9701 (A4.2, table A3).

#### 5.2.6.6 Cloning of *pimpMiII[mse_EG-kni2intron-FP]*

The construct *pimpMiII[mse_EG-kni2intron-FP]* was created through the following steps:

(i) ***pCRII_kni2intron***: A putative *Ph kni2* intron containing fragment was isolated via PCR (5.2.3), using the oligonucleotides 1864 and 1865 (A4.2, table A3) and the BAC plasmid L23 #152 (courtesy of N. Patel) as template. The fragment was T/A cloned into *pCRII* (done by Jonas Schwirz; 5.2.6.1), creating *pCRII_kni2intron*.

(ii) ***pCRII_EG-kni2intron***: A fragment encompassing the 5' part of the EGFP ORF (42 bp) and the complete *Ph kni2* intron was generated via PCR, using the oligonucleotides 8250 and 8251 (A4.2, table A3) and the plasmid *pCRII_kni2intron* as template. In detail, the 42 bp of the 5' EGFP ORF were added 5' to the intron sequence via the oligonucleotide 8250 (A4.2, table A3). The obtained PCR fragment was T/A cloned in *pCRII* (Jonas Schwirz; 5.2.6.1), creating *pCRII_EG-kni2intron*.

(iii) ***pCRII_-FP***: A fragment covering the 3' part of the EGFP ORF was generated via PCR, using the oligonucleotides 8258 and 8259 (A4.2, table A3) and *pMiII* (courtesy of Nipam Patel) as template. This fragment was T/A cloned in *pCRII* (Jonas Schwirz; 5.2.6.1), creating *pCRII_-FP*.

(iv) ***pCRII_EG-kni2intron-FP***: The fragment containing the 3' part of the EGFP ORF ("-FP") was generated by *Sac*I/*Bam*HI restriction of *pCRII_-FP*. It was cloned in *pCRII_EG-kni2intron via SacI/BamHI,* creating *pCRII_EG-kni2intron-FP*.

(v) ***pimpMiII[mse_EG-kni2intron-FP]***: The fragment containing the composite *EG-kni2intron-FP* sequence was generated via *Pac*I/*Sbf*I restriction of *pCRII_EG-kni2intron-FP*. It was cloned in *pimpMiII[mse_EGFP]* (5.2.6.5) via *Pac*I/*Sbf*I, thereby substituting the original EGFP ORF of *pimpMiII[mse_EGFP]*. As a result, *pimpMiII[mse_EG-kni2intron-FP]* was created. *pimpMiII[mse_EG-kni2intron-FP]* was used to test the functionality of the putative *Ph kni2* intron (3.5.3).

## 5.3 Sequence analysis

### 5.3.1 Sequence alignments

For the generation of degenerate primers, protein sequence alignments were done using MultAlin (Corpet, 1988).

For visualization, recovered sequences and sequence fragments of *Parhyale hawaiensis* homologs of candidate genes, e. g. from vsdPCR or RACE, were aligned using MultAlin (Corpet, 1988), as well.

For phylogenetic calculations and analyses, protein sequence alignments were done using the CLUSTALW2 algorithm (Larkin et al., 2007). Unless mentioned otherwise, standard parameters were used.

### 5.3.2 Phylogenetic calculations

Phylogenetic calculations were done using MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). To this end, alignments (5.3.1) of all protein sequences to

be included in the phylogenetic calculation were generated using CLUSTALW2 (Larkin et al., 2007). Unless specifically mentioned otherwise, complete protein sequences were used.

Typically, MrBayes calculations were with the following parameters set: mixed amino acid models ("prset aamodelpr=mixed"); no outgroup set, 100,000 generations ("mcmc ngen=100000") and a (default) sample frequency of 100. In all other regards, the programmers' instructions and recommendations were followed (Ronquist et al., 2005; http://mrbayes.csit.fsu.edu/mb3.1_manual.pdf). In case alternative parameters were used, they are described in detail in the respective chapter.

### 5.3.3 Evaluation of gene sequences

Recovered gene sequences were evaluated with regard to (i) sequence novelty, (ii) homology to known genes from other species, (iii) paralogy, more precisely the question whether they derive from a single or from several Parhyale hawaiensis loci, (iv) the existence of polymorphisms, alleles and isoforms of the corresponding *Parhyale* gene and (v) occurrence of artefacts resulting from the molecular techniques in use.

#### 5.3.3.1 Test of sequence novelty and assessment of identity and homology of gene sequences

To test if recovered sequences represented novel *Parhyale hawaiensis* genes or, in contrast, derived from contaminations by material from other species, BLAST evaluation using the tblastn algorithm (searching translated nucleotide databases using a protein query, Altschul et al., 1997) was performed.

Identity and putative homology of primary *Parhyale hawaiensis* gene sequences, recovered from vsdPCR, was assessed by using blastx (searching of protein databases using a translated nucleotide query, Altschul et al., 1997).

To verify initial suggestions of homology, translated protein sequences of complete *Parhyale hawaiensis* cDNA sequences, more precisely the corresponding reference sequences (5.2.1), were subjected to blastp searches (searching protein databases using a protein query, Altschul et al., 1997). In particular, the searches were performed on the protein databases of

*Drosophila melanogaster* (taxid 7227, Appendix) and *Mus musculus* (taxid 10090, Appendix). In addition, they were subjected to tblastn searches (see above) to avoid biased homology assessment resulting from *Drosophila*- and mouse-specific gene repertoires. In case these tblastn search results appeared ambiguous (e.g. because the retrieved sequences of highest similarity represented unverified gene predictions), they were 'reBLASTed' against *Drosophila* or mouse databases (blastp, see above).

#### 5.3.3.2 Evaluation of sequence polymorphisms and transcript isoforms

An individual site of nucleotide exchange was considered polymorphic, if each of the different occurring nucleotide residues was found in more than one recovered sequence of the same gene. E.g., a site of the nucleotide exchange (A/T/C) was considered polymorphic, if at least two recovered sequences had an adenosine at this position, two other had a thymidine and two further ones had a cytosine. If none of the different nucleotides of a polymorphic site altered the translated protein sequence (in case the site lay within the ORF of the gene of interest), the site was considered to represent an allelic variation of transcripts deriving from the same gene locus. These "allelic polymorphisms" were excluded from assessments of paralogy (5.3.3.3).

In case polymorphisms resulted in protein sequence changes or premature stop codons, the corresponding sequences were analysed in more detail. In particular, they were examined for putatively deriving from a different gene locus and, therefore, representing a paralogous gene (5.3.3.3). If the frequency of occurrence of such non-silent polymorphisms was too low to suggest paralogy, they were considered artefacts. However, although they might also present naturally occurring mutant or hypomorphic alleles, this was not followed further.

Nucleotide exchanges that occurred only in one corresponding sequence were subjected to paralogy assessment (5.3.3.3). If the corresponding sequence was not evaluated as potentially paralogous to the sequence it was compared to, these nucleotide exchanges were considered artefacts.

#### 5.3.3.3 Assessment of paralogy

In case a nucleic acid sequence revealed less than 1% non-polymorphic nucleotide exchanges compared to the respective reference sequence (or, if no reference sequence was available, compared to all other sequences of the same gene isolation approach), it was considered the same gene as represented by the reference sequence or derived from of the same gene locus as the other sequences, respectively.

In case the difference was more than 10%, the examined gene sequence was considered deriving from another, supposedly paralogous gene.

In the remaining cases (between 1-10% differences in the nucleic acid sequence), the respective sequence was tested in more detail, e.g. via a series of BLAST searches (Altschul et al., 1997), on the base of its translated protein sequence and the like. The results of these tests were then described in detail in the respective paragraph.

## 5.4 Embryology and Histology

### 5.4.1 Harvesting of *Parhyale hawaiensis* embryos

Mating *Parhyale* couples were collected from culture stocks. Directly after release from the mating posture, females were separated from males and kept for 1h to allow successful completion of the transfer of the embryos into the brood pouch (2.5.4). Females were then transferred on a silica pad, fixed with preparation needles and dissected in order to obtain the embryos. The entire batch of embryos was immediately transferred to a single well of 24-well plates that contained 1 ml filtered 3% sea salt water. The embryos were used directly or were incubated at 28°C until they reached the desired age.

### 5.4.2 Fixation and dissection of *Parhyale hawaiensis* embryos

For fixation, *Parhyale* embryos of the desired age were transferred to welled glass slides and submerged in 5% formaldehyde in filtered 3% sea salt water. Embryos were dissected

manually and individually at room temperature using tungsten wire needles and binoculars. The needles were made by inserting a 3 cm piece of tungsten thread (0.125 mm diameter) into the tip of a Pasteur pipette, sealing the composition by melting the borosilicate over a Bunsen burner and sharpening the tungsten tip by holding it into the burner flame for 2 min. In order to allow thorough permeation of embryonic tissue by the fixative, the embryo's egg shell and, when required, the embryonic cuticle (stages 16 and later) were removed. According to their age, incubation time was varied. Embryos aged between the onset of gastrulation until the completion of segmentation (stage 6-15) were incubated in the fixative for 30 min. Subsequently, the egg shell was easily removed. Embryos of stages 16-20 were incubated for 1h, punctured by a tungsten needle and repeatedly dragged through the liquid surface in order to strip off egg shell and embryonic cuticle. Embryos of stage 21 and later were incubated for 2 h in the fixative. The embryo was then everted out of egg shell and cuticle from the dorsal side of the embryo, using two tungsten needles. After dissection, fixative and debris were removed. Fresh 7.5% formaldehyde in 3% sea salt water was added and the dissected embryos were incubated for another 1 h. The glass slides holding the embryos were covered in order to prevent drying-out. After the second fixation, embryos were put into PT and gradually, over several 10 min washings, transferred into methanol (50% PT/50% methanol, 25% PT/75% methanol, 10% PT/ 90%methanol, 5% PT/ 95 % methanol). Embryos were stored in methanol at -20°C until further use.

### 5.4.3 Whole-mount *in situ* hybridization (WMISH)

Whole-mount *in situ* hybridization (WMISH) was performed as reported in Finnerty et al. (2003b), with modifications as described in Browne et al. (2006). Digoxigenin- or Fluorescein- ("Fluorescein RNA Labeling Mix", Roche Applied Science) labelled RNA probes were produced using the "DIG RNA Labeling Kit (SP6/T7)" (Roche Applied Science), according to the manufacturer's protocol. DNA templates were created via PCR (5.2.3). It was seen to it that RNA probes were at least 500bp in length. In general, full-length ORF or, if available, cDNA plasmids of the genes of interest were preferred as templates. Digoxigenin-labelled RNA probes were used in a final concentration of 1 ng/µl. Fluorescein-labelled RNA probes were used in a final concentration of 3 ng/µl.

Finalizing the WMISH procedure, the embryos were subjected to colour reactions until specific signal arose or until weak unspecific staining became visible. After that, the embryos were treated as described in Browne et al. (2006). In detail, the visualization of WMISH signals via BCIP/NBT and Fast Red colour reactions was customized as follows: during the first three days, colour reactions were done at room temperature with the BCIP/NBT or Fast Red reaction solutions being exchanged and renewed every 2 h. Overnight, the colour reactions were done at 4°C without renewing the reaction solutions. If required, the colour reactions were continued over the subsequent four weeks (room temperature during day and 4°C over night), renewing the reaction solutions once per day. After that, the reactions were stored at 4°C and the reaction solutions were renewed only if needed.

WMISH was performed on embryos of stages S6 to S23. Embryos of later stages than S23 were not suitable for WMISH due to their fully developed cuticle.

### 5.4.4 WMISH targeting two different transcripts (double *in situ*)

WMISH targeting two different transcripts in one embryo (double *in situ*) were performed as standard WMISH (MM), with the following modifications: (i) During hybridization steps, both respective RNA probes were used at the same time, one Digoxigenin- and the other Fluorescein-labelled. (ii) After the Digoxigenin-labelled RNA probe was visualized, the Digoxigenin-specific antibody was inactivated through the following steps: well stained embryos were washed three times for 5 min with PT at RT; then, embryos were incubated at 65°C with SDS-Hyb for 15 min and, after that, again with fresh SDS-Hyb for 60 min at 65°C; finally, embryos were washed at RT three times for 5 min with PT and 4 times for 15 min with PT. (iii) After inactivation of the Digoxigenin-specific antibody, the Fluorescein-labelled RNA probe was visualized analogously as the Digoxigenin-labelled probe, using specifications for Fluorescein labelling (specific antibody, reaction buffer and FastRed substrate), following the manufacturer's instructions.

### 5.4.5 Nuclear labelling and mounting of *Parhyale hawaiensis* embryos

Prior to microscopy, *Parhyale* embryos that had been subjected to WMISH were incubated with 1:2000 SYTOX® Green Nucleic Acid Stain (invitrogen, Darmstadt, Germany) in pre-mounting medium (50% glycerol, 50% PT) for at least 12 h in order to ensure thorough penetration of the embryo's nuclei with SYTOX®. Afterwards, they were incubated for 4 h minimum in straight pre-mounting medium to remove unspecific SYTOX® signal. Single embryos were then transferred in mounting medium (70% glycerol, 30% PT). Until further use, they were stored individually in mounting medium at 4°C.

For microscopy, individual embryos of choice were transferred onto a custom-built glass slide, submerged in a small droplet of mounting medium, set up in a desired orientation and covered with a cover slip. By shifting the cover slip, the orientation of the embryo could be adjusted or altered as required. Mounted WMISH embryos with nuclear SYTOX® labelling were then photographed using the inverse microscope Observer Z1 (Carl Zeiss MicroImaging, Germany, 5.6).

The custom-built glass slide was created by fixing stacks of 1-3 cover slips onto two sides of the glass slide, leaving a gap between them. Within this gap, the embryo was mounted. This set-up allowed unfolding the embryo as well as determining and repeatedly adjusting its orientation without flattening and damaging it.

### 5.4.6 Injection of *Parhyale hawaiensis* embryos

*Parhyale hawaiensis* embryos were injected at 18°C. Embryos of the desired stage were lined up on the linear niche of a step-shaped agar block. Excessive water was removed by using tissue paper, retaining a fine film of water to cover the embryos. This procedure strongly decreased the chance of yolk leakage directly following injection. Because *Parhyale* embryos exhibited a high sensitivity to drying, they were never exposed more than 5 minutes to air-dry, water-free conditions. Also, a maximum of ten embryos was lined up and used for injection. Embryonic injections were performed under the microscope Axiostar (Carl Zeiss AG, Germany) using Femtotips® (Eppendorf, Hamburg, Germany) mounted to a micromanipulator set at an angle of 10°. A Femtojet ® microinjector (Eppendorf, Hamburg,

Germany) was used as an injection device. The micromanipulator was held in a fixed position. Therefore, injections were carried out moving the microscope plate.

Injection and threshold pressure parameters were adjusted anew prior to every set of injections in order to ensure an injection volume of 50 pl, independent of properties of the individual needles as well as the varying viscosity of different injection mixes. To this end, the injection needle tip was submerged in Voltalef H10S (Arkema GmbH, Düsseldorf, Germany) and the Femtojet® parameters were adjusted as needed. As a general guide line, injection pressure was initially set to 2500 hPa and threshold pressure to 1800 hPa.

### 5.4.7 Injection of *Parhyale hawaiensis* adult females

*Parhyale* females were sedated with 0.1% clove oil (*Syzygium aromaticum*) in 3% sea salt water and transferred to a furrow that was shaped using plasticine on a welled glass slide. For injections, females were arranged in a ventral posture. Excessive water was removed using tissue paper. The females were punctured laterally into the soft tissue that connects the tergites of segments T4 and T5, using individually pulled and cut quartz needles (P-2000 micropipette puller, Sutter Instrument Company, Novato, USA). The injection mix was released into the space of the dorsal tube where the female's ovaries are situated. After injection, females were immediately transferred to 6-well plates containing fresh filtered 3% sea salt water. Their constitution was documented on consecutive days post-injection.

## 5.5 Genetics

### 5.5.1 Transgenesis of *Parhyale hawaiensis* embryos

For stable transgenesis, S1 or S2 *Parhyale* embryos were injected (5.4.6) with a mix of capped transposase mRNA, purified plasmid carrying the transgene construct of choice and nuclease-free water as reported in Pavlopoulos and Averof (2005). Capped transposase mRNA was created using the "mMESSAGE mMACHINE® High Yield Capped RNA Transcription Kit" (Applied Biosystems/Ambion, Austin, USA), following the manufacturer's instructions. Typically, the concentrations of transposase mRNA were set to 100 ng/µl and of the plasmid construct to 50-200 ng/µl. Hatched embryos were screened for marker gene expression.

### 5.5.2 Conducting siRNA-mediated RNAi in *Parhyale hawaiensis*

In this work, siRNA-mediated RNAi was approached for *Ph six3* and *Ph kni1*. To this end, three individual STEALTH siRNA duplex molecules (Invitrogen, Darmstadt, Germany) were designed for each gene, using the "BLOCK-iT™ RNAi Designer" platform (https://rnaidesigner.invitrogen.com/rnaiexpress/rnaiExpress.jsp), following the provider's instructions and recommendations. To circumvent off-target effects that could arise from affecting transcripts with high sequence similarities, only non-conserved parts of the *Ph six3* and *Ph kni1* cDNA sequence were used as basis for the generation of siRNA duplex molecules (A4.3, table A4).

To mediate RNAi, *Parhyale* embryos were injected with solutions containing single siRNA molecule species (100 µM) or combinations of all three species targeting the same gene (final concentrations 100 µM and 300 µM, corresponding to 33 µM and 100 µM each, respectively). For addressing behavioural effects in hatched animals, embryos were also injected with all siRNA molecule species targeting both *Ph six3* and *Ph kni1* (final concentration 600 µM, corresponding to 100µM each). Degradation of the targeted mRNA after injection of siRNA was addressed via WMISH (5.4.3).

### 5.5.3 Monitoring of injected embryos

Injected embryos were immediately transferred into individual wells of a sterile 96-well plate containing 350 µl of filtered 3% sea salt water. The embryos were arranged according to desired parameters (e.g. batch, age, injection components etc.). This arrangement was documented. In cases where a fluorescent tracer was used, embryos were checked 2 h after injection at room temperature for presence, intensity and localisation of the injected reagents as well as for immediate rupture and damage resulting from mechanical strain. Subsequently, they were incubated at 28°C. In all other cases, the plates containing the injected embryos were directly transferred to 28°C. The development of all injected embryos was monitored and documented on consecutive days, with the categories being (i) wild-type development, (ii) aberrant development and (iii) lethality. If considered appropriate, salt water was exchanged. Embryos that had reached the desired age (e.g. for dissection) were removed and taken to further use. Hatched embryos were transferred into 24 well plates and provided with fresh filtered 3% sea salt water.

### 5.5.4 Analysing pereopod phenotypes of hatched *Parhyale* after siRNA-mediated RNAi targeting *Ph kni1*

Injected animals were allowed to hatch. Preparative of phenotype examination, 3-day-old adult animals were checked for vitality. In detail, they were examined for heart beat, movement, coordinated locomotion and muscle reaction to stimulus to rule out unspecific appendage phenotypes caused by muscle paralysis, tissue degradation or death.

Vital animals with functional appendage musculature were sedated with 0.1% clove oil (*Syzygium aromaticum*) in 3% sea salt water (see also 5.1). Sedated animals were set up ventrally and laterally. The arrangement of individual pereopods relative to the body axis as well as to the other segmental appendages was examined under the epifluorescence stereomicroscope MZ16FA (Leica Microsystems, Germany, 5.6), using indirect illumination by a cold light source and no filter, and compared to the wild-type arrangement. In wild-type animals, the appendages of segments Mn, Mx1, Mx2 and T1-T3 are not visible from dorsal view. The pereopods of T4 and T5 are bent anterior, those of T6-T8 are bent posterior (amphipodal arrangement, 2.5.3). The pereopods of T7 are visible as the longest ones. In

lateral view, the pereopods of T7 and T8 conclude an angle of about 60°. Also, both T7 and T8 pereopods are bent dorsally. Apart from their anterior positions, visible in dorsal view, the first antennae can be easily distinguished by their slight inflection in contrast to the second antenna, which project straight in sedated animals.

## 5.6 Microscopy

Images of embryos that had been subjected to WMISH were taken using the inverse microscope Observer Z1 (Carl Zeiss MicroImaging, Germany). For standard images, the x10 air objective was used. Detailed images were usually taken with the x20 air objective. To visualise the 3D-morphology, the complete WMISH signal with high contrast and the position of the SYTOX®-labelled nuclei of any given embryo, DIC, bright field and fluorescence images (mercury-vapor lamp and FITC filter set 44; excitation 455-495 nm, emission 505-555 nm, beam splitter FT500) were taken, respectively.

Injected embryos were checked for rupture and presence of injection tracer dye using the epifluorescence stereomicroscope MZ16FA (Leica Microsystems, Germany). This microscope was also used for examining adult *Parhyale* morphology and for expression of transgenesis markers (EGFP-LP filters, excitation 405-425 nm, emission > 460 nm).

# 6 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.

Acampora, D., Avantaggiato, V., Tuorto, F., Simeone, A., 1997. Genetic control of brain morphogenesis through Otx gene dosage requirement. Development 124, 3639-3650.

Acampora, D., Gulisano, M., Simeone, A., 2000a. Genetic and molecular roles of Otx homeodomain proteins in head development. Gene 246, 23-35.

Acampora, D., Postiglione, M.P., Avantaggiato, V., Di Bonito, M., Simeone, A., 2000b. The role of Otx and Otp genes in brain development. Int J Dev Biol 44, 669-677.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.

Angelini, D.R., Kikuchi, M., Jockusch, E.L., 2009. Genetic patterning in the adult capitate antenna of the beetle Tribolium castaneum. Dev Biol 327, 240-251.

Arendt, D., Nübler-Jung, K., 1994. Inversion of dorsoventral axis? Nature 371, 26.

Arendt, D., Nübler-Jung, K., 1999. Comparison of early nerve cord development in insects and vertebrates. Development 126, 2309-2325.

Arnosti, D.N., Gray, S., Barolo, S., Zhou, J., Levine, M., 1996. The gap protein knirps mediates both quenching and direct repression in the Drosophila embryo. Embo J 15, 3659-3666.

Averof, M., Cohen, S.M., 1997. Evolutionary origin of insect wings from ancestral gills. Nature 385, 627-630.

Averof, M., Patel, N.H., 1997. Crustacean appendage evolution associated with changes in Hox gene expression. Nature 388, 682-686.

Bally-Cuif, L., Boncinelli, E., 1997. Transcription factors and head formation in vertebrates. Bioessays 19, 127-135.

Bebenek, I.G., Gates, R.D., Morris, J., Hartenstein, V., Jacobs, D.K., 2004. sine oculis in basal Metazoa. Dev Genes Evol 214, 342-351.

Beermann, A., Jay, D.G., Beeman, R.W., Hulskamp, M., Tautz, D., Jurgens, G., 2001. The Short antennae gene of Tribolium is required for limb development and encodes the orthologue of the Drosophila Distal-less protein. Development 128, 287-297.

Beermann, A., Schröder, R., 2004. Functional stability of the aristaless gene in appendage tip formation during evolution. Dev Genes Evol 214, 303-308.

Belles, X., 2010. Beyond Drosophila: RNAi in vivo and functional genomics in insects. Annu Rev Entomol 55, 111-128.

Ben-David, J., Chipman, A.D., 2010. Mutual regulatory interactions of the trunk gap genes during blastoderm patterning in the hemipteran Oncopeltus fasciatus. Dev Biol 346, 140-149.

Biffis, C., Alwes, F., Scholtz, G., 2009. Cleavage and gastrulation of the dendrobranchiate shrimp Penaeus monodon (Crustacea, Malacostraca, Decapoda). Arthropod Struct Dev 38, 527-540.

Boyer, J.C., Bebenek, K., Kunkel, T.A., 1996. Analyzing the fidelity of reverse transcription and transcription. Methods Enzymol 275, 523-537.

Brand-Saberi, B., Christ, B., 1999. Genetic and epigenetic control of muscle development in vertebrates. Cell Tissue Res 296, 199-212.

Brown, A.E., Crisanti, A., Catteruccia, F., 2003. Comparative analysis of DNA vectors at mediating RNAi in Anopheles mosquito cells and larvae. J Exp Biol 206, 1817-1823.

Browne, W.E., Price, A.L., Gerberding, M., Patel, N.H., 2005. Stages of embryonic development in the amphipod crustacean, Parhyale hawaiensis. Genesis 42, 124-149.

Browne, W.E., Schmid, B.G., Wimmer, E.A., Martindale, M.Q., 2006. Expression of otd orthologs in the amphipod crustacean, Parhyale hawaiensis. Dev Genes Evol 216, 581-595.

Bucher, G., Farzana, L., Brown, S.J., Klingler, M., 2005. Anterior localization of maternal mRNAs in a short germ insect lacking bicoid. Evol Dev 7, 142-149.

Bucher, G., Scholten, J., Klingler, M., 2002. Parental RNAi in Tribolium (Coleoptera). Curr Biol 12, R85-86.

Budd, G.E., 2008. The earliest fossil record of the animals and its significance. Philos Trans R Soc Lond B Biol Sci 363, 1425-1434.

Burke, R.D., Angerer, L.M., Elphick, M.R., Humphrey, G.W., Yaguchi, S., Kiyama, T., Liang, S., Mu, X., Agca, C., Klein, W.H., Brandhorst, B.P., Rowe, M., Wilson, K., Churcher, A.M., Taylor, J.S., Chen, N., Murray, G., Wang, D., Mellott, D., Olinski, R., Hallbook, F., Thorndyke, M.C., 2006. A genomic view of the sea urchin nervous system. Dev Biol 300, 434-460.

Burroughs-Garcia, J., Sittaramane, V., Chandrasekhar, A., Waters, S.T., 2011. Evolutionarily conserved function of Gbx2 in anterior hindbrain development. Dev Dyn 240, 828-838.

Burton, P.M., 2008. Insights from diploblasts; the evolution of mesoderm and muscle. J Exp Zool B Mol Dev Evol 310, 5-14.

Campbell, G., Weaver, T., Tomlinson, A., 1993. Axis specification in the developing Drosophila appendage: the role of wingless, decapentaplegic, and the homeobox gene aristaless. Cell 74, 1113-1123.

Casillas, S., Negre, B., Barbadilla, A., Ruiz, A., 2006. Fast sequence evolution of Hox and Hox-derived genes in the genus Drosophila. BMC Evol Biol 6, 106.

Catteruccia, F., Levashina, E.A., 2009. RNAi in the malaria vector, Anopheles gambiae. Methods Mol Biol 555, 63-75.

Cerny, A.C., Grossmann, D., Bucher, G., Klingler, M., 2008. The Tribolium ortholog of knirps and knirps-related is crucial for head segmentation but plays a minor role during abdominal patterning. Dev Biol 321, 284-294.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., Prasher, D.C., 1994. Green Fluorescent Protein as a Marker for Gene-Expression. Science 263, 802-805.

Chen, C.K., Kuhnlein, R.P., Eulenberg, K.G., Vincent, S., Affolter, M., Schuh, R., 1998. The transcription factors KNIRPS and KNIRPS RELATED control cell migration and branch morphogenesis during Drosophila tracheal development. Development 125, 4959-4968.

Chen, S., Wang, Q.L., Nie, Z., Sun, H., Lennon, G., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Zack, D.J., 1997. Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. Neuron 19, 1017-1030.

Chiang, C., Young, K.E., Beachy, P.A., 1995. Control of Drosophila tracheal branching by the novel homeodomain gene unplugged, a regulatory target for genes of the bithorax complex. Development 121, 3901-3912.

Clark, I.B., Boyd, J., Hamilton, G., Finnegan, D.J., Jarman, A.P., 2006. D-six4 plays a key role in patterning cell identities deriving from the Drosophila mesoderm. Dev Biol 294, 220-231.

Cobos, I., Broccoli, V., Rubenstein, J.L., 2005. The vertebrate ortholog of Aristaless is regulated by Dlx genes in the developing forebrain. J Comp Neurol 483, 292-303.

Cohen, C., 2010. Darwin on woman. C R Biol 333, 157-165.

Cohen, S.M., Bronner, G., Kuttner, F., Jürgens, G., Jäckle, H., 1989. Distal-less encodes a homoeodomain protein required for limb development in Drosophila. Nature 338, 432-434.

Cohen, S.M., Jürgens, G., 1990. Mediation of Drosophila head development by gap-like segmentation genes. Nature 346, 482-485.

Colas, J.F., Schoenwolf, G.C., 2001. Towards a cellular and molecular understanding of neurulation. Dev Dyn 221, 117-145.

Copf, T., Schroder, R., Averof, M., 2004. Ancestral role of caudal genes in axis elongation and segmentation. Proc Natl Acad Sci U S A 101, 17711-17715.

Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16, 10881-10890.

Crozatier, M., Valle, D., Dubois, L., Ibnsouda, S., Vincent, A., 1999. Head versus trunk patterning in the Drosophila embryo; collier requirement for formation of the intercalary segment. Development 126, 4385-4394.

Cui, X., Doe, C.Q., 1995. The role of the cell cycle and cytokinesis in regulating neuroblast sublineage gene expression in the Drosophila CNS. Development 121, 3233-3243.

Curtiss, J., Heilig, J.S., 1995. Establishment of Drosophila imaginal precursor cells is controlled by the Arrowhead gene. Development 121, 3819-3828.

Curtiss, J., Heilig, J.S., 1997. Arrowhead encodes a LIM homeodomain protein that distinguishes subsets of Drosophila imaginal cells. Dev Biol 190, 129-141.

Damen, W.G., Hausdorf, M., Seyfarth, E.A., Tautz, D., 1998. A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. Proc Natl Acad Sci U S A 95, 10665-10670.

Damen, W.G., Saridaki, T., Averof, M., 2002. Diverse adaptations of an ancestral gill: a common evolutionary origin for wings, breathing organs, and spinnerets. Curr Biol 12, 1711-1716.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.

Davis, R.J., Tavsanli, B.C., Dittrich, C., Walldorf, U., Mardon, G., 2003. Drosophila retinal homeobox (drx) is not required for establishment of the visual system, but is required for brain and clypeus development. Dev Biol 259, 272-287.

De Robertis, E.M., 2008. Evo-devo: variations on ancestral themes. Cell 132, 185-195.

De Robertis, E.M., Sasai, Y., 1996. A common plan for dorsoventral patterning in Bilateria. Nature 380, 37-40.

de Rosa, R., Prud'homme, B., Balavoine, G., 2005. Caudal and even-skipped in the annelid Platynereis dumerilii and the ancestry of posterior growth. Evol Dev 7, 574-587.

Diederich, R.J., Merrill, V.K., Pultz, M.A., Kaufman, T.C., 1989. Isolation, structure, and expression of labial, a homeotic gene of the Antennapedia Complex involved in Drosophila head development. Genes Dev 3, 399-414.

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.

Domazet-Loso, T., Tautz, D., 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature 468, 815-818.

Donnell, D.M., Strand, M.R., 2006. Caste-based differences in gene expression in the polyembryonic wasp Copidosoma floridanum. Insect Biochem Mol Biol 36, 141-153.

Doyle, H.J., Harding, K., Hoey, T., Levine, M., 1986. Transcripts encoded by a homoeo box gene are restricted to dorsal tissues of Drosophila embryos. Nature 323, 76-79.

Driever, W., Nüsslein-Volhard, C., 1988. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. Cell 54, 95-104.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sorensen, M.V., Haddock, S.H., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745-749.

Edgecombe, G.D., 2010. Arthropod phylogeny: an overview from the perspectives of morphology, molecular data and the fossil record. Arthropod Struct Dev 39, 74-87.

El-Hodiri, H.M., Qi, X.L., Seufert, D.W., 2003. The Xenopus arx gene is expressed in the developing rostral forebrain. Dev Genes Evol 212, 608-612.

Elgar, S.J., Han, J., Taylor, M.V., 2008. mef2 activity levels differentially affect gene expression during Drosophila muscle development. Proc Natl Acad Sci U S A 105, 918-923.

Erwin, D.H., Davidson, E.H., 2002. The last common bilaterian ancestor. Development 129, 3021-3032.

Extavour, C.G., 2005. The fate of isolated blastomeres with respect to germ cell formation in the amphipod crustacean Parhyale hawaiensis. Dev Biol 277, 387-402.

Finkelstein, R., Boncinelli, E., 1994. From fly head to mammalian forebrain: the story of otd and Otx. Trends Genet 10, 310-315.

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, T.M., Spradling, A.C., 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, J.R., 2003. The origins of axial patterning in the metazoa: how old is bilateral symmetry? Int J Dev Biol 47, 523-529.

Finnerty, J.R., Paulson, D., Burton, P., Pang, K., Martindale, M.Q., 2003. Early evolution of a homeobox gene: the parahox gene Gsx in the Cnidaria and the Bilateria. Evol Dev 5, 331-345.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531-1545.

Franch-Marro, X., Martin, N., Averof, M., Casanova, J., 2006. Association of tracheal placodes with leg primordia in Drosophila and implications for the origin of insect tracheal systems. Development 133, 785-790.

Fröbius, A.C., Seaver, E.C., 2006. Capitella sp. I homeobrain-like, the first lophotrochozoan member of a novel paired-like homeobox gene family. Gene Expr Patterns 6, 985-991.

Fuss, B., Meissner, T., Bauer, R., Lehmann, C., Eckardt, F., Hoch, M., 2001. Control of endoreduplication domains in the Drosophila gut by the knirps and knirps-related genes. Mech Dev 100, 15-23.

Galliot, B., de Vargas, C., Miller, D., 1999. Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class. Dev Genes Evol 209, 186-197.

Gallitano-Mendel, A., Finkelstein, R., 1997. Novel segment polarity gene interactions during embryonic head development in Drosophila. Dev Biol 192, 599-613.

Gantier, M.P., Williams, B.R., 2007. The response of mammalian cells to double-stranded RNA. Cytokine Growth Factor Rev 18, 363-371.

Gayon, J., 2010. Sexual selection: Another Darwinian process. C R Biol 333, 134-144.

Gerberding, M., Browne, W.E., Patel, N.H., 2002. Cell lineage analysis of the amphipod crustacean Parhyale hawaiensis reveals an early restriction of cell fates. Development 129, 5789-5801.

Ghanbari, H., Seo, H.C., Fjose, A., Brandli, A.W., 2001. Molecular cloning and embryonic expression of Xenopus Six homeobox genes. Mech Dev 101, 271-277.

Gilbert, S.F., 2006. Developmental Biology Eighth Edition ed. Sinauer Associates, Inc., Sunderland, Massachusetts.

Giribet, G., 2002. Current advances in the phylogenetic reconstruction of metazoan evolution. A new paradigm for the Cambrian explosion? Mol Phylogenet Evol 24, 345-357.

Glenner, H., Thomsen, P.F., Hebsgaard, M.B., Sorensen, M.V., Willerslev, E., 2006. Evolution. The origin of insects. Science 314, 1883-1884.

Gonzalez-Gaitan, M., Rothe, M., Wimmer, E.A., Taubert, H., Jackle, H., 1994. Redundant functions of the genes knirps and knirps-related for the establishment of anterior Drosophila head structures. Proc Natl Acad Sci U S A 91, 8567-8571.

Goriely, A., Mollereau, B., Coffinier, C., Desplan, C., 1999. Munster, a novel paired-class homeobox gene specifically expressed in the Drosophila larval eye. Mech Dev 88, 107-110.

Gould, S.J., 1977. Ontogeny and Phylogeny. Harvard University Press.

Graham, A., Begbie, J., McGonnell, I., 2004. Significance of the cranial neural crest. Dev Dyn 229, 5-13.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., Brown, J.B., Cherbas, L., Davis, C.A., Dobin, A., Li, R., Lin, W., Malone, J.H., Mattiuzzo, N.R., Miller, D., Sturgill, D., Tuch, B.B., Zaleski, C., Zhang, D., Blanchette, M., Dudoit, S., Eads, B., Green, R.E., Hammonds, A., Jiang, L., Kapranov, P., Langton, L., Perrimon, N., Sandler, J.E., Wan, K.H., Willingham, A., Zhang, Y., Zou, Y., Andrews, J., Bickel, P.J., Brenner, S.E., Brent, M.R., Cherbas, P., Gingeras, T.R., Hoskins, R.A., Kaufman, T.C., Oliver, B., Celniker, S.E., 2011. The developmental transcriptome of Drosophila melanogaster. Nature 471, 473-479.

Grbic, M., 2003. Polyembryony in parasitic wasps: evolution of a novel mode of development. Int J Dev Biol 47, 633-642.

Grifone, R., Demignon, J., Houbron, C., Souil, E., Niro, C., Seller, M.J., Hamard, G., Maire, P., 2005. Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. Development 132, 2235-2249.

Grigoriou, M., Tucker, A.S., Sharpe, P.T., Pachnis, V., 1998. Expression and regulation of Lhx6 and Lhx7, a novel subfamily of LIM homeodomain encoding genes, suggests a role in mammalian head development. Development 125, 2063-2074.

Grossniklaus, U., Cadigan, K.M., Gehring, W.J., 1994. Three maternal coordinate systems cooperate in the patterning of the Drosophila head. Development 120, 3155-3171.

Hahn, M., Jäckle, H., 1996. Drosophila goosecoid participates in neural development but not in body axis formation. Embo J 15, 3077-3084.

Hammond, S.M., 2005. Dicing and slicing: the core machinery of the RNA interference pathway. FEBS Lett 579, 5822-5829.

Hanken, J., Gross, J.B., 2005. Evolution of cranial development and the role of neural crest: insights from amphibians. J Anat 207, 437-446.

Hanson, I.M., 2001. Mammalian homologues of the Drosophila eye specification genes. Semin Cell Dev Biol 12, 475-484.

Harzsch, S., 2004. The tritocerebrum of Euarthropoda: a "non-drosophilocentric" perspective. Evol Dev 6, 303-309.

Havemann, J., Muller, U., Berger, J., Schwarz, H., Gerberding, M., Moussian, B., 2008. Cuticle differentiation in the embryo of the amphipod crustacean Parhyale hawaiensis. Cell Tissue Res 332, 359-370.

Heidenreich, K.A., Linseman, D.A., 2004. Myocyte enhancer factor-2 transcription factors in neuronal differentiation and survival. Mol Neurobiol 29, 155-166.

Hertzler, P.L., Wang, S.W., Clark, W.H., Jr., 1994. Mesendoderm cell and archenteron formation in isolated blastomeres from the shrimp Sicyonia ingentis. Dev Biol 164, 333-344.

Hirth, F., 2010. On the origin and evolution of the tripartite brain. Brain Behav Evol 76, 3-10.

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.

Hobert, O., Westphal, H., 2000. Functions of LIM-homeobox genes. Trends Genet 16, 75-83.

Holland, P.W., Takahashi, T., 2005. The evolution of homeobox genes: Implications for the study of brain development. Brain Res Bull 66, 484-490.

Horn, C., Schmid, B.G., Pogoda, F.S., Wimmer, E.A., 2002. Fluorescent transformation markers for insect transgenesis. Insect Biochem Mol Biol 32, 1221-1235.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Hughes, C.L., Kaufman, T.C., 2002. Hox genes and the evolution of the arthropod body plan. Evol Dev 4, 459-499.

Hughes, R.N., D'Amato, M.E., Bishop, J.D., Carvalho, G.R., Craig, S.F., Hansson, L.J., Harley, M.A., Pemberton, A.J., 2005. Paradoxical polyembryony? Embryonic cloning in an ancient order of marine bryozoans. Biol Lett 1, 178-180.

Irish, V., Lehmann, R., Akam, M., 1989. The Drosophila posterior-group gene nanos functions by repressing hunchback activity. Nature 338, 646-648.

Jaeger, J., 2011. The gap gene network. Cell Mol Life Sci 68, 243-274.

Jouve, C., Iimura, T., Pourquie, O., 2002. Onset of the segmentation clock in the chick embryo: evidence for oscillations in the somite precursors in the primitive streak. Development 129, 1107-1117.

Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M., Tomancak, P., 2010. Gene expression divergence recapitulates the developmental hourglass model. Nature 468, 811-814.

Kammermeier, L., Reichert, H., 2001. Common developmental genetic mechanisms for patterning invertebrate and vertebrate brains. Brain Res Bull 55, 675-682.

Karpala, A.J., Doran, T.J., Bean, A.G., 2005. Immune responses to dsRNA: implications for gene silencing technologies. Immunol Cell Biol 83, 211-216.

Kawakami, K., Sato, S., Ozaki, H., Ikeda, K., 2000. Six family genes--structure and function as transcription factors and their roles in development. Bioessays 22, 616-626.

Kennerdell, J.R., Carthew, R.W., 2000. Heritable gene silencing in Drosophila using double-stranded RNA. Nat Biotechnol 18, 896-898.

Kenyon, K.L., Yang-Zhou, D., Cai, C.Q., Tran, S., Clouser, C., Decene, G., Ranade, S., Pignoni, F., 2005. Partner specificity is essential for proper function of the SIX-type homeodomain proteins Sine oculis and Optix during fly eye development. Dev Biol 286, 158-168.

Kimmel, C.B., 1996. Was Urbilateria segmented? Trends in Genetics 12, 329-331.

Kirby, R.J., Hamilton, G.M., Finnegan, D.J., Johnson, K.J., Jarman, A.P., 2001. Drosophila homolog of the myotonic dystrophy-associated gene, SIX5, is required for muscle and gonad development. Curr Biol 11, 1044-1049.

Klingler, M., 1990. The organization of the antero-posterior axis. Semin Cell Biol 1, 151-160.

Knecht, A.K., Bronner-Fraser, M., 2002. Induction of the neural crest: A multigene process. Nature Reviews Genetics 3, 453-461.

Koenemann, S., Jenner, R.A., Hoenemann, M., Stemme, T., von Reumont, B.M., 2010. Arthropod phylogeny revisited, with a focus on crustacean relationships. Arthropod Struct Dev 39, 88-110.

Kotkamp, K., Klingler, M., Schoppmeier, M., 2010. Apparent role of Tribolium orthodenticle in anteroposterior blastoderm patterning largely reflects novel functions in dorsoventral axis formation and cell survival. Development 137, 1853-1862.

Kozmik, Z., 2008. The role of Pax genes in eye evolution. Brain Res Bull 75, 335-339.

Kurokawa, D., Kiyonari, H., Nakayama, R., Kimura-Yoshida, C., Matsuo, I., Aizawa, S., 2004a. Regulation of Otx2 expression and its functions in mouse forebrain and midbrain. Development 131, 3319-3331.

Kurokawa, D., Takasaki, N., Kiyonari, H., Nakayama, R., Kimura-Yoshida, C., Matsuo, I., Aizawa, S., 2004b. Regulation of Otx2 expression and its functions in mouse epiblast and anterior neuroectoderm. Development 131, 3307-3317.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Laudet, V., Hanni, C., Coll, J., Catzeflis, F., Stehelin, D., 1992. Evolution of the nuclear receptor gene superfamily. Embo J 11, 1003-1013.

Lavado, A., Lagutin, O.V., Oliver, G., 2008. Six3 inactivation causes progressive caudalization and aberrant patterning of the mammalian diencephalon. Development 135, 441-450.

Lawrence, P.A., Morata, G., 1994. Homeobox genes: their function in Drosophila segmentation and pattern formation. Cell 78, 181-189.

Lecuit, T., 2004. Junctions and vesicular trafficking during Drosophila cellularization. J Cell Sci 117, 3427-3433.

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

Liubicich, D.M., Serano, J.M., Pavlopoulos, A., Kontarakis, Z., Protas, M.E., Kwan, E., Chatterjee, S., Tran, K.D., Averof, M., Patel, N.H., 2009. Knockdown of Parhyale

Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. Proc Natl Acad Sci U S A 106, 13892-13896.

Loehle, C., 1995. Social Barriers to Pathogen Transmission in Wild Animal Populations. Ecology 76, 326-335.

Lunde, K., Biehs, B., Nauber, U., Bier, E., 1998. The knirps and knirps-related genes organize development of the second wing vein in Drosophila. Development 125, 4145-4154.

Lynch, J., Desplan, C., 2003. 'De-evolution' of Drosophila toward a more generic mode of axis patterning. Int J Dev Biol 47, 497-503.

Lynch, J.A., Brent, A.E., Leaf, D.S., Pultz, M.A., Desplan, C., 2006. Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp Nasonia. Nature 439, 728-732.

Maire, P., Niro, C., Demignon, J., Vincent, S., Liu, Y.B., Giordani, J., Sgarioto, N., Favier, M., Guillet-Deniau, I., Blais, A., 2010. Six1 and Six4 gene expression is necessary to activate the fast-type muscle gene program in the mouse primary myotome. Dev Biol 338, 168-182.

Mallatt, J., Giribet, G., 2006. Further use of nearly complete, 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. Mol Phylogenet Evol 40, 772-794.

Manabe, T., Tatsumi, K., Inoue, M., Makinodan, M., Yamauchi, T., Makinodan, E., Yokoyama, S., Sakumura, R., Wanaka, A., 2007. L3/Lhx8 is a pivotal factor for cholinergic differentiation of murine embryonic stem cells. Cell Death Differ 14, 1080-1085.

Manabe, T., Tatsumi, K., Inoue, M., Matsuyoshi, H., Makinodan, M., Yamauchi, T., Makinodan, E., Yokoyama, S., Sakumura, R., Okuda, H., Wanaka, A., 2008. Knockdown of the L3/Lhx8 gene suppresses cholinergic differentiation of murine embryonic stem cell-derived spheres. Int J Dev Neurosci 26, 249-252.

Martindale, M.Q., Pang, K., Finnerty, J.R., 2004. Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). Development 131, 2463-2474.

Mazza, M.E., Pang, K., Reitzel, A.M., Martindale, M.Q., Finnerty, J.R., 2010. A conserved cluster of three PRD-class homeobox genes (homeobrain, rx and orthopedia) in the Cnidaria and Protostomia. Evodevo 1, 3.

McGinnis, W., Krumlauf, R., 1992. Homeobox genes and axial patterning. Cell 68, 283-302.

McGregor, A.P., 2005. How to get ahead: the origin, evolution and function of bicoid. Bioessays 27, 904-913.

McGregor, A.P., Hilbrant, M., Pechmann, M., Schwager, E.E., Prpic, N.M., Damen, W.G., 2008. Cupiennius salei and Achaearanea tepidariorum: Spider models for investigating evolution and development. Bioessays 30, 487-498.

McManus, C.J., Graveley, B.R., 2011. RNA structure and the mechanisms of alternative splicing. Curr Opin Genet Dev.

Meijlink, F., Beverdam, A., Brouwer, A., Oosterveen, T.C., Berge, D.T., 1999. Vertebrate aristaless-related genes. Int J Dev Biol 43, 651-663.

Melkman, T., Sengupta, P., 2005. Regulation of chemosensory and GABAergic motor neuron development by the C. elegans Aristaless/Arx homolog alr-1. Development 132, 1935-1949.

Merrill, V.K., Diederich, R.J., Turner, F.R., Kaufman, T.C., 1989. A genetic and developmental analysis of mutations in labial, a gene necessary for proper head formation in Drosophila melanogaster. Dev Biol 135, 376-391.

Miller, D.J., Ball, E.E., 2009. The gene complement of the ancestral bilaterian - was Urbilateria a monster? J Biol 8, 89.

Millet, S., Campbell, K., Epstein, D.J., Losos, K., Harris, E., Joyner, A.L., 1999. A role for Gbx2 in repression of Otx2 and positioning the mid/hindbrain organizer. Nature 401, 161-164.

Minelli, A., 2001. A three-phase model of arthropod segmentation. Dev Genes Evol 211, 509-521.

Miura, H., Yanazawa, M., Kato, K., Kitamura, K., 1997. Expression of a novel aristaless related homeobox gene 'Arx' in the vertebrate telencephalon, diencephalon and floor plate. Mech Dev 65, 99-109.

Miyawaki, K., Inoue, Y., Mito, T., Fujimoto, T., Matsushima, K., Shinmyo, Y., Ohuchi, H., Noji, S., 2002. Expression patterns of aristaless in developing appendages of Gryllus bimaculatus (cricket). Mech Dev 113, 181-184.

Morris, S.C., 2000. Evolution: bringing molecules into the fold. Cell 100, 1-11.

Myat, M.M., Lightfoot, H., Wang, P., Andrew, D.J., 2005. A molecular link between FGF and Dpp signaling in branch-specific migration of the Drosophila trachea. Dev Biol 281, 38-52.

Nauber, U., Pankratz, M.J., Kienlin, A., Seifert, E., Klemm, U., Jackle, H., 1988. Abdominal segmentation of the Drosophila embryo requires a hormone receptor-like protein encoded by the gap gene knirps. Nature 336, 489-492.

Negre, B., Casillas, S., Suzanne, M., Sanchez-Herrero, E., Akam, M., Nefedov, M., Barbadilla, A., de Jong, P., Ruiz, A., 2005. Conservation of regulatory sequences and gene expression patterns in the disintegrating Drosophila Hox gene complex. Genome Res 15, 692-700.

Noll, M., 1993. Evolution and role of Pax genes. Curr Opin Genet Dev 3, 595-605.

Nüsslein-Volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature 287, 795-801.

Oliver, G., Mailhos, A., Wehr, R., Copeland, N.G., Jenkins, N.A., Gruss, P., 1995. Six3, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. Development 121, 4045-4055.

Olson, E.N., Perry, M., Schulz, R.A., 1995. Regulation of muscle differentiation by the MEF2 family of MADS box transcription factors. Dev Biol 172, 2-14.

Oro, A.E., Ong, E.S., Margolis, J.S., Posakony, J.W., McKeown, M., Evans, R.M., 1988. The Drosophila gene knirps-related is a member of the steroid-receptor gene superfamily. Nature 336, 493-496.

Otto, S.P., Yong, P., 2002. The evolution of gene duplicates. Adv Genet 46, 451-483.

Ozhan-Kizil, G., Havemann, J., Gerberding, M., 2009. Germ cells in the crustacean Parhyale hawaiensis depend on Vasa protein for their maintenance but not for their formation. Dev Biol 327, 230-239.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40, 1413-1415.

Panganiban, G., Rubenstein, J.L., 2002. Developmental functions of the Distal-less/Dlx homeobox genes. Development 129, 4371-4386.

Papillon, D., Telford, M.J., 2007. Evolution of Hox3 and ftz in arthropods: insights from the crustacean Daphnia pulex. Dev Genes Evol 217, 315-322.

Parchem, R.J., Poulin, F., Stuart, A.B., Amemiya, C.T., Patel, N.H., 2010. BAC library for the amphipod crustacean, Parhyale hawaiensis. Genomics 95, 261-267.

Passalacqua, K.D., Hrycaj, S., Mahfooz, N., Popadic, A., 2010. Evolving expression patterns of the homeotic gene Scr in insects. Int J Dev Biol 54, 897-904.

Pavlopoulos, A., Averof, M., 2005. Establishing genetic transformation for comparative developmental studies in the crustacean Parhyale hawaiensis. Proc Natl Acad Sci U S A 102, 7888-7893.

Pavlopoulos, A., Kontarakis, Z., Liubicich, D.M., Serano, J.M., Akam, M., Patel, N.H., Averof, M., 2009. Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. Proc Natl Acad Sci U S A 106, 13897-13902.

Peel, A.D., 2008. The evolution of developmental gene networks: lessons from comparative studies on holometabolous insects. Philos Trans R Soc Lond B Biol Sci 363, 1539-1547.

Porcher, A., Dostatni, N., 2010. The bicoid morphogen system. Curr Biol 20, R249-254.

Posnien, N., 2009. Function and Evolution of highly conserved head genes in the red flour beetle Tribolium castaneum, Department of Developmental Biology. Georg-August-University, Göttingen.

Posnien, N., Koniszewski, N., Bucher, G., 2011. Insect Tc-six4 marks a unit with similarity to vertebrate placodes. Dev Biol 350, 208-216.

Posnien, N., Schinko, J.B., Kittelmann, S., Bucher, G., 2010. Genetics, development and composition of the insect head--a beetle's view. Arthropod Struct Dev 39, 399-410.

Pourquie, O., 2001. Vertebrate somitogenesis. Annu Rev Cell Dev Biol 17, 311-350.

Presgraves, D.C., 2010. Darwin and the origin of interspecific genetic incompatibilities. Am Nat 176 Suppl 1, S45-60.

Price, A.L., Modrell, M.S., Hannibal, R.L., Patel, N.H., 2010. Mesoderm and ectoderm lineages in the crustacean Parhyale hawaiensis display intra-germ layer compensation. Dev Biol 341, 256-266.

Price, A.L., Patel, N.H., 2008. Investigating divergent mechanisms of mesoderm development in arthropods: the expression of Ph-twist and Ph-mef2 in Parhyale hawaiensis. J Exp Zool B Mol Dev Evol 310, 24-40.

Prpic, N.M., Telford, M.J., 2008. Expression of homothorax and extradenticle mRNA in the legs of the crustacean Parhyale hawaiensis: evidence for a reversal of gene expression regulation in the pancrustacean lineage. Dev Genes Evol 218, 333-339.

Raff, R.A., 2000. Evo-devo: the evolution of a new discipline. Nat Rev Genet 1, 74-79.

Raible, F., Brand, M., 2004. Divide et Impera--the midbrain-hindbrain boundary and its organizer. Trends Neurosci 27, 727-734.

Ramani, A.K., Calarco, J.A., Pan, Q., Mavandadi, S., Wang, Y., Nelson, A.C., Lee, L.J., Morris, Q., Blencowe, B.J., Zhen, M., Fraser, A.G., 2011. Genome-wide analysis of alternative splicing in Caenorhabditis elegans. Genome Res 21, 342-348.

Ranade, S.S., Yang-Zhou, D., Kong, S.W., McDonald, E.C., Cook, T.A., Pignoni, F., 2008. Analysis of the Otd-dependent transcriptome supports the evolutionary conservation of CRX/OTX/OTD functions in flies and vertebrates. Dev Biol 315, 521-534.

Regier, J.C., Shultz, J.W., Ganley, A.R., Hussey, A., Shi, D., Ball, B., Zwick, A., Stajich, J.E., Cummings, M.P., Martin, J.W., Cunningham, C.W., 2008. Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Syst Biol 57, 920-938.

Regier, J.C., Shultz, J.W., Kambic, R.E., 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc Biol Sci 272, 395-401.

Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463, 1079-1083.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., Patel, N.H., 2009a. Antibody staining of Parhyale hawaiensis embryos. Cold Spring Harb Protoc 2009, pdb prot5129.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., Patel, N.H., 2009b. The crustacean Parhyale hawaiensis: a new model for arthropod development. Cold Spring Harb Protoc 2009, pdb emo114.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., Patel, N.H., 2009c. In situ hybridization of labeled RNA probes to fixed Parhyale hawaiensis embryos. Cold Spring Harb Protoc 2009, pdb prot5130.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., Patel, N.H., 2009d. Injection of Parhyale hawaiensis blastomeres with fluorescently labeled tracers. Cold Spring Harb Protoc 2009, pdb prot5128.

Reichert, H., Simeone, A., 2001. Developmental genetic evidence for a monophyletic origin of the bilaterian brain. Philos Trans R Soc Lond B Biol Sci 356, 1533-1544.

Remane, A., Storch, V., Welsch, U., 1975. Systematische Zoologie. Gustav Fischer Verlag, Stuttgart.

Rida, P.C., Le Minh, N., Jiang, Y.J., 2004. A Notch feeling of somite segmentation and beyond. Dev Biol 265, 2-22.

Rieger, S., Wang, F., Sagasti, A., 2011. Time-lapse imaging of neural development: Zebrafish lead the way into the fourth dimension. Genesis.

Rogers, B.T., Kaufman, T.C., 1996. Structure of the insect head as revealed by the EN protein pattern in developing embryos. Development 122, 3419-3432.

Rogers, B.T., Kaufman, T.C., 1997. Structure of the insect head in ontogeny and phylogeny: a view from Drosophila. Int Rev Cytol 174, 1-84.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Ronquist, F., Huelsenbeck, J.P., van der Mark, P., 2005. MrBayes 3.1 Manual, Draft 5/26/2005.

Rothe, M., Nauber, U., Jackle, H., 1989. Three hormone receptor-like Drosophila genes encode an identical DNA-binding finger. Embo J 8, 3087-3094.

Rothe, M., Pehl, M., Taubert, H., Jackle, H., 1992. Loss of gene function through rapid mitotic cycles in the Drosophila embryo. Nature 359, 156-159.

Rothe, M., Wimmer, E.A., Pankratz, M.J., Gonzalez-Gaitan, M., Jackle, H., 1994. Identical transacting factor requirement for knirps and knirps-related Gene expression in the anterior but not in the posterior region of the Drosophila embryo. Mech Dev 46, 169-181.

Rubenstein, J.L., Martinez, S., Shimamura, K., Puelles, L., 1994. The embryonic vertebrate forebrain: the prosomeric model. Science 266, 578-580.

Rushlow, C., Frasch, M., Doyle, H., Levine, M., 1987. Maternal regulation of zerknüllt: a homoeobox gene controlling differentiation of dorsal tissues in Drosophila. Nature 330, 583-586.

Ryan, J.F., Burton, P.M., Mazza, M.E., Kwong, G.K., Mullikin, J.C., Finnerty, J.R., 2006. The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, Nematostella vectensis. Genome Biol 7, R64.

Sambrook, J., MacCallum, P., Russell, D., 2000. Molecular Cloning: A Laboratory Manual, Third Edition ed.

Sander, K., 1976. Specification of the basic body pattern in insect embryogenesis, Adv. Insect Physiol, p. 125−138

Sander, K., 1983. The evolution of patterning mechanisms: gleanings from insect embryogenesis and spermatogenesis, in: Goodwin, B.H., N.; Wylie, C. (Ed.), Development & Evolution. Cambridge Univ. Press, UK, p. 137−161.

Schaeper, N.D., Prpic, N.M., Wimmer, E.A., 2010. A clustered set of three Sp-family genes is ancestral in the Metazoa: evidence from sequence analysis, protein domain structure, developmental expression patterns and chromosomal location. BMC Evol Biol 10, 88.

Schetelig, M.F., Schmid, B.G., Zimowska, G., Wimmer, E.A., 2008. Plasticity in mRNA expression and localization of orthodenticle within higher Diptera. Evol Dev 10, 700-704.

Schilling, T.F., 1997. Genetic analysis of craniofacial development in the vertebrate embryo. Bioessays 19, 459-468.

Schilling, T.F., Knight, R.D., 2001. Origins of anteroposterior patterning and Hox gene regulation during chordate evolution. Philos T Roy Soc B 356, 1599-1613.

Schinko, J.B., Kreuzer, N., Offen, N., Posnien, N., Wimmer, E.A., Bucher, G., 2008. Divergent functions of orthodenticle, empty spiracles and buttonhead in early head patterning of the beetle Tribolium castaneum (Coleoptera). Dev Biol 317, 600-613.

Schlosser, G., Ahrens, K., 2004. Molecular anatomy of placode development in Xenopus laevis. Dev Biol 271, 439-466.

Schneitz, K., Spielmann, P., Noll, M., 1993. Molecular genetics of Aristaless, a prd-type homeo box gene involved in the morphogenesis of proximal and distal pattern elements in a subset of appendages in Drosophila. Genes Dev 7, 911.

Scholtz, G., Dohle, W., 1996. Cell lineage and cell fate in crustacean embryos - A comparative approach. International Journal of Developmental Biology 40, 211-220.

Scholtz, G., Edgecombe, G.D., 2006. The evolution of arthropod heads: reconciling morphological, developmental and palaeontological evidence. Dev Genes Evol 216, 395-415.

Schoppmeier, M., Fischer, S., Schmitt-Engel, C., Lohr, U., Klingler, M., 2009. An ancient anterior patterning system promotes caudal repression and head formation in ecdysozoa. Curr Biol 19, 1811-1815.

Schröder, R., 2003. The genes orthodenticle and hunchback substitute for bicoid in the beetle Tribolium. Nature 422, 621-625.

Schröder, R., Beermann, A., Wittkopp, N., Lutz, R., 2008. From development to biodiversity--Tribolium castaneum, an insect model organism for short germband development. Dev Genes Evol 218, 119-126.

Seo, H.C., Curtiss, J., Mlodzik, M., Fjose, A., 1999. Six class homeobox genes in drosophila belong to three distinct families and are involved in head development. Mech Dev 83, 127-139.

Seo, H.C., Drivenes, O., Fjose, A., 1998. A zebrafish Six4 homologue with early expression in head mesoderm. Biochim Biophys Acta 1442, 427-431.

Shimamura, K., Hartigan, D.J., Martinez, S., Puelles, L., Rubenstein, J.L., 1995. Longitudinal organization of the anterior neural plate and neural tube. Development 121, 3923-3933.

Shoubridge, C., Fullston, T., Gecz, J., 2010. ARX spectrum disorders: making inroads into the molecular pathology. Hum Mutat 31, 889-900.

Shultz, J.W., Regier, J.C., 2000. Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean plus hexapod clade. P Roy Soc B-Biol Sci 267, 1011-1019.

Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A., Boncinelli, E., 1992. Nested expression domains of four homeobox genes in developing rostral brain. Nature 358, 687-690.

Simeone, A., D'Apice, M.R., Nigro, V., Casanova, J., Graziani, F., Acampora, D., Avantaggiato, V., 1994. Orthopedia, a novel homeobox-containing gene expressed in the developing CNS of both mouse and Drosophila. Neuron 13, 83-101.

Simeone, A., Puelles, E., Acampora, D., 2002. The Otx family. Curr Opin Genet Dev 12, 409-415.

Slack, J.M., Holland, P.W., Graham, C.F., 1993. The zootype and the phylotypic stage. Nature 361, 490-492.

Smith, J.L., Schoenwolf, G.C., 1997. Neurulation: coming to closure. Trends Neurosci 20, 510-517.

St Johnston, D., Nüsslein-Volhard, C., 1992. The origin of pattern and polarity in the Drosophila embryo. Cell 68, 201-219.

Stauber, M., Jackle, H., Schmidt-Ott, U., 1999. The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene. Proc Natl Acad Sci U S A 96, 3786-3789.

Steinmetz, P.R., Kostyuchenko, R.P., Fischer, A., Arendt, D., 2011. The segmental pattern of otx, gbx, and Hox genes in the annelid Platynereis dumerilii. Evol Dev 13, 72-79.

Steinmetz, P.R., Urbach, R., Posnien, N., Eriksson, J., Kostyuchenko, R.P., Brena, C., Guy, K., Akam, M., Bucher, G., Arendt, D., 2010. Six3 demarcates the anterior-most developing brain region in bilaterian animals. Evodevo 1, 14.

Stierwald, M., Yanze, N., Bamert, R.P., Kammermeier, L., Schmid, V., 2004. The Sine oculis/Six class family of homeobox genes in jellyfish with and without eyes: development and eye regeneration. Dev Biol 274, 70-81.

Strausfeld, N.J., 2005. The evolution of crustacean and insect optic lobes and the origins of chiasmata. Arthropod Struct Dev 34, 235-256.

Strausfeld, N.J., 2009. Brain organization and the origin of insects: an assessment. Proc Biol Sci 276, 1929-1937.

Streit, A., 2007. The preplacodal region: an ectodermal domain with multipotential progenitors that contribute to sense organs and cranial sensory ganglia. Int J Dev Biol 51, 447-461.

Sunmonu, N.A., Li, K., Li, J.Y., 2011. Numerous isoforms of Fgf8 reflect its multiple roles in the developing brain. J Cell Physiol 226, 1722-1726.

Tautz, D., 2004. Segmentation. Dev Cell 7, 301-312.

Telford, M.J., Thomas, R.H., 1995. Systematics - Demise of the Atelocerata. Nature 376, 123-124.

Telford, M.J., Thomas, R.H., 1998. Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. Proc Natl Acad Sci U S A 95, 10671-10675.

Timmermans, M.J., Roelofs, D., Marien, J., van Straalen, N.M., 2008. Revealing pancrustacean relationships: phylogenetic analysis of ribosomal protein genes places Collembola (springtails) in a monophyletic Hexapoda and reinforces the discrepancy between mitochondrial and nuclear DNA markers. BMC Evol Biol 8, 83.

Tomoyasu, Y., Denell, R.E., 2004. Larval RNAi in Tribolium (Coleoptera) for analyzing adult development. Dev Genes Evol 214, 575-578.

Tomoyasu, Y., Miller, S.C., Tomita, S., Schoppmeier, M., Grossmann, D., Bucher, G., 2008. Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in Tribolium. Genome Biol 9, R10.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E., Desplan, C., 1989. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell 59, 553-562.

Treisman, J., Harris, E., Wilson, D., Desplan, C., 1992. The homeodomain: a new face for the helix-turn-helix? Bioessays 14, 145-150.

Tuschl, T., 2001. RNA interference and small interfering RNAs. Chembiochem 2, 239-245.

Ungerer, P., Scholtz, G., 2008. Filling the gap between identified neuroblasts and neurons in crustaceans adds new support for Tetraconata. Proc Biol Sci 275, 369-376.

Urbach, R., 2007. A procephalic territory in Drosophila exhibiting similarities and dissimilarities compared to the vertebrate midbrain/hindbrain boundary region. Neural Dev 2, 23.

Urbach, R., Technau, G.M., 2003. Molecular markers for identified neuroblasts in the developing brain of Drosophila. Development 130, 3621-3637.

Valentine, J.W., 2006. Ancestors and urbilateria. Evol Dev 8, 391-393.

van der Zee, M., Berns, N., Roth, S., 2005. Distinct functions of the Tribolium zerknullt genes in serosa specification and dorsal closure. Curr Biol 15, 624-636.

Vieira, C., Pombero, A., Garcia-Lopez, R., Gimeno, L., Echevarria, D., Martinez, S., 2010. Molecular mechanisms controlling brain development: an overview of neuroepithelial secondary organizers. Int J Dev Biol 54, 7-20.

Vincent, A., Blankenship, J.T., Wieschaus, E., 1997. Integration of the head and trunk segmentation systems controls cephalic furrow formation in Drosophila. Development 124, 3747-3754.

W., G., 1922. "ontogeny does not recapitulate phylogeny, it creates it".

Walldorf, U., Kiewe, A., Wickert, M., Ronshaugen, M., McGinnis, W., 2000. Homeobrain, a novel paired-like homeobox gene is expressed in the Drosophila brain. Mech Dev 96, 141-144.

West-Eberhard, M.J., 2005. Developmental plasticity and the origin of species differences. Proc Natl Acad Sci U S A 102 Suppl 1, 6543-6549.

Williams, T.A., 1994. The Nauplius Larva of Crustaceans: Functional Diversity and the Phylotypic Stage. Integrative and Comparative Biology 34, 562-569.

Wimmer, E.A., Cohen, S.M., Jackle, H., Desplan, C., 1997. buttonhead does not contribute to a combinatorial code proposed for Drosophila head development. Development 124, 1509-1517.

Wimmer, E.A., Jackle, H., Pfeifle, C., Cohen, S.M., 1993. A Drosophila homologue of human Sp1 is a head-specific segmentation gene. Nature 366, 690-694.

Wolpert, L., Beddington, R., Jessell, T., Lawrence, P., Meyerowitz, E., J., S., 2002. Principles of Development, second edition ed. Oxford University Press.

Wurst, W., Bally-Cuif, L., 2001. Neural plate patterning: upstream and downstream of the isthmic organizer. Nat Rev Neurosci 2, 99-108.

Xiong, W.C., Okano, H., Patel, N.H., Blendy, J.A., Montell, C., 1994. repo encodes a glial-specific homeo domain protein required in the Drosophila nervous system. Genes Dev 8, 981-994.

Xu, J., Tan, A., Palli, S.R., 2010. The function of nuclear receptors in regulation of female reproduction and embryogenesis in the red flour beetle, Tribolium castaneum. J Insect Physiol 56, 1471-1480.

Zhang, Y., Miki, T., Iwanaga, T., Koseki, Y., Okuno, M., Sunaga, Y., Ozaki, N., Yano, H., Koseki, H., Seino, S., 2002. Identification, tissue expression, and functional characterization of Otx3, a novel member of the Otx family. J Biol Chem 277, 28065-28069.

Zhao, Y., Marin, O., Hermesz, E., Powell, A., Flames, N., Palkovits, M., Rubenstein, J.L., Westphal, H., 2003. The LIM-homeobox gene Lhx8 is required for the development of many cholinergic neurons in the mouse forebrain. Proc Natl Acad Sci U S A 100, 9005-9010.

# A Appendix

All paragraphs of the appendix are enclosed electronically on compact disc (CD) as supplement. The paragraphs A2-A6 are found exclusively on the enclosed CD. The corresponding directory structure of A2-A6 will be given in the following.

## A1 Published data

Parts of the data of this dissertation were published in "Expression of *otd* orthologs in the amphipod crustacean, *Parhyale hawaiensis*" as the result of a collaborative approach by William E. Browne, Bernhard G. M. Schmid, Ernst A. Wimmer and Mark Q. Martindale (3.2.1.5). The complete publication and its supplements follow.

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
© Springer-Verlag 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

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

## 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

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 Pre-amplification 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).

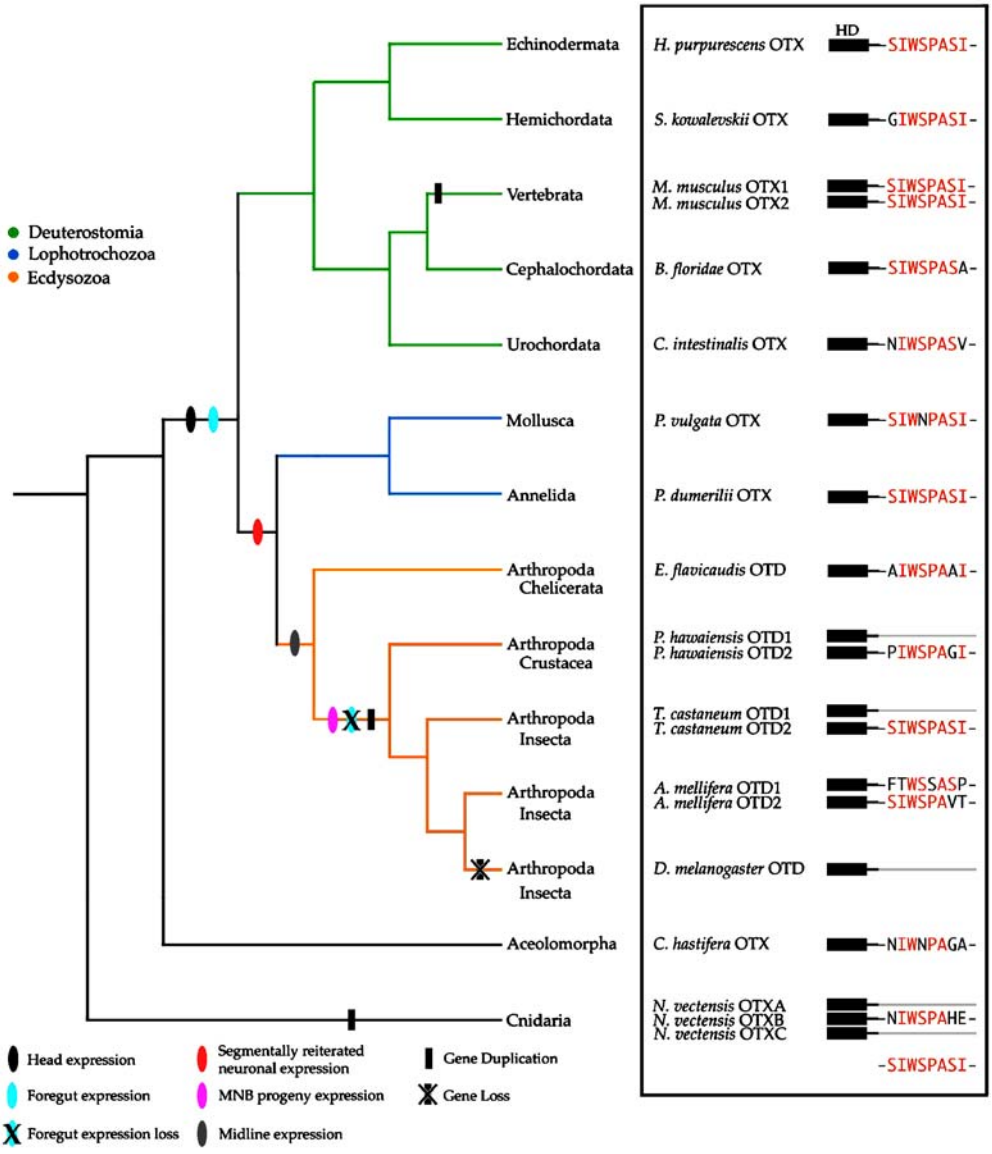



**Fig. 1** A model proposing evolutionary duplication and loss of *orthodenticle* paralogs in metazoans. Character legend at the *left bottom* of figure. *Green* associated with deuterostome lineages, *blue* associated with lophotrochozoan lineages, *orange* associated with ecdysozoan lineages. Cladogram (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

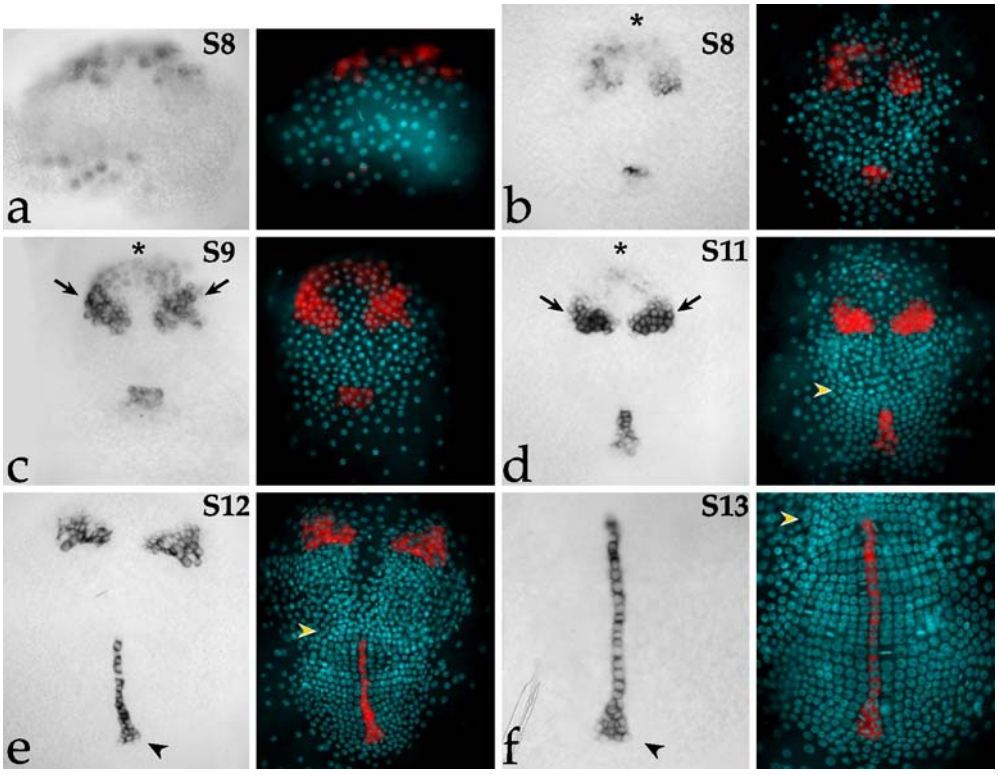



**Fig. 2** Expression of *P. hawaiensis otd1*, gastrulation through germband formation. All embryos are mounted ventral side up with the anterior end oriented towards the top. For each panel, a brightfield image is on the *left* (*Photd1* label is *black*), and a Hoechst label of the same embryo is on 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).

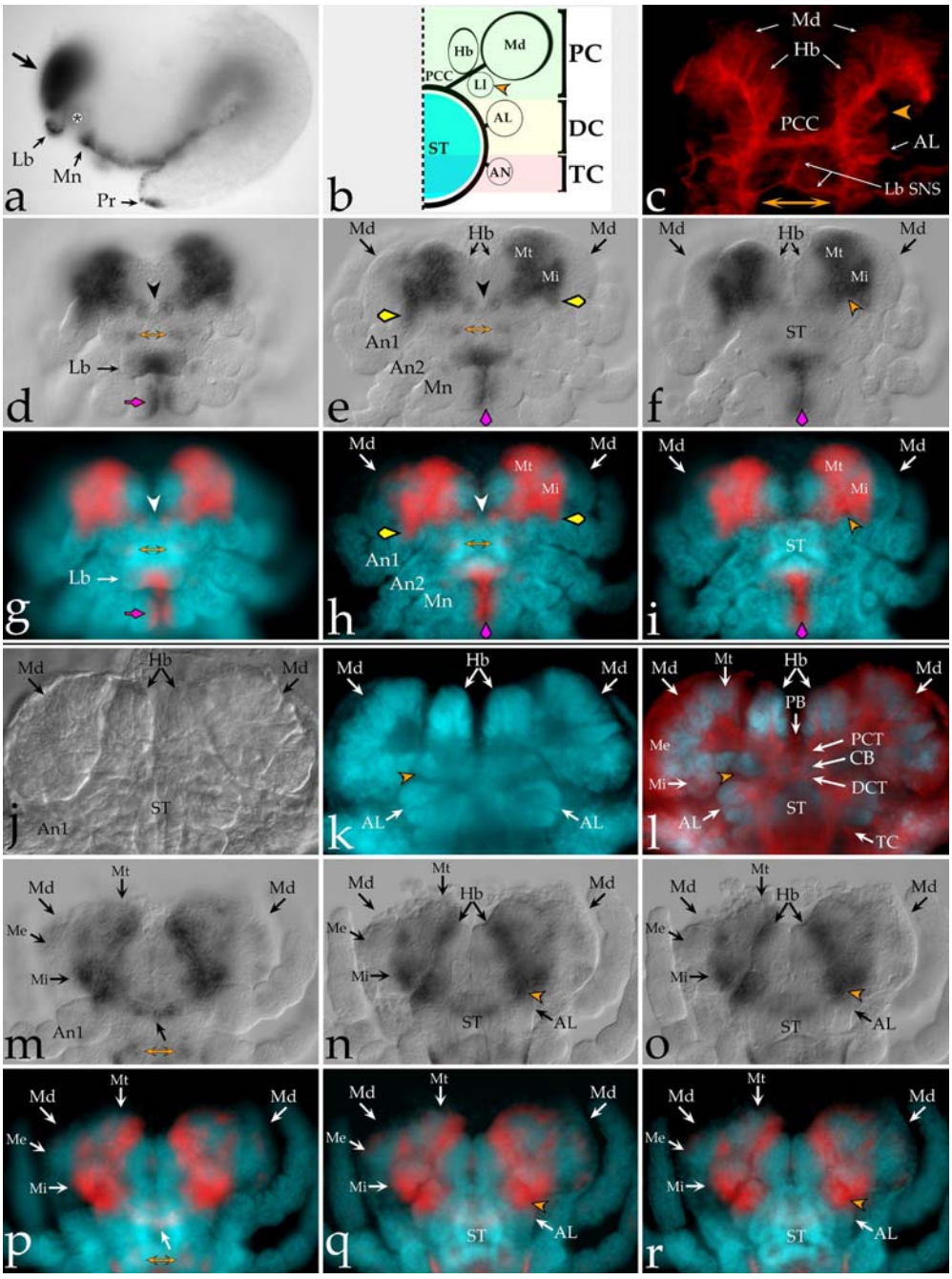
By stage 21, segmentation is complete and the stomodeum is visible medially in the anterior region of the An2 segment (Fig. 3f,i). Expression of *Photd1* appears strongly in the anterior head ectoderm and developing brain, medially in the labrum, and *Photd1* expression in the ventral midline has expanded 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

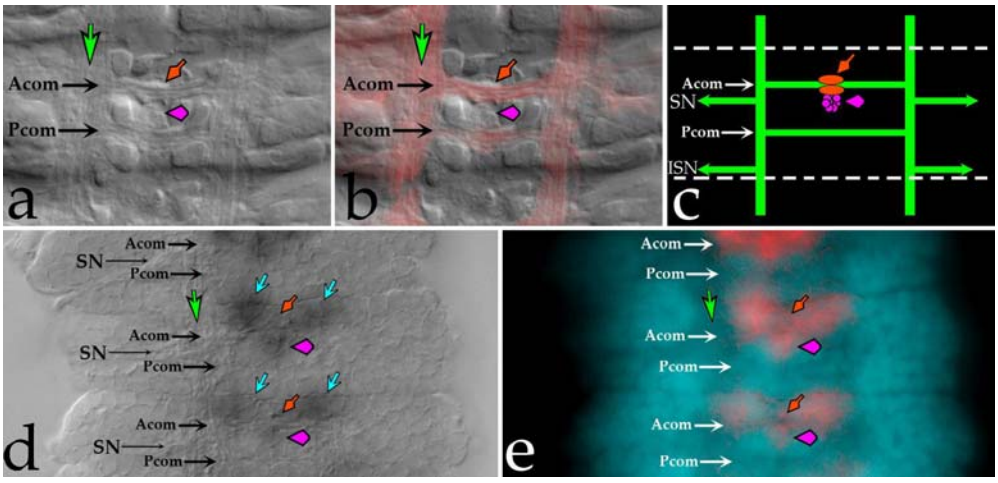



**Fig. 4** Ventral nerve cord expression of *P. hawaiensis otd1*. All panels are dorsal views bisecting a stage 23 embryonic VNC and are oriented anterior towards the top. **a**–**c** Nomarski (**a**), matching planes for nomarski and acetylated tubulin (**b**, acetylated tubulin in *red*), and a VNC 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

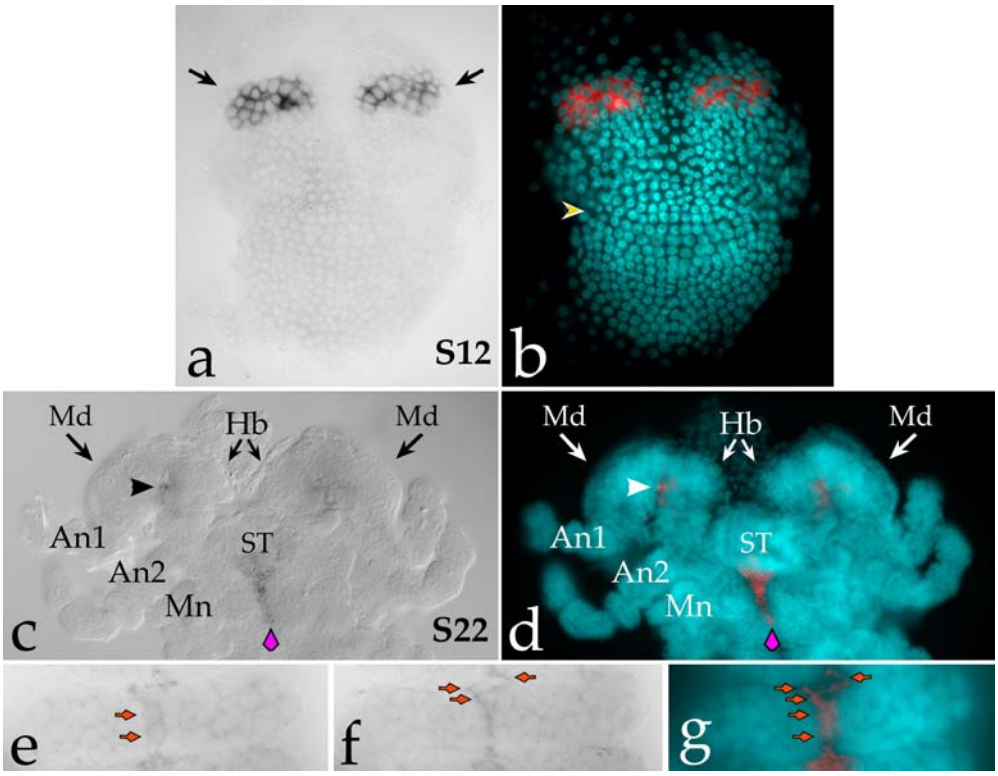



**Fig. 5** Expression of *P. hawaiensis otd2*. All embryos are mounted with the anterior end oriented towards the top. **a**, **b** A stage 12 embryo is mounted ventral side up with a nomarski image on the left and *Photd2* signal in *black*. The image on the *right* is overlay of matching focal planes 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 manifestion 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

# Supplement to *Parhyale hawaiensis otd2*

## Supplementary Figure A

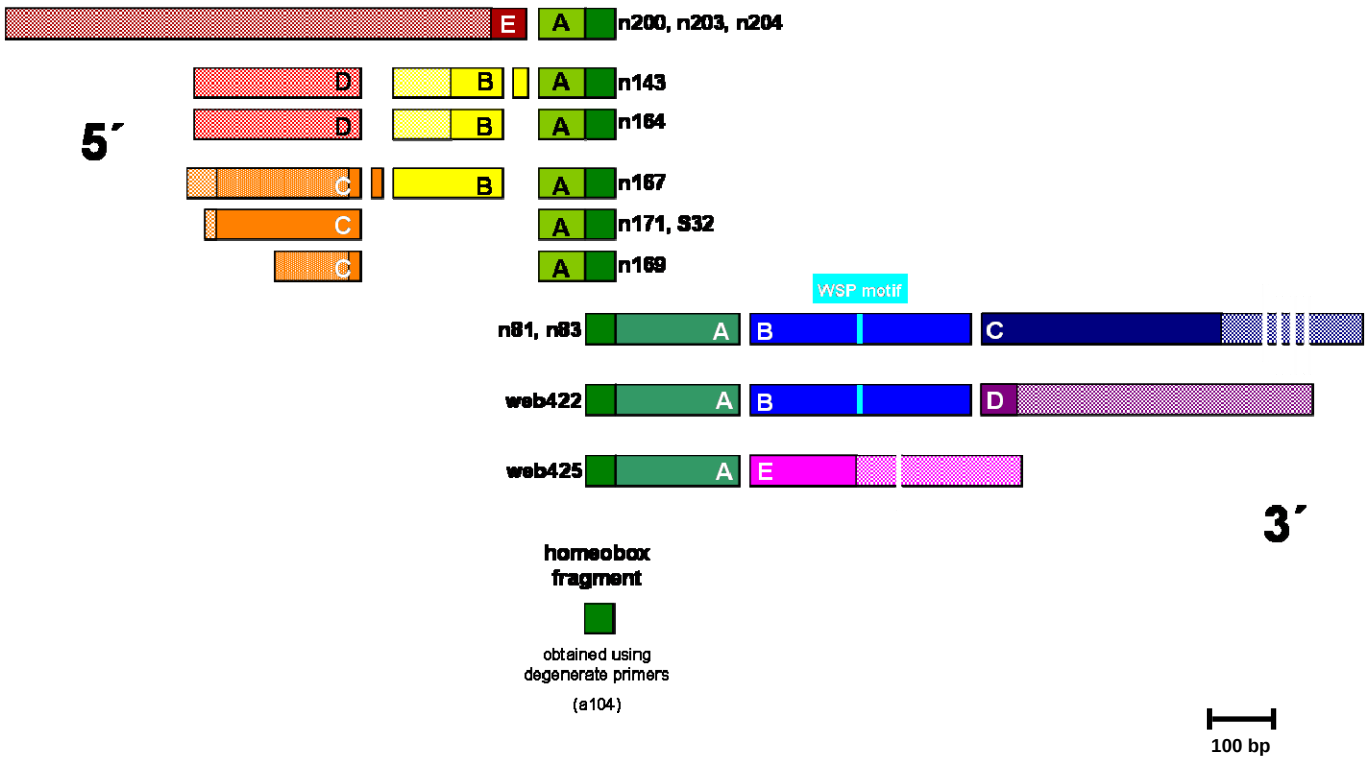



**Figure A. Schematic overview over 5' and 3' RACE clones of *Parhyale hawaiensis otd2* and their putative splicing structure.** A schematic alignment of the nucleic acid sequences of 5' RACE clones (top of panel), 3' RACE clones (middle of panel) and the short *otd*2 homeobox fragment obtained by PCR 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      HHHTTPNNNNNNNNSNSSNSSTTLTSPLTTSHGARDASPPSPPPSVSSGSS
web422   HHHTTPNNNNNNNNSNSSNSSTTLTSPLTTSHGARDASPPSPPPSVSSGSS
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  gattatttaacagcgactgaaaatttccagtgtgaaaatacatataaatatgagcagtgt  537

538  agggcagtgactttttaacattaggtctttgattgtactgtgcagataggtatgcggttg  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  tcagcagcagcagccgctggggtctatctcaaaacccccgcccccctacaccATGAACGGA  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  gggggcagttgcggggctgggggtcagcagcagcagccgctggggtctatctcaaaacc  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                       gaattccagcaagtttaATGCCGTGGGAGTCACATCTA   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  gcccgcttactcttgaagagtttctgtaaaactattgcgatttttccagtaatattagc  1140

1141  ttatgactgttcaaatattatcgaacagtaattaatgtggctgatattcaatacgaataa  1200

1201  tttttgtagatgaacttgggaagcaacacagtgtaacctacacgatcacggaggacagtt  1260

1261  ttccatatggtaggctaaatttggaaatctaacgcggatctgggatacgtcacgtggaac  1320

1321  acgagggaaattcattagccttagccatatataacgtccagggggtctaatagggaacgt  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 aaaatcagttttctttttcttcatcagcagcgactggtttccttctttgatgtttctgat 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  agaatgaaatttttggctgtagaacaactattttggaacagagggggtataaaggaatgg  1080

1081  cacgtggacctcattgcattacatttattgcaaagttgactatggagataaactctgaaa  1140

1141  cgtttttcgtctaagaaaaaaaaaaaaaaaaa                              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 cacagtaagcaccacggaaaacataaggagaaagagggggaaagtaaaagacaaagaaggt   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.

# A2-A6 Directory structure of the enclosed electronic appendix

| | |
|---|---|
| **A2** | **PhD_A2_Parhyale_goi_sequences (folder)** |
| A2.1 | PhD_A2.1_Ph_otd1 (folder) |
| *A2.1.1* | *PhD_A2.1.1_Ph_otd1_sequences (folder)* |
| *A2.1.2* | *PhD_A2.1.2_Ph_otd1_alignments (folder)* |
| A2.2 | PhD_A2.2_Ph_otd2 (folder) |
| A2.3 | PhD_A2.3_Ph_hbn (folder) |
| *A2.3.1* | *PhD_A2.3.1_Ph_hbn_sequences (folder)* |
| *A2.3.2* | *PhD_A2.3.2_Ph_hbn_alignments (folder)* |
| A2.4 | PhD_A2.4_Ph_al1 (folder) |
| *A2.4.1* | *PhD_A2.4.1_Ph_al1_sequences (folder)* |
| *A2.4.2* | *PhD_A2.4.2_Ph_al1_alignments (folder)* |
| A2.5 | PhD_A2.5_Ph_al2 (folder) |
| *A2.5.1* | *PhD_A2.5.1_Ph_al2_sequences (folder)* |
| *A2.5.2* | *PhD_A2.5.2_Ph_al2_alignments (folder)* |
| A2.6 | PhD_A2.6_Ph_pby1_pby2 (folder) |
| *A2.6.1* | *PhD_A2.6.1_Ph_pby1_pby2_sequences (folder)* |
| *A2.6.2* | *PhD_A2.6.2_Ph_pby1_pby2_alignments (folder)* |
| A2.7 | PhD_A2.7_Ph_awh (folder) |
| *A2.7.1* | *PhD_A2.7.1_Ph_awh_sequences (folder)* |
| *A2.7.2* | *PhD_A2.7.2_Ph_awh_alignments (folder)* |
| A2.8 | PhD_A2.8_Ph_six3 (folder) |
| *A2.8.1* | *PhD_A2.8.1_Ph_six3_sequences (folder)* |
| *A2.8.2* | *PhD_A2.8.2_Ph_six3_alignments (folder)* |
| A2.9 | PhD_A2.9_Ph_six4 (folder) |
| *A2.9.1* | *PhD_A2.9.1_Ph_six4_sequences (folder)* |
| *A2.9.2* | *PhD_A2.9.2_Ph_six4_alignments (folder)* |
| A2.10 | PhD_A2.10_Ph_gbx (folder) |
| *A2.10.1* | *PhD_A2.10.1_Ph_gbx_sequences (folder)* |
| *A2.10.2* | *PhD_A2.10.2_Ph_gbx_alignments (folder)* |
| A2.11 | PhD_A2.11_Ph_kni1 (folder) |
| *A2.11.1* | *PhD_A2.11.1_Ph_kni1_sequences (folder)* |
| *A2.11.2* | *PhD_A2.11.2_Ph_kni1_alignments (folder)* |
| A2.12 | PhD_A2.12_Ph_kni2 (folder) |
| *A2.12.1* | *PhD_A2.12.1_Ph_kni2_sequences (folder)* |
| *A2.12.2* | *PhD_A2.12.2_Ph_kni2_alignments (folder)* |

**A3** **PhD_A3_list_of_sequence_clones (folder)**

PhD_A3_List_of_Parhyale_goi_sequence_clones.pdf

**Table A1:** **List of *Parhyale hawaiensis* gene of interest sequence clones.**

**A4** **PhD_A4_oligonucleotides (folder)**

A4.1 PhD_A4.1_degenerate_oligonucleotides.pdf

**Table A2:** **List of degenerate oligonucleotides.**

A4.2 PhD_A4.2_specific_oligonucleotides.pdf

**Table A3:** **List of specific oligonucleotides.**

A4.3 PhD_A4.3_siRNA_molecules.pdf

**Table A4:** **List of siRNA molecules.**

**A5** **PhD_A5_plasmid_sequences (folder)**

**A6** **PhD_A6_phylogeny_database (folder)**

A6.1 PhD_A6.1_List_of_species.pdf

**Table A5:** **List of species included in phylogenetic analyses.**

A6.2 PhD_A6.2_List_of_non-Parhyale_sequences.pdf

**Table A6:** **List of non-*Parhyale* protein sequences included in phylogenetic analyses.**

A6.3 PhD_A6.3_phylogeny_ClustalW2_alignments

## 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 |

# Curriculum Vitae

Dipl.-Biol.
**Bernhard Georg Michael Schmid**
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

## Education

2005-2011 PhD at the Göttingen Centre for Molecular Biosciences (GZMB), Georg-August-University Göttingen, Germany, supervised by Prof. Dr. Ernst A. Wimmer.

Dissertation topic: "Molecular Studies on Head Development of the Amphipod Crustacean *Parhyale hawaiensis*"

2004–2005 Member of the scientific staff
Department of Developmental Biology, Georg-August-Universität Göttingen, Germany

2003–2004 Member of the scientific staff
Department of Genetics, Universität Bayreuth, Germany

1997-2003 Study of Biology (Diploma), University Bayreuth, Germany
Grade: 1.2 (sehr gut/A)
*Major exam subject*: genetics
*Minor exam subjects*: zoology, biochemistry

Diploma thesis:
„Isolierung und Klonierung von zu *otd* sequenzähnlichen Genen verschiedener Arthropoden" (Grade: sehr gut/A)

1994 Abitur (final secondary-school examinations qualifying for university entrance, Deutschherren-Gymnasium, Aichach, Germany)

## Additional skills

| | |
|---|---|
| Languages | German (native), English (fluently written and spoken), Kiswahili (written and spoken) |
| Software | image processing (Adobe Photoshop, Adobe Illustrator) |

## Invited and contributing Speaker

ZooNet Meeting, Naples, 2005 (contributed)

*Parhyale* symposium, Tübingen, 2006 (invited)

Regional *Tribolium* Meeting, Göttingen, 2007 (contributed)

ZooNet Meeting, Heraklion, 2007 (contributed)

Euro Evo Devo (EED) Conference, Paris, 2010 (contributed)

## Publications

Schetelig, M. F.*, **Schmid, B. G.***, Zimowska, G., Wimmer, E. A. (2008). "Plasticity in mRNA expression and localization of *orthodenticle* within higher Diptera." Evolution & development 10(6): 700-704.

Browne, W. E., **Schmid, B. G.**, Wimmer, E. A., Martindale, M. Q. (2006). "Expression of *otd* orthologs in the amphipod crustacean, *Parhyale hawaiensis*." Development genes and evolution 216(10): 581-595.

Horn, C., **Schmid, B. G.**, Pogoda, F. S., Wimmer, E. A. (2002). "Fluorescent transformation markers for insect transgenesis." Insect biochemistry and molecular biology 32(10): 1221-1235.

*co-first author