## Evolution of Bicoid-dependent *hunchback* Regulation in Diptera

## **PhD Thesis**

in partial fulfillment of the requirements for the degree "Doctor of Philosophy (PhD)" in the Molecular Biology Program at the Georg August University Göttingen, Faculty of Biology

## submitted by

Steffen Joachim Lemke

## born in

Wuppertal

2006

Herewith I declare that I prepared the PhD Thesis "Evolution of Bicoid-dependent *hunchback* Regulation in Diptera" on my own and with no other sources and aids than quoted.

Chicago, 2006

Parts of this work were published in the following meeting presentations :

Lemke SJ, Rafiqi AM, Prell AH, Stauber M, Schmidt-Ott U. Evolution of transcriptional control of the *Drosophila* gap gene *hunchback*. 45th Annual *Drosophila* Research Conference, March 24-28, 2004. Washington, DC USA. Poster.

Lemke SJ, Prell AH, Stauber M, Schmidt-Ott U. Evolution of transcriptional control of the *Drosophila* gap gene *hunchback*. 44th Annual *Drosophila* Research Conference, March 5-9, 2003. Chicago, Illinois USA. Poster.

## **Table of Contents**

	Abstract	1
1	Introduction	2
1.1	Comparative embryology of Diptera	2
1.2	Pattern formation in Drosophila	4
1.3	How did <i>hunchback</i> regulation in dipterans evolve?	7
1.3.1	hunchback in Drosophila	7
1.3.2	hunchback regulation in Drosophila	7
1.3.3	hunchback regulation in dipterans and other insects	9
1.3.4	Complementary approaches to explore the evolution of <i>hunchback</i> regulation	12
2	Material and Methods	14
2.1	Fly culture and egg collection	14
2.2	Cloning	15
2.2.1	Preparation of genomic phage libraries	15
2.2.2	Preparation of cDNA templates	15
2.2.3	Isolation of homeobox genes	16
2.2.4	Isolation of hunchback homologues	17
2.2.5	Isolation of hunchback genomic DNA	17
2.3	In situ hybridization, immunocytochemistry and microscopy	21
2.3.1	Embryo fixation	21
2.3.2	RNA in situ probes	21
2.3.3	Whole mount RNA in situ hybridization	21
2.4	cDNA library from embryonic pole cytoplasm	22
2.4.1	cDNA preparation from pole cytoplasm	22
2.4.2	Preparation of bacterial libraries from amplified cDNA	23
2.4.3	Hybridization of the libraries and subtractive screening	23
2.4.4	Virtual northern hybridization	24

3	Results	24
3.1	Identification of <i>bicoid</i> orthologues from <i>Platypeza</i> and <i>Lonchoptera</i>	24
3.2	Subtractive screening for <i>bicoid</i> -like genes: a new method	25
3.3	Subtractive screening for <i>bicoid</i> -like genes in <i>Clogmia</i>	27
3.4	Cloning of dipteran <i>hunchback</i> genes	29
3.5	Functional comparison of early dipteran <i>hunchback</i> regulation	31
3.6	Expression studies in <i>Episyrphus</i>	33
3.6.1	<i>Episyrphus hunchback</i> shares expression characteristics of cyclorrhaphan and non-cyclorrhaphan dipterans	33
3.6.2	The extraembryonic anlage of <i>Episyrphus</i> extends to the anterior pole	34
3.6.3	Episyrphus orthodenticle is not expressed in pre-blastoderm embryos	36
4	Discussion	37
4.1	Do all cyclorrhaphan dipterans have <i>bicoid</i> ?	37
4.2	Does <i>Episyrphus</i> reflect the primitive patterning	
	mechanism of cyclorrhaphan flies?	38
4.3	How did early anterior <i>hunchback</i> regulation change in dipteran evolution?	39
	Summary and Conclusions	40
A	Appendix	41
A.1	Figures and Tables	41
A.1.1	Figures	41
A.1.2	Tables	59
A.2	Supplemental Figures	64
A.3	Sequences	66
	Bibliography	79

## List of Figures

- Figure 1. Phylogenetic relationships of taxa mentioned in the text.
- Figure 2. Homeodomain alignment and percent sequence similarity relative to the homeodomains of *Platypeza* Bicoid and *Lonchoptera* Bicoid.
- Figure 3. Protein alignment of *bicoid* homologues.
- Figure 4. cDNA pools from anterior and posterior cytoplasm of the same *Megaselia* embryo differ in composition.
- Figure 5. Identification of *Megaselia bicoid* by subtractive screening of a spotted cDNA library made from anterior egg cytoplasm.
- Figure 6. Classification and abundance of *Clogmia* clones.
- Figure 7. Testing for differential expression of *Clogmia* candidates in virtual Northern Blots.
- Figure 8. Quartet Puzzling analysis of the newly identified Hunchback homologues.
- Figure 9. Alignment of predicted protein sequences of the newly identified *hunchback* homologues.
- Figure 10. Genomic organization of Dipteran *hunchback* homologues.
- Figure 11. Putative *nanos* response element (NRE) sequences from dipteran *hunchback* homologues.
- Figure 12. Functional analysis of *hunchback* regulatory DNA by reporter gene expression in *Drosophila* embryos.
- Figure 13. Effects of bicoid RNAi on hunchback expression in Megaselia.
- Figure 14. Expression of Episyrphus hunchback.
- Figure 15. In *Episyrphus, hunchback* and *zerknüllt* are co-expressed along the dorsal midline.
- Figure 16. Expression of Episyrphus zerknüllt.
- Figure 17. Protein alignment of *Episyrphus* Orthodenticle with two isoforms of *Drosophila* Orthodenticle/Ocelliless.
- Figure 18. Expression of *Episyrphus orthodenticle*.

## **List of Tables**

- Table 1.cDNA isolation of dipteran *bicoid* and *orthodenticle* homologues: templates, primers and products.
- Table 2.
   cDNA isolation of dipteran *hunchback* homologues: templates, primers and products.
- Table 3. RNA probes for whole mount *in situ* hybridization.
- Table 4.
   Sequence comparison of Clogmia candidates with Drosophila and Anopheles genes.
- Table 5.Reporter gene constructs to analyze hunchback regulatory DNA in transgenic Dro-<br/>sophila embryos.

## Acknowledgements

The presented thesis has been carried out at the Max-Planck-Institut für biophysikalische Chemie in Göttingen, in the Department Molekulare Entwicklungsbiologie, and at the University of Chicago, in the Department of Organismal Biology and Anatomy. First of all I would like to thank my advisor, Urs Schmidt-Ott for his years of patient teaching and supervision, for his help, advice, and for his encouragement to follow up my own ideas. I would like to thank my committee members Herbert Jäckle and Reinhard Jahn for their critical interest and support. I thank Ab. Matteen Rafiqi, Sean Ferguson, Yu-Chiun Wang, Olivia Casanueva and Chip Ferguson, and of the former Göttinger group especially Michael Stauber, Alexander Prell and Ralf Pflanz for help and discussions. A special thanks goes to my parents for support and trust in me, and Annette, for her affection and patience.

## Contributions

**cDNA:** *Platypeza bicoid* was cloned by Ab. Matteen Rafiqi (MR). *Lonchoptera tristis bicoid*, *Episyrphus orthodenticle*, and a zygotic transcript of *Empis-hunchback* were cloned by Michael Stauber (MS). A full-length cDNA of *Clogmia hunchback* and a zinc finger fragment of *Haematopota hunchback* were cloned by Alexander Prell (AP). **Genomic DNA:** The *Megaselia hunchback* locus was cloned by MS. The *Clogmia hunchback* locus was cloned by AP. **Reporter constructs:** Transgenic *Drosophila* lines carrying the *Megaselia hunchback* locus were established by MS. Transgenic *Drosophila* lines carrying the reporter constructs with *Clogmia hunchback* regulatory DNA were established and analyzed by AP. **Fly work:** Technical assistance with RNAi in *Megaselia*, removal of cytoplasm, and fixation of embryos was provided by Sean Ferguson (SF). **cDNA library from anterior cytoplasm:** The library was established as a glycerol stock in collaboration with MR; MR, SF, Urs Schmidt-Ott, and Irene Hsiao helped with colony picking. The library was spotted by MR in collaboration with Professor Helmut Bloecker at the Department of Genome Analysis, German Research Center for Biotechnology, Braunschweig, Germany

## Abstract

An early segmentation gene of Drosophila melanogaster, hunchback, with an evolutionarily conserved function but diverging regulation was used as an entry point to explore the evolution of early patterning mechanisms in true flies (Diptera). In Drosophila, a gradient of bicoid protein activates the transcription of *hunchback* in the anterior blastoderm and thereby initiates patterning of the thorax. Very similar hunchback expression has been reported for other dipterans but a correlation with the occurrence of *bicoid* could not be established. Therefore, one or several hunchback regulators may have been exchanged in dipteran evolution. To map this transition in the regulation of *hunchback* expression, I expanded previous screens for *bicoid* orthologues using low stringency PCR and cDNA subtraction as technical approaches, and compared the results to the response of *hunchback* promoters from the same species using reporter constructs in transgenic Drosophila. Reporter expression in the anterior blastoderm of transgenic Drosophila was recorded only when the promoter was taken from a species with a *bicoid* orthologue. The reporter constructs of the hunchback promoters of all other species (five out of eight) were expressed in the posterior (2) or extraembryonic blastoderm (1), or were not expressed at all (2). These experiments enabled me to identify a lower cyclorrhaphan fly (*Episyrphus balteatus*; Syrphidae) with an early patterning mechanism likely to be fundamentally different from Drosophila and potentially similar to lower dipterans. To explore the possibility that Episyrphus shares developmental traits with lower dipterans, I studied the expression of Episyrphus hunchback, Episyrphus zerknüllt, and Episyrphus orthodenticle and compared the expression of these genes to their direct homologues in Megaselia abdita, Drosophila and Clogmia albipunctata. I found that Episyrphus combines expression characteristics of cyclorrhaphan and non-cyclorrhaphan dipterans indicating that this species might use a patterning mechanism that is an intermediate between lower and higher flies.

## 1 Introduction

The genetic basis of morphological evolution has received much attention in recent years (reviewed by Orr, 2005). Yet many genetic interactions change in the course of evolution without affecting morphology in an obvious way (reviewed e.g. by Raff, 1996). These changes might reflect neutral evolution of the developmental gene network (Raff, 1996), or alternatively they could be an adaptation to the developmental process itself (e.g. Bullock *et al.*, 2004). In flies (Diptera), the segmentation gene *hunchback* provides a striking example for this phenomenon: early zygotic expression of this gene is very similar across dipterans, while the regulation of *hunchback* expression has undergone fundamental changes. To understand the evolutionary significance of this transition, I have explored the evolution of *hunchback* regulation in its phylogenetic context.

## 1.1 Comparative embryology of Diptera

In Dipterans, as in most other insects, the zygote nucleus divides without cell division (reviewed in Anderson, 1966; Anderson, 1972). After a series of four to ten nuclear divisions, most nuclei migrate to the periphery and form a monolayer around the yolk (Anderson, 1966). This layer of nuclei is referred to as syncytial blastoderm (Anderson, 1966). When the plasmamembrane folds inwards between the nuclei, the syncytial blastoderm turns into a cellular blastoderm (Anderson, 1966), although in *Drosophila melanogaster*<sup>1</sup> at least, the cells do not pinch off completely from the underlying yolky cytoplasm until early gastrulation (Foe and Alberts, 1983).

By the onset of gastrulation, most of the dipteran blastoderm has been specified to become embryonic tissue, which is also referred to as germband (Johannsen and Butt, 1941); the remaining portion of the blastoderm will give rise to extraembryonic cell layers (Anderson, 1972; Johannsen and Butt, 1941). Shortly after the onset of gastrulation, the germband begins to extend from the posterior pole to the dorsal and then anteriorly such that the cells destined to form the most posterior larval structures are located transiently directly behind the future head region (Anderson, 1966; Campos-Ortega and Hartenstein, 1997). During germband re-

<sup>&</sup>lt;sup>1</sup> Referred to as *Drosophila* in the remaining text. In the same way I will refer to other species by only their genus name after the first introduction in the main text.

traction, this process is reversed (Anderson, 1966; Campos-Ortega and Hartenstein, 1997). In the retracting germband, segmental grooves form, which demarcate, in anterior-posterior sequence, the head, thoracic, and abdominal segments of the embryo (Anderson, 1966; Campos-Ortega and Hartenstein, 1997). After germband retraction, the epidermis closes dorsally. At this stage, all organs are established and the epidermis of the embryo secretes the exoskeleton (cuticle) of the larva (Anderson, 1966; Campos-Ortega and Hartenstein, 1997).

This developmental blueprint varies in some aspects among dipterans. Most higher dipterans (Cyclorrhapha, Figure 1) and also culicomorphan mosquitoes (Culicomorpha, Figure 1) such as Anopheles gambiae develop according to an extreme long-germ mode of insect development (Anderson, 1972; Sander, 1976). In these taxa, all segments are specified prior to gastrulation (Bullock et al., 2004; Goltsev et al., 2004b), the germ band extends to the anterior pole, and the extraembryonic tissue originates from dorsal blastoderm only (Anderson, 1972; Sander, 1976). By contrast, some (probably most) lower dipterans retain a more ancestral mode of development. Similar to the intermediate or short-germ development of most holometabolous insects, the posterior-most segments of these lower dipterans are specified in a posterior "growth zone" after the onset of gastrulation, and the extraembryonic anlage extends to the anterior pole (Anderson, 1972; Sander, 1976). Apart from the size of the extraembryonic anlage, cyclorrhaphan and non-cyclorrhaphan dipterans also differ in the organization of extraembryonic tissue. The extraembryonic tissue of non-cyclorrhaphan dipterans, as in most holometabolous insects, differentiates into two cell layers, the amnion and the serosa (Anderson, 1966; Anderson, 1972; Schmidt-Ott, 2000). The amnion remains linked to the embryo and covers the ventral side of the embryo after germband retraction, whereas the serosa detaches from the embryonic tissue to completely close around the embryo and the yolk (Handel et al., 2000; Schwalm, 1987). By contrast, the extraembryonic tissue of higher cyclorrhaphans (Schizophora, Figure 1) is a derived character and consists of only a single cell layer that covers the yolk sac dorsally, the amnioserosa (Anderson, 1966; Anderson, 1972).

## 1.2 Pattern formation in Drosophila

The molecular basis of dipteran segmentation has been studied primarily in *Drosophila*, where many segmentation genes have been discovered through saturating genetic screens for female sterile or embryonic lethal mutations, many of which cause phenotypes in the larval cuticle (Gans *et al.*, 1975; Jürgens *et al.*, 1984; Mohler, 1977; Nüsslein-Volhard *et al.*, 1987; Nüsslein-Volhard and Wieschaus, 1980; Nüsslein-Volhard *et al.*, 1984; Perrimon *et al.*, 1986; Schüpbach and Wieschaus, 1989; Wieschaus *et al.*, 1984). Depending on the cuticle phenotypes, the maternal genes were classified into four distinct maternal systems of anterior, posterior, terminal, and dorsal-ventral genes (Nüsslein-Volhard *et al.*, 1987; St Johnston and Nüsslein-Volhard, 1992). The zygotic genes were classified according to their mutant phenotypes in the cuticle as gap genes, pair-rule genes, and segment polarity genes (Nüsslein-Volhard and Wieschaus, 1980). Loss-of-function mutations in these zygotic genes cause either missing blocks of segments in the cuticle (gap genes), defects in every other segment (pair-rule genes), or an altered polarity of each segment (segment polarity genes) (Nüsslein-Volhard and Wieschaus, 1980). The *Hox* genes are involved in giving the embryonic segments their individual identity and were discovered independently (Lewis, 1978).

The anterior-posterior body axis is established by the anterior, the posterior, and the terminal maternal system (Nüsslein-Volhard *et al.*, 1987). The anterior system is required for head and thorax development (Nüsslein-Volhard *et al.*, 1987). The key gene is *bicoid*<sup>2</sup> (Berleth *et al.*, 1988; Frohnhöfer and Nüsslein-Volhard, 1986). *bicoid* transcripts become enriched at the anterior pole of the oocyte during oogenesis (Berleth *et al.*, 1988; Cha *et al.*, 2001). Translation of the localized *bicoid* transcripts and assumed diffusion of the protein establish a Bicoid gradient along the anterior-posterior axis of the embryo (Driever and Nüsslein-Volhard, 1988). Bicoid binds to the ubiquitous transcript of *caudal* (Macdonald and Struhl, 1986; Mlodzik *et al.*, 1985; Mlodzik and Gehring, 1987a; Rivera-Pomar *et al.*, 1996). Thus, Bicoid induces a Caudal gradient complementary to the Bicoid gradient (Macdonald and Struhl, 1986; Mlodzik and Gehring, 1987b; Rivera-Pomar *et al.*, 1996). In addition, Bicoid binds DNA (Driever and Nüsslein-Volhard, 1989; Struhl *et al.*, 1989) and directly activates the tran-

<sup>&</sup>lt;sup>2</sup> Nomenclature of genes and gene products is according to Drysdale *et al.* (2005).

scription of the gap gene *hunchback* (Driever and Nüsslein-Volhard, 1989; Driever *et al.*, 1989; Lehmann and Nüsslein-Volhard, 1987a; Struhl *et al.*, 1989; Tautz *et al.*, 1987) as well as a number of other gap and pair-rule genes (reviewed in Pankratz and Jäckle, 1993; reviewed in Rivera-Pomar and Jäckle, 1996).

The posterior system controls the establishment of a germ line and abdominal segmentation (Nüsslein-Volhard *et al.*, 1987). The key genes are *oskar* (Ephrussi *et al.*, 1991; Kim-Ha *et al.*, 1991; Lehmann and Nüsslein-Volhard, 1986) and *nanos* (Lehmann and Nüsslein-Volhard, 1991; Nüsslein-Volhard *et al.*, 1987; Wang and Lehmann, 1991). Oskar recruits factors for the germ line, including the *nanos* transcript, to the posterior pole (Ephrussi *et al.*, 1991; Ephrussi and Lehmann, 1992). Translation of the localized *nanos* transcript generates, possibly by diffusion, a Nanos gradient opposite to the Bicoid gradient (Gavis and Lehmann, 1992). Nanos is essential for abdominal segmentation as it suppresses the translation of maternal *hunchback* transcript in the posterior half of the early embryo (Hülskamp *et al.*, 1989; Irish *et al.*, 1989; Struhl, 1989; Tautz, 1988).

The terminal maternal system is required for the formation of the terminal body parts (Nüsslein-Volhard *et al.*, 1987). The key gene is *torso* (Casanova and Struhl, 1989; Schüpbach and Wieschaus, 1986; Sprenger *et al.*, 1989), which codes for a receptor tyrosine kinase (Sprenger *et al.*, 1989; Sprenger *et al.*, 1993). Torso is expressed evenly on the surface of the blastoderm embryo (Casanova and Struhl, 1989), but the receptor tyrosine kinase signaling pathway is activated only at the anterior and the posterior pole (Gabay *et al.*, 1997). At both poles, the activated signaling pathway leads to the de-repression of *tailless* (Jiménez *et al.*, 2000; Paroush *et al.*, 1997; Pignoni *et al.*, 1990), which is required for setting up the terminal structures of the larval cuticle (Jürgens *et al.*, 1984).

The dorsal-ventral body axis of the embryo is established independently from the anterior-posterior body axis (St Johnston and Nüsslein-Volhard, 1992). The key gene is *dorsal* (Nüsslein-Volhard, 1979; Steward, 1987; Steward *et al.*, 1984), which codes for a transcription factor (Thisse *et al.*, 1991). Dorsal is ubiquitously distributed in the cytoplasm of the freshly laid egg (Roth *et al.*, 1989; Rushlow *et al.*, 1989; Steward, 1989). In response to an extracellular signaling cascade (reviewed in Moussian and Roth, 2005), Dorsal enters the nuclei on the prospective ventral side of the embryo (Roth *et al.*, 1989; Rushlow *et al.*, 1989; Steward, 1989). Due to its asymmetry, the signaling cascade creates a nuclear Dorsal gradient in the ventral half of the embryo, with highest levels of Dorsal in the ventral-most nuclei (Moussian and Roth, 2005; Roth *et al.*, 1989; Rushlow *et al.*, 1989; Steward, 1989). This Dorsal gradient subdivides the embryo axis into three main regions – ventral (presumptive mesoderm), lateral (presumptive neuroectoderm) and dorsal (presumptive ectoderm) – by triggering threshold responses from a number of zygotic patterning genes (reviewed by Stathopoulos and Levine, 2002). In the dorsal half of the blastoderm embryo, *decapentaplegic (dpp*; Padgett *et al.*, 1987; Spencer *et al.*, 1982) is expressed due to the absence of nuclear Dorsal (Ray *et al.*, 1991). *dpp* encodes a transforming growth factor- $\beta$  (TGF- $\beta$ ) protein (Padgett *et al.*, 1987), which, together with the TGF- $\beta$  protein Screw (Arora *et al.*, 1994; Nüsslein-Volhard *et al.*, 1984), establishes a signaling center along the dorsal midline (reviewed in Ashe, 2005; reviewed in Raftery and Sutherland, 2003). Screw and peak levels of Dpp along the dorsal midline are required to specify the extraembryonic anlage (Arora *et al.*, 1994; Ferguson and Anderson, 1992), which is established by Dpp dependent activation of *zerknüllt* (Doyle *et al.*, 1986; Rushlow *et al.*, 2001; Wakimoto *et al.*, 1984).

Deviations from the *Drosophila* paradigm of early pattern formation have been reported for several non-cyclorrhaphan dipterans (Bullock *et al.*, 2004; Goltsev *et al.*, 2004b; Rohr *et al.*, 1999; Stauber *et al.*, 1999; Stauber *et al.*, 2002). In *Coboldia fuscipes* and *Clogmia albipunctata* (Figure 1), for example, the onset of posterior pair-rule gene expression is delayed (Rohr *et al.*, 1999), which correlates with the observation of a posterior "growth zone" in these species (Anderson, 1972). Furthermore, unlike in *Drosophila, hunchback* in non-cyclorrhaphan dipterans is expressed in the presumptive extraembryonic anlage (Goltsev *et al.*, 2004a; Rohr *et al.*, 1999). This expression indicates a potential role of *hunchback* in extraembryonic development that could relate morphological differences in extraembryonic development between cyclorrhaphan and non-cyclorrhaphan dipterans. Most intriguingly, however, *bicoid* is absent from the *Anopheles* genome (Zdobnov *et al.*, 2002), and, in addition to other *Drosophila* species (Drysdale *et al.*, 2005), *bicoid* homologues have been found only in cyclorrhaphan flies (Schröder and Sander, 1993; Sommer and Tautz, 1991; Stauber *et al.*, 1999). These and other studies led to the hypothesis that *bicoid* evolved only recently and is confined to cyclorrhaphan flies (Schröder Cott, 2000; Stauber *et al.*, 2002). Sequence data sug-

gest that *bicoid* and *zerknüllt* are sister genes, which most likely emerged from a *Hox3* gene duplication in the stem lineage of Cyclorrhapha (Stauber *et al.*, 1999; Stauber *et al.*, 2002). This postulation is supported by the finding that the closest homologue of *bicoid* in the *Anopheles* genome is *zerknüllt* (own observation). Thus, the anterior patterning mechanism of flies must have changed with the emergence of *bicoid*. In particular, the regulation of its *Drosophila* target gene *hunchback* must have changed accordingly.

## 1.3 How did hunchback regulation in dipterans evolve?

## 1.3.1 hunchback in Drosophila

hunchback codes for a C<sub>2</sub>H<sub>2</sub> zinc finger-type transcription factor (Tautz et al., 1987). In addition to its role as a gap gene during early embryogenesis, hunchback is also required during the development of the central nervous system (Grosskortenhaus et al., 2005; Isshiki et al., 2001; Kambadur et al., 1998; Lehmann and Nüsslein-Volhard, 1987a; Novotny et al., 2002). Here, I will focus on *hunchback* expression and regulation during early embryogenesis. At the onset of zygotic gene activity, hunchback protein is expressed throughout the anterior half of the embryo while being repressed in the posterior half (Tautz, 1988). In the anterior half, Hunchback is required to initiate development of the head and thorax: the cuticle patterns of mutant embryos devoid of any hunchback protein display a mirror image of abdominal segments in the anterior half of the embryo (Lehmann and Nüsslein-Volhard, 1987a). In the posterior half, the absence of Hunchback is required to allow for the development of a segmented abdomen: if Hunchback is prematurely expressed in the posterior half of the blastoderm, abdominal segmentation is severely affected or completely missing (Hülskamp et al., 1989; Struhl, 1989). Possibly because of its critical role in Drosophila patterning, hunchback is regulated, in part redundantly, by both the posterior and the anterior maternal systems (reviewed in Dearden and Akam, 1999).

#### 1.3.2 hunchback regulation in Drosophila

hunchback is transcribed from two different promoters (Tautz et al., 1987); but both transcripts produce the same protein (Tautz, 1988; Tautz et al., 1987). During oogenesis, hunchback is transcribed from its distal promoter (P1) (Schröder et al., 1988; Tautz et al.,

1987), and the P1 transcripts are evenly loaded into the egg (Margolis *et al.*, 1994; Tautz *et al.*, 1987). This maternal expression is driven by an enhancer that is located close to P1 (Lukowitz *et al.*, 1994; Margolis *et al.*, 1994). Nanos, together with Pumilio (Lehmann and Nüsslein-Volhard, 1987b; Macdonald, 1992), is required to repress translation of the uniformly distributed maternal *hunchback* mRNA in the posterior half of the embryo (Hülskamp *et al.*, 1989; Irish *et al.*, 1989; Struhl, 1989; Tautz, 1988). Sequences in the 3' untranslated region (UTR) of the *hunchback* mRNA (Nanos response elements: NREs) have been shown to recruit a complex with Nanos and Pumilio and thereby mediate the translational repression (Murata and Wharton, 1995; Sonoda and Wharton, 1999; Wharton and Struhl, 1991). As a result, maternal *hunchback* transcripts are translated only in the anterior half of the embryo and are degraded in the posterior half (Tautz, 1988; Tautz and Pfeifle, 1989).

At the onset of zygotic transcription, *hunchback* is transcribed from its proximal promoter (P2) (Driever and Nüsslein-Volhard, 1989; Schröder et al., 1988; Struhl et al., 1989; Tautz et al., 1987), resulting in strongly increased Hunchback levels throughout the anterior half of the embryo (Schröder et al., 1988; Tautz et al., 1987). This early zygotic hunchback expression is driven by a Bicoid-binding enhancer (Driever and Nüsslein-Volhard, 1989; Driever et al., 1989; Schröder et al., 1988; Struhl et al., 1989), which is about 250 bp long and located immediately upstream of P2 (Driever and Nüsslein-Volhard, 1989; Driever et al., 1989; Struhl et al., 1989). Although Bicoid is required, it appears to be not sufficient to drive hunchback expression throughout the anterior half of the embryo (Simpson-Brose et al., 1994): in mutant embryos that lack functional hunchback protein, expression of hunchback mRNA is restricted to the anterior-most 20% of the embryo (Simpson-Brose et al., 1994), indicating that Hunchback activates its own transcription synergistically with Bicoid (Simpson-Brose et al., 1994). This interpretation is supported by the presence of a Hunchback-binding site in the minimal Bicoid-binding enhancer (Treisman and Desplan, 1989). Whether, in addition to Bicoid and Hunchback, additional factors are required to sharpen the posterior boundary of early zygotic hunchback expression is the subject of a current debate: staufen (Schüpbach and Wieschaus, 1986; St Johnston et al., 1991), which is required to anchor bicoid transcripts to the anterior pole (St Johnston *et al.*, 1989) and to localize *oskar* transcripts to the posterior pole (Ephrussi et al., 1991; Kim-Ha et al., 1991), has been suggested to regulate the posterior *hunchback* boundary independent of *bicoid* (Houchmandzadeh *et al.*, 2002), but this hypothesis has been called into question again by a recent study (Crauk and Dostatni, 2005).

Shortly before the onset of gastrulation, a second zygotic *hunchback* enhancer located upstream of P1 drives the expression of P1 and P2 transcripts in two circumferential stripes (Lukowitz *et al.*, 1994; Margolis *et al.*, 1995). The anterior stripe is expressed in the presumptive thorax (parasegment four), and the posterior stripe is expressed in the presumptive abdomen (parasegment 13) (Lukowitz *et al.*, 1994; Margolis *et al.*, 1995; Schröder *et al.*, 1988; Tautz *et al.*, 1987). This posterior stripe is under the control of the terminal system and directly activated by the terminal gap gene *tailless* (Margolis *et al.*, 1995).

Both the maternal and the early zygotic regulation of *hunchback* are to a certain degree redundant. Maternal *hunchback* expression is not essential for development (Lehmann and Nüsslein-Volhard, 1987a): mutants without maternal *hunchback* are viable and do not display a distinct phenotype (Lehmann and Nüsslein-Volhard, 1987a). However, maternal *hunchback* expression can partly compensate for the loss of zygotic, Bicoid-dependent Hunchback contribution: dependent on the dose of maternal *hunchback*, the zygotic *hunchback* phenotype can be partly rescued (Lehmann and Nüsslein-Volhard, 1987a; Wimmer *et al.*, 2000). In the absence of Bicoid-dependent *hunchback* activation, high amounts of maternal *hunchback* (four copies) can, in combination with a reduction of the *hunchback* repressor *knirps* (one copy), rescue all thoracic segments (Wimmer *et al.*, 2000). Head segments, however, are not rescued in the absence of *bicoid* activity (Wimmer *et al.*, 2000), indicating that even higher *hunchback* levels or Bicoid-targets other than *hunchback* are required for this body part (Wimmer *et al.*, 2000).

#### 1.3.3 *hunchback* regulation in dipterans and other insects

The early zygotic expression of *hunchback* throughout the anterior half of the embryo is highly conserved in dipterans (Bonneton *et al.*, 1997; Goltsev *et al.*, 2004a; McGregor *et al.*, 2001a; Rohr *et al.*, 1999; Sommer and Tautz, 1991; Stauber *et al.*, 2000; Treier *et al.*, 1989). Within Cyclorrhapha, regulation of *hunchback* expression was investigated in *Drosophila virilis*, *Musca domestica*, *Calliphora vicina*, and *Lucilia sericata* (Bonneton *et al.*, 1997; Lukowitz *et al.*, 1994; McGregor; McGregor *et al.*, 2001b; Shaw *et al.*, 2001). A *bicoid* homologue has been identified from all four flies (MacDonald, 1990; Schröder and Sander, 1993; Shaw *et al.*, 2001; Sommer and Tautz, 1991), and Bicoid-binding sites in the regulatory DNA of all respective *hunchback* homologues have been mapped within 800 bp upstream of the putative P2 transcription start sites (Bonneton *et al.*, 1997; Lukowitz *et al.*, 1994; McGregor *et al.*, 2001b). For *Drosophila virilis, Musca*, and *Calliphora*, *hunchback* regulatory DNA including these mapped Bicoid-binding sites has also been analyzed in transgenic *Drosophila* embryos, and *hunchback* regulatory DNA of all three species drives reporter gene expression throughout the anterior half of *Drosophila* blastoderm embryos (Bonneton *et al.*, 1997; Lukowitz *et al.*, 1994; McGregor). Knockdown of *hunchback* by RNA interference (RNAi) in *Musca* and *Megaselia abdita* also suggests a conserved function of early zygotic Hunchback among dipterans (McGregor *et al.*, 2001b; Stauber *et al.*, 2000).

However, it is currently unclear how the anterior domain of *hunchback* expression is established in non-cyclorrhaphan dipterans and other insects without a bicoid homologue (for a recent review, see Liu and Kaufman, 2005; Stauber et al., 2002). Several lines of evidence suggest that in non-cyclorrhaphan dipterans, a gene with properties very similar to *bicoid* is responsible for *hunchback* activation and thus anterior patterning. In the non-cyclorrhaphan dipterans Chironomus spec., Smittia spec., and Bradysia tritici, a symmetrical double abdomen, reminiscent of a combined loss of bicoid and hunchback in Drosophila (Hülskamp et al., 1990), can be induced by UV ablation of the anterior cortex (Kalthoff, 1983; Kalthoff and Sander, 1968; Perondini et al., 1987; Yajima, 1964). In Smittia, this double abdomen phenotype has also been induced by removal of anterior cytoplasm (Schmidt et al., 1975) and by applying RNase to the anterior pole (Kandler-Singer and Kalthoff, 1976), while in *Chironomus*, the UV induced double abdomen has been reportedly rescued by fractions of poly(A)<sup>+</sup> RNA (Elbetieha and Kalthoff, 1988). These and additional experiments in other insects led to the prediction that a localized transcript is essential for patterning the anterior of all dipterans and possibly other insects (reviewed by Kalthoff, 1979; Kalthoff, 1983; reviewed by Sander, 1976).

Recent studies have suggested that *orthodenticle* (Finkelstein *et al.*, 1990), an evolutionarily conserved *Hox* gene (reviewed by Reichert and Simeone, 1999), acts synergistically with *hunchback* to partially substitute for the anterior determinant *bicoid* in the flour beetle Tribolium castaneum and the jewel wasp Nasonia vitripennis (Lynch et al., 2006; Schröder, 2003). If orthodenticle and hunchback are depleted by RNAi, Tribolium and Nasonia embryos, lack head, thorax, and most abdominal segments (Lynch et al., 2006; Schröder, 2003). Orthodenticle, like Bicoid, carries a lysine at position 50 of its homeodomain (Finkelstein et al., 1990). In Bicoid, this residue is critical for the selective binding of the protein to its natural enhancer targets (Hanes and Brent, 1989; Hanes et al., 1994; Treisman et al., 1989). Most homeodomain proteins carry a glutamine at this position (reviewed by Gehring et al., 1994) and differ significantly from Bicoid in their DNA-binding affinities (Hanes and Brent, 1989; Treisman et al., 1989). In Nasonia, maternal orthodenticle transcripts are localized to the anterior pole, similar to bicoid transcripts in Drosophila (Lynch et al., 2006). In Tribolium, maternal orthodenticle transcripts are evenly distributed in the embryo (Li et al., 1996), but translation is repressed in the posterior so that the protein is expressed in an anterior to posterior gradient (Schröder, 2003). Thus, it seems possible that Bicoid substitutes for maternal orthodenticle activity. In Nasonia, however, zygotic hunchback of Nasonia is still expressed in the anterior third of embryos that have been depleted of orthodenticle activity by parental RNAi (Bucher et al., 2002; Lynch et al., 2006), and in Anopheles orthodenticle is not expressed maternally (Goltsev et al., 2004a). Thus, unlike bicoid, orthodenticle is most likely not a primary (Anopheles) or not the only primary (Nasonia) determinant responsible for anterior zygotic hunchback activation.

An alternative model, based on Nanos-mediated translational repression of maternal *hunchback* transcripts in *Drosophila*, could explain anterior *hunchback* expression without anterior input (Curtis *et al.*, 1995; Irish *et al.*, 1989; Simpson-Brose *et al.*, 1994). Enrichment of *nanos* transcripts at the posterior pole is conserved throughout Diptera (Calvo *et al.*, 2005; Curtis *et al.*, 1995), dipteran *nanos* homologues can substitute for *nanos* function in *Drosophila* (Curtis *et al.*, 1995), and conserved NRE sequences have been identified in the 3' UTRs of *hunchback* homologues from *Tribolium* (Wolff *et al.*, 1995), *Nasonia* (Pultz *et al.*, 2005), and the grasshoppers *Schistocerca americana* and *Locusta migratoria* (Patel *et al.*, 2001). Thus, Nanos-dependent translational repression of maternal *hunchback* transcripts in the posterior half of the embryo might be conserved in many insects (Curtis *et al.*, 1995). Maternal *hunchback* activity in the anterior half of the embryo may then initiate an auto-regulatory loop,

which would explain zygotic up-regulation of *hunchback* in the anterior of lower dipterans (Curtis *et al.*, 1995; Simpson-Brose *et al.*, 1994).

In addition to Nanos, Caudal has also been suggested as a key regulator of early hunchback expression in insects without bicoid (reviewed in Dearden and Akam, 1999; Liu and Kaufman, 2005). This model is based on the analysis of Tribolium hunchback regulatory sequences in transgenic Drosophila embryos (Wolff et al., 1998). In transgenic Drosophila embryos, hunchback regulatory sequences of Tribolium drive reporter gene expression in a Caudal-dependent manner (Wolff et al., 1998), and, consistently, Caudal-binding sites have been mapped to Tribolium hunchback regulatory DNA (Wolff et al., 1998). Recent functional studies of *caudal* in *Tribolium* (Copf *et al.*, 2004) and in the cricket *Gryllus bimaculatus* (Shinmyo et al., 2005) are consistent with a Caudal-dependent hunchback activation. In both species, knockdown of *caudal* by RNAi results in embryos with only a few head segments (Copf et al., 2004; Shinmyo et al., 2005). Furthermore, caudal RNAi in Gryllus leads to a significant decrease in hunchback expression and a posterior shift of the expression domain (Shinmyo et al., 2005). Since putative NRE sequences have been identified in the 3' UTR of Tribolium hunchback (Wolff et al., 1995), the studies in Tribolium and Gryllus suggests a regulatory mechanism, where hunchback transcription is activated via Caudal and translation of the mRNA is repressed by Nanos (Wolff et al., 1998).

However, the Nanos/Caudal models do not explain how early zygotic *hunchback* expression in *Tribolium* is activated in the serosal anlage, since the onset of this expression is independent of *caudal* (Wolff *et al.*, 1998; Wolff *et al.*, 1995). Furthermore, in *Anopheles* neither *hunchback* nor *caudal* appear to be maternally expressed (Goltsev *et al.*, 2004a). Both taxa, therefore, apparently use alternative means for *hunchback* regulation. Thus, currently available data strongly suggest that an unidentified anterior maternal system regulates *hunchback* expression in non-cyclorrhaphan dipterans and possibly in other insects.

#### 1.3.4 Complementary approaches to explore the evolution of *hunchback* regulation

To explore how and when during dipteran evolution *hunchback* regulation changed from a Bicoid-independent to a Bicoid-dependent mechanism, I first tested the hypothesis that *bicoid* emerged at the transition from non-cyclorrhaphan to cyclorrhaphan dipterans. In an at-

tempt to map the emergence of *bicoid* to this transition, previous studies have covered a variety of non-cyclorrhaphan dipterans (Stauber et al., 2002), but only one of several families from the basal and most likely paraphyletic ashizans (Phoridae; Figure 1) (Stauber et al., 1999). These studies have been extended to other ashizan families and new *bicoid* homologues have been identified for *Platypeza consobrina* (Platypezidae, Figure 1) and *Lonchoptera lutea* (Lonchopteridae, Figure 1). To extend this screen to the predicted anterior determinant of noncyclorrhaphan dipterans, I developed a new screening method for anterior localized transcripts and explored the non-cyclorrhaphan *Clogmia albipunctata* (Psychodidae, Figure 1). This direct approach failed to identify an anterior localized transcript in Clogmia, and a bicoid homologue in the cyclorrhaphan *Episyrphus balteatus* (Syrphidae, Figure 1). Thus, to complement this quest for the potential hunchback activator in non-cyclorrhaphan dipterans and Episyrphus, I have also generated reporter constructs to directly compare the early regulation of dipteran hunchback homologues in transgenic Drosophila. This approach also extends previous work, in which the regulation of hunchback homologues from Drosophila virilis, Musca, Calliphora, and Tribolium had been studied in transgenic Drosophila embryos (Bonneton et al., 1997; Lukowitz et al., 1994; McGregor; Wolff et al., 1998). In this way, I studied the hunchback homologues from four basal cyclorrhaphans (Episyrphus, Megaselia, Platypeza and Lonchoptera), and four non-cyclorrhaphan dipterans (Empis livida, Haematopota pluvialis, Clogmia, and Anopheles), which represent mostly paraphyletic dipteran branches (Figure 1). The results provide additional support for Bicoid-dependent hunchback regulation in flies with bicoid, they provide support for Bicoid-independent hunchback regulation in non-cyclorrhaphan dipterans, and they also support for the initial observation that *Episyrphus* might not contain a bicoid homologue. This unexpected and peculiar position of Episyrphus among cyclorrhaphans was further explored by studying expression of a set of early patterning genes. Also in these expression analyses, *Episyrphus* displayed intermediate characters between non-cyclorrhaphan and cyclorrhaphan dipterans.

## 2 Material and Methods

#### 2.1 Fly culture and egg collection

Megaselia abdita Schmitz (Phoridae; scuttle or humpbacked flies) were reared as described (Schmidt-Ott et al., 1994) with modifications: The generation time at 25 °C with a 14/10-hour light/dark cycle was 18-20 days. The flies were reared in plastic stock bottles (diameter: 5.5 cm, height: 13 cm, Genesee) on wet cotton sprinkled with 4-5 grams of crushed aquarium fish food (Aquatic EcoSystems, Spirulina Flake) per bottle. For egg collection, adults (2500-3000) were placed in a cylindrical Plexiglas cage (diameter: 8 cm, height: 10 cm). Prior to collecting eggs, adult flies were starved on a water-agar plate for the duration of a light cycle. Eggs were collected on moistened filter paper supplemented with a streak of moistened fish food. A peak in egg deposition was observed shortly after the beginning of the dark cycle. Clogmia albipunctata Williston (Psychodidae; moth flies) were reared as described (Schmidt-Ott et al., 1994). The generation time at 25 °C with a 14/10-hour light/dark cycle was 22-26 days. Eggs were collected as described (Schmidt-Ott et al., 1994). Adults of Episyrphus balteatus Degeer (Syrphidae; hover flies) were collected in the surroundings of Göttingen (Germany); embryos were obtained from P. Katz (Katz Biotech AG, Baruth, Germany). Platypeza consobrina Zetterstedt (Platypezidae; flat-footed flies), Lonchoptera lutea Panzer (Lonchopteridae; pointed-wing flies), Empis livida L. (Empididae; dance flies), and Haematopota pluvialis L. (Tabanidae; horse flies) were collected in the surroundings of Göttingen. Females of Anopheles gambiae Giles, PEST strain (Culicidae; African malaria mosquitoe) were a gift from Frank H. Collins (University of Notre Dame, IN, USA). Drosophila was of the wild-type Oregon-R strain. Sample specimen of *Platypeza consobrina* larvae were classified by Peter Chandler (Slough, UK), samples of adult Lonchoptera lutea were classified by Urs Schmidt-Ott according to Smith (1969), and samples of adult Empis livida and Haematopota pluvialis were classified by Andreas Stark (Halle, Germany) and Marcel Leclercq (Beyne Heusay, Belgium), respectively.

## 2.2 Cloning

#### 2.2.1 Preparation of genomic phage libraries

Genomic Lambda-Fix II phage libraries were available for Megaselia, Lonchoptera, Haematopota and Clogmia (Schmidt-Ott, unpublished). For this work genomic Lambda-Fix II libraries (Stratagene) were established for Episyrphus and Platypeza. Several attempts to establish a genomic Lambda-Fix II library for Empis failed for unknown reasons. For the Episyrphus library, genomic DNA was prepared from a single adult female; for the Platypeza library, genomic DNA was prepared from 0.5 ml of larvae. Genomic DNA was isolated by SDS lysis as described (Andres and Thummel, 1994) followed by a digest with DNase-free RNase. To generate genomic fragments of 15-20 kb length, the genomic DNA was digested partially by MboI (NEB) for 60 minutes at 37 °C using 0.03 to 0.05 units of MboI per µg DNA. The libraries were constructed according to the Lambda-Fix II library manual using the Gigapack III XL-11 packaging extract (Stratagene). The primary libraries were titered (Episyrphus balteatus: 710'000 primary clones; Platypeza consobrina: 600'000 primary clones) and amplified; aliquots were stored at -80 °C. For screening, Hybond-N<sup>+</sup> nylon membranes (Amersham) were used to prepare plaque lifts. Probes were labeled radioactively using the Rediprime II Random Prime Labeling System (Amersham) with the following modifications: prior to the initial denaturation step, the reaction mix was supplemented with random hexanucleotides to a final concentration of 200 nM. After the denaturation step, instead of snap cooling the DNA, the hexanucleotides were allowed to anneal at 37 °C for 5 minutes. These steps significantly increased labeling efficiency for short probes (130-150 bp). Isolated phages were amplified according to the manual, and phage DNA was prepared using the Lambda Midi Kit (Qiagen).

#### 2.2.2 Preparation of cDNA templates

cDNA templates were prepared with the SMART RACE cDNA Amplification Kit (Clontech) and the Marathon cDNA Amplification Kit (Clontech), respectively. For cDNA preparation, 120-150  $\mu$ g of total RNA was extracted from 50 to 100  $\mu$ l fly tissue with RNAwiz (Ambion) according to the manual. Poly A<sup>+</sup> RNA was enriched using the Oligotex mRNA midi kit (Qiagen), with an average yield of about 2  $\mu$ g poly A<sup>+</sup> RNA. Enriched poly A<sup>+</sup> RNA was used to prepare the cDNA according to the user manuals. The *Anopheles* 5' SMART

RACE cDNA template was prepared from 1.2  $\mu$ g of total RNA, which was isolated from three adult females. The sources of the RNA material and the respective cDNA Amplification Kits used for each of the other species are listed in Table 1 and Table 2, respectively.

#### 2.2.3 Isolation of homeobox genes

*Platypeza bicoid* was amplified by PCR on genomic DNA with the degenerate primer pair 5'-YTGGGYMMAGCYCAGGTSAARATWTGGTT/5'-TYTTBGGYGTYAAHGGYT-CRTAGAC, corresponding to positions 367-395 and 805-830 in *Megaselia bicoid* (GenBank entry AJ133024, Stauber *et al.*, 1999). The product was cloned into pCRII-TOPO (Invitrogen) and sequenced. To obtain *Platypeza bicoid* cDNA, 5' and 3' rapid amplification of cDNA ends (RACEs) were performed (Table 1), and the products were cloned into pCRII-TOPO. The RACE products did not cover the open reading frame (ORF) completely; therefore, an additional PCR with specific primers was performed on cDNA (Table 1), and the product was cloned into pCR2.1-TOPO. The cDNA sequence of *Platypeza bicoid* (SEQ01 in the Appendix A.3) is derived from all three clones.

Lonchoptera bicoid was initially isolated from Lonchoptera tristis by PCR on genomic DNA with the degenerate primer pair 5'-TNGTNATGMGNMGNMGNMGNAC/5'-CKNCKRTTYTTRAACCA, corresponding to positions 239-260 and 391-407 in *Megaselia* bicoid. However, due to limited availability of Lonchoptera tristis, Lonchoptera lutea was eventually used in this study. To test for the presence of bicoid in this species, Lonchoptera bicoid was also isolated from Lonchoptera lutea. Using specific primers derived from the bicoid homologue of Lonchoptera tristis, 5' and 3' RACEs were performed (Table 1). The RACE products did not cover the homeobox completely; therefore, an additional PCR with specific primers was performed on cDNA (Table 1). All PCR products were cloned into pCRII-TOPO. The sequence of the Lonchoptera lutea bicoid (SEQ02 in the Appendix A.3) is derived from all three clones.

*Episyrphus orthodenticle* was amplified by PCR on genomic DNA with the same degenerate primer pair that was used to isolate *Lonchoptera bicoid*. cDNA was prepared by RACEs, and the products were cloned into pCRII-TOPO (Table 1) and sequenced (SEQ03 in the Appendix A.3).

#### 2.2.4 Isolation of hunchback homologues

hunchback fragments, encoding 133 bp of the conserved first zinc-finger domain (Sommer et al., 1992; Stauber et al., 2000; Tautz et al., 1987), were amplified by PCR from genomic DNA of *Platypeza*, *Lonchoptera*, *Episyrphus*, *Empis*, and *Haematopota* as described previously (Stauber et al., 2000). For each of these hunchback homologues, as well as for *Megaselia hunchback* and *Anopheles hunchback*, RACEs were performed and cloned into pCRII-TOPO (Table 2). For *Episyrphus*, *Lonchoptera*, and *Haematopota*, respectively, the 3' RACE products did not cover the hunchback ORFs completely; therefore, additional PCRs with specific primers based on cDNA and genomic DNA sequence (Material and Methods, 2.2.6) were performed on cDNA (Table 2), and the products were cloned into pCR2.1-TOPO. Primers to isolate zygotic *Megaselia hunchback* (Stauber et al., 2000); primers to isolate *Anopheles hunchback* (Stauber et al., 2000); primers to isolate *Anopheles hunchback* (Stauber et al., 2000); primers to isolate *Anopheles hunchback* (Stauber et al., 2000); primers to isolate *Anopheles hunchback* (Stauber et al., 2000); primers to isolate *Anopheles hunchback* (Stauber et al., 2000); primers to isolate *Anopheles gambiae* (Zdobnov et al., 2002). A cDNA clone of *Clogmia hunchback*, spanning the entire ORF, was isolated from a maternal Lambda-ZAP cDNA library (Schmidt-Ott, unpublished) using a partial *Clogmia hunchback* cDNA (Rohr et al., 1999) as a probe (Table 2).

#### 2.2.5 Isolation of hunchback genomic DNA

*Episyrphus hunchback* genomic DNA was isolated by screening a genomic Lambda-Fix II library, using the 133 bp fragment obtained by degenerate PCR (see Material and Methods 2.2.4) as probe, and by PCR. A phage (Eba-hb ph10) spanning 14 kb of genomic DNA, including 2.1 kb upstream of the ORF, was isolated. The region upstream of the ORF, together with 0.9 kb of the ORF, was amplified by PCR from phage Eba-hb ph10 using a gene-specific primer (5'-CCGACGAGTGTGACTTCCGGTGGGAGTTCAAC) and a T7 primer specific for the phage-internal MCS. The product (3.0 kb) was cloned into pGEM-T Easy (Promega). A second, partially overlapping fragment was amplified by long range PCR from independently prepared genomic DNA using a primer specific for the first exon of the P1 transcript (5'-GGGAATATTAATTCTGTAAACGGAGA) and a primer specific for the second exon of the transcript at the beginning of the ORF (5'-CTGCATTGAATCCCAGTTCTGC). This and other long range PCRs were performed using TaKaRa La Taq (Takara). The product (5.4 kb) was cloned into pGEM-T Easy, yielding plasmid C616. The genomic *Episyrphus hunchback* sequence (SEQ07 in the Appendix A.3) is derived from both plasmids and phage Eba-hb ph10. The insert of C616 was cloned as *Not*I fragment in front of the *Drosophila* hsp43 basal promotor of the P-element transformation vector pCaSpeR-hsp43-*lacZ* (Thummel and Pirrotta, 1992), yielding plasmid C681.

*Megaselia hunchback* genomic DNA was isolated by screening a genomic Lambda-Fix II library, using a *Megaselia hunchback* 5' RACE product (Stauber *et al.*, 2000) as probe. A phage (Mab-hb ph2a) spanning 15 kb of genomic DNA, including 8 kb upstream of the ORF, was isolated. Two partially overlapping fragments of the phage insert (a 4.5 kb *SpeI*-fragment and an 8.0 kb *XbaI*-fragment) were subcloned into pBluescript (Stratagene) and partially sequenced. The genomic *Megaselia hunchback* sequence (SEQ09 in the Appendix A.3) is derived from both plasmids. The ORF, together with 8 kb upstream and 1 kb downstream of the ORF, was amplified by long range PCR from the phage Mab-hb ph2a, using a primer specific to the region 3' of the ORF (5'-CCGTAACATTAACCGTAAC) and a T7 primer specific for the phage-internal multiple cloning site (MCS). The product (11 kb) was cloned into pGEM-T Easy, then excised with *Not*I and cloned into the *Not*I site of the P-element transformation vector pCaSpeR 4 (Thummel and Pirrotta, 1992), yielding plasmid C220.

*Platypeza hunchback* genomic DNA was isolated by screening a genomic Lambda-Fix II library, using the 133 bp fragment obtained by degenerate PCR (see Material and Methods 2.2.4) as probe. A phage (Pco-hb ph1) spanning 16 kb of genomic DNA, including 9 kb upstream of the ORF, was isolated. The phage insert was subcloned into the *Not*I site of the vector pZErO-1 and partially sequenced (SEQ12 in the Appendix A.3), yielding plasmid C690. 6.2 kb upstream of the ORF were amplified by long range PCR from plasmid C690, using the primer pair (5'-ATAATCCAGGTGTTGCATCAGG/5'-CTCGTAGCTAGCTGGC-TGAAGTGC). The product was cloned into pGEM-T Easy, then excised with *Not*I and cloned into the *Not*I site of pCaSpeR-hsp43-*lacZ*, yielding plasmid C622.

*Lonchoptera hunchback* genomic DNA was isolated by screening a genomic Lambda-Fix II library, using the 133 bp fragment obtained by degenerate PCR (see Material and Methods 2.2.4) as probe, and by PCR. A phage (Llu-hb ph2) spanning 16 kb of genomic DNA, including 1.9 kb upstream of the ORF, was isolated. The region upstream of the ORF, together with 0.8 kb of the ORF, was amplified by PCR from phage Llu-hb ph2 using a gene specific primer (5'-CGGCACAACGATACTGATACACAGAAG) and a T3 primer specific for the phage-internal MCS. The product (2.7 kb) was cloned into pGEM-T Easy. The phage insert was subcloned into the *Not*I site of the vector pZErO-1. A third and partially overlapping fragment was amplified by long range PCR from independently prepared genomic DNA using a primer specific for the first exon of the P1 transcript (5'- GACGCGTTCCGATTAACGGA-TATAA) and a primer specific for the second exon of the transcript immediately upstream of the ORF (5'-TTCAAATTTAACTGCGATGGAGAGC). The product (4.6 kb) was cloned into pGEM-T Easy, yielding plasmid C514. The genomic *Lonchoptera hunchback* sequence (SEQ14 in the Appendix A.3) is derived from all three plasmids. The insert of C514 was cloned as *Not*I fragment into the P-element transformation vector pCaSpeR-hsp43-*lacZ*, yielding plasmid C515.

*Empis hunchback* genomic DNA was isolated by long range PCR from genomic DNA using a primer specific for the first exon of the P1 transcript (5'-GTACGCGGGA-GTCATGTCTGATGTCTTATA) and a primer specific for the second exon of the transcript at the start of the ORF (5'-ACTATTAATTGCTGTTTGTGGGTTCA). The product (6.0 kb) was cloned into pGEM-T Easy and sequenced (SEQ16 in the Appendix A.3). The fragment was then excised with *Not*I and cloned into the *Not*I site in front of the minimal *even-skipped* promotor of the P-element transformation vector pCaSpeR-E2G-*lacZ* (Markstein *et al.*, 2002), yielding plasmid C681.

*Haematopota hunchback* genomic DNA was isolated by screening a genomic Lambda-Fix II library, using the 133 bp fragment obtained by degenerate PCR (see Material and Methods 2.2.4) as probe. A phage (Hpl-hb phB) spanning 15 kb of the locus, including 9 kb upstream of the ORF, was isolated. The phage insert was subcloned into the *Not*I site of the vector pZErO-1 and sequenced (SEQ18 in the Appendix A.3). 9 kb upstream of the ORF were amplified by long range PCR on phage Hpl-hb phB using a gene-specific primer (5'- TCCATT-GATGGGTATGTTGTAG) and a T7 primer specific for the phage-internal MCS. A smaller fragment (1.8 kb) comprising the intron sequence of the P1 transcript was amplified by PCR from the same phage using the primer pair (5'-ATTTTGTGAAAATTATGAAATAATTTGGACGC/5'- TCCATTGATGGGTATGTTGTAG). Both PCR products were cloned into pGEM-T Easy, then excised with *Not*I and subcloned in the *Not*I site of pCaSpeR-hsp43-*lacZ*, yielding plasmids C423 (9 kb insert; H1) and C688 (1.8 kb insert; H2), respectively.

*Clogmia hunchback* genomic DNA was isolated by screening a genomic Lambda-Fix II library using a partial Clogmia hunchback cDNA (Rohr et al., 1999) as probe. A phage (Cal-hb ph1) spanning 15 kb of genomic DNA, including 6.9 kb upstream of the ORF, was isolated. The region upstream of the ORF, together with 0.7 kb of the ORF, was amplified by long (5'-PCR from phage DNA using gene-specific range the a primer TTGATGTGGATCCTATTGTGCT) and a T7 primer specific for the phage-internal MCS. The product (7.6 kb) was cloned into pGEM-T Easy, yielding plasmid C213. 6.9 kb upstream of the ORF were amplified by long range PCR from plasmid C213 using a gene specific primer with an added XhoI site (5'-ATCTCGAGTGACTGAAAGAATAGAAA) and a T7 primer specific for the phage-internal MCS. The product was cloned into pGEM-T Easy, yielding plasmid C214. The insert of C214 was then subcloned as NotI fragment into pCaSpeR-hsp43*lacZ*, yielding plasmid C215 (K2). In addition, a 4.3 kb fragment was amplified by long range PCR with specific primers (5'-TGGCTTAGATATAGTCATTACC/5'-ATCTCGAGTGA-CTGAAAGAATAGAAA) from C213, cloned into pGEM, excised with NotI and cloned into the NotI site of pCaSpeR-hsp43-lacZ, yielding clone C305 (K4). The insert of C305 was then digested with SacII/AgeI, the overhangs were blunted with Mung Bean Nuclease, and the vector was religated yielding clone C305 (K13) with 2.4 kb of the Clogmia hunchback intron in pCaSpeR-hsp43-lacZ. The genomic Clogmia hunchback sequence (SEQ20 in the Appendix A.3) is derived from plasmids C213 and C214.

Genomic DNA of *Anopheles hunchback* PEST strain was isolated by long range PCR with specific primers (5'-TGTGAGCATTTGCA-TGAGGCTGATTA/5'-CCATCGCCATTA-CGGAGTCAAAGTTC) based on the sequence of GenBank entry AAAB01008979. The 5.2 kb fragment, including the intron sequence of the P1 transcript, was cloned into pGEM-T Easy, then excised with *Not*I and subcloned into the *Not*I site of pCaSpeR-hsp43-*lacZ*, yield-ing plasmid C683.

## 2.3 In situ hybridization, immunocytochemistry and microscopy

#### 2.3.1 Embryo fixation

Embryos were dechorionated and fixed as described (Kosman *et al.*, 2004; Rohr *et al.*, 1999; Stauber *et al.*, 2002), with the following modifications. All embryos were fixed for 25 minutes in 500  $\mu$ l fixation buffer consisting of 50 mM EGTA (pH 8.0), 8% (*Megaselia, Clogmia*) or 4% (*Drosophila, Episyrphus*) formaldehyde in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4), and 500  $\mu$ l n-heptane. As required, vitelline membranes were manually removed using a pair of tungsten needles.

#### 2.3.2 RNA in situ probes

RNA antisense probes for whole mount *in situ* hybridization were prepared as described (Lehmann and Tautz, 1994) with modifications. The template vector was linearized at the 5' end of the insert to avoid run-off transcripts, and only RNA probes larger than 1.2 kb were carbonate-treated (Lehmann and Tautz, 1994). RNA probes were labeled using an NTP mix including either digoxigenin- (DIG), fluorescein- (FITC), or biotin- (BIO) conjugated UTP analogues (Roche) as the substrate for RNA synthesis. The yield of the probe synthesis was determined on an agarose gel in comparison with a DNA standard. The RNA probes used for whole mount *in situ* hybridization, the UTP analogues used to label each probe, and the templates for the probes are listed in Table 3.

#### 2.3.3 Whole-mount RNA in situ hybridization

Whole-mount *in situ* hybridizations were performed as described (Kosman *et al.*, 2004) with modifications. Postfixation after the xylene washes was omitted. Probes were used at a final concentration of 1-2 ng/ $\mu$ l. For histochemical probe detection, 5% goat serum in PBT was used as blocking reagent, and alkaline phosphatase conjugated Fab fragments against DIG, FITC or BIOTIN (Roche), depending on the modification of the UTP analogue in the probe. Staining was performed as described (Tautz and Pfeifle, 1989); all embryos of a developmental series were stained equally long. For fluorescent probe detection, FITC-labeled *Episyrphus hunchback* was detected using a rabbit anti-FITC (1:300 diluted, Molecular Probes) as the primary and an A488 conjugated goat anti-rabbit (1:400, Molecular Probes) as a secondary

antibody. BIOTIN-labeled *Episyrphus zerknüllt* was detected using a mouse anti-BIOTIN (1:400, Roche) as the primary and a Cy3 conjugated goat anti-mouse (1:400, Jackson ImmunoResearch) as secondary antibody. The embryos were mounted in 70% glycerol in PBS, supplemented with 4% N-propyl gallate for fluorescent microscopy. Confocal scans were taken on a Leica SP2 AOBS Spectral Confocal Microscope. For 3D projection of image stacks ImageJ software (Wayne Rasband, National Institute of Mental Health, Bethesda, Maryland, USA) was used. Images were finished in Photoshop 7 (Adobe).

## 2.4 cDNA library from embryonic pole cytoplasm

#### 2.4.1 cDNA preparation from pole cytoplasm

Synthesis and amplification of cDNA from pole cytoplasm were performed as previously described (Brady and Iscove, 1993; Dulac and Axel, 1995; Kramer, 2000) with modifications. Pole cytoplasm was isolated from the embryo using a Narishige XYZ micromanipulator system with a Narishige IM-300 Microinjector (version 8.2A). Needles were prepared from glass capillaries (A-M Systems: 615000; glass, filament, thin-wall, 1.0 mm x .75 mm, 4") using a Flaming/Browning Micropipette puller (Sutter Instrument: Model P-87) with a trough-style heating element (pulling parameters: pressure: 505; heat: 560; pull: 100; velocity: 40; time: 100). Needles were ground using a Narishige's EG-44 capillary grinder at a speed of 8.0 at 30° for 40-45 seconds, which produced a pore small enough to allow for control of capillary forces. Prior to use, the needles were UV-irradiated. They were loaded from the tip with approximately 0.1  $\mu$ l of cDNA lysis buffer (1X MMLV buffer [Invitrogen] with 0.5% NP40 [USB], containing 24 µM pd(T)<sub>24</sub> [IDT], 0.2 U/µl SuperRNaseIn [Ambion], 0.3 U/µl RNAguard [Amersham], and 20 µM each of dATP, dCTP, dGTP, dTTP [Roche]). Approximately 0.5% of the total egg volume was taken up from anterior cortical egg cytoplasm by carefully releasing the balance pressure. The contents of the entire needle were then cleared from the needle into thin-welled 0.5 ml microfuge tubes (Costar) with 4.5  $\mu$ l of cDNA lysis buffer. The isolated cytoplasm was subsequently dissociated for one minute at 65 °C followed by an annealing step for the pd(T)<sub>24</sub> oligonucleotide at room temperature (22 °C) for 2 minutes. Firststrand cDNA synthesis, terminal transferase, and cDNA amplification were then performed exactly as described (Kramer, 2000). To amplify the cDNA, however, a modified AL1 primer was used in order to introduce a *Not*I site within the linker region (AL1mod: 5'-AGCGGCCGCGAATCC(T)<sub>24</sub>). To avoid a bias towards smaller transcripts during the PCR-based cDNA amplification, the first strand synthesis conditions were chosen to generate cDNAs of around 100-700 bases regardless of the size of the original RNA template (Brady and Iscove, 1993). Following cDNA amplification, the PCR reactions were stored in aliquots at -80 °C.

#### 2.4.2 Preparation of bacterial libraries from amplified cDNA

The PCR products were size-selected on a 2% agarose gel (300-1000 bp), eluted with QIAquick Gel Extraction Kit (Qiagen), and ligated into pCRII-TOPO vector. Transformation was performed by electroporation using a MicroPulser (BioRad). Random colonies were hand-picked into 384-well plates (Genetix) and grown overnight in 2YT media supplemented with 7% glycerol as described (Dunham *et al.*, 1997). Libraries were stored at -80°C. The bacterial libraries were spotted in multiple replicates onto Hybond-N<sup>+</sup> nylon membrane using a QBot spotting robot (Genetix). For the *Megaselia* library, a low-density spotting scheme (4 twin spots per 3x3 mm) was used; for the *Clogmia* libraries, the spotting density was doubled (8 twin spots per 3x3 mm). Spotted filters were processed as described (Dunham *et al.*, 1997).

#### 2.4.3 Hybridization of the libraries and subtractive screening

For subtractive screening, different filter replicates of a spotted library were hybridized with radioactively labeled cDNA from the anterior and the posterior pole, respectively. cDNA pools were radioactively labeled by PCR with 1  $\mu$ Ci/ $\mu$ l  $\alpha$ [P<sup>32</sup>] dCTP (Amersham, specific activity: 3000 Ci/mmol) exactly as described (Kramer, 2000). Hybridizations were performed in Rapid-hyb buffer (Amersham) at 65 °C for 2-3 hours. Excess radioactivity was removed by washing the filters in 0.2x SSC (3 mM sodium citrate; 30 mM NaCl, pH 7) and 0.1% SDS at 65 °C for 30 minutes. Following hybridization, the filter replicates were exposed overnight to Storage Phosphor Screens (Molecular Dynamics). The screens were read with a Storm860 Scanner (Molecular Dynamics). The read-outs were exported as grayscale images using ImageQuant version 1.2 for Macintosh (Molecular Dynamics). Brightness and contrast were uniformly adjusted in Photoshop 7. The images were aligned in Freehand 11 (Macromedia) and

imported in Photoshop using the red and the green channel of an RGB images for hybridization experiments with anterior and posterior cDNA, respectively. Red signals indicated hybridization predominantly with anterior cDNA and were identified by eye.

#### 2.4.4 Virtual northern hybridization

Samples of 10  $\mu$ l of PCR amplified cDNA were separated on a 2% agarose gel together with the 1 kb DNA ladder (Invitrogen) as size standard. Separated cDNA was transferred to Hybond-N<sup>+</sup> nylon membranes (Sambrook and Russel, 2001) and hybridized in Rapid-hyb buffer at 65 °C for 2-3 hours to radioactively labeled probes. Probes were labeled radioactively using the Rediprime II Random Prime Labeling System. Nonspecifically bound probe molecules were removed by washing in 0.2x SSC and 0.1% SDS at 65 °C for 30 minutes.

## 3 Results

# 3.1 Identification of *bicoid* orthologues from *Platypeza* and *Lonchoptera*

The presence of a functionally conserved *bicoid* gene in *Megaselia* and its apparent absence in the lower dipterans indicates that this activator of zygotic *hunchback* transcription originated in the stem lineage or early radiation of cyclorrhaphan flies (see Introduction). To more precisely determine the occurrence of *bicoid* in lower cyclorrhaphans, a PCR-based screen for *bicoid* orthologues was performed in *Lonchoptera*, *Platypeza*, and *Episyrphus* (Material and Methods 2.2.3). The exact phylogenetic relationship of these taxa has not been firmly established, but they constitute a very broad sample of lower cyclorrhaphans and are probably of paraphyletic origin (Figure 1). For each taxon, amplified homeobox fragments were recovered and compared to the GenBank database (Benson *et al.*, 2006) using the BLAST algorithm (Altschul *et al.*, 1997). The results suggest that the homeobox sequences from *Platypeza* and *Lonchoptera* are orthologous to *bicoid* (data not shown). Homeobox fragments from *Episyrphus* that were isolated using degenerate *bicoid* primers proved to be an *orthodenticle* orthologue (see Results 3.6.3) or a *zerknüllt* orthologue, which will be described elsewhere (Rafiqi *et al.*, in preparation). Despite multiple attempts, a *bicoid*-like sequence could not be recovered from *Episyrphus* (see Materials and Methods 2.2.3 for details on the cloning strategy and Results 3.5 for additional evidence supporting the absence of bicoid from Episyrphus). For the putative bicoid homologues of Lonchoptera and Platypeza, the corresponding cDNAs were cloned (Material and Methods 2.2.3), and the predicted homeodomain sequences were aligned with the Bicoid homeodomains of Drosophila and Megaselia (Figure 2). The homeodomains from Lonchoptera and Platypeza carry a lysine (K) at position 50 and an arginine (R) at position 54, which are both characteristic amino acids for Bicoid homeodomains and essential for their binding specificity to nucleic acids (Dave et al., 2000; Hanes and Brent, 1989; Niessing et al., 2000; Treisman et al., 1989). The homeodomains from Lonchoptera and *Platypeza* share significantly higher similarities with the homeodomains of *Drosophila* and Megaselia Bicoid than with the homeodomains of Zerknüllt or Orthodenticle (Figure 2). These observations strongly suggest that the newly identified homeobox sequences are orthologous to bicoid. An alignment of the entire open reading frames (Figure 3) reveals sequence conservation not only in the functional homeodomain but also in all additional domains that are known to be required for Bicoid function (reviewed in McGregor, 2005). Together with the functional analysis of Megaselia bicoid (Stauber et al., 2000), this high degree of sequence conservation indicates that the newly identified homologues are also functionally similar to bicoid in Drosophila. Since previous searches for a bicoid orthologue in Anopheles (Zdobnov et al., 2002), Empis, Haematopota and Clogmia (Stauber et al., 2002) were negative, I conclude that of the eight species compared in this study, at least three (Megaselia, Platypeza, Lonchoptera) – and probably only these three – contain bicoid.

## 3.2 Subtractive screening for *bicoid*-like genes: a new method

To screen *Episyrphus* and lower dipterans for genes that encode anterior localized transcripts, I developed a new method based on protocols to prepare cDNA libraries from single cells (Brady and Iscove, 1993; Dulac and Axel, 1995; Kramer, 2000). I tested this protocol in *Megaselia* using *Megaselia bicoid* as a positive control. cDNA was synthesized and PCR amplified from RNA that was isolated from the anterior and posterior pole cytoplasm of an hourold *Megaselia* embryo, respectively (Figure 4 A, Materials and Methods 2.4.1). At this intravitteline cleavage stage, the embryo contains only maternal mRNAs. The amplified cDNA pools were blotted and analyzed by hybridization against a radioactively labeled *bicoid* probe (virtual northern, Material and Methods 2.4.4). As expected, *Megaselia bicoid* cDNA was detected only in cDNA pools prepared from anterior cytoplasm (Figure 4 B). Next, a cDNA library was prepared from the anterior cDNA pool (Material and Methods, 2.4.2). To determine the relative abundance of *Megaselia bicoid* clones in this library, an estimated 5,000 bacterial clones of the library were hybridized with a labeled *Megaselia bicoid* probe. 48 of these colonies hybridized to the probe, indicating that approximately one in a hundred clones contain a *Megaselia bicoid* cDNA (data not shown).

To test whether *Megaselia bicoid* could be reliably identified by subtractive screening, 1,536 bacterial clones were spotted, grown, and lysed on nylon filters according to a predetermined twin spot scheme (Figure 5 A, Material and Methods 2.4.2). One filter replica of the library was hybridized with a radioactively labeled pool of cDNAs, which had been prepared from posterior cytoplasm. Signals obtained from this hybridization were shaded green. A second filter was hybridized against a pool of cDNA prepared from anterior cytoplasm, and signals resulting from this hybridization were shaded red. Both images were then merged (Figure 5 B). In these merged images, green and yellow twin signals indicated hybridization with the posterior cDNA pool, and the corresponding clones were excluded from further analysis. Clones that hybridized only with the anterior cDNA pool can be identified as red twin spots. A sample of 14 clones was sequenced. Of this sample, only the clones with a strong red signal contain Megaselia bicoid (see boxes in Figure 5 B). This result indicates that the subtraction of non-localized transcripts is efficient and provides a filter for undesired but presumably abundant cDNAs from housekeeping genes in each cDNA sample. 'False positive' clones (red twin spots of clones, which do not contain Megaselia bicoid) were not observed. To test for 'false negative' clones (Megaselia bicoid-containing clones which remain hidden in the subtractive screen), a third filter replicate was hybridized with labeled Megaselia bicoid cDNA. This control reveals additional putative *Megaselia bicoid* clones (Figure 5 C), indicating that the subtractive screen detects only a subset of the Megaselia bicoid clones. Out of 28 clones positive for Megaselia bicoid (1.8 % of all spotted clones), only four were detected after subtraction, indicating a relatively high rate (86%) of 'false negative' clones.

The results demonstrate that a gene encoding a strictly anterior localized transcript, such as *Megaselia bicoid*, can be cloned using the described cDNA subtraction screen. Furthermore, the results suggest that only *bicoid* transcripts are abundantly localized to the anterior pole of *Megaselia* embryos.

## 3.3 Subtractive screening for *bicoid*-like genes in *Clogmia*

The subtractive screen was designed to screen either for a potentially missed *bicoid* homologue in *Episyrphus* (see Results 3.1) or to screen non-cyclorrhaphan dipterans without a *bicoid* homologue for anterior localized transcripts (see Introduction 1.3.3). Because only the *Clogmia* culture was readily available then, I decided to test the hypothesis that *Clogmia*, a non-cyclorrhaphan dipteran (Figure 1), uses localized mRNA as an anterior determinant.

It has been shown that *bicoid* transcripts diffuse slightly during early development (Stauber *et al.*, 2000). Without knowing where exactly transcripts might be localized in the anterior cortex of *Clogmia* embryos, I therefore decided to prepare cDNA libraries from anterior poleplasm of *Clogmia* embryos from two consecutive intravitelline cleavage stages prior to the onset of zygotic transcription. A first cDNA library from anterior pole cytoplasm was prepared from a one hour-old embryo, with the chance to isolate cytoplasm highly enriched in anterior localized transcripts, but with the risk that the wrong portion of cytoplasm would be chosen or that the transcripts might still be too tightly localized to be efficiently removed. A second library was prepared from a three hour-old embryo, with a higher chance of retrieving relevant cytoplasm, but with the risk that the transcripts would be too dilute to be successfully amplified and enriched during the cloning process.

Roughly 14,000 clones of each library were spotted and screened as described for *Megaselia*. A total of 161 clones were selected and sequenced. Of those clones, 6% (9 out of 161) contained empty vectors, and 9% (15/161) contained multiple and therefore unreadable inserts. 45% (72/161) of the clones contained ribosomal subunits or riboproteins, and 17% (27/161) of the clones contained human genes (Figure 6 A). The remaining 23% (38/161) of the sequenced clones comprised ten different cDNAs (Figure 6 B) which could not be placed into any of the former categories. In order to assess possible functions of these ten cDNAs, their sequences were compared with genes from two virtual GenBank databases (Cummings *et* 

al., 2002) that either contained only *Drosophila* genes or all Arthropoda genes (Table 4). Significant support for sequence homology (BLAST Expect values < 0.01) to Drosophila genes was found in six of the ten cDNAs. The corresponding six Drosophila genes are the gene Odorant-binding protein 99a (Obp99a, (Galindo and Smith, 2001)), which encodes a member of a large family of proteins that bind lipophilic odorant molecules (Vogt et al., 1991), the gene exuperantia (exu, (Marcey et al., 1991; Schüpbach and Wieschaus, 1986)), which encodes a protein required for bicoid mRNA localization (Berleth et al., 1988), the putative Histone 3-encoding gene His3:CG31613 (Drysdale et al., 2005), the gene Decondensation factor (Df31, (Crevel and Cotterill, 1995)), encoding a chromatin associated component (Crevel et al., 2001), the gene CG14764 with currently unknown function (Drysdale et al., 2005), and the gene CG1967, encoding a putative p24 protein (Liang and Biggin, 1998), which is involved in intracellular post-golgi transporter activity (reviewed in Carney and Bowen, 2004). Insignificant homology scores to Drosophila genes were obtained for the remaining four of the ten cDNAs (BLAST Expect values > 2.5). The corresponding four *Drosophila* genes are CG6459, which encodes a putative component of the mitochondrial matrix (Drysdale et al., 2005), jing (Karpen and Spradling, 1992; Liu and Montell, 2001), which encodes a C<sub>2</sub>H<sub>2</sub> zinc finger transcription factor (Liu and Montell, 2001), the Amylase distal gene (Amy-d (Boer and Hickey, 1986)), and the Graf gene (Drysdale et al., 2005), which encodes a product with putative Rho GTPase activator activity (Drysdale et al., 2005).

Based on their functions and their high similarity to *Anopheles* homologues, seven cDNAs were selected for further analyses (Table 4). To test whether these cDNAs were enriched in the anterior pole of the embryo, the expression profiles of *Obp99a like* (Figure 7 A), *CG14761 like* (Figure 7 B), *Df31 like* (Figure 7 C), *exu like* (Figure 7 D), *jing like* (Figure 7 E), *CG6459 like* (Figure 7 F), and *CG1967 like* (Figure 7 G), respectively, were analyzed in virtual northern blots. A radioactively labeled probe of each cDNA was hybridized to pools of amplified cDNA from anterior and posterior cytoplasm of four *Clogmia* embryos (Material and Methods 2.4.4). None of the seven cDNAs always and exclusively hybridized with the cDNA pools from anterior poleplasm. Six of the analyzed cDNAs were excluded from further analyses because they hybridized to amplified cDNA from anterior and posterior and posterior and posterior cytoplasm. *CG6459 like* hybridized to the anterior fraction of one preparation but not at all to the cDNA pools
from other embryos. Expression of this candidate was tested by whole-mount *in situ* hybridization of *Clogmia* embryos. Using an antisense RNA probe of *CG6459 like*, staining was detected weakly but ubiquitously throughout the embryo; using a respective sense probe of *CG6459 like*, staining was not observed (data not shown).

Although most cDNAs hybridized to amplified cDNA from both anterior and posterior cytoplasm, some gave a stronger signal with the cDNA pool from which the library was made (e.g. *jing like*, Figure 7 E, embryo 3; or *CG1967 like*, Figure 7 G, embryo 1). These observations suggest that the cDNAs of evenly distributed transcripts may be under- or overrepresented in individual cDNA pools. Such biologically insignificant artifacts could be caused by a bias in the amount of removed cytoplasm or PCR amplification of cDNAs after reverse transcription. Together, the results suggest that *Clogmia* might lack abundant, strictly localized maternal transcripts at the anterior pole of early developing eggs.

### 3.4 Cloning of dipteran hunchback genes

The restricted occurrence of *bicoid* in dipterans suggests differences in early *hunchback* regulation between *Episyrphus* and lower dipterans on one hand and other cyclorrhaphans on the other. To test this hypothesis, I compared the regulation of eight dipteran hunchback homologues in transgenic Drosophila embryos. For these investigations, hunchback cDNA containing 5' UTR with putative leader sequence was mapped onto genomic DNA. cDNAs including 5' UTRs were newly isolated from Episyrphus, as well as from Megaselia, Lonchoptera, Platypeza, Empis, Haematopota, Clogmia, and Anopheles (Table 2 and Materials and Methods 2.2.5). Genomic hunchback DNA was isolated from Episyrphus, Megaselia, Platypeza, Lonchoptera, Empis, Haematopota, and Clogmia, employing genomic phage libraries and/or PCR on genomic DNA (Materials and Methods 2.2.6). Genomic DNA sequence of Anopheles hunchback was obtained directly from the sequenced Anopheles genome (Holt et al., 2002). Protein trees based on the predicted amino acid sequences of the N-terminal zinc finger domain (amino acids 243-349 in the Drosophila protein, Figure 8) together with the alignment of the predicted amino acid sequences of the entire open reading frames (Figure 9) strongly suggest that the newly identified genes are *hunchback* orthologues. The alignment reveals sequence conservation not only in the functional zinc finger domains but also in several additional motifs that are thought to be specific for the *hunchback* protein, such as the A-, C-, D-, E, and F-boxes, the molecular functions of which, however, are still unknown (Hülskamp *et al.*, 1994; McGregor *et al.*, 2001a; Tautz *et al.*, 1987; Figure 9).

In Megaselia, a second transcript was isolated in addition to the previously identified maternal transcript (Stauber et al., 2000). Both transcripts differ in their first exon. The maternal transcript derives from the distal promotor (P1, Figure 10); the newly identified transcript isolated from early embryos is probably zygotic (see Material and Methods 2.2.5) and derives from the proximal promotor (P2, Figure 10). These findings suggest that the genomic organization of hunchback is conserved between Megaselia and Drosophila and that the P1 and P2 transcripts of both species are directly homologous. Two alternative transcripts with differing 5' UTRs were also identified in *Platypeza* (Figure 10). These splicing variants, however, were obtained from larval tissue (embryos were not available), and it is unclear whether they are homologues to the maternal and zygotic hunchback transcripts of Drosophila and Megaselia. In Lonchoptera, a single maternal transcript was detected in adult females (embryos were not available; Figure 10). In Episyrphus, three splice variants with alternative 5' UTRs were identified from pools of 0-5 hours old embryos (Figure 10). In all lower dipterans (*Empis, Haema*topota, Clogmia, Anopheles), only one splice variant was isolated (Figure 10). In Empis, the occurrence of only a single splice variant was confirmed by comparing cDNAs that were isolated from ovarian and embryonic cDNA templates. In Clogmia, the occurrence of only a single splice variant was confirmed by comparing cDNAs from 0-2 hour-old and 5-6 hour-old embryonic libraries (onset of zygotic transcription at about 4 hours of development) and by developmental Northern analysis (Prell and Schmidt-Ott, unpublished; supplemental Figure S1). Together, the data suggest that, unlike cyclorrhaphans, non-cyclorrhaphan dipterans use the same *hunchback* splice variant during oogenesis and early embryogenesis.

In the putative 3' UTRs of *Megaselia*, *Platypeza*, *Lonchoptera*, *Episyrphus*, *Haema-topota* and *Clogmia hunchback* sequences, I identified putative NRE sequences (Figure 11), which all reside within 0.6 kb downstream of the ORF (Figure 10). The presence or absence of NRE sequences in *Empis* could not be determined due to limited sequence information. In *Anopheles*, NRE sequences could not be identified within 8.0 kb downstream of the ORF, which is consistent with the reported absence of maternal *hunchback* expression in this species

(Goltsev *et al.*, 2004a). These findings support previous studies, which suggested that translational repression of maternal *hunchback* by Nanos is conserved in dipterans (Curtis *et al.*, 1995).

### 3.5 Functional comparison of early dipteran hunchback regulation

To functionally compare the transcriptional regulation of the *hunchback* homologues, I cloned reporter constructs with putative regulatory DNA of each *hunchback* homologue (Figure 10, Table 5) and compared their expression in transgenic *Drosophila* embryos (Figure 12). In the case of *Megaselia*, the entire *hunchback* locus was tested. All other constructs include a strong basal *Drosophila* promoter (eve or hs43, respectively) and the *lacZ* gene as reporter (see Materials and Methods 2.2.6). For each construct, two to four independent stable transgenic *Drosophila* lines were established by P-element mediated germline transformation (Rubin and Spradling, 1982; Table 5). The transgenic expression patterns of the reporter genes were compared to endogenous *hunchback* expression in *Drosophila* (Figure 12 A, B; Tautz *et al.*, 1987), *Megaselia* (Figure 12 C, D; Stauber *et al.*, 2000), and *Clogmia* (Figure 12 E, F; Rohr *et al.*, 1999).

At the onset of blastoderm cellularization, the reporter of the *Megaselia* construct is activated in the anterior half of the embryo. During cellularization, anterior reporter expression disappears from the anterior-most portion of the embryo, while a new domain appears at the posterior pole (Figure 12 G, H). This pattern resembles the endogenous expression of *hunchback* in *Drosophila* and *Megaselia* (Figure 12 A-D).

The *Platypeza* construct is also expressed in the anterior half of syncytial blastoderm embryos, excluding, however, the anterior 20% of the embryo. During cellularization, this expression extends ventrally towards the anterior pole, and reporter expression also appears at the posterior pole of the embryo (Figure 12 I, J).

The *Lonchoptera* construct drives reporter expression in an anterior stripe from 90-75% EL (egg length; 0% at the posterior pole) in syncytial blastoderm embryos but, unlike the *Megaselia* and *Platypeza* constructs, expression during cellularization is less dynamic and is absent in the posterior half of the embryo (Figure 12 K, L).

The *Episyrphus* construct, in contrast to the other cyclorrhaphan constructs, is expressed exclusively in the posterior half of the syncytial blastoderm embryo. During cellularization, the expression disappears from the posterior pole, resulting in a broad stripe of expression from 50-20% EL (Figure 12, M, N). Thus, this reporter expression is roughly complementary to the endogenous early *hunchback* expression of *Drosophila* (Figure 12 A, B).

For *Haematopota*, two constructs were analyzed. The larger construct includes 9 kb of genomic DNA upstream of the ORF (H1). This construct initially drives posterior reporter gene expression in the syncytial blastoderm, which is then cleared from the posterior pole (Figure 12 O, P) during cellularization. This expression is similar to both the early expression of the *Episyrphus* construct (Figure 12 M, N) and the expression of a comparable *Tribolium* construct (Wolff *et al.*, 1998). Subsequently, the *Haematopota* construct is also expressed in a weak anterior stripe, which appears as a second domain towards the end of cellularization (Figure 12 P). Similar expression has also been reported for the *Tribolium* construct (Wolff *et al.*, 1998). A shorter *Haematopota* construct which included only the 1.8 kb intron sequence of *Haematopota hunchback* (H2) was not expressed in pregastrular embryos (data not shown).

For *Clogmia*, both a larger construct (K2), comprising 6.9 kb upstream of the ORF, and a shorter construct (K13), spanning only intron sequence, drive reporter gene expression in a dorsal domain and in a weak transverse stripe at 65-60 % EL of the blastoderm and subsequently in the developing amnioserosa of the gastrulating embryo (Figure 12 Q, R). Expression of these constructs in the dorsal blastoderm and the amnioserosa is reminiscent of the endogenous extraembryonic expression of *Clogmia hunchback* shortly before the onset of gastrulation (Figure 12 E, F).

The constructs with genomic DNA from *Empis hunchback* and *Anopheles hunchback* (Figure 10) are not expressed in pregastrular *Drosophila* embryos (data not shown).

Although the results of these enhancer analyses are heterogeneous, at least two aspects deserve attention. First, reporter expression of the *Megaselia* and *Platypeza* constructs in transgenic *Drosophila* is similar to endogenous *hunchback* expression patterns in *Drosophila* (Tautz *et al.*, 1987) and *Megaselia* (Stauber *et al.*, 2000). This finding does not exclude the possibility of substitutions among *hunchback* regulators between these species, but it is more parsimonious to explain the results with an essentially conserved regulatory network for early

*hunchback* activation between *Drosophila*, *Megaselia*, and *Platypeza*. This conclusion is supported by RNAi knockdown of *bicoid* in *Megaselia*, which causes a duplicated posterior *hunchback* expression at the anterior pole (Figure 13), and by the presence of Bicoid-binding sites within the P1 intron upstream of P2 of *Megaselia hunchback* (Shaw and Schmidt-Ott, unpublished; supplemental Figure S3). The *Lonchoptera* data are consistent with this hypothesis but more difficult to interpret because expression in an anterior head stripe may occur as an artifact (Klingler *et al.*, 1996). The second finding of special interest is that expression of the *Episyrphus* construct is confined to the posterior blastoderm. The expression of this construct is significantly different from the expression of all other cyclorrhaphan reporter constructs and resembles the posterior expression patterns that were obtained with the *Haematopota* and *Tribolium* constructs (Wolff *et al.*, 1998). This raises the question, whether these species use, at least in part, a similar Bicoid-independent mechanism of *hunchback* regulation. The mechanism could be dependent on the transcription factor Caudal as has been suggested previously for *Tribolium* (Wolff *et al.*, 1998).

### 3.6 Expression studies in *Episyrphus*

To explore whether *Episyrphus*, besides a putative lack of Bicoid-dependent *hunchback* regulation, also shares characteristics in early pattern formation with non-cyclorrhaphan dipterans and other holometabolous insects, I decided to study the expression of *Episyrphus hunchback*, *zerknüllt* as a marker for extraembryonic tissue, and *orthodenticle* as a potential alternative to Bicoid as *hunchback* activator. The results suggest that early pattern formation in *Episyrphus* is a mosaic of pattern formation in cyclorrhaphan and non-cyclorrhaphan dipterans.

3.6.1 *Episyrphus hunchback* shares expression characteristics of cyclorrhaphan and non-cyclorrhaphan dipterans

Cyclorrhaphans express *hunchback* in a posterior domain (Bonneton *et al.*, 1997; McGregor *et al.*, 2001a; Sommer and Tautz, 1991; Stauber *et al.*, 2000; Tautz *et al.*, 1987; Treier *et al.*, 1989), while only lower dipterans express *hunchback* in the extraembryonic blastoderm (Goltsev *et al.*, 2004a; Rohr *et al.*, 1999). The loss of this expression in cyclorrhaphan

dipterans correlates with the occurrence of Bicoid (Berleth et al., 1988; Gregor et al., 2005; Schröder and Sander, 1993; Seeger and Kaufman, 1990; Shaw et al., 2001; Sommer and Tautz, 1991). I studied the expression of *Episyrphus hunchback* throughout early development and compared it to both lower and higher dipterans. In the freshly laid egg, the maternal transcripts of *Episyrphus hunchback* are evenly distributed (Figure 14 A) but disappear from the posterior half during blastoderm formation (Figure 14 B, C). A distinct increase of Episyrphus hunchback expression throughout the anterior half of the syncytial blastoderm embryo marks the onset of zygotic expression (Figure 14 D). During cellularization, a second expression domain appears at the posterior pole (Figure 14 E), and expression in the anterior half resolves into a prominent stripe from about 60-55% EL (Figure 14 E, F). This pattern closely resembles hunchback expression in other cyclorrhaphans (Figure 12 A-D; Bonneton et al., 1997; McGregor et al., 2001a; Sommer and Tautz, 1991; Stauber et al., 2000; Tautz et al., 1987; Treier et al., 1989), but differs from pregastrular hunchback expression in lower dipterans, which lack the posterior domain (Figure 12 E, F; Goltsev et al., 2004a; Rohr et al., 1999). However, Episyrphus hunchback is also expressed in a mid-dorsal stripe of the blastoderm, which expands from anterior to posterior (Figure 14 G-L). Dorsal hunchback expression is absent in other cyclorrhaphans but reminiscent of hunchback expression in the extraembryonic anlage of lower dipterans (Figure 12 E, F; Goltsev et al., 2004a; Rohr et al., 1999), Tribolium (Wolff et al., 1995), and Nasonia (Pultz et al., 2005). With the onset of gastrulation, the dorsal expression of Episyrphus hunchback broadens and transcripts are predominantly detected in a narrow band along the margins of the extraembryonic primordium (Figure 14 I, L). Similar expression dynamics have been reported for *Tribolium hunchback* (Wolff *et al.*, 1995). Thus, *Episyrphus hunchback* expression in the early embryo shares characteristics specific for cyclorrhaphans (posterior expression domain), for lower dipterans (dorsal/extraembryonic expression domain), and with all dipterans it shares early expression in an anterior cap.

#### 3.6.2 The extraembryonic anlage of Episyrphus extends to the anterior pole

To further explore the hypothesis that *Episyrphus hunchback* is expressed in the extraembryonic anlage, I compared the dorsal *hunchback* expression with the expression of *zerknüllt*, a conserved marker for the extraembryonic anlage (Falciani *et al.*, 1996). During the onset of gastrulation, dorsal *hunchback* expression (Figure 15 A, B) and *zerknüllt* expression (Figure 15 C, D) perfectly overlap (Figure 15 E, F). This result not only indicates that *Episyrphus hunchback* is expressed in the extraembryonic anlage, but it also suggests that in *Episyrphus* the extraembryonic anlage extends to the anterior pole. In *Drosophila* and *Megaselia*, both of which use *bicoid* as an anterior determinant, the extraembryonic anlage is restricted to the dorsal-most blastoderm while the anterior blastoderm gives rise to embryonic structures (Campos-Ortega and Hartenstein, 1997; Rushlow and Levine, 1990; Stauber *et al.*, 1999). In many insects that lack *bicoid*, the extraembryonic (serosal) primordium extends to the anterior tip of the blastoderm (e.g. *Clogmia* (Rohr *et al.*, 1999), the honey bee *Apis mellifera* (Fleig and Sander, 1988), and *Tribolium* (Wolff *et al.*, 1995)).

To test whether the extraembryonic anlage of *Episyrphus* is structurally more closely related to non-cyclorrhaphan dipterans than to Drosophila, Episyrphus zerknüllt expression was further analyzed. In pre-blastoderm embryos, Episyrphus zerknüllt transcripts could not be detected by whole-mount in situ hybridization (Figure 16 A), suggesting that, like in other cyclorrhaphans, zerknüllt is not maternally expressed in Episyrphus. During the early blastoderm stage, zygotic transcripts appear in a broad dorsal domain with an enrichment of transcripts at the anterior pole (Figure 16 B). During cellularization of the blastoderm, anterior zerknüllt expression extends in a mid-dorsal stripe, while all other expression disappears (Figure 16 C-I). At the onset of gastrulation, Episyrphus zerknüllt is exclusively expressed in dorsal stripe, which extends from the anterior pole to about 15% EL (Figure 16 E, H). This expression domain marks, probably precisely, the anlage of the prospective serosa (Rafiqi et al., in preparation). At the onset of germband extension, the expression follows the spreading of the serosa (Figure 16 F, I, J-L). Unlike Clogmia zerknüllt, Episyrphus zerknüllt is not expressed maternally. However, the zygotic expression of *Episyrphus zerknüllt* is very similar to zygotic expression of *Clogmia zerknüllt*, which appears in a slightly broader domain and does not extend quite as far to the posterior pole as *Episyrphus zerknüllt* at a comparable stage (supplemental Figure S2 A-D; Stauber et al., 2002). Other cyclorrhaphans, such as Drosophila and Megaselia, share with Episyrphus the absence of maternal zerknüllt transcripts in early embryos, but they differ in that their zygotic *zerknüllt* expression domains do not extend to the anterior pole (supplemental Figure S2 E-H; Doyle et al., 1986; Stauber et al., 1999). Thus, the expression of *Episyrphus zerknüllt* shares similarities with both lower dipterans such as *Clogmia*, as well as other cyclorrhaphans, and might best be described as an intermediate.

#### 3.6.3 *Episyrphus orthodenticle* is not expressed in pre-blastoderm embryos

Recent studies propose that maternal *orthodenticle* activity substitutes for *bicoid* functions in *Tribolium* and *Nasonia* (Lynch *et al.*, 2006; Schröder, 2003). A putative *orthodenticle* homologue was cloned from *Episyrphus* in an attempt to isolate *bicoid*. In an alignment of the predicted amino acid sequence with *Drosophila orthodenticle/ocelliless* (Finkelstein *et al.*, 1990) (Figure 17), the putative *Episyrphus* homologue shows over 75% sequence similarity with one of the putative *Drosophila orthodenticle/ocelliless* protein isoforms, suggesting that the newly identified gene is *Episyrphus orthodenticle*.

To test whether *Episyrphus orthodenticle* is expressed maternally, I studied the expression of this gene. In pre-blastoderm embryos, *Episyrphus orthodenticle* transcripts could not be detected by whole-mount *in situ* hybridization (data not shown), suggesting that, like in *Drosophila, orthodenticle* is not maternally expressed in *Episyrphus* (Finkelstein *et al.*, 1990). However, zygotic *Episyrphus orthodenticle* transcript is expressed in the anterior 20% of the embryo (Figure 18 A, B). During cellularization of the blastoderm, these transcripts disappear from the anterior-most blastoderm, and later also from the ventral-most region of the remaining anterior stripe (Figure 18 C-F). During cellularization of the blastoderm, *orthodenticle* expression begins to clear from the dorsal region, resulting in a dorsal stripe free of expression at the onset of gastrulation (Figure 18 D, F). Thus, expression of *Episyrphus orthodenticle* is reminiscent of the expression of this gene in *Drosophila* and *Anopheles* (Finkelstein *et al.*, 1990; Goltsev *et al.*, 2004a). Provided that *Episyrphus does* not contain additional *orthodenticle cle* genes, these data suggest that early zygotic *Episyrphus hunchback* expression throughout the anterior half of the embryo is under the control of a different gene.

### 4 Discussion

### 4.1 Do all cyclorrhaphan dipterans have *bicoid*?

It has been proposed that *bicoid* evolved in the stem lineage of cyclorrhaphan dipterans (Schmidt-Ott, 2000; Stauber et al., 2002). The identification of bicoid orthologues in *Platypeza* and *Lonchoptera* (Figures 2, 3) demonstrates that *bicoid* is in fact widely conserved in basal cyclorrhaphans. In addition, conserved sequence motifs of in the newly isolated bicoid homologues (Figure 3), the analyses of *hunchback* reporter gene expression in transgenic Drosophila embryos (Figure 12), bicoid RNAi data from Megaselia (Figure 13) (Stauber et al., 2000) and Bicoid binding sites upstream of the Megaselia hunchback P2 promoter (supplemental Figure S3) support the hypothesis of a conserved early patterning role of this gene, not only in higher (Bonneton et al., 1997; Driever and Nüsslein-Volhard, 1989; Driever et al., 1989; Lukowitz et al., 1994; McGregor; McGregor et al., 2001b; Shaw et al., 2001; Struhl et al., 1989) but also in lower cyclorrhaphans. However, the PCR-based screen for bicoid homologues in dipterans also suggests that Episyrphus lacks this gene. In Episyrphus, low stringency PCR with degenerate *bicoid* primers only yielded homeobox genes that are phylogenetically (zerknüllt) or functionally (orthodenticle) related to bicoid (Finkelstein et al., 1990; Lynch et al., 2006; Schröder, 2003; Stauber et al., 1999). This result raises the question how zygotic expression of zygotic hunchback throughout the anterior half of the early Episyrphus embryo (Figure 14 C-E) is activated. The question has been approached by the functional comparison of the *Episyrphus hunchback* enhancer with the *hunchback* enhancers from other cyclorrhaphans (in which *bicoid* has been identified) and non-cyclorrhaphan dipterans, respectively. In this comparison, regulatory DNA of Episyrphus hunchback (Figure 12 M, N) differs significantly from regulatory DNA of the other cyclorrhaphan hunchback homologues (Figure 12 G-L). Instead, it shares characteristics with the hunchback regulatory DNA of the noncyclorrhaphan insects Haematopota (Figure 12 Q, P) and Tribolium (Wolff et al., 1998). I cannot exclude the possibility that a Bicoid-response element in Episyrphus is located outside the sequence analyzed (complete intron of the P1 transcript) and was missed. However, all Bicoid-response elements of hunchback genes that have been characterized until now have been mapped to the intron of the P1 transcript (Bonneton et al., 1997; Driever and Nüsslein-Volhard, 1989; Driever et al., 1989; Lukowitz et al., 1994; Schröder et al., 1988; Shaw et al.,

2001; Struhl *et al.*, 1989; Treier *et al.*, 1989). More importantly, the congruence between the screening data for *bicoid* orthologues and the transgenic data with regulatory DNAs of *hunchback* homologues provides independent support for the hypothesis that a *bicoid* orthologue is absent in *Episyrphus*.

# 4.2 Does *Episyrphus* reflect the primitive patterning mechanism of cyclorrhaphan flies?

Episyrphus may have lost bicoid or may primarily lack this gene. Unfortunately, the position of syrphids within Cyclorrhapha is still unclear. Although there is agreement about assigning syrphids to lower cyclorrhaphans (Aschiza; Figure 1), the position within the Aschiza has been subject to controversy (reviewed in Collins and Wiegmann, 2002; reviewed in Yeates and Wiegmann, 1999). In recent studies, taxonomists have favored the hypothesis that Aschiza are paraphyletic and that syrphids, together with a second family (Pipunculidae), constitute the sister-group of Schizophora, which comprises all higher cyclorrhaphans (e.g. Moulton and Wiegmann, 2004, and references therein). This phylogenetic hypothesis is consistent with the fossil record (Grimaldi and Engel, 2005). Currently, the oldest putative syrphid fossil has been described in 80 million year old (myo) amber, while fossils of other basal ashizan taxa (Lonchopteridae, Platypezidae, Phoridae) have been found in 115-130 myo amber (Grimaldi and Engel, 2005). Considering this phylogeny, however, the outcome of my investigations is very unexpected. Not only the loss of an important developmental regulator (bicoid) has to be explained but also the anterior specification of extraembryonic blastoderm (Figure 16 D-L) and the extraembryonic expression of Episyrphus hunchback (Figure 14 J-L and Figure 15), all of which have been reported for non-cyclorrhaphan but not for cyclorrhaphan dipterans (Goltsev et al., 2004a; Rohr et al., 1999; Stauber et al., 2002). Alternatively, syrphids might be an outgroup to the cyclorrhaphans studied. Under this assumption, pattern formation in Episyrphus might reflect the primitive condition in cyclorrhaphans and could be considered as intermediate to lower and higher cyclorrhaphan dipterans.

# 4.3 How did early anterior *hunchback* regulation change in dipteran evolution?

Expression of zygotic hunchback in the anterior half of the embryo is conserved throughout the insect order Diptera (Bonneton et al., 1997; Goltsev et al., 2004a; McGregor et al., 2001a; Rohr et al., 1999; Sommer and Tautz, 1991; Stauber et al., 2000; Tautz et al., 1987; Treier et al., 1989). However, the blastoderm fate-map changed in dipteran evolution (Anderson, 1972) and it has been recently pointed out that the extreme expansion of the embryonic blastoderm to the anterior pole in cyclorrhaphans and culicomorphan mosquitoes may have been accompanied by the independent evolution of localized transcripts with a role in head specification (Schmidt-Ott, 2005). In both cyclorrhaphans with bicoid and culicomorphan mosquitoes, where a *bicoid*-like mRNA has been predicted (see Introduction), the extraembryonic anlage is restricted to dorsal blastoderm and the embryonic primordium extends to the anterior pole, while in other dipterans, the embryonic blastoderm is slightly smaller relative to the egg and the extraembryonic anlage extends to the pole (Anderson, 1972; Sander, 1976). The localization of a head inducing transcription factor to the anterior tip of the developing egg may have caused this shift in the fate map by repressing the extraembryonic anlage at the anterior pole. The high concentration of this transcription factor at the anterior pole may have also gradually shifted the balance between the ancestral activators of anterior hunchback expression in favor of the most abundant activator – Bicoid in case of cyclorrhaphan flies and a protein X in the case of culicomorphan mosquitoes.

The reverse argument could explain the results for *Episyrphus* and *Clogmia*: in both flies, the extraembryonic anlage extends to the anterior pole and in neither species an anterior localized transcript could be isolated. The implication of this argument is that dipterans may use three different modes of early *hunchback* activation: the *bicoid* dependent mechanism of most cyclorrhaphans, the ancestral dipteran mechanism, which might involve Nanos-dependent posterior repression and an auto-regulatory loop in the anterior, and a third mechanism in culicomorphan mosquitoes involving a localized transcript encoded by gene X.

## **Summary and Conclusions**

I have shown that the lower cyclorrhaphan fly *Episyrphus balteatus* (Syrphidae) combines patterning elements of lower and higher flies. Similarities between *Episyrphus* and lower flies/insects include the expression of *zerkniillt* at the anterior pole of the blastoderm and the expression of *hunchback* in the *zerkniillt* domain (extraembryonic anlage). In addition, I showed that the response of *hunchback* regulatory DNA of *Episyrphus* (5.4 kb upstream of the ORF) in transgenic *Drosophila* is functionally comparable to lower insects rather than higher flies with a *bicoid* gene. Similarities between *Episyrphus* and higher flies include the absence of maternal *zerkniillt* expression, and the expression of *hunchback* in a posterior stripe of the blastoderm. This mosaic of developmental traits from lower and higher dipterans suggests that early pattern formation in syrphids resembles the ancestral cyclorrhaphan patterning mechanism shortly before the emergence of *bicoid*.

The correlating occurrence of *bicoid* and the expansion of the embryonic blastoderm to the anterior pole suggests that both aspects of development are contingent on each other. A cDNA subtraction screen, which was developed during the course of this work, can be used to test this hypothesis in mosquitoes.

### A Appendix

### A.1 Figures and Tables

A.1.1 Figures



Figure 1. Phylogenetic relationships of taxa mentioned in the text. The insect order Diptera emerged 250 million years ago and about 125,000 comprises to 150,000 species (Grimaldi and Engel, 2005; Yeates and Wiegmann, 1999). The dipterans constitute a strongly supported monophyletic group (Yeates and Wiegmann, 1999). A common character, shared by all dipterans, is the transformation of the second wing pair into halteres (McAlpine, 1989). The Cyclorrhapha emerged about 150 million years ago and comprises about 65,000 species (Grimaldi and Engel, 2005; Yeates and Wiegmann, 1999). Cyclorrhaphans differ from non-cyclorrhaphan or "lower"

Diptera in a number of derived characters, which are shared by cyclorrhaphan flies only (e.g. invaginated head capsule of the larva – for a list see McAlpine, 1989). The suborder Cyclorrhapha has been further subdivided into the monophyletic Schizophora and Aschiza, which are probably paraphyletic (Yeates and Wiegmann, 1999). Species that have been analyzed in this work are shown in bold. Abbreviation: Cm, Culicomorpha. Quotes indicate paraphyletic sub-orders. Branch lengths are not to scale.



Figure 2. Homeodomain alignment and percent sequence similarity relative to the homeodomains of *Platypeza* Bicoid and *Lonchoptera* Bicoid. Abbreviations: Pco BCD, Bicoid of *Platypeza* (this work); Llu BCD, Bicoid of *Lonchoptera* (this work); Mab BCD, Bicoid of *Megaselia* (GenBank entry AJ133024, Stauber *et al.*, 1999); Dme BCD, Bicoid of *Drosophila* (GenBank entry X07870, Berleth *et al.*, 1988); Mab ZEN, Zerknüllt of *Megaselia* (GenBank entry AJ133025, Stauber *et al.*, 1999); Dme ZEN, Zerknüllt of *Drosophila* (GenBank entry X68347, Rushlow *et al.*, 1987); Dme OTD, Orthodenticle of *Drosophila* (GenBank entry X58983, Finkelstein *et al.*, 1990). Numbers refer to amino acid position, percentage at the right indicate the similarity of the noted homeodomain with the Bicoid homeodomain of *Platypeza* (Pco) and *Lonchoptera* (Llu), respectively. Amino acids identical with the homeodomain of Llu BCD are shaded red; amino acids identical with the homeodomain of Llu BCD, are shaded yellow.

PCO	BCD		94
Llu	BCD		69
Mab	BCD	MAOPPPPI/CDTSAVFHPVHHAPAHPHPPPPPHPOMOIPSOFLNPFEMLYDDRTGTLNVNVMRPVIPSOTOLPD	82
Dme	BCD	MAOPPPDONFYHPI, PHTHTHPHSHPHSHPHSHPHSHPHSHPHOHDOLOLPPOPRNPERII.FDERTGAINYNYTRPYLPNOMPKPDVFPSRFLPDSLVM	94
Dine	000		
Pco	BCD	RETERL RTTFT0001.0FLE0FF01NKYVTALFLADTSRLNLANAOVKIWFKNRRKHKTEEARMKELKG-TLPLGLNVSIPNLNGSLTSNSLDSSISESAPPSFTKSESPPLPL	208
L1u	BCD	REPRETETTET SAOTSKLEOVENES KVVNAS RLAFTSCKLNI GNAOVKIWFKNRRRED FOLKTKELNGSNDTTPAVSVSKDI CLALP	159
Mab	BCD	RE-RRTRTTFTSSOIAELEEYFROGKYLNNIRLSELTGRLNLGOAOVKIWFKNRRRFKIEOTKLNDSASFDMPLOLKDVKVPVGELTPSS	172
Dme	BCD	REPRETETTETSSOIABLEOHFLOGRYLTAPRIADLSAKLALGTAOVKIWFKNRRRHKIOSDOHKDOSYEGMPLSPGMKOSDGDBPSLOTISLGGGA	192
		Homeodomain PEST	
Pco	BCD	TPNPLTPSPTPSATSTPSASDKQSDNSNYGNQFYYNNNNNQMPQYYQTPPATSNQQQFEFPTKVQQQNETRYNNNNNNFSQQQQFNRL	296
Llu	BCD	TPTTLTPSPSLTPTSTPNISDQYSENYTYNPYTYNPYVQQHAYEQQVRAQPMATQYYQQPSAISQQLTRSAISQQLTRSAISQQLTR	229
Mab	BCD	TPSSAASSPAPPTTTTSSYIGNEIPSQPDTPNCFASGYFFNHNFPSHSYIGNEIPSQ	220
Dme	BCD	$\label{eq:transform} TP NALTPSPT-PSTPTAHMTEHYSESFNAYYNYNGGHNHAQANRHMMQYPSGGGPGPGSTNVNGGQFFQQQQVHNHQQQLHQGNHVPHQMQQQQQQAQQQQYHHFQHAQANRHMMQYPSGGPGPGSTNVNGGQFFQQQQVHNHQQQLHQGNHVPHQMQQQQQQAQQQQYHHFQHAQANRHMMQYPSGGPGPGSTNVNGGQFFQQQQVHNHQQQLHQGNHVPHQMQQQQQQAQQQQYHHFQHAQANRHMMQYPSGGPGPGSTNVNGGQFFQQQQVHNHQQQLHQGNHVPHQMQQQQQQAQQQQYHHFQHAQANRHMMQYPSGGPGPGSTNVNGGQFFQQQQVHNHQQQLHQGNHVPHQMQQQQQQAQQQQYHHFQHAQANRHMMQYPSGGPGPGSTNVNGGQFFQQQQVHNHQQQLHQGNHVPHQMQQQQQQAQQQQYHHFQHAQANRHMMQYPSGGPGPGSTNVNGGQFFQQQQVHNHQQQLHQGNHVPHQMQQQQQQAQQQQYHHFQHAQANRHMMQYPSGGPGPGSTNVNGGQFFQQQQQHNHQQQQQHANRHMQYPSGGPGPGSTNVNGQQFFQQQQQDAQQQQYHHFQHAQANRHMMQYPSGGPGPGSTNVNGGQFFQQQQQHNHQQQQQDAQQQQYHHFQHAQANRHMMQYPSGGPGPGSTNVNGGQFFQQQQQHNHQQQQQQDAQQQQYHHFQHAQANRHMQYPSGGPGPGYGTNVNGGQFFQQQQQNHNQQQQQQQDAQQQQYHHFQHAQANRHMQYPSGGPGPGYGTNVNGGQFFQQQQQHYNGGYGPGGYGTNVNGQQFFQQQQYHNYQQQQQQQDAQQQQYHHFQHAQANRHMQYPSGGPGPGYGTNVNGGQFFQQQQYHNQQQQQQQAAQQQQYHHFQHAQANRHMQYPSGGPGFGYGTNVNGGQFFQQQQYHNYQQQQQDAQQQQYHHFQTNVQQQQQQAAQQQQYHHFQTNVQQQQQQAAQQQQYHHFQTNVQQQQQQAAQQQQYHYQQQQQAAQQQQYHYQQQQQQAAQQQQYHYQQTNVQQQQQAAQQQQYHYQQQQQQAAQQQYYNQQQQQQAAQQQQYHYQQQQQQAAQQQYYNQQQQQQQAAQQQQYHYQQQQQQQQAAQQQQYHYQQQQQQQQQQ$	298
		Q-rich	
Pco	BCD	ASQEKLAEFAKQLKIKSEMADFNSAHSGHSDEIDET	354
Llu	BCD	DFLTSIKTEPDFNYNSTPYMRMPAAETMVNYTKIPTKNCYLPELSPNSEVYEPLTPKTEGRGSPKMANTSDEISNTHL	306
Mab	BCD	PYPTPPTDPAFDLSTHHGFSYGSNPLWRIAPQTPSSTSSEPSPTTVADVYEPLTPKNEDS-SPKIRAPDEIEDKSSLLK	298
Dme	BCD	DFQQKQASACRVLVKDEPEADYNFNSSYYMRSGMSGATASASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDE-SPSLCGIGIGGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDE-SPSLCGIGIGGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDE-SPSLCGIGIGGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDE-SPSLCGIGIGGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDE-SPSLCGIGIGGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDE-SPSLCGIGIGGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGPCAIAVGETEAADDMDDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGPCAIAVGETEAADDMDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGPCAIAVGETEAADDMDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGPCAIAVGETEAADDMDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGPCAIAVGETEAADDMDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGPCAIAVGETEAADDMDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGFCAIAVGETEAADDMDGTSKKTTLQILEPLKSASAVARGAASPGSEVYEPLTPKNDESPSLCGIGIGFCAIAVGETEAADDMDGTSKKTTLQILEPLKSASAVAA	406
		A-rich Acidic	
Pco	BCD	SQPMYNNNSNRRCGDEPQPDAASAAYQGQPMYNNNSNRRCGDEQMFGYRYN	405
Llu	BCD	-VDAKAYNQHRNTNTNNGYNTQFAFCFN	365
Mab	BCD	-VDCSQSTVDTILQ-AYSTHRATNAGGQFAYCFN	338
Dme	BCD	<u>GLDKSCDDGSSDDMSTGIRALAGTGNRGAAFAKFGKPSPPQGPQPPLGMGGVALGESN-QYQCTMDTIMQ-AYNPHRNAAGNSQEAYCFN</u>	494

**Figure 3. Protein alignment of** *bicoid* **homologues.** The predicted amino acid sequences of *bicoid* from *Platypeza* (Pco BCD), *Lonchoptera* (Llu BCD), *Megaselia* (Mab BCD), and *Drosophila* (Dme BCD) are shown. Amino acids that are conserved in at least 3 sequences (75%) are shaded in grey; dashes denote gaps. The numbers to the right refer to the last amino acid in each row. The homeodomain is boxed; other conserved domains and motifs of *bicoid* proteins, which have been functionally characterized (reviewed in McGregor, 2005), are underlined. For GenBank entry numbers of *Drosophila* and *Megaselia* Bicoid, see Figure 2.



Figure 4. cDNA pools from anterior and posterior cytoplasm of the same *Megaselia* embryo differ in composition. (A) PCR amplified cDNA from anterior (a) and posterior (p) pole cytoplasm of three *Megaselia* embryos (1-3); a mock cDNA preparation (0), made in the absence of *Megaselia* cytoplasm, served as negative control. Single bands in the negative control have possibly been amplified from minute DNA remnants in the enzyme solutions. (B) Southern Blot of the gel shown in (A), hybridized to *Megaselia bicoid*. Note that only cDNA pools from anterior cytoplasm hybridize with *Megaselia bicoid*. The smear in these lines is expected due to a truncated reverse transcription reaction during cDNA preparation. As size marker, a 1 kb DNA Ladder was used (Invitrogen); fragment sizes are given in on the left kilo base pairs.



Figure 5. Identification of *Megaselia bicoid* by subtractive screening of a spotted cDNA library made from anterior egg cytoplasm. (A) 1536 bacterial colonies were spotted onto nylon filters according to a twin-spotting scheme. Within each 3 mm square (see grid in B and C), four different clones were spotted in the indicated patterns. The remaining eight positions in each square were left blank. (B) Filter replicates of the spotted library were hybridized independently to radioactively labeled cDNA pools prepared from anterior or posterior pole cytoplasm. The signals of the filter replica, which was hybridized with anterior cDNAs, were color-coded in red; the signals of the filter replica, which was hybridized with posterior cDNAs, were color-coded in green. The color-coded images were aligned and merged. Green signals indicate hybridization with the posterior cDNA pool, while red signals indicate hybridization to anterior cDNA only. (C) A third replica of the spotted library was hybridized with a radioactively labeled *Megaselia bicoid* probe. For four clones (corresponding to the boxed twin spots in B and C), the presence of *Megaselia bicoid* was verified by sequencing.



**Figure 6. Classification and abundance of** *Clogmia* **clones.** (A) By subtractive screening, 161 *Clogmia* clones were isolated from the *Clogmia* cDNA library of anterior pole cytoplasm. The clones were classified according to their sequence similarity with genes in the GenBank database; the abundance, with which each clone was isolated, is indicated in parentheses. (B) The remaining 38 candidates were compared to *Drosophila*- and arthropod-specific gene databases, revealing sequences of 10 distinctive transcripts. For details see text and Table 4.



Figure 7. Testing for differential expression of Clogmia candidates in virtual Northern Blots. cDNA was prepared from anterior and posterior pole cytoplasm of Clogmia embryos and amplified by PCR. Of each candidate, a radioactively labeled probe was hybridized to anterior (a) and posterior (p) cDNA of four different Clogmia embryos (1-4). As negative control, mock cDNA preparations were used, which were prepared in the absence of *Clogmia* cytoplasm (0). In each panel, asterisks demark those pools of amplified cDNA that have been used to construct and screen the particular anterior cDNA library, from which the tested candidate was isolated. The smear or multiple lanes in the probe hybridizations are expected due to a truncated reverse transcription reaction during cDNA preparation. (A) Obp99a like, hybridized to cDNA of three-hour old embryos. (B) CG14764 like, hybridized to cDNA of one-hour old embryos. (C) Df31 like, hybridized to cDNA of one-hour old embryos. The cDNA pools, which have been used to construct and screen the library from which Df31 like was isolated, have not been included on the blot due to limited cDNA material. (D) exu like, hybridized to cDNA of three-hour old embryos. (E) jing like, hybridized to cDNA of one-hour old embryos. (F) CG6459 like, hybridized to cDNA of one-hour old embryos. The cDNA pools, which have been used to construct and screen the library from which CG6459 like was isolated, have not been included on the blot due to limited cDNA material. (G) CG1967 like, hybridized to cDNA of three-hour old embryos. The same cDNA preparations have been used for the blots in panels A, D, and G, in panels B and E, and in panels C and F, respectively. Size markers on the left of each panel are given in kilo base pairs. For details, see text.



Figure 8. Quartet Puzzling analysis of the newly identified Hunchback homologues. 108 amino acids comprising the conserved N-terminal zinc finger domain (corresponding to amino acids 242-349 of the Drosophila protein) of the predicted Hunchback protein sequences were compared with the zinc finger domains of the four most closely related hunchback paralogs of Drosophila using the Quartet Maximum-Likelihood Method of Strimmer and von Haeseler (Strimmer and von Haeseler, 1996). Numbers refer to reliability values of the branching pattern in percent; branch lengths indicate the average number of amino acid changes per position (see scale). Abbreviations are Dme HB, Hunchback of Drosophila (GenBank entry Y00274, Tautz et al., 1987); Eba HB, Hunchback of *Episyrphus* (this work); Mab HB, Hunchback of Megaselia (GenBank entry AJ295635, Stauber et al., 2000); Pco HB, Hunchback of Platypeza (this work); Llu HB, Hunchback of Lonchoptera (this work); Eli HB, Hunchback of Empis (this work); Hpl HB, Hunchback of Haematopota (this work); Cal HB, Hunchback of Clogmia (this work); Aga HB, Hunchback of Anopheles (Zdobnov et al., 2002); and of the Hunchback paralogues in Drosophila, GL, Glass (GenBank entry X15400, Moses et al., 1989); KR, Krüppel (GenBank entry X03414, Rosenberg et al., 1986); SNA, Snail (GenBank entry Y00288, Boulay et al., 1987); WOR, Worniu (GenBank entry AF118857, Ashraf et al., 1999).

Dme Mab Pco Llu Eba Eli Hpl Cal Aga	HB HB HB HB HB HB HB	MQNWETTATTNYEQHNAWYNSMF-AANIKQBPGHHLDGN-SVASSPRQSPIPSTNHLEQFLKQQQQ MQNWES	64 53 47 49 51 77 63 76 99
Dme Mab Pco Llu Eba Eli Hpl Cal Aga	HB HB HB HB HB HB HB	QLQQQPMDTLCAMTPSPSQNDONSLOHYDANLQQQLLQQQYQOHFQAAQQQHHHHHHLMCGFNPLTPPGLPNP-MOHFYGGNL 	147 108 125 113 123 162 108 109 150
Dme Mab Pco Llu Eba Eli Hpl Cal Aga	HB HB HB HB HB HB HB HB	RPSPQPTPTSASTIAPVAVATGSSE       KIQAL-TPPMDVTPPKSPAKS       SQSAIEPEKHD_QMSNSSEDMKYMAESED         SQLAASANN       TPTPTSTPPMDVTPPKSP       SSEDMKYMAESED         PENINAQ       TQSITPRNTPPMDVTPPKSPF       SKADS         GIHANNPTSPLAQQSEG       TQSITPRNTPPMDITPPKSPKP       SKADS         HQSVSTGHVEXVKMOQST       NNNSLTPRNTPPMDITPPKSPKLS       EKDHD         LSIQQPSSSIGQLNHNQQDINITKLDNNSLTPRNTPPIDITPPKSPKS       IMISTSS       ELDODVMSSNSSEDMKY-LESE         NVDQSFLQQLNHQQDINITKLDNNSLTPRNTPPIDITPPKSPKH       INSSDTKF-LESE       INSSDTKF-LESE         NVDQSFLQUININQDINITKLDNNSLTPRNTPPIDITPPKSPKH-       INSSDTKF-LESE       INSSDTKF-LESE         NVDQSFLQUININQDINITKLDNNSLTPRNTPPIDITPPKSPKH-       DKTPFKSFKFQUIPTPKS       DKDQE-ILSNSSD-         NVDQSFLQUININGSIGFAFAKNDGSNPSITFSHTPHDVTPFKSPKFVDIPTPEK       DKDQE       DNDLNSNY-NDSIDTRSL-         FRSFTKGLGIFFG	222 168 181 179 197 248 184 201 225
Dme Mab Pco Llu Eba Eli Hpl Cal Aga	HB HB HB HB HB HB HB	DDTNIRMPIYNSHGKMKNYKČKTČGUVAITKUDFWAÄTRTÄNKPDKILQČPKČPPVTEFKHHLEYHIRKÄKNOKPFQČDKČSVTCUNKSMLNSÄRKSÄSSU DDEAIMMPIYNSHGKMKNYKCKSCGTAITKUSFWTHMRSHNKPEKILQCPKCPFVTELKHHLEYHIRKKNIKPFQCDKCNYSCUNKSMLNSHRKSHSSU DDSIKTPIYNSHGKMKNYKCKSCGTAITKIGFWQHARTHNKPEKILQCPKCPFVTELKHHLEYHIRKKNIKPFQCDKCNYSCUNKSMLNSHRKSHSSU DDSIKTPIYNSHGKMKNYKCKSCGYMAUTKUAFWEHASSHMKPEKILQCPKCPFVTELKHHLEYHIRKKNIKPFQCDKCNYSCUNKSMLNSHRKSHSSU DDSIKTPIYNSHGKMKNYKCKSCGYMAUTKUAFWEHASSHMKPEKILQCSKCPFVTELKHHLEYHIRKKNIKPFQCDKCNYSCUNKSMLNSHRKSHSSU DDSIKTPIYNSHGKMKNYKCKSCGYMAUTKUAFWEHASSHMKPEKILQCSKCPFVTELKHHLEYHIRKKNIKPFQCDKCNYSCUNKSMLNSHRKSHSSU DDSIKTPIYNSHGKMKNYKCKSCGYMAUTKUAFWEHASSHMKPEKILQCSKCPFVTELKHHLEYHIRKKNIKPFQCDKCNYSCUNKSMLNSHKSHSSU DDEIIKMPIYNSHGKMKNYKCKSCGYMAUTKUAFWEHASSHKPEKILQCSKCPFVTELKHHLEYHIRKKNIKPFQCDKCNYSCUNKSMLNSHKSHSSU DDEIIKMPIYNSHGKMKNYKCKSCGYMAUTKIAFWOHRSHNKPEKILQCSKCPFVTELKHHLEYHIRKKNIKPFQCUNCONSUNSMLNSHKSHSSU DDEIIKMPIYNSHGKMKYKCKCQCPFIAVTKLSFWEHRKGHIKPEKMLKCYCCPFVTEYKHHLEYHIRKKNIKPFQCUNCSUNSSMLNSHLSSH DGADVDDENDAEDDD-ITFVKINSHGKMKYKCQCDFIAVTKLSFWEHRKGHIKPEKHLKCYCCPFVTEYKHHLEYHIRNGKSKPFQCVNKSSMLNSHKSHSNI DGGADVDDENDAEDDD-ITKFKNNSHGKMKYKKCQCDFIAVTKLSFWEHRKHIHIFKHKINTCPKCPVTEYKHHLEYHLRNINGSKPFQCSVNSSMLNSHKSHSNI DGGADVDDENDAEDDD-ITKFKNSHGKKNKKKCQCDFIAVTKLSFWEHRKHIKTSHTKSHKNKKSKSVCNKSMLNSHKSHSNI DFGADVDDENDAEDDD-ITKFYNNSHGKMKTKKCQCDFIAVTKLSFWEHRKHIKFKNIKPKKNKCSCCPFVTEYKHLEYHLRNINGSKPFQCNCNSSUNSSMLNSHKSHSNI	322 268 281 279 297 345 288 315 324
Dme Mab Pco Llu Eba Eli Hpl Cal Aga	HB HB HB HB HB HB HB HB	YQYRCADCDYATKYCHSFKLHLRKYCHKPGMVLDEGTPNPSLVIDVYGTRRGPKSKNGGPIASGSGSGSGSKSNVAAVAPQQQQSQPAQFVATSQLSAALQGFPLVQGNSAPPA YQYRCADCDYATKYCHSFKLHLRKYDHKFGMVLDEGIPNPSIVIDVYGTRRGPKMKGGISTFSVSHRTUPDQKPSLSDLKIP	437 356 365 379 387 345 375 401 422
Dme Mab Pco Llu Eba Eli Hpl Cal Aga	HB HB HB HB HB HB HB	C-BOX ASPVLPLPASPAKSVASVEQTPSLPSPAN_LPPL_ASLLQQNRNMAFFPYWNLNLQMLAAQQQALPTSPAKSTTSSNSEVNTPPSSANQMKPNGOISNLLPPLVQSMLQQQQQMSQQCMAQQQAQQQQQQQQQQQQQQQQQQQQQQQQ	500 425 425 484 477 345 438 454 493
Dme Pco Llu Eba Eli Hpl Cal Aga	HB HB HB HB HB HB HB HB	AVLAOLSPRINE      OOODNEEE      EQDDEYE      RSSVDSAMDLSQGTPVKEDE      OOODPQQ-PL        LAOLSPSNRES-LQ      NDSNREPUVEDEEEDEHDQKEEHUAAAIDLSAQASPFIKDEEE      SPVLP        VLAOMSPRINEATLONLINGGGSUIENDKDSVQDFEC      SAMDLSQGS_PLKNE      ITNPNQDTFOMSTNVSNVLADSGMQQ        VLAOMSPRINETTIONLONGGSQVIENDKDSVQDFEC      ETDDEFNRSN-GSAIDLSQSNGTPTK      ITNPNQDTFOMSTNVSNVLADSGMQQ         AVLAOLSPRIRETALONLOKO	563 490 491 573 540 345 500 496 588
Dme Mab Pco Llu Eba Eli Hpl Cal Aga	HB HB HB HB HB HB HB	AMNLKVEEEA-TP-LMSSSNAS-RRKGRVLKLDTTLLQLRSEAMTSPEQLKVPSTPMPTASSPIAGRKPMPEEHCSGTSSADESMETAHVPQANT-SASSTA AKE-EETSSNTP-TVSTTPIS-RRKGRVLKLDTTTNTQSQVDEDSDRQSPSSS-FEEFKETAATSTPSPAPAPAPAPAPAPAPAPAPA YNLKM-TEEEVNTP-TISSSSS-RRKGRVLKLDTSTQHLPVAEEIAVPEPIRSTESPSSSSFEEFKLVQS	660 562 573 679 614 345 591 545 662
Dme Mab Pco Llu Eba Eli Hpl Cal Aga	HB HB HB HB HB HB HB	F-Box SSSGNSSNASSNASSNSNGNSSSNGSTSAVAAPPSGTPAAAGAIYFCKVCDIFFKDAVLYTIHMGYHSCDDVFKCNMCGEKCDGPVGLFVHMARNAHS SSSGNSSNASSNSNGNSSSNGSTSAVAAPPSGTPAAAGAIYFCKVCDIFFKDAVLYTIHMGYHSCDDVFKCNMCGEKCDGPVGLFVHMARNAHS TTPSPPAVV	758 620 642 777 705 345 681 631 722

**Figure 9.** Alignment of predicted protein sequences of the newly identified *hunchback* homologues. Amino acids conserved in at least 5 sequences (56%) are shaded in grey; dashes denote gaps. Asterisks mark the relevant cysteine and histidine residues of the zinc fingers (Tautz *et al.*, 1987). The numbers to the right refer to the last amino acid in each row. The highest similarity is seen in the zinc finger domains ZFD1 and ZFD2 (boxed). Outside of the zinc finger domains, the sequence is conserved in previously defined regions (dashed boxes) (Hülskamp *et al.*, 1994; McGregor *et al.*, 2001a; Tautz *et al.*, 1987), suggesting that these regions are of structural or functional importance. Abbreviations as in Figure 9. The predicted protein sequences is known.



**Figure 10. Genomic organization of Dipteran** *hunchback* **homologues.** For each species, *hunchback* cDNA (light-blue: untranslated region, dark-blue: ORF) has been mapped to genomic DNA (black: sequenced, grey: not sequenced). All sequences were aligned relative to nucleotide position +1 as the start of the ORF; one vertical bar equates to 1 kb of sequence. Putative promoters are indicated as vertical lines in front of the leading exons. The stretch of genomic DNA, which has been analyzed in transgenic *Drosophila* embryos, is indicated below the genomic organization (red). Black wedges indicate putative NRE sequences (Wharton and Struhl, 1991), for an alignment of the putative NREs see Figure 11; blue arrows denote the sequence AATAAA as putative polyadeny-lation signal (reviewed in Birnstiel *et al.*, 1985). The P1 exon of *Platypeza* is not positioned in scale. The dotted lines for *Megaselia* and *Haematopota* indicate additional upstream regulatory sequence. All sequences have been documented in the Appendix A.3. For details, see text.

	T.GTTGTC.AATTGTA.ATA
1.1.1	
Mao I	TTTTTTCGTTGCTTTGAATTGTAAATAATTA
Mdo 2	AGTGAATCGTTGTCATGAATTGTAAATATGAA
Dme 1	ATATAATCGTTGTCCAGAATTGTATATATTCG
Dme 2	ATTATTTTGTTGTCGAAAATTGTACATAAGCC
Dvi 1	CATATTTCGTTGTCCAGAATTGTAAATACTCG
Dvi 2	<b>TTGATTTTGTTGTCGAGAATTGTACATAAGCC</b>
Mab	САААААСТСТТСТСАААСАТТСТАСАТАТСАА
Dec 1	
FCO I	TATTATTIGTIGTCAAAGATTGTACATATGAA
Pco 2	TAAGAAAAGTTGTCAAGGATTGTACATAAAAA
Llu	AGCACAATGTTGTC-ATAATTGTACATAAAAA
Eba 1	AGAGTTTCGTTGTCAAAGATTGTAAATATTAA
Eba 2	AAAATACTGTTGTCCAAAATTGTACATACTAT
Hpl	AGCGCTTTGTTGTGTAGAATTCAACTTGAAAT
Cal	
Cal	ATTIGATOGTIGIATA-GATTGTTGTTATATT

Figure 11. Putative nanos response element (NRE) sequences from dipteran hunchback homologues. Nucleotides shared by at least ten sequences (71%) are shaded in grey and given as consensus. In addition to the NRE sequences in hunchback of Drosophila melanogaster (Dme), putative NRE sequences in dipteran hunchback homologues have been previously identified in Drosophila virilis (Dvi) (GenBank entry X15359, Hancock et al., 1999; GenBank entry X15359, Treier et al., 1989) and Musca domestica (Mdo) (GenBank entry Y13050, Bonneton et al., 1997). During the course of this study, putative NRE sequences have been identified in hunchback homologues of Megaselia, Platypeza, Lonchoptera, Episyrphus, Haematopota, and Clogmia. For the positions of the NRE sequences within the respective hunchback loci, see Figure 8 and in the sequences noted in the Appendix A.3. For further abbreviations, see legend of Figure 9.



(E, F) *Clogmia*, endogenous *hunchback* expression during late cellularization, lateral (E) and dorsal (F) view, respectively. (G, H) Expression driven by 10 kb of the *Megaselia hunchback* locus. (I, J) Expression driven by 6.2 kb of *Platypeza hunchback* regulatory DNA. (K, L) Expression driven by 4.6 kb of *Lonchoptera hunchback* regulatory DNA. (M, N) Expression driven by 5.4 kb of *Episyrphus hunchback* regulatory DNA. (O, P) Expression driven by 9 kb of *Haematopota hunchback* regulatory DNA. (Q) Expression driven by 6.9 kb of *Clogmia hunchback* regulatory DNA (K2) during cellularization and (R) at the onset of gastrulation. Expression driven by a 2.4 kb sub-fragment of the 6.9 kb fragment (K13) shows the same pattern (data not shown). *lacZ* was used as reporter gene in all constructs, except for *Megaselia*, where *Megaselia hunchback* was used. For positions of the tested fragments within the respective *hunchback* loci, see Figure 10 and Table 5. Scale bar: 215 µm in A, B, G-R; 225 µm in C, D; 180 µm in E, F. Panels Q, R: courtesy of Alexander Prell.



Figure 13. Effects of *bicoid* RNAi on *hunchback* expression in *Megaselia*. Whole-mount *in situ* hybridizations of *Megaselia hunchback*. *bicoid* RNAi was performed as previously described (Stauber *et al.*, 2000), embryos are shown during cellularization.

(A) In the wild-type embryo *hunchback* expression has started to clear from the anterior pole. (B) A strong *bicoid* RNAi phenotype is shown. The posterior *hunchback* expression is duplicated at the anterior pole. Anterior is to the left, dorsal is up. Suppression of anterior clearance at the onset of cellularization and reduction of the anterior expression domain was observed in 68% of the RNAi embryos (n=56). Buffer injected embryos (n=16) did not show a phenotype.



**Figure 14. Expression of** *Episyrphus hunchback.* Whole-mount *in situ* hybridizations showing embryos at (A, B) pre-blastoderm, (C-F) syncytial blastoderm, (G) cellular blastoderm, (H) the onset of gastrulation, and (I) the beginning of germband extension. Embryos are shown in lateral view. (J-L) Dorsal views of the same embryos shown in (G), (H) and (I), respectively. Anterior is to the left. See text for details.



**Figure 15.** In *Episyrphus, hunchback* and *zerknüllt* are co-expressed along the dorsal midline. *hunchback* and *zerknüllt* expression in *Episyrphus* embryos at the onset of gastrulation were analyzed in a single embryo by fluorescent whole-mount *in situ* hybridization. (A, B) Lateral and dorsal view of *Episyrphus hunchback* expression. (C, D) Lateral and dorsal view of *Episyrphus zerknüllt* expression. (E) Merged image of A and C, (F) merged image of B and D. *hunchback* expression was false-colored in red; *zerknüllt* expression was false-colored in green; embryos are oriented with anterior to the left.



**Figure 16. Expression of** *Episyrphus zerknüllt.* Whole-mount *in situ* hybridizations showing embryos at (A) pre-blastoderm, (B-D) syncytial blastoderm, (E) the onset of gastrulation, (F) the beginning of germband extension, and (J-L) during germband extension. Embryos are shown in lateral orientation. (G-I) Dorsal views of the same embryos shown in (D), (E) and (F), respectively. Anterior is to the left.

Dme	OTD	MAAGFLKSGDLGPHPHSYGGPHPHHSVPHGPLPPGMPMPSLGPFGLPHGLEAVGFSQGMWGDLCYPGVNTRK <mark>DRRERTTFTRAQLDVLEALFGKTRYPDIFMREEVALKINLPE</mark>	114
Dme	OTD '	MAAGFLKSGDLGPHPHSYGGPHPHHSVPHGPLPPGMPMPSLGPFGLPHGLEAVGFSQGMWGVNTRK <mark>DRRERTTFTRAQLDVLEALFGKTRYPDIFMREEVALKINL</mark> PE	108
Eba	OTD	MAAGFLKSGDLGPHPHSYGGPHPHHSVPHGPLPPGMPMPSLGPFGLPHGLEAVGFSQGMWGVNTRK <mark>DRRERTTFTRAQLDVLESLFGKTRYPDIFMREEVALKINL</mark> PE	108
Dme Dme Eba	OTD OTD ' OTD	Homeodomaln SRVQVWFKNRRAKCRQQIQQQQSNSLSSSKNASGGGSGNSCSSSSANSRSNSNNNGSSSNNNTQSSGGNNSNKSSQKQGNSQSSQQGGSSGGNNSNNNSAAAAASAAAAVAA SRVQVWFKNRRAKCRQQIQQQQSNSLSSSKNASGGSGNSCSSSSANSRSNSNNNGSSSNNTQSSGGNNSNKSSQKQGNSQSSQQGGSSGGNNSNNNSAAAAAAAAAAAAA SRVQVWFKNRRAKCRQQIQQQQSNSLNSSKGNSGNVGSGGNSGSSRNSNSNNNSCNANNQNSSSGSCNSGTNSSNT	228 222 187
Dme	OTD	AQSIKTHHSSFLSAAAAASAQSIKTHHSSFLSAAAAASGGTNQSANNNSNNNNQGNSTPNSSSSGGG-SQAGGHLSAAAAAAANVTAAHQNSSPLLPTPATSVSPVSIVCKK	340
Dme	OTD '	AQSIKTHHSSFLSAAAAAASGGTNOSANNNSNNNNOGNSTPNSSSSGGGGSQAGGHLSAAAAAAANVTAAHONSSPLLPTPATSVSPVSIVCKK	318
Eba	OTD		206
Dme	OTD	EHLSGGYGSSVGGGGGGGGGGGGGGSSGGLNLGVGVGVGVGVGVGVGVGVQSQDLLRSPYDQLKDAGGDIGAGVHHHRSIYGSAAGSNPRLLQPGGNITPMDSSSSITTPSPPITPMSPQSA	454
Dme	OTD'	EHLSGGYGSSVGGGGGGGG-ASSGGLNLGVGVGVGVGVGVGVGVGQUGQQDLLRSPYDQLKDAGGDIGAGVHHHRSIYGSAAGSNPRLLQPGGNITPMDSSSSITTPSPPITPMSPQSA	430
Eba	OTD		215
Dme Dme Eba	OTD OTD ' OTD	PQRPMPPNRPSPFTILPPIRPPICPINIRITSGTISTSNIRITMPRRPATTHRWSTLAIRIRSTTTWAIRATRPPILVCRHRHPSRAPCPRRPSPRTAWITCRRRISTRIWCRI AAAAHASAAHSAAYMSNHDSYNFWHNO AVAHAAAASAAHSAAYMSNHDSYNFWHNO AVAHAAAASAAAHSAYMSNHDSYNFWHNO	568 466 253
Dme	OTD	YSSNTAAVAATTTVQRGQVVRVRVRVRVRVRVRVRVRVRVRVRVRVLVLVLVLVLURGAIVLPSWSSTIISTTSTSYSSISITRIITRINTRITTAIIIISSNTIMMMNSDRI	672
Dme	OTD '	YQQYPNNYAQAPSYY-5QMEYFSNQNQVNYNMGHSGYTASNFGLSPSPSFTGTVSAQAFSQNSLYYMSPQDKYANMV	543
Eba	OTD	YNQYPNNY-QTPSYY-SQMEYFSNQNQVNYNMGHSGYSASNFGLSPSSSFTGTMSAQAFSQNGLDYMSPQDKYVNMV	329

Figure 17. Protein alignment of *Episyrphus* Orthodenticle with two isoforms of *Drosophila* Orthodenticle/Ocelliless. The predicted amino acid sequences of *orthodenticle* from *Episyrphus* (Eba OTD), and two *Drosophila orthodenticle/ocelliless* transcripts are shown (Dme OTD, GenBank entry X58983; Dme OTD', Genbank entry BT011185). Amino acids that are identical with Eba OTD are shaded in grey; dashes denote gaps. The numbers to the right refer to the last amino acid in each row. The homeodomain is boxed.



**Figure 18. Expression of** *Episyrphus orthodenticle.* Whole-mount *in situ* hybridizations showing embryos (A, B) at the beginning of blastoderm cellularization, (C, D) at mid-cellularization, and (E, F) at the onset of gastrulation. Embryos are shown in lateral (A, C, E) and in dorsal view (B, D, F). Anterior is to the left. For details, see text.

Homologue	Ter	nplate*	Primer pair <sup>†</sup>	Length (bp)	Sequences <sup>§</sup>
	5,	Pco-SMART 5' cDNA (I)	GTGGCGCTGATTCCGAAAGTGAG/10xUPM <sup>MR</sup>		
Diatunaza hirorid	Q,	Pco-SMART 5' cDNA (I)	GTGGCGCTGATTCCGAAAGTGAG/10xUPM he: CTGGAACTCTTGCTCGAGCTCC/NUP <sup>MR</sup>	1568	SEO01
n laithea bicoin	ώ	Pco-SMART 3' cDNA (I)	CGAGTTCCCGACCAAGGTTCAAC/10xUPM <sup>MR</sup>	200	
	Ω	Pco-SMART 5' cDNA (I)	CAAAAATAGGCGCTAGTGCAG/ TGTTGAGTGGATGATTTCCTTC		
	5,	Llu-SMART 5' cDNA (a)	GTTTACCTGATAGCTCAGCTAGACG/10xUPM ne: GCGTTAACATATTTACTTTCGTTG/NUP		
Lonchonters biroid	ς	Llu-SMART 3' cDNA (a)	AGAACAACATTTACAAGTGCACAAA/10xUPM ne: AACGCTTCACGTCTAGCTGAG/NUP	0238	SECOS
FUILVIUDIEIA DICUIA	ώ	<i>Llu</i> -SMART 3' cDNA (a)	TCAGCAACCATCAGCGATTAGTCA/10xUPM	0000	
	õ	Llu-SMART 3' cDNA (a)	AGAACAACATTTACAAGTGCACAAA/ ATCATATTGTCTAAGCCTC		
م ام تو مع ما هم ما مع ما مع مع ما مع مع ما مع ما مع ام مع ما مع	ດ	<i>Eba</i> -SMART 5' cDNA (e: 0-5 hrs) <sup>MS</sup>	CATCTAATTGCGCTCGTGTGAATG/10xUPM <sup>MS</sup>	000	
Episyrphus orthodenticle	'n	<i>Eba</i> -SMART 3' cDNA (e: 0-5 hrs) <sup>MS</sup>	CATTCACACGAGCGCAATTAGATG/10xUPM <sup>MS</sup>	1003	SEGUS

## A.1.2 Tables

 

 Table 2. cDNA isolation of dipteran hunchback homologues: templates, primers and products. cDNA has been isolated by PCR on cDNA pre-pared with SMART RACE cDNA Amplification Kit, or on cDNA prepared with Marathon cDNA Amplification Kit, respectively. The cDNA of Clog 
 mia hunchback has been isolated from a maternal Lambda-ZAP cDNA library.

Homologue	Ter	iplate*	Primer pair <sup>t</sup> L	ength (bp)	Sequences <sup>§</sup>
	2	Eba-SMART 5' cDNA (e: 0-5 hrs) <sup>MS</sup>	GATACACCGACGAGGGTGTGGCTTCC/10xUPM ne: AGCCCTGGTGGAGTAAGTGGATTA/NUP		
	Ω	Eba-SMART 5' cDNA (e: 0-5 hrs) <sup>MS</sup>	GGGAATATTAATTCTGTAAACGGGGGA AGTATGTACAATTTTGGACAACAGTATTTT	2876	SEQ04 (P1)
Episyrphus hunchback	õ	Eba-SMART 3' cDNA (e: 0-5 hrs) <sup>MS</sup>	GCACAAGAATTTAAAGCCATTCCA/10xUPM ne: CTATGTTGAACTCCCACCGGAAG/NUP		
	Ω	Eba-SMART 5' cDNA (e: 0-5 hrs) <sup>MS</sup>	GATACACCGAGGGGGGGGGGGGGCTTCC/10xUPM ne: AGCCCTGGTGGGGGGTAAGTGGATTA/NUP	372	SEQ05 (P2)
	2	<i>Eba</i> -SMART 5' cDNA (e: 0-5 hrs) <sup>MS</sup>	CCGACGAGTGTGACTTCCGGTGGGGGGGTCAAC/UPM ne: GATACACCGACGAGTGTGACTTCC/NUP	1052	SEQ06 (P3)
Megaselia hunchback	Q	<i>Mab</i> -SMART 5' cDNA (e: 0.5-4 hrs) <sup>MS</sup>	ATCACAATCAGCACAACGGTATTGG/10xUPM <sup>ws</sup>	873	SEQ08 (P2)
	2,	Pco-SMART 5' CDNA (I)	AATCGGAGCAACGGTACTGGTAGA/10xUPM ne: CGCATTGGAATGGCTTCAAGTTCTT/NUP		
	Ω	Pco-SMART 3' cDNA (I)	AGCACAAGAACTTGAAGCCATTCC/10xUPM ne: AACGACGACAATAACGCTGAAGAC/NUP	2106	SEQ10 (P1)
гацурега пилспраск	õ	Pco-SMART 3' cDNA (I)	AGCACAAGAACTTGAAGCCATTCC/ AGATGTTGGGTGGATGTGTCTAGC		
	õ	Pco-SMART 5' cDNA (I)	AATCGGAGCAACGGTACTGGTAGA'10xUPM ne: AGCGAATTCATCAACGTTTTGC/NUP	485	SEQ11 (P2)
	Ω	Llu-SMART 5' cDNA (a)	TAATTGTGATGTAGTTGGCGAATGAGTC/10xUPM		
Lonchoptera hunchback	õ	Llu-SMART 3' cDNA (a)	AGATCGCTCTCCATCGCAGTTAAA/10xUPM	3119	SEQ13 (P1)
	Ω	Llu-SMART 5' cDNA (a)	GACGCGTTCCGATTAACGGATATAA TGTACAAAAGATGACAAGGCAGAAAA		
-	Q,	<i>Eli</i> -SMART 5' cDNA (e: <24 hrs) <sup>MS</sup>	CGGTATTGATAAACTGATGA/10xUPM <sup>MS</sup>		
Empis hunchback	ũ	<i>Eli</i> -Marathon cDNA (o) <sup>(Stauber et al., 2002)</sup>	TGGTGGACCATTATCATTACTA/AP1 ne: ACTATTAATTGCTGTTTGTGGTTCA/AP2	1477	SEQ15 (P1)
	Q	<i>Hpl</i> -Marathon cDNA (o) <sup>(Stauber et al., 2002)</sup>	GGTTTCGAGCCATCATGATTTCGTAAA/AP1		
Haematopota hunchback	õ	Hpl-Marathon cDNA (o) <sup>(Stauber et al., 2002)</sup>	TTTACGAAATCATGATGGCTCGAAACC/AP1	2451	SEQ17 (P1)
	വ്	Hpl-Marathon cDNA (o) <sup>(Stauber et al., 2002)</sup>	ATTTGTGAAAATTATGAAATAATTTGGACGC/ AGCGCTTGCGTTTGTACT		
Clogmia hunchback	ı	Lambda-ZAP library (e: 0-2 hrs) <sup>(Schmidt-Ott, unpublished)</sup>		2802	SEQ19 (P1) <sup>AP</sup>
Anopheles hunchback	5,	Aga-SMART 5' cDNA (a)	TACCATTGCCGTACATTCGGTTGG/10XUPM CCATCGCCATTACGGAGTCAAAGTTC/NUP	78	SEQ21 (P1)
* Origin of the tissue for mRNA isola 5'-3' direction. ne, nested RACE; AI <sup>MS</sup> Cloning/preparation by Michael Str UTRs: The transcript with the first exo	tion: adh P1/AP2, auber. <sup>Al</sup>	ult females (a), ovaries (o), larvae (l) and embryos (e). The age , adaptor primers of Marathon Kit; 10xUPM/NUP, adaptor 1 <sup>p</sup> Cloning by Alexander Prell. The position of the first exon re proximal to the start of the ORF has been assigned to "P1", th	ge of the embryos in hours at 25 °C is indicated following primers of SMART RACE Kit. <sup>§</sup> All sequences have by the ative to the ORF is indicated in parentheses for cDNAs he next proximal "P2", etc.	the colon. <sup>†</sup> Priesen listed in the colon the second second the second s	imer sequences in he Appendix A.3. with alternative 5'

Antisense RNA probe	Label*	Template
hunchback	DIG	2.4 kb genomic <i>Xbal</i> fragment comprising the region of -14 to +2422 relative to the first nucleotide of the ORF (Tautz <i>et al.</i> , 1987).
Episyrphus hunchback	DIG/ FITC	1.1 kb P3 5' RACE product, comprising 163 bp of UTR and the adjacent nucleotides 1 to 889 of the ORF (this work)
Megaselia hunchback	DIG	1.1 kb 3' RACE product, comprising nucleotides 797 to 1863 of the ORF and 53 bp of the adjacent UTR (Stauber <i>et al.</i> , 2002).
Clogmia hunchback	FITC	2.1 kb cDNA fragment, comprising nucleotides 433 to 1896 of the ORF and 593 bp of the adjacent 3' UTR (Rohr <i>et al.</i> , 1999).
zerknüllt	DIG	1.4 kb cDNA (ps60-7, gift from Siegfried Roth).
Episyrphus zerknüllt	DIG/ BIO	1 kb 3' RACE product, comprising nucleotides 120 to 993 of the ORF and 56 bp of the adjacent UTR (Rafiqi <i>et al.</i> , in preparation).
Megaselia zerknüllt	DIG	0.8 kb, complete ORF (Stauber et al., 1999).
Episyrphus orthodenticle	BIO	0.9 kb 3' RACE product, comprising nucleotides 219 to 769 of the ORF and 94 bp of the adjacent UTR (this work).
lacZ	DIG	lacZ ORF (pBST-lacZ, gift from Ronald Kühnlein)

Table 3. RNA probes for whole mount *in situ* hybridization.

\* Independently prepared probes with differently conjugated UTP analogues are not listed separately. Instead, both label types are listed, separated by a slash. Abbreviation: BIO, biotin; DIG, digoxigenin; FITC, fluorescein.

**Table 4. Sequence comparison of** *Clogmia* **candidates with** *Drosophila* **and** *Anopheles* **genes.** The putative functions of the proteins encoded by ten distinct *Clogmia* cDNAs were assessed by sequence comparison to *Anopheles* and *Drosophila* genes. *Clogmia* cDNAs were named after the putatively closest related sequence in *Drosophila*. The Expect value describes for a given query sequence, how often an equally good or better alignment could have been found in the database by chance (Altschul *et al.*, 1994). The Expect value is often written as x to the power of *e*; here it is converted to x to the power of 10 as a more comprehendible tool to assess homology. Expect values lower than 0.01 were considered as reasonable support for homology, whereas Expect values higher than 1 were not considered as support for homology. For all *Clogmia* candidates, the identified *Anopheles* sequences were themselves homologues to the *Drosophila* genes. The degree of conservation serves as a visualization of the Expect value, while the abundance indicates, how often a particular cDNA was isolated among the 161 sequenced clones. The origin from one-hour and three-hour old embryos, respectively, is indicated in parentheses.

cDNA	Putative molecular function	Expect value ( <i>Anopheles</i> )	Expect value ( <i>Drosophila</i> )	Degree of conservation	Abu in s	indance creen
Chosen for analyses of tran	script localization, based on putative molecul	ar function and high deg	ree of conservation			
Obp99a like	pheromone binding	2·10 <sup>-13</sup>	5·10 <sup>-10</sup>	+++	12	(5/7)
exu like	RNA localization	6·10 <sup>-6</sup>	2·10 <sup>-3</sup>	++	4	(-/4)
Df31 like	chromatin remodeling	5·10 <sup>-2</sup>	7·10 <sup>-3</sup>	++	1	(1/-)
CG14764 like	unknown	5·10 <sup>-12</sup>	7·10 <sup>-10</sup>	+++	1	(1/-)
CG1967 like	post-Golgi transport	2·10 <sup>-40</sup>	3·10 <sup>-35</sup>	+++	1	(1/-)
CG6459 like	mitochondrial	7·10 <sup>-3</sup>	2.8	+	7	(7/-)
jing like	transcription factor	1.7	4.5	-	1	(1/-)
Putative house-keeping functions and/or lack of conservation in lower dipterans and other insects						
His3:CG31613 like	histone	2·10 <sup>-32</sup>	1·10 <sup>-31</sup>	+++	1	(-/1)
Graf like	Rho GTPase	no hit	4.9	-	4	(3/1)
Amy-d like	sugar metabolism	no hit	4.9	-	6	(6/-)

**Table 5. Reporter gene constructs to analyze** *hunchback* regulatory DNA in transgenic *Drosophila* embryos. Putative regulatory DNA of the newly isolated *hunchback* homologues was cloned in front of a reporter gene and integrated by P-element mediated germline transformation into the genome of *Drosophila*. Except for *Platypeza*, the DNA fragments cloned from each species included the intron of the respective P1 transcript. For *Haematopota* and *Clogmia*, two reporter gene constructs were analyzed. The positions of the P1 leader, the P1 intron and the region tested in transgenic *Drosophila* embryos are given relative to the start of the ORFs (+1). The number of established transgenic lines and the number of lines that showed the same reporter expression (in parentheses) are listed.

Species	P1 lead	er	P1 intron	hunchback f	ragment	Lines	Blastoderm expression
Episyrphus	-5839 to -5277	(563 bp)	-5276 to -39	-5382 to +24	(5406 bp)	4 (4)	yes
Megaselia	-1117 to -907	(211 bp)	-906 to -20	-8000 to +2868	(11 kb)*	2 (2) <sup>MS</sup>	yes
Platypeza	-2035 to -1765	(272 bp, P2)	-1764 to -12 (P2)	-6173 to +39	(6212 bp)	3 (4)	yes
Lonchoptera	-5119 to -4565	(555 bp)	-4564 to -88	-4640 to -58	(4583 bp)	2 (3)	yes
Empis	-5882 to -5450	(433 bp)	-5449 to -10	-5892 to +105	(5997 bp)	3	no
Haematopota	-1937 to -1646	(292 bp)	-1645 to -15	-9000 to +105 -1752 to +105	(9 kb)* (1857 bp)	3 (2) 5	yes no
Clogmia	-3050 to -2730	(321 bp)	-2729 to -9	-6872 to -3 -2440 to -3	(6870 bp) (2438 bp)	3 (3) <sup>ap</sup> 3 (3) <sup>ap</sup>	yes yes
Anopheles	-2464 to -2393	(72 bp)	-2392 to -2	-5205 to +31	(5236 bp)	4	no

\* estimate, fragment was not completely sequenced. <sup>MS</sup> Cloning and fly lines established by Michael Stauber. <sup>AP</sup> Cloning and fly lines established by Alexander Prell

## A.2 Supplemental figures



**Figure S1. Northern blot analysis of** *Clogmia hunchback.* The sampling covers the early embryogenesis up to the extended germ band stage. Each lane, 300 ng of poly  $A^+$  RNA were loaded. Poly  $A^+$  RNA was prepared from pools of 0-1 hour-old embryos, 3-6 hours-old embryos, and 6-11 hours-old embryos, respectively. After separation and transfer onto Hybond-N<sup>+</sup> nylon membrane, the RNA was hybridized to a radioactively labeled *Clogmia hunchback* cDNA probe, which covered the leader of the maternal transcript (351 bp) and parts of the second exon (747 bp). In all three lanes, a band is detected that corresponds to a transcript of about 2.8 kb. As size standard, the 0.24-9.5 kb RNA Ladder (Gibco BRL) was used; fragment sizes are given in on the left kilo bases. The experimental data are a courtesy of Alexander Prell.



Figure S2. Comparison of dipteran zerknüllt expression. Expression of zerknüllt homologues was compared at the onset of gastrulation by whole-mount in situ hybridization in Episyrphus (A, B), Clogmia (C, D), Drosophila (E, F) and Megaselia (G, H). Each embryo is shown in a lateral view (left panel) and in a dorsal view (right panel), respectively. Anterior is to the left. (A, B, C, D) In Episyrphus and Clogmia, zerknüllt is expressed in a stripe along the dorsal midline and extends to the anterior pole. Similar to dorsal hunchback expression (Figure 12 E, F), in Clogmia the dorsal zerknüllt expression is slightly broader and extends less far to the posterior than in Episyrphus. (E, F, G, H) In Drosophila and Megaselia, zerknüllt is expressed in a stripe along the dorsal midline. Expression is absent at the anterior pole (arrow indicates anteriormost zerknüllt expression. Arrows hint at the anteriormost expression, which in Megaselia extends slightly more to the anterior than in Drosophila. Scale bar: 450 μm in A, B; 180 μm in C, D; 220 μm in E, F; 240 μm in G, H.


Figure S3. DNaseI footprint mapping of Bicoid binding regions in Megaselia hunchback regulatory DNA. A Megaselia Bicoid-GST fusion protein spanning amino acid residues 78-159 (including the complete homeodomain, Stauber et al., 1999) was expressed and purified as described previously (McGregor et al., 2001b). The concentration of active protein was estimated by gel-shift assays using a double-stranded oligonucleotide with a single Bicoid binding site (5'-ATCTAATCCC) as described previously (Shaw et al., 2002; Zhao et al., 2000). (A) The genomic region 838-1151 of Megaselia hunchback (SEQ09, Appendix A.3) was analyzed by DNaseI footprinting for the antisense strand as described previously (Bonneton et al., 1997). 0.5 ng of labeled DNA were titrated. The triangle represents increasing concentrations of purified Megaselia bicoid protein (5, 0.5, 0.05, 0.005 nM). The negative control lane (no protein added) is indicated by "-"; bars indicate protected regions and the asterisk marks a hypersensitive site. The TATA box of the P2 transcript as well as exon sequence (P1 leader and P2 leader) are marked. The region between the 3' end of P1 exon 1 and the 5' end of P2 exon 1 corresponds to nucleotides 1040-1103 in with SEO09. (B) Summary of the protected (footprinted) regions in front of Megaselia hunchback P2. Protected sites are underlined and shown in bold, the hypersensitive site is marked with an asterisk. The experimental data are a courtesy of Philip Shaw.

# A.3 Sequences

The sequences are color-coded. Genomic DNA is set in black letters. Sequences that belong to the putative open reading frame (ORF) are marked red. Putative untranslated regions (UTRs) of transcripts are marked blue. Numbers to the left of the sequences indicate the position of the first nucleotide in the row.

## SEQ01 Platypeza bicoid, cDNA.

1	AGTTTAAGGC	CGGCGGCAAA	AATAGGCGCT	AGTGCAGTAG	GCAAAATGGC	GCAACACCCG	GACCAGAATT	TCTACACCCA	TCAACAACAG	TACGGGTTTA
101	ACAATAACCA	TCAACAAATG	CAATTTCCAC	CGCATTTCCG	GACGCCGTAC	GATTTTGTCA	AAATGTTTGA	CGAACGCGCG	GTGGCTCTGA	ATTACAACTA
201	TATGCGACCG	TATATGGCTC	ATCAGATGCA	GCAGATGCAG	ATGCAACAAA	TGCAGCAGCA	AATGCAACAA	GGTTACCATG	ATATGAACAA	TTCGATGCAC
301	GACATGTTGT	CCGAGTCGCT	AGTCATGCGG	CGTACGCGTC	GGTTGCGCAC	GACGTTTACC	CAACAACAAC	TGCAGGAGCT	CGAGCAAGAG	TTCCAGATCA
401	ACAAATATGT	AACAGCGCTC	CGCTTAGCGG	ACATTACAAG	CAGATTGAAT	TTGGCAAACG	CTCAGGTGAA	GATCTGGTTT	AAAAATCGGC	GGCGAAAGCA
501	TAAAATCGAA	GAGGCTCGCA	TGAAAGAGCT	CAAGGGCACA	CTCCCACTTG	GGTTGAATGT	GTCGATTCCC	AATTTGAATG	GGTCCCTCAC	CTCAAACAGT
601	CTGGACAGCT	CACTTTCGGA	ATCAGCGCCA	CCTAGCGAAA	CGAAAAGCGA	ATCGCCACCG	CTGCCGCTTA	CACCAAATCC	ACTAACACCG	TCGCCAACCC
701	CGTCTGCTAC	CTCAACACCA	AGTGCGTCTG	ATAAACAGTC	GGACAATTCC	AACTACGGCA	ATCAGTTCTA	TTACAACAAC	AATAACAACC	AAATGCCGCA
801	GTATTACCAA	ACACCGCCGG	CCACAAGCAA	CCAACAACAG	TTCGAGTTCC	CGACCAAGGT	TCAACAACAA	AACGAAACAA	GATACAACAA	CAATAACAAC
901	AACTTCAGCC	AGCAACAGCA	ATTCAACCGA	TTGGCATCCC	AGGAGAAGCT	CGCCGAGTTT	GCCAAACAAC	ТАААААТСАА	ATCGGAAATG	GCCGATTTTA
1001	ATTCGGCGGA	ATTGTCGCCA	AATTCTGAAG	TGTACGAACC	ACTGACACCC	CGAACTGACA	CGAGCCCACA	TTCCGGGCAT	TCAGACGAGA	TCGATGAAAC
1101	TCTAAAGTCA	AATCACGCTC	ACACTCCGAC	TGCAGCGGAG	TTAAACGGCG	ACGAGCCGCA	ACCCGATGCT	GCGTCCGCTG	CCTACCAGGG	CCAACCGATG
1201	TACAACAACA	ACTCGAATAG	AAGATGTGGC	GACGAACAGA	TGTTCGGCTA	CAGATACAAC	TAAACGAGTT	GTTCTCTAAT	TACCGTTATA	AAATTGTTTA
1301	TAATCTCAGT	GATTAGGTTT	CCGACCTAGT	ACATGTTTAG	TTGATAAGCG	CTTAGCCACA	TAAGTTTAGT	TTTAGTAACG	GTTTCCATCT	GCTAGTGATT
1401	TTTCGCTTCG	CTAGTCTCTT	CCGGTTTTCG	GTCCATGAAT	TCTGAAGAGC	CTACCGAAGC	CCATGGACCA	TTATCGCTAC	CAGATCGAAA	CAAATTCACG
1501	ATTTTCGCAA	ATTATGAAAA	ATCGCAAAGA	AAACAAAATG	AAGGAAATCA	TCCACTCAAC	AAAGACGG			

Source: Two independent 5' RACE products (1..374; 49..608), 3' RACE product (866..1568) and an additional PCR product (38..1539), all amplified from a larval cDNA template.

## SEQ02 Lonchoptera bicoid, cDNA.

1	TTGAACGATT	TGCGATAGTG	ATAGTTGTGA	TGGTCAAGTC	AGGTTTTGAT	ТАААААТАТА	AGAAAAGAAA	AAGTTTGTGA	ATAATATTTG	ATTGAATTTT
101	AAGTGATTTT	ATGTTTATAG	TAGTGAATAA	TTTAAGGAAT	AGTAAAATTA	AGGAATAATT	TACATAAAAC	TTATGTGAAA	GATATGTGAA	ATTTTTGTTT
201	TTTTTACATC	AAAGCAACTC	GCACATTCCA	TACAAAATG	GCGCAACCGC	CTGATCAAAA	TTTCTATCAC	CATCCGCAAC	TACAGCAACT	ACAGCTGCCT
301	ACGCAATTTC	GGAATCCATT	CGATTTGTTA	TTTGACGAAA	GAACTGGAGG	TTTAAACTAC	AATTATATTC	GGCCATATAT	ACCAACTCAA	CCAGTGGTAC
401	CAGATGTTCG	AAATGAAGCA	GTACGTGCTG	ATCCACTTGT	TATGCGAAGA	CCACGACGTA	CTCGAACCAC	ATTTACAAGT	GCGCAAATTT	CAAAACTTGA
501	ACAGTACTTC	AACGAAAGTA	AATACGTAAA	CGCTTCACGT	CTAGCTGAGC	TATCTGGTAA	ACTTAATCTT	GGAAATGCGC	AAGTAAAAAT	TTGGTTTAAA
601	AATCGTAGAC	GTCGATTGAG	AATTGAACAA	CTAAAACTGA	AGGAACTAAA	TGGATCAAAT	GATACAACAC	CAGCAGTCAG	TGTTTCTAAG	GATTTGTGTC
701	TTGCGTTGCC	ATTAACTCCA	ACAACTTTAA	CACCTTCGCC	ATCTTTAACA	CCGACTAGTA	CACCAAATAT	AAGCGATCAG	TACAGCGAGA	ATTATACGTA
801	CAATCCGTAT	ACTTATAATC	CATATGTACA	GCAGCATGCA	TATGAGCAAC	AAGTCAGAGC	ACAACCAATG	GCAACGCAAT	ATTATCAGCA	ACCATCAGCG
901	ATTAGTCAAC	AGCTTACAAG	AGATTTTCTA	ACATCAATTA	AAACGGAACC	GGATTTCAAT	TACAATAGTA	CTCCTTATAT	GCGAATGCCG	GCCGCAGAAA
1001	CTATGGTGAA	TTACACTAAA	ATTCCTACTA	AAAATTGCTA	TTTACCCGAA	CTGTCACCCA	ATTCTGAAGT	CTACGAACCG	TTAACACCAA	AAACTGAAGG
1101	CAGAGGAAGC	CCAAAAATGG	CAAATACATC	AGATGAAATT	AGCAATACAC	ATTTAGTTGA	TGCTAAACCA	GAAGTTTCGT	CGGATACAGC	ATCACAGATA
1201	TATGAAATGA	CTAAGTCAGT	ACCCGAAGGT	GGATACCAAT	GCACCATGGA	TTCGATATTG	CAAGCATACA	ATCAACATCG	CAATACCAAT	ACCAATAATG
1301	GTTACAATAC	TCAGTTTGCA	TTTTGCTTTA	ATTAAGTAAA	CAATCAAAAT	TATATTAATA	ACAAAATATA	ATTAGTTATT	AAGTATTAAT	TATAAAATT
1401	ATAATGTCTC	AGTGAGGTTT	GTTAGTTATT	AGCTTAAGTT	ATCGTTTAGA	AAATAGGCAC	TTACACCAAT	TTTTTTTCTCT	TTTTTTGCAG	GGTATTTTGT
1501	AACCAATTTA	GCTCTAAGTT	AAGATATACA	TATTATTATT	ATTTTTTTTT	TTTATAGTTA	TTTAAATGAT	TTGATAATAA	TTAGCTACGA	AACTTAATCC
1601	AAATATTAGG	TCCATGATAT	TTGAATGAAG	TTTTGTAGCT	CATGGACCAA	TATTATTATT	TCACTGCTGG	GGTATATATT	TGCTTAGCAA	AAACATAAAA
1701	AAGAAATAAA	CAAATTATGA	ATCTTTTTTA	AGAATTATCA	AAAATCTGAT	AGATACTCTG	AACAATTGCA	GATTATTTCT	CACAATTTTT	TTTATTTGTT
1801	AAGATCCCAA	CAATGCATTT	ATTTAAAACA	ATTGCAATGA	AACTATATTT	TTCGAAATAT	TTAAATTATA	AATATAATCT	GTAATTAAAA	CAGCGTTCCT
1901	CTTGCGAAAA	ATTCAAAAAT	GTTAATTTCT	CGAAAAGATT	GAGGCTTAGA	CAATATGATA	AAAAATTTTT	ATAATGTGAA	TTAATTTCAA	AAAACTTCAC
2001	CGAAATTTTT	AAATTTCTCT	ТАААТТАААА	TTATGTAATT	TCGGATGAAG	TTTTAAGAAA	TATCTTGGAA	TTGTTTTATT	TAGATACATT	TTTGATAAAG
2101	ATTTTTTGTT	TGTTTTTTTT	TTTTTTTAAT	TTTTTTTTGT	TTAATTTAGT	TATATTAAGA	TTACCAAATT	TAGAATAAGT	TATAATTCAA	GTTGTTGTTG
2201	CAAGAATTAA	ATGAATAGGT	ACCTAAAACT	CGTAAGGAAA	ATTGTTTTTT	TTTCAATATT	ACTTTGTACA	TTTAGTAATT	TAGTATTAAG	TTTTTCTATT
2301	TTTCTTTCTT	TTTTTGTAAT	ATTAATATGT	TGATTAAT						

Source: 5' RACE product (1..509), two independent 3' RACE products (550..1192; 909..2338) and an additional PCR product (488..1940), all amplified from a cDNA template of adult females.

### SEQ03 Episyrphus orthodenticle, cDNA.

1 GACTTCTGTA CGTCTTGGGG TCTTGGTCTA GCTTCTAGCA TATATAAATA TAAATTATAA GGTGTAATGT TATGATTGTG ATAAACTTTA AGAACTGTAA 101 ATTTATAGAC ATTTTATATG GCCCGCTCAG TAGCCTGTTG TTGATAAAGC TTGATAAATT ACTATTTCCA TTCAATTCCA TTGATAAAGT TGTTGAAGAT 201 AATTCACAGT ATTCTTACAT CACAATATTC CGCCGCAGCC GTTTAACCAST TGTCGAAGATA TATACAAAGT TTTTCTTTAG TCGATTGTGA AATAAAAAA 301 TAAAAAAAGT GCAAGAAAAT CATAGTTCTT CAAAACTTGA TTCGAAGCTG TTCCTCCAAA AAAATTCTCA TCTTTTCTGA TAATCAAACAT ATCGAGTGTT 401 AATAATAAT AATAATAATC ATCATCATAA CAACAGCAGA AGCAATATAA CAGCAGCAAC AAATCAGGAAA CACAAGTAAA ATAAATTCAT TCCAACGGCGC 501 GCTGCATTCA ACAGAAGCC TCATGGCCGC GGGCTTTTTA AAATCTGGTG ATTTAGGACC ACATCCGCAT AGTTATGGTG GTCCGCATCC 601GTACCACATGGACCATTACCACCGGGCATGCCAATGCCATCATTAGGACCCTTTGGGTTACCTCACGGTTAGAAGCTGTTGGGTTCTCCCAAGGTAGT701GGGGTGTAAATACCCGCAAACAAAGACCGAACGTACAACATTCACACGAAGGTATTAGATGGTATTGGGGCATATCCGGCAAAACACGATACCTGA801TATTTTATGCGTGAAGAAGTTGGTTAAAAATAAATCAACCCAACGACGAGTACAACAATCGGCACCGCCAAATCGA901CAACAACAAAACAACCAATTGGTTCAACAATTAATCACCCAACGACGGCAATAGGAACTCCGGCGACCCACGACAT1001GCAATAATAACACGGCCAACGCCAACAATCAAATAACTCTAGTGGCACCGGCAAATCCAAATCGACAACACAGGGCAATC1011CAACAATAACAACAACAAATCATCGGCGAATTCACCACCGGCCAATCAGGACCAAACCAGTGCTCACACGGGCAATCA1011CAACAATAACAACAACAAATCATCGGCGAATTCAGCGCCCGTCACACCGATGCGCCACACGGGCACATCGGGCACATCA1011CAACAATACAACAACAAATCATCGGCGCACACTCTGCCCACATGTCCAAAATCAGGTCAGACCAATCAGAGCACCCAGGCACAATCAGAGCACCCA1201GCTCAATCGCCCACCACTCAGCTCGCGCCCACTCTGCCAACATGTCCAAAATCAGGTCAATACAGGTCAGACAATCAGACCAATCAGACCACACA1301ACTACCAGACACCCACGCATGCGCGCCAGCGCAATGTCCGGGCACTCAGGCACATCAGACATGCACCCCAAGACACGCCACAGACCCCCAAGACACGCCACAGACCGCCAAGACACGACATACCGGCCACACACGCCACAACACGACCACACACGACCCACACGCCACAACAC</t

Source: 5' RACE product (1..763) and 3' RACE product (740..1603), amplified from an embryonic cDNA template.

#### SEQ04 Episyrphus hunchback, cDNA, P1 transcript.

1	AGTCGTGTTC	GAGACATCAC	AACGCAAAGG	ATGCGTGCGA	TCACGGTTTA	GATATTTATA	TACAAATAAC	AAAATTAAAA	TTTTAATAAT	ACATGCATTT
101	TTTCGTCTTC	TATAAACAAA	CAAATATAAT	TATTATTATT	GCTCTCTCAC	AACTGTCAAA	TCTTGGAAAA	TAAATTTAAC	AAGTGTATTT	ATTGAGACAT
201	TACAAAAACG	AATCAACTGG	ATTACCATTA	TATTTTTTT	AATGAAATCA	TATGATTTTG	TAGATTTATA	AACAATAATT	AACTGTGAAA	TAAAATATTG
301	CCAACAAAAT	CAGAACAGAT	CAAAGTGTGA	ATAAAAGAGT	ТАААААААА	TCTTTTTGTT	GTTGTAAGTG	ATGATGCTGG	CGCGTATTCG	AAAATAACGG
401	ATATTTAGTC	AGAATAAAAG	AGAAAACTGT	TTCATATTAA	ACTAAAAGTT	AATATTTAAA	TTTCGAAAAC	GGGGAATATT	AATTCTGTAA	ACGGAGAGAA
501	GTCTGAGAGA	GGGCAGTCGC	AGAGAAGCAA	AGAAAACCGC	TTTGAAACAA	AATTATTATA	AAGTTTCTAA	TTCAAATTAT	ATACCCAGAA	GCGGCTCCAA
601	GATGCAGAAC	TGGGATTCAA	TGCAGCCAGC	AGCCAATTAC	GAGCACAATT	GGTACAGCAA	CATGTTCCAT	CAGACAATCA	AGCAAGAGCC	TTCACAATCC
701	ACCACCCCA	CCACAAATCA	ACTGGAGCAT	TATCTCAACA	TGAAACAGCA	GGAGCTGTCA	TCGGCGATGA	CTCCTTCGCC	ACGAGTTCCC	GACTCAAATG
801	TCAATTCGGC	GATGATAGGG	GGCGATGTTG	GTAACAATAC	ACAGCATTAC	TTTGACAGCT	CAACGGGAAT	GTTGCATCAA	CATCATCCAC	TCGGATTTAA
901	TCCACTTACT	CCACCAGGGC	TGCCAAATGC	TGTCTTGCCT	TCGATGTCAC	ATTTTTATCA	ACAGAATACC	CATCAAAGTG	TCAGTACAGG	ACATCCTGTT
1001	GAATCTGTGA	CCAAAATGGA	TCAACAGTCT	ACGAATAACA	ACTCACTTAC	ACCGAGAAAT	ACTCCACCGA	TGGATGTGAC	TCCACCCAAG	TCACCCAAAC
1101	TATCACTAAT	GATTTCCACT	TCGAGTGAGT	TAGATCAAGA	TGTGATGTCA	TCTAATTCTA	GTGAGGACAT	GAAATACTTA	GAAAGCGAGG	ATGACGAAAG
1201	TATCCGCTTG	CCAATCTATA	ATTCACATGG	AAAGATGAAG	AATTACAAAT	GCAAGACTTG	TGGATTTCTG	GCTATAACAA	AAGTGGCCTT	TTGGGAACAT
1301	GCACGTTGTC	ATATGAAGCC	TGAAAAAACA	CTTCAATGTT	CCAAATGCCC	TTTTGTCACC	GAATTAAAGC	ACCACCTTGA	ATATCACATT	AGGAAGCACA
1401	AGAATTTAAA	GCCATTCCAA	TGCGACAAGT	GTAACTACAG	CTGTGTTAAC	AAGTCTATGT	TGAACTCCCA	CCGGAAGTCA	CACTCGTCGG	TGTATCAGTA
1501	TCGTTGTTCT	GACTGTGATT	ACGCCACAAA	ATACTGCCAT	TCTTTTAAAT	TACATTTAAG	AAAGTACGAC	CATAAGCCCG	GCATGGTTTT	AGATGAAGAT
1601	GGTGCACCCA	ATCCATCTGT	TGTAATTGAT	GTCTATGGAA	CACGACGTGG	GCCAAAAATG	AAGTCCGGCA	GGAAAGAATC	AACTGGGGCG	TCAAAGATGC
1701	CACAATTGAG	TGCAGCCTTG	CAAGGATTTG	CCCTTAATCA	GCACAACAAT	CAACACAACA	TGCCTGCTTC	TCCAGCCAAA	AGTACTGCTT	CGTCATCGTC
1801	AGAGATAGTT	CCAAATACCC	AGTCGAATAT	TCAAGCTAAC	CACCAGCATT	TACAGCAGTC	AACTTCCACT	CAACAAGAAC	AAATCGAACA	ACACCATCAT
1901	CAGCAACAAC	AGCAGCAGCT	TTCTAATCTC	ATCCCCTCAT	CGTTGTCAGC	GATTCTGCAA	CAACAAAGAA	ACATACCATT	TTTTCCTTAT	TGGAACCTTA
2001	ATCTTCAAAT	GTTGGCTGCC	CAACAACAAG	CTGCTGTTTT	GGCCCAGTTA	TCGCCAAGGC	TTCGTGAAAC	AGCATTACAA	AATCTGCAAG	ATAAGCAAGA
2101	AGACAAAAAT	GCAAGCTTTA	TAAATGAAGA	TGACGAAGAG	CATGATGAAA	GTTGTGACGG	AACAGCTATG	GATCTCACGG	CCGCTACTCC	AACTAAAAAC
2201	AACGAAGACA	TCAATTCTTC	TATTGTAAAC	CTTAAATTGA	AAGAGGATGA	CCAGCATGAG	ACTCCTCTTA	TAAGTTCGTC	CAATCAGTTC	CGTCGCAAAG
2301	GTCGAGCACT	AAAACTTGAT	GCAGCTCTTC	AAGCTAAGGA	AAATTCCTTG	AGTCCAGAAG	AGAAGCCTCG	ACTTAGCCCG	AATCCAAATG	AGGTTCCGTC
2401	ATCTTCATCA	TTCGATGATC	CATCAAAAGA	GAGTGAAACC	TCACAATCTA	ATGAAGACTC	GTCAAGGCCC	TCCAACAGCA	CCTCAAATCC	AACATCAACT
2501	CCCACAACCC	ACTCCACGTC	CACAACCAAC	AAATCACCTC	CATCRGGCGC	CATTTTCGAA	TGCAAATATT	GTGATATATA	TTTCCGTGAT	GCAGTCCTTT
2601	ACACTATTCA	TATGGGCTAT	CACAGTTGTG	ATGATGTATT	CAAGTGCAAC	ATGTGTGGTG	AAAAATGTGA	TGGTCCCGTC	GGATTGTTCG	TCCACATGGC
2701	ACGCAATGCT	CACTCTTAAA	AGAGAGTTCA	CTTGATAAAC	CCCTTTTTTG	TTATTTCATA	GCTCGTTTTT	GGGAGATATT	CAAAATCAAC	CAAAAACAGA
2801	GTTTCGTTGT	CAAAGATTGT	AAATATTAAC	ATAATAAAT	TTATATTTTG	TTATGAAAAA	ATTAAATATA	AGTAAA		

Source: 5' RACE product (1..897), 3' RACE product (1478..1885), and an additional PCR product (498..2876), all amplified from an embryonic cDNA template.

#### SEQ05 Episyrphus hunchback, cDNA, partial P2 transcript.

1 GAAATCCTTA AAATGTGTTT CTTGGTGCAC AATAAAAATT TCTAATTCAA ATTATATACC CAGAAGCGGC TCCAAGATGC AGAACTGGGA TTCAATGCAG 101 CCAGCAGCCA ATTACGAGCA CAATTGGTAC AGCAACATGT TCCATCAGAC AATCAAGCAA GAGCCTTCAC AATCCACCAC CCCCACCACA AATCAACTGG 201 AGCGTTATCT CAACATGAAA CAGCAGGAGC TGTCATCGGC GGTGACTCCT TCGCCACGAG TTCCCGACTC AAATGTCAAT TCGGCGATGA TAGGGGGCGA 301 TGTTGGTAAC AATACACAGC ATTACTTGGA CAGCTCGACG GGAATGTTGC ATCAACATCA TCCACTCGGA TT

Source: 5' RACE product, amplified from an embryonic cDNA template.

#### SEQ06 Episyrphus hunchback, cDNA, partial P3 transcript.

1	AGTTCGTGAT	CGGTATCTTG	TTGTAAAACA	CCAAATCTTC	GTTCCGTTAT	CAGGAAAACA	ATTTAAATAA	CGAAATTTGA	TGTCAAAGTG	TATTTCTGGC
101	CAAGGAACAA	АТААТАТААА	AAAAGTTTCT	AATTCAAATT	ATATACCCAG	AAGCGGCTCC	AAGATGCAGA	ACTGGGATTC	AATGCAGCCA	GCAGCCAATT
201	ACGAGCACAA	TTGGTACAGC	AACATGTTCC	ATCAGACAAT	CAAGCAAGAG	CCTTCACAAT	CCACCACCCC	CACCACAAAT	CAACTGGAGC	ATTATCTCAA
301	CATGAAACAG	CAGGAGCTGT	CATCGGCGAT	GACTCCTTCG	CCACGAGTTC	CCGACTCAAA	TGTCAATTCG	GCGATGATAG	GGGGCGATGT	TGGTAACAAT
401	ACACAGCATT	ACTTTGACAG	CTCAACGGGA	ATGTTGCATC	AACATCATCC	ACTCGGATTT	AATCCACTTA	CTCCACCAGG	GCTGCCAAAT	GCTGTCTTGC
501	CTTCGATGTC	ACATTTTTAT	CAACAGAATA	CCCATCAAAG	TGTCAGTACA	GGACATCCTG	TTGAATCTGT	GACCAAAATG	GATCAACAGT	CTACGAATAA
601	CAACTCACTT	ACACCGAGAA	ATACTCCACC	GATGGATGTG	ACTCCACCCA	AGTCACCCAA	ACTATCACTA	ATGATTTCCA	CTTCGAGTGA	GTTAGATCAA
701	GATGTGATGT	CATCTAATTC	TAGTGAGGAC	ATGAAATACT	TAGAAAGCGA	GGATGACGAA	AGTATCCGCT	TGCCAATCTA	TAATTCACAT	GGAAAGATGA
801	AGAATTACAA	ATGCAAGACT	TGTGGATTTC	TGGCTATAAC	AAAAGTGGCC	TTTTGGGAAC	ATGCACGTTG	TCATATGAAG	CCTGAAAAAA	CACTTCAATG
901	TTCCAAATGC	CCTTTTGTCA	CCGAATTAAA	GCACCACCTT	GAATATCACA	TTAGGAAGCA	CAAGAATTTA	AAGCCATTCC	AATGCGACAA	GTGTAACTAC
1001	AGCTGTGTTA	ACAAGTCTAT	GTTGAACTCC	CACCGGAAGT	CACACTCGTC	GG				

Source: 5' RACE product, amplified from an embryonic cDNA template.

# SEQ07 Episyrphus hunchback, genomic.

1	GAAGTCTGAG	AGAGGGCAGT	CGCAGAGAAG	CAAAGAAAAC	CGCTTTGAAA	СААААТТАТТ	ATAAAGGTAA	GTTTTTGATT	ATTGGAAAAA	GTTATTTTT
101	ጥር ጥጥጥ እ እ ጥ እ	ጥጥጥጥ <u>አ</u> ጥጥጥ አ	ጥጥ <b>አ</b> ርጥጥጥጥል ጥ	CCCTCCATCA	ልጥሮርጥጥጥጥሮል	ሮሞልልልጥጥጥዋል	TTCCAACCTC	ርጥጥጥልጥልልጥጥ	ጥልጥልጥጥጥጥሮር	CTCACTTTT
201	СЛССЛПЛПСЛ	777711111111 7777777777777777777777777				CULLUITI	ACTICANACTIC		ACCUACUAAC	CTAACCCCCA
201	GAGCATAIGA	AAAIIIAIIC	AGCIAIIIGA	ATTCGAATCT	ITAAGACATT	GIIIAGCAAI	ACICAAAGIC	IGACATITAA	ACCIAGIAAC	GIAACCGGGA
301	CTCTTAGAGT	TGATCTGTGC	ATCT TGAAGC	TGAATGGGCA	ATTITACAATA	ATCGTATAGC	ACT TTTTTTG	TTGATTCTCG	'I''I'AGAAG'I''I'C	AGGTACCTAC
401	CCTAAATAAT	TGTTTACAAA	ACCTGAAATT	AAATTTGTAT	CTTACAAAAA	AGTTGTGCTG	TAAAATAGAC	TGACTTTCCC	TTTTCATCAG	TAATTTAGGA
501	ACAATGAAAG	ATTCTAATAA	AGTTTTCAAA	ACCTTTCTTT	TTTGTTTACT	TTAACATTAT	AATAATTTAC	GCTATTTAAA	GTTTTTAAAG	AGTGGTTCTG
601	TTATCGTGGA	TGCAACATTC	AAACTGATCA	GGTAGAATAA	TAGGTAATAA	GCCTTGAAAG	AATAATTTCA	TCAACAAGTT	TATTTTGTTT	TTTCGTACCT
701	ааатааатт	САААТАТСТА	CAAATGTTGC	ACCCGAGATA	ACTGAATCCC	TAATTTAGCT	GTTTGCTAAA	АТСТСАААТТ	GATACTTTA	ACTATAATTC
001				пссополіті			CACCECAACA	CUMCACACCA	CINCCOMAINCO	
001	TATACTTAAC	AIGCIGAIII	TIAAICIIAI	ICCATAAAAA	AIIGCIIAAA	AAAIAAAAAI	CACGIGAAGA	CIIGAGAGCA	CICGCIAIGC	AGAAIAIAAI
901	GGAACACCTG	CITTITATATAG	T'TTCAATTGC	TAAAGCTACA	TAAATCCCAC	GAAA'I'A'I''I'AA	ATAAAACGAT	GATCTAATCT	TCCTTTTAAG	ATGCACCATT
1001	TATACATTAA	TAGATATATG	TACACACATA	TGTATATATC	GAAAACAAAC	TTATAAGAAT	GTTAAAATAA	ATCCCAAAAA	CCATTAAAAA	CAAATAAACA
1101	AACAAAAAAA	AACAAAGATT	CTCACAGAAA	AATTTCTATT	CTTGTGGTTA	AAAAATTTTA	CAAAATTACA	GTTTTTTTTT	TAATTTGCAA	AATTTTATTT
1201	TAAAATCATT	ACTCACAAAC	TAAAATGCAG	CGTGTATAAC	CTGCCCCTAC	CTACCTATTA	CCTTCCATGC	CTCCACATGT	AAAACAATTT	TTTGTTGCTG
1301	CACAAAAAAG	TCTGCGGAAT	ATAGTTTTTG	TCTCACAACA	AAATTAGAAG	CAGCCATAAT	TCAAAGTCAC	ТААААААСТТ	GGCGCAAAGC	CTCTGTATTC
1401	ACTICACA ATC	ልጥልጥሮልጥጥልሮ	ССТСАТСАТС	GTCAAGATTG	ጥርጥጥጥጥርርጥል	Շափանանան	TTTTCCTTTCTC	AAATCCCTAC	ΔͲͲϹΔΔͲͲႺϹ	ΔΔͲϹͲͲϹͲϹͲ
1501	ACTORORATC	ATAICATIAC	TCTCCATCATC	CAAAMMMAAC	CAACACAACA		TITCCITCIC	AMMOMMMMM	MICANACAA	
1501	AAAAICACCG	ACGGCAAAAG	IGIGCAICAI	CAAAIIIAAG	GAACACAAGA	AAAAAAIAAA	ICIGGCIIIA	AIIIIIIA	IAGAAAACAA	AGAIIIICAA
1601	AAATACGCAG	ATACAATTI'T	GCAGCGTGTG	TATTITTCAG	ACGTTTTTTC	CICCCCGCCC	ACCGAGAACT	TGACCGGAAC	ACAACCAAAC	'I'AGAAAAAAA
1701	AGATGAAAGA	GAAAAAGATT	TTTGCACTAA	AATAGTTTAA	AAATATATAA	AATATAAAAA	CAAAAAGAAA	AAACAAAACT	AAAATGTATC	TTCTACAGGA
1801	AAACTTTTTC	ATATAACTAA	AAGAATGTAT	CTTAGATTTT	TTAAAATACA	TTTTTTTGTTT	ATGCATCGTT	TGATGCATTT	TTCGTTTATA	TGGTTGCAAG
1901	GCAGTATCGG	TGAAGCAACG	ATGAGAAATG	AAGATTCAAC	AAGTCGATTT	AGATGCAAGA	AGATTCTCTG	ATAGATAAAT	TTTTCGAAAA	TTGCTTCATT
2001	AATTTTTATA	АААААААТАG	AAAAAAAGAT	АТААААТСТА	CACTTGGCCT	TTGAGATCAG	ATGTGCACAT	GTGGGCGTAG	TCGGCAATTT	TTTTTTCTTG
2101	CAATAAGGAA	AGTATCCACA	ጥልጥጥጥጥጥርር	CTTTCACAAC	ΔͲΔͲϹϹϹΔΔΔ	<u>አ</u> ሞአ አጥጥጥጥ አጥ	ΔΔΔΔΤΔΔΔΔC	<u>አ</u> ሞርጥጥል አጥጥጥ	ጥልልሮሞልልልልጥ	
2101	AMOCOMMAA	CONCINENCACA	1AIIIIIICC	GIIIGACAAC	ATAICGCAAA		MANALAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	MONORMONN		
2201	AATCGCTTAA	GICATTIGCA	AAATACATGG	CTAIGTAAAC	ATAGATGCGG	TTAACTAAAG	TGGTGAAAAA	TCACATCAAA	CITTAAACCG	AATTCTTTTA
2301	CTTTATAGAC	CAAGTCATTC	TTAATTGAAT	AGGTGTAATT	TAAAATCTCT	TCTTTTTTTT	TGTTGTTGTG	TCAGCTTTCA	GGTTAAAGGC	TGGCGCTGGG
2401	TTAAACTATA	AAATTTAACC	TAGCGTTAGC	CTTTAACCTG	CTGCTGCTGG	TTATTGGTAT	GGTGATGTTG	AATTCAGTTC	ATTATTTAAG	CGAGATTTAG
2501	ATTTGATTAT	TGGTAAGGTG	AGTTTTATTT	TTTACCAACA	CTCAAATAGA	TATTTTACAT	AGTTTCCAAG	TCCTGTCATA	TGAAAGTAAT	TGATATTTAT
2601	GAAATTAAAT	ТСТТААСТСА	TTGCATCTGG	ТТААТАССАА	GTACTTCACT	GTTTTGTTTT	TTCTTGTTAA	CCGTTATCTT	ΑΤΑΤΤΑΤΑΤΑ	TAGCTTTAAC
2701	ልጥል ልጥጥል ጥርጥ		ΔΔΔCCͲΔͲͲͲ		ͲͲͲϪͲϹͲͲϹͲ	TCCCACCTAA	CCCAATTCTA	ACTTCAAACC	ጥልሮጥጥጥጥጥጥ	ͲͲͲͲΔΔΩͲͲϠ
2001	COMPANY CC				ANGGGARTES	AMA ACCENT	TOOCOCO	ANDARA		A DEPACTOR
2801	CGTATGTAGC	TTAAAA'I'A'I'T	GCTGAA'I''I''I'A	AATTCATACT	AAGCGGA'I'AG	ATAACGTAAG	TGGGCGATCG	AATAAAAAAA	AACTAAGAAA	AATTACCTGA
2901	CACATAAAGG	CTAAATATAG	AATGAATTAA	GAAATTATTG	TCTTAGACCA	ATTTCGAGAA	CGACGCTTAT	ATTATATAAT	AAAATCCACT	TTTTCCAGTG
3001	AACTCTTCAT	AAAATTGTTA	TCCTGATAAT	TCGAGCGAAT	CAAGGAAAAG	AAACTTCTTT	TGATTATTTT	TCATGAACAA	TTCTAATACA	TTAAAAAAAT
3101	CCTTGTGTTA	TTTAGACAAA	AAATATGGTT	TTTTAAGTAG	TTGAAATTAT	ATGTAAACTT	AAACTGAAAG	CCATCTAACT	GCAACCTGTT	AGTTCGACAT
3201	TATAATGATT	TCGATCATTT	ΑΤΑΑΤΑΑΤΑ	CAATCAGCTT	тсааатстаа	ΑΤΑΑΤΑΤΤΤΤ	TACATTACAA	ATAATATCGT	ACTACATTTA	TGTACACACC
3301	<u>አ</u> መል አ አመምርምር	TATCTA A A CC	አመምርጥጥምርጥጥ	እእእመአመሮአመአ	ሚአአመመመርመር	እርሞርእር እርሞሞ	CATCTCATT	TOTACOTOTO	TTA A A A A CCT	አ አ መመመመመመር
2401	ATAAATIGIG		MIICIIIGII	AAAIAICAIA	AMAIIIICIC	CTCCAGACIT	COMMONDATI	TCIACCIGIG		
3401	AATAATCCTT	CTATATTGC	TTTTGCTCAA	AAATCAATGA	ATATTGGGGGT	GIGCACTITIC	GITTITCTT	TIGTAAGTAT	GATGTACTTT	TTTGTGAAAA
3501	CCGTGACGAT	ATATCTTGAA	AATTCTTGCA	TACACACTAT	AAATAACAAA	TAATATGAAA	ATTTATAGGA	AATCCTTAAA	ATGTGTTTCT	TGGTGCACAA
3601	TAAAAAGGAT	GTTGATTTTT	CAAACCGATT	TGCATTGCAA	GATGTTTGAC	TTTTACTTTT	ATGGTTACCT	ATGCATACTT	TTTTATTTT	TTTGTTTTTG
3701	TGTTCGTGTT	GACCCTAAGT	CATGAAACTT	AGTAATTTTG	AAAGTAATTT	TTGACTGGTC	ATTGGGTTTT	TATGAGTAGT	GAGTTGATGT	CTTTTGAAAA
3801	AGATTTATTT	TTTCTCAAAT	AATGGTTTTA	AATTCTGAGA	AAGAAATTAA	CAAAGGAGAC	AAGTCATAAA	GGTTTTAAAC	AAATATTCAT	ATTTGTAAAT
3901	ΔGΔTΔTGTΔΔ	GGAGTATTAG	AACCAAATGC	ጥጥጥጥጥጥርርጥር	GTCATTTTT	ጥጥጥጥጥርልልልል	ACAGGTAGGT	ΔGGTAAAATT	TCACTGGACA	ΔͲͲͲGΔΔGΔΔ
4001	110111111011111	0010111110	10100101100	1111110010	0101111111	TTTTTOTTTT	1101100111001	TICOTTURNET	10101001001	111 1 1 01 01 01 01 01
		mammmmmm		CCMACCMAMA					ananan nan	
4001	CAGAAATTAG	TCTTTTTTTA	ATATGATTTA	GGTAGCTATA	TATCACAAAC	ACATAATATT	TTTTTTTAAG	TAAAATAAGT	CAGACAAACA	CCTGATTAAA
4001 4101	CAGAAATTAG AAATTTTGGA	TCTTTTTTTA ATACAAAATG	ATATGATTTA TGGTAATCCA	GGTAGCTATA CTATAGAAGC	TATCACAAAC CTTCAAGAAT	ACATAATATT ATATTAAAGT	TTTTTTTAAG TGAGCTATTT	TAAAATAAGT TTCGAGGAAA	CAGACAAACA AGCTGTGCTT	CCTGATTAAA CCAGCAATCA
4001 4101 4201	CAGAAATTAG AAATTTTGGA GTCATACTTA	TCTTTTTTA ATACAAAATG TTTCATTTAA	ATATGATTTA TGGTAATCCA AAAACTTTAT	GGTAGCTATA CTATAGAAGC ATTTTTAAGA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG	ACATAATATT ATATTAAAGT TATGTATCTA	TTTTTTTAAG TGAGCTATTT AGTACCTCAC	TAAAATAAGT TTCGAGGAAA CAAATTATAG	CAGACAAACA AGCTGTGCTT CCATAAATAT	CCTGATTAAA CCAGCAATCA TTTTCTCATT
4001 4101 4201 4301	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG	TTTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA
4001 4101 4201 4301 4401	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA	TTTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTTAACATT	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC
4001 4101 4201 4301 4401 4501	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTTAC	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTTGGCTT GAATGTCAAT	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT	TTTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAAATC TTTTTAACATT AGTTCGTGAT	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGCTATCTTG	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC
4001 4101 4201 4301 4401 4501	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTTAC	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAACA TCCTGCCTTT	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT GAATGTCAAT	АСАТААТАТТ АТАТТАААGТ ТАТGТАТСТА АСАСАТАААG ССАСАGААТА GATATGACTT	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT AGTTCGTGGTAT	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC
4001 4101 4201 4301 4401 4501 4601	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTTAC GTTCCGTTAT	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGGAAAACA	АТАТGАТТТА ТGGTAATCCA ААААСТТТАТ СТGTAAGCCT АGAAAAAACA TCCTGCCTTT АТТТАААТАА	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT GAATGTCAAT TGTCAAAGTG	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAAA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTTTC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA
4001 4101 4201 4301 4401 4501 4601 4701	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTTAC GTTCCGTTAT TAATAAAATG	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGGAAACA GAATGTTATT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATTCTA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA	TTTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTA	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAAA AAAACCTCAA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA CGCTTGCTTC
4001 4101 4201 4301 4401 4501 4601 4701 4801	CAGAAATTAG AAATTTTGGA GTCATACTTG ACCTACATTG ATTGCCATAA TGATTTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGGAAAACA GAATGTTATT GAATTGTGCT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATTCTA ATATACATAT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT	TTTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTA ACTTTCCTGA	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAAACCGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCATCA AATAAATCAA CGCTTGCTTC AAAAACAATAA
4001 4101 4201 4301 4401 4501 4601 4701 4801 4901	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGGAAAACA GAATGTTATT GAATTGTGCT TTGCTTCCCA	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAAAA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATTCTA ATATACATAT AAAAAACATA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTA ACTTCCTGA AAAAATATC	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGG CGGTATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCAAAA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA CGCTTGCTTC AAAACAATAA AGAACTATAT
4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001	CAGAAATTAG AAATTTTGGA GTCATACATTG ACTACATTG ATTGCCATAA TGATTTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT	TCTTTTTTA ATACAAAATG CTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGGAAAACA GAATGTTATT GAATGTTCCCA ATAGTAATCG	ATATGATTTA TGGTAATCCA AAAAATTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCTA ATATACATAT ATAACATA CTTCTTCTTA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA TGTATGAAGT	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCCAGA	TTTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTA AATACCATGTA AAAAAATATC AACTACATAC	ТААААТААGТ ТТСGAGGAAA САААТТАТАG ААСТСТАСТТ ТАССТGCGCG СGGTATCTTG АТААТАТААА ААААССТСАА АТТТТАТТТGA АССАААААС АGGGGGAGAG	CAGACAAACA AGCTGTGCTT CCATAAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGT TATTTCAAAA ATTTCCATGA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA CGCTTGCTTC AAAACAATAA AGAACTATTA ATTCCACCCT
4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGGAAAACA GAATGTTATT GAATTGTGCT TTGCTTCCCA ATAGTAATCG TGCTTTATAG	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCGGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGGT CGAAATTTGA AAATATTCTA ATATACATAT CATATACATAT CTCTCTTTA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA TGTATGAAGT GACGTTTTTT	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCCAGA CCTGGACACG	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTATAAATC TTTTAACATT AGTTCGTGAT ATACCATGTA ACTTTCCTGA AAAAAATATC AACTACATAC	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAA AAAACCTCAA TTTTATTTGA ATCCAAAAAC AGGGGAGAG GTGCCAAAAA	САGАСАААСА АGCTGTGCTT ССАТАААТАТ ААGAAACCAA ААААССGGGAT ТТGTAAAACA АЛААGGTACA АЛGGATTTTT АAGCAGGAGA ТАТТТСАААА АТТТССАТGA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCTC
4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101	CAGAAATTAG AAATTTTGGA GTCATACATTG ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT AAAGGGGGTT	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGGAAAACA GAATGTAATTG GAATTGTGCT TTGCTTCCCA ATAGTAATCG TGCTTTATAG	АТАТGАТТТА ТGGTAATCCA ААААСТТТАТ СТGTAAGCCT АGAAAAACA TCCTGCCTTT АТТТААТАА ТССGATACTA ТАGCGCGTGG АААСАТАААА СТТТТСАGTA АТТАТТАСАТ	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTCA AAATATTCTA AAATATCTA AAAAACATA CTTCTTCTTA ACGGAGGGGT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA TGTATGAAGT GACGTTTTTT AAACCATAAA	ACATAATATT ATATTAAAGT CACATAAAG CCACAGAATA GATATGACTT TATTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCCAGA CGTGGACCAG	TTTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTA AACTTCCTGA AAAAATATC AACTACATAC CGCAGAACGT	ТААААТААGT ТТССАGGAAA САААТТАТАG ААСТСТАСТТ ТАССТGCGCG СGGTATCTTG АТААТТАТАА ААААССТСАА АТТТАТТТGA АТССАААААС АGGGGGGAGAG GTGCCAAAAA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCAAAA ATTTCCATGA AAAACAAACA	ССТGАТТААА ССАGСААТСА ТТТТСТСАТТ ССААТСАGGА ТGААААТТАС ССАААТСТСС ААТАСАТСА СGСТТGСТТС ААЛАСААТАА АGAACTATAT АТТССАСССТ АТТССАССТ
4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101 5201	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT AAAGGGGGTT	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGCAAAACA GAATGTTATT GAATTGTCCT TTGCTTCCCA ATAGTAAATCG TGCTTTATAG CAAATGATAA	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCCAGTA ATTATTACAT AAAATGGCGT	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATAATCATA ATATACATAT AAAAAACATA CTTCTTCTTA TCGGAGGGGT AGCTTAAAAA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA TGTATGAAGT GACGTTTTTT AAACCAATAAA	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTCTGGC ATCTGCTTCA AGGACCAAAC CCACCCCAGA CGTGGACACG AAATGAACTC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTA ACTTTCCTGA AAAAAATATC AACTACATAC CGCAGAACGT ACAAAATGGC	ТААААТААGT ТТССАGGAAA ААСТСТАСТТ ТАССТGССGС СGGTATCTTG АТААТАТАТА АТААТАТТАА АТААССТСАА ТТТТАТТГGА АТССАААААС GGGGGAGAG GTGCCAAAAA CCACATTTAT	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAAAGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCCATGA AATTCCATGA AAAACAAACA TCATCTATAT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATTAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCCT ATTGCAGATA ACTTTTCTCTC
4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101 5201 5301	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATGCCATAA TGATTTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT AAAGGGGGTT TTATCTTTAC	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTGA ATAATTGAGT CTGCACTTTT CAGGAAAACA GAATGTTATT GAATTGTGCT TTGCTTCCCA ATAGTAATCG CAAATGATAA AGTTTCTTATAG	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAATTATA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATTCTA ATATACATAT CTATCTTCTTA AGGAGGGGT AGCTTAAAA TACCCAGAAG	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA GGACGTTTTT AAACCATAAA CGGCTCCAAG	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCCAGA CCTGGACACG AAATGAACTC ATCCAGAACTC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTATAAATC TTTTAACATT AGTTCGTGAT AATACAATGTA ACTTTCCTGA AAAAAATGTC CGCAGAACGT ACAAAATGGC GGGATTCATT	ТААААТААGT ТТССАGGAAA ААСТСТАСТТ ТАССТGСGСG СGGTATCTTG АТААТАТААА АЛАССТСАА ТТТТАТТТGА АСССААААА СGCGCAAAAA GCACATTTAT GCACCCACAA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA AAGGAGTTTTT AAGCAGGAGA TATTTCCAAA ATTTCCAAAA ATTTCCATGA AAAACAAACA TCATCTATAT GCCAATTACG	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATATA ATTCCACCT ACTCTTCCTC AGCACAATTG
4001 4101 4201 4301 4401 4501 4601 4701 4801 4801 5001 5101 5201 5301	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TAATAACATA AAAGGGGGTT TTATCTTTAC GTACAGCAAC	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTAA CTTCCAGTTTG ATAATTGAGT CTGCACATTT CAGGAAAACA GAATGTTACTT GAATGTGCT TTGCTTCCCA ATAGTAATCG TGCTTTATAG CAAATGATAA AGTTTCTAAT ATGTTCCAC	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCCAGTA ATTATTACAT AAAATGGCGT TCAAATTATA AGACCATCAA	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTCA AAATATACATAT ATATACATAT ATATACATAT TCGGAGGGGT AGCTTAAAAA TACCCAGAAG GCAAGAGCCT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA TGTATGAAGT GACGTTTTTT AAACCATAAA CGGCTCCAAG TCACAATCCA	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAC CGTGGACAG AAATGAACTC ATGCAGAACT CCACCCCAC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTAAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTA AACTTCCTGA AAAAAATATC AACTACAATAC GGCAGAACGT GGGATTCAAT CACAAATCAA	ТААААТААGT ТТССАGGAAA САААТТАТАG ААСТСТАСТТ ТАССТGСGСG СGGTATCTTG АТААТАТААА ААААССТСАА АТАТАТААА АТТТАТТТ	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA ATGGATTTT AAGCAGGAGAA TATTTCAAAA ATTTCCAAAA ATATTCCATGA AAAACAAACA TCATCTATAT GCCAATTACG ATCTCCAACAT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCATCA CGAAATCATCA CGCTTGCTTC AAAACAATAA AGAACTATAT ATTGCAGATA ACTTTTCTCT CACACAATTG GAAACAGCAG
4001 4101 4201 4301 4401 4501 4601 4701 4801 5001 5101 5201 5301 5501	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT TATATACCAT TATATACCAT GAACGGGGTT TTATCTTTAC GAACTGTCAT	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGCAAAACA GAATGTTATT GAATTGTCCT TTGCTTCCCA ATAGTAATCG CAAATGATAA AGTTTCTAAT CGGCGATGAC	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA AATATACAT AAAATGGCGT TCAAATTATA AGACAATCAA TCCTTCGCCA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATTCTA ATATACATAT CTTCTTAT TCGGAGGGGT AGCTTAAAAA TACCCAGAAG GCAAGAGCCT CGAGTTCCCG	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAC GACGTTTTTT AAACCATAAA CGGCTCCAAG TCACAATCCA	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTC AGGACAAACT CCACCCCACA ATGCAGAACTT CCACCCCCCC CAATTCGGCG	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT CAAGGAACAA ATACCATGTA ACTTCCTGA AACACATGTA ACTATCCTGA AACTACATAC GGCAGAACGT ACAAAATGGC GGGATTCAAT CACAAATCAA ATGATAGGGG	TAAAATAAGT TTCGAGGAAA AACTCTACGT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACAAAAAC GCGCAAAAA GCACCAACA GCGACCAGCA GCGACCAGCA GCGATGTTGG	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAAAGGATAT TTGTAAAACA ATAGGATTTTT AAGCAGGAGA ATTTCCATGA AATTCCATGA AAAACAAACA TCATCTATAT GCCAATTACG ATCTCAACAT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATC AATAAATTAC GCTTGCTTC AAAACAATAA AGAACTATAA ATTCCACCCT ATTGCAGATA ACTTTTCTCT AGCACAATTG GAAACAGCAG CAGCATTACT
4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101 5201 5301 5401 5501	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT AAAGGGGGTT TTATCTTTAC GTACAGCAAC GAGCTGTCAT TTGACAGCAC	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTAA CTTCCAGTTTG ATAATTGAGT CTGCACTTTT GAATGTGTCTTATT GAATTGTGCT TTGCTTTATTAG CAAATGATAATCG CAAATGATAA AGTTTCCAATC CGGCGATGAC AACGGGAATG	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCCAAATTATA AGACAATCAA TCCTTCGCCA TTGCATCAAC	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATTCTA ATATACATAT CATATACATAT CTCTTCTTA TCGGAGGGG AGCTTAAAAA TACCCAGAAG GCAAGACCCT CGAGTCCCG ATCATCCACT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAG GACGTTTTT GACGTTTTT AAACCATAAA CGGCTCCAAG TCACAATGT CGGATTTAAT	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTY AGGACCAAAC CCACCCAAG CAGTGGACACG AAATGAACTC CCACCCCAC CAATTCGGC CCACTTACTC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAAATC TTTTAAATC CAAGGAACAA ATACCATGTA ACTTTCCTGA AAAAATATC AACTACATAC GGCAGAACGT GGCAATCAA ATGATAGGG GACCAGGGCT	ТААААТААGТ ТТССАGGAAA ААСТСТАСТТ ТАССТGСGСG СGG7AACTTG АТААТАТААА АЛААССТСАА ТТТТАТТТGА АТАССТСАА АТССАААААА СТССАААААА GCACATTTAT GCAGCCAGCA CTGGCAAAAA GCCAATTTAG GCCAATGCT GCCAATGCT	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA AAGGAGGAGA TATTTCAAAA ATTTCCAAAA ATTTCCATGA AAAACAAACA TCATCTATAT GCCAATTACG ATCTCCAACAT TAACAATACA GTCTTGCCTT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA GGCTGGCTTC AAAACAATAA AGAACTATAT ATTGCACCAT ACTTTTCTCT AGCACAATTG GAAACAGCAG CAGCATTACT CGATGTCACA
4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101 5201 5401 5501 5601 5701	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT TATATACCAT GTACAGCAAC GAGCTGTCAT TTGACAGCTC TTTTTATCAA	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTAA CTTCCATTTG ATAATTGAGT CTGCACATTT CAGGAAAACA GAATGTTATTG GAATGTTACTT TTGCTTCCCA ATAGTAATCG TGCTTTATAG TGCTTTCTAAT ATGTTCCATC CGCCGATCAC AACGGAATGA AACGGAATGA	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCCAGTA AATTATTACAT AAAATGGCGT TCCAAATTATA AGACAATCAA TTGCATCAAC	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTCA AAATATACATAT AAAAAACATA CTCCTCTTCTA AGCTTAAAAA TCCGGAGGGCT GGCAAGAGCCT CGAGTTCCAG ATCATCCAGGA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGTAC ACTTTGGCTT GAATGTCAAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA TGTATGAAGT GACGTTTTTT AAACCATAAA CGGCTCCAAG TCACAATGCA ACTCAAATGT CGGATTTAAT CACCCTGTTG	ACATAATAT ATATTAAAGT CACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCCAGA AATGAACTC CAACCCCAC CAATTCGGGG CCACTTACTC AATCTGTGAC	TTTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTGA ACTTTCCTGA AAAAATATC AACTACATAC GGCAGAACGT CACAAATGGAT CACAATGGAT	ТААААТААGТ ТТССАGGAAA САААТТАТАG ААСТСТАСТТ ТАССТGСGCG СGGTATCTG АТААТАТААА ААААССТСАА АТТТАТТТGА АТССААААА АССССАААА СССАСАТТТТ ССАСССАСАСА СТGGAGCATT ССАССАСТТТАТ ССАССАСТТТАТ ССАССАСТТТАТ	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA ATGGATTTT AAGCAGGAGA TATTTCAAAA ATTTCCATGA AAAACAAACA TCATCTATAT GCCAATTACG ATCTCAACAT TAACAATACA GTCTTGCCTT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCATCA CGATTGCTTC AATAAATCAA CGCTTGCTTC AAAACAATAA AGAACTATAT ATTGCAGATA ACTTTTCTCT GAAACAGCAG CAGCATTACT CGATGTCACA
4001 4101 4201 4301 4401 4501 4701 4801 4701 5001 5101 5201 5301 5401 5501 5701 5701	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATACCCC GCATACATAT TATATACCAT TATATACCAT GAACGGGGTT TTGACAGCTC TTTTATCACA CCGGGAAATA	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT GAATGTAATG GAATGTTATT GAATTGTCCT TTGCTTCCCAT CGCGATGAC ACGGGATGAC AACGGGATGAC CAGAATACCC CGCGCATGAC	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAAATTAA AGACAATCAA TTCCTTCGCCA TTCCATCAAC ATCAAAGTGT	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATTCTA ATATACATAT CTACTTCTTA TCGGAGGGGGT AGCTTAAAAA TACCCAGAAG GCAAGAGCCT CGAGTTCCCG ATCATCCACT CAGTACAGGA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACG GACGTTTTTT AAACCATAAA CGCCTCCAAA CCACAATCCA CGCATTTAAT CACCATTAAT	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCAGA CGTGGACACG AAATGAACTC CCACCCCCCC CAATTCGGCG CCACTTACTC ATCCTGTGAC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAAATC TTTTAACATT CAAGGAACAA ATACCATGTA ACTTCCTGA AACAACATGTA ACTTTCCTGA AAAAATATCC GGCAGAACGT ACAAAATGGC GGCATTCAAT CACAAATCAA ATGATAGGGG CACCAGGGCT CAAAATGGCA	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGTATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA ATCCAAAAAC GGGGGGAGAG GCCCAAAAA CCACATTTAT GCAGCCAGCA CTGGAGCATT GCGAATGTTG GCCACAATCCT CAACAGTCTA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCAATAA TCATCTATAT GCCAATTACA ATCTCAACAT TAACAATACA GTCTTGCCTT CGAATAACA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTCC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCCT ATTGCAGATA ACTTTTCTCT GAAACAATTG GAAACAATTG GAAACAATTG GAAACAGCAG CAGCATTACA CTCACTTACA
4001 4101 4201 4301 4501 4501 4701 4801 4701 4801 5101 5201 5201 5301 5501 5501 5501	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTTAC GTTCCGTTAT TCATTACAGAATATG TCATTCTCGA AAATATCCCC GCATACATAT TATATAACATA AAAGGGGGTT TTATCTTTAC GTACAGCAAC GAGCTGTCAT TTGACAGCAC CTGACATCTCAC	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTGA ATAATTGAGT CTGCACTTTT CAGGAAAACA GAATGTTATT GAATTGTGCT TTGCTTTATAG CAAATGATAACG CAAATGATAAA AGTTTCCAATC CGGCGATGAC AACGGGAATG CAGAATACCC CTCCACGAT	ATATGATTTA TGGTAATCCA AAAACTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCCAAATTATA AGACAATCAA TCCTTCGCCA TTGCATCAAC ATCAAAGTGT GGATGTGACT	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AATATACATAT AAAAACATA TCGCAGGGG AGCTTAAAAA TACCCAGAAG GCAAGAGCCT CGAGTCCCG ATCATCCACT CAGTACAGGA CCACCAAGA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA GTATATGAAAT GACGTTTTT AAACCATAAA CGGCTCCAAG TCACAAATGT CGGATTTAAT CACCCAAACT CACCCAAACT	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTY AGGACCAAAC CCACCCAAG CGTGGACACG AAATGAACTC CCACTCCCAC CAATTCGGC CCACTTACTC ATCTGTGAC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAAATC TTTTAACATT AGTTCGTGAT ATTCCATGTA ACTTTCCTGA AAAAATATC AACTACATAC GGCAGAACGT GGCATTCAAT CACAAATCGC GGCATTCAAT CACAAATGGC CACCAGGGCT CACAAGGGT ATTTCCACTT	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA ATCCAAAAAAC AGGGGAGAGG GTGCCAAAAA GCACATTTAT GCAGCCAGCA CTGGAGCATTGG GCCAAAATGCT CAACAGTCTA CGAGTGACTA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCCAAA ATTTCCATGA ATCATCTATAT GCCAATTACG ATCTCCAACAT TAACAATACA GTCTTGCCTT CGAATCAAGAT AGATCAAGAT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA GGCTGGCTTC AAAACAATAA AGAACTATAT ATTGCAGCAT ATTGCAGATA ACTTTTCTCT AGCACAATTG GAAACAGCAG CAGCATTACA CGATGTCACA GTGATGTCACA
4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101 5201 5201 5501 5501 5601 5701 5801	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACATT TATATACCAT TATATACCAT GTACAGCAAC GAGCTGTCAT TTGACAGCTC TTTTTATCAA CCGAGAAATA CCAAGAATA	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGCAAAACA GAATGTTATT GAATTGTCCT TTGCTTCCCA ATAGTAATCG TGCTTTATAG CAAATGATCA AGTTCCAAC CGGCGATGAC AACGGAATG CCACGAAT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCCAGTA ATTATTACAT AAAATGGCGT TCAAATTATA AGACAATCAA TTGCATCAAC ATCAAAGTGT GGATGTGACT AAATACTTAG	GGTAGCTATA CTATAGAAGC ATTTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTCA AAATATACATAT AAAAAACATA CTCCTCTTCTA AGCTTAAAAA TCCGGAGGGCT GGCAAGAGCCT CGAGTTCCAG ATCATCCAGGA CCACCCAAGT AAAGCGAGGA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAA TGTATGAAGT GACGTTTTTT AAACCATAAA CGGCTCCAAG TCACAATCCA ACTCAAATGT CGGATTTAAT CACCCTGTTG CACCCAAACT TGACGAACT	ACATAATATT ATATTAAAGT CACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCCAGA AATGAACTC CAACCCCAC CAATTCGGCG CCACTTACTC AATCTGTGAC ATCCCGCTTGC	TTTTTTAAG TGAGCTATTT AGTACTCAC TTTTAAATC TTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTGA ACTTTCCTGA AAAAATATC AACATACATAC GGCAGAACGT ACAAAATGGC GGCATTCAAT CACAAATGGAT ATTCCACTT CAAAATGGAT	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAAA AAAACCTCAA ATTTTATTGA ATCCAAAAAC AGGGGAGAGG GTGCCAAAAA CTGGAGCATTTTG CCACCATCAA CTGGAGCATTT CCAACAGTCTA CGAGTCAGTCT TTCACATGGA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA ATGGATTTT AAGCAGGAGA TATTTCAAAA ATTTCCATGA AAAACAAACA TCATCTATAT GCCAATTACG ATCTCAACAT TAACAATAACA GTCTTGCCTT CGAATAACAA AGATCAAGAT AAGATGAAGA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCATCA CGATACATCA CGCTTGCTTC AAAACAATAA AGAACTATAT ATTGCAGATA ACTTTTCCAC GAAACAGCAG GAAACAGCAG CAGCATTACT CGATGTCACA GTGATGTCAT ATTACAAATG
4001 4101 4201 4301 4401 4501 4701 4801 4701 5001 5201 5301 5301 5501 5701 5801 5801 5801	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATACCAT AAAGGGGGTT TTATATACCAT AAAGGGGGTT TTATATCATTACA GAGCTGTCAT TTGACAGCTC TTTTATCAA CCGAGAAATA CCGAGAAATA	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT GAATTGAGT GAATGTTATT GAATTGTGCT TTGCTTTATAG CAAATGATAATG CAAATGATAA AGTTTCTAAT ATGTTCCAAT CGCCGATGAC AACGGGAATG CAGAATACCC CTCCACCGAT TGAGGACATG GGATTTCTGAG	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAATTATA AGACAATCAA TTGCATCAAC ATCATCGCCA TTGCATCAAC ATGATGTGACT AAATACTTAG	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATTCTA ATATACATAT CTACTTCTTA TCGGAGGGGT AGCTTAAAAA TACCCAGAAG GCAAGAGCCT CGGATTCCCC ATCATCCACT CGGACTACAGA CCACCAAGA CCACCAAGA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACGA GGACGTTTTTT AAACCATAAA CGGCTCCAAG TCAACAATCA CGGATTTAAT CACCAAATGT CACCCAAATGT CACCCAAATGT TGACGAAAGT TGACGAAAGT	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACAAAC CCACCCAGA CGTGGACACG AAATGAACTC ATGCAGAACT CCACCCCAC CCACTTACTC ATCTGTGAC ATCCTGTGCA ATCCGCTTGC CACGTTGCAC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAAATC TTTTAACATT CAAGGAACAA ATACCATGTA ACTTCCTGA AAAAAATATC CGCAGAACGT ACAAAATGGC GGGATTCAAT ATGATAGGGG CACCAGGGCT CAAAATGGAT ATTTCCACTT CAACTTATAA	TAAAATAAGT TTCGAGGAAA AACTCTACGT TACCTGCGCG CGTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACTAAAAAA GCGCGAGGAGAG GTGCCAAAAA GCACATTTAT GCACCCACCA CTGGAGGCATT GCGAATGCTTGG GCCAAATGCT CACAGTCTA GCGCCTTGG GCCAAATGCT TCACATGAGTT	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCAAAG ATTTCCATGA ATATCCATGA ATCTCAATACA GCCAATTACA GCCATTACA GTCTTGCCTT CGAATAACAA AGATCAAGAT AAGATGAAGA TTCAATGACA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTCC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCTT AATGCAGATA ACTTTTCTCT GGAACAATTG GAAACAGCAG CAGCATTACA GTGATGTCACA ATTACAAATG CAAATGCCCT
4001 4101 4201 4301 4401 4501 4701 4801 5001 5101 5301 5401 5501 5501 5501 5501 5501 5601 5901 6001 6101	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATGCCATAA TGATTTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT AAAGGGGGTT TTATCTTTAC GTACAGCAAC GAGCTGTCAT TTGACAGCAC CCAGACATTG TTATCTTAC	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTGA ATAATTGAGT CTGCACATTT CAGGAAAACA GAATGTTATT GAATTGTGCT TTGCTTTATAG CAAATGATAATG CAAATGATAAA AGTTTCCAATC CGGCGATGAC AACGGGAATG CAGAATACCC CTCCACGAAT TGAGGACATG CAGACATCTTGG AATTAAAGCA	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTG AAAATGCGCT TCAAATTATA AGACATCAA TCCTTCGCCA TTGCATCAAC ATCAAAGTGT GGATGTGACT AAAATACTTAG CTATAACAAA CCACCTTGAA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCTA ATATACATAT AAAAACATA TCGTCTTCTTA AGCTTAAAAA TACCCAGAAG GCAAGAGCCT CGAGTTCCAC ATCATCCACT AAAGCGAGA AGTGCCTTT TATCACATTA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACGA GCATATGTAT GACGTTTTT AAACCATAAA CGGCTCCAAG TCACAATCCA ACTCAAATGT CGGATTTAAT CACCCAAACT TGACGAACATG GGAAGCACAA	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCAGA CGTGGACACG AAATGAACTC CAACTCGCGC CCACTTACTC ATCCACTAATG ATCCGCTTGC CACGTTGTCA GAATTTAAAG	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTA ACTTTCCTGA AAAAATATC AACTACATAC CGCAGAACGT GGGATTCAAT CACAAATCGA ATGATAGGG CACCAGGGCT CAAAATGGAT ATTTCCACTT CAATCTATAA TATGAAGCCT CCATTCCAAT	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA ATCCAAAAAC AGGGGGAGAG GTGCCAAAAA GCACATTAT GCGACCAGCA CTGGAGCATT CAACAGTCT CAACAGTCT CAACAGTCT TTCACAGGG GCAAAAACAC GCGACAAGTG	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCCAAAA ATTTCCATGA AAAACAAACA TCATCTATAT GCCAATTACG ATCTCCAACAT TAACAATACA GTCTTGCCTT CGAATCAAGAT AAGATCAAGAT AAGATGAAGA TTCAATGTCC TAACTACAGC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCT ATTGCAGATA ACTTTTCTCT AGCACAATTG GAAACAGCAG CAGCATTACT CGATGTCACA GTGATGTCACA TGTGTTAACA
4001 4101 4201 4301 4401 4501 4601 4701 4801 4701 5101 5201 5301 5501 5501 5601 5601 5601 6001 6101 6201	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT TATATACCAT TTATCTTAC GTACAGCAAC GAGCTGTCAT TTGACAGCTC TTTTATCAA CCGAGAAATA CTAATTCTAG CAAGACTTGT TTTGTCACGGT	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGCAAATCA GAATGTTATT GAATTGTCTT TTGCTTCCCA ATAGTAATCG TGCTTTATAG CAAATGATCAA AGTTTCTAAT AGGTTCCAAT CGGCGATGAC AACGGAATG GAATTCCCAC	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA AATAATGCGT TCAAATTACAA TCCATCAGCA TTCAAAGTCAA TCCTTCGCCA ATCAAAGTGT GGATGTGGACT AAATACTTAG CTATAACAAA CCGCATGAAC	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATTCTA ATATACATAT CTCCTTAT TCGGAGGGGGT AGCTTAAAAA TACCCAGAG GCAAGAGCCT CGAGTTCCCG ATCATCCAGT CAGTACAGGA GCACGAGGAGC	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAG GACGTTTTTT AAACCATAAA CGGCTCCAAG TCACAATCCA ACTCAAATGT CGCACTCAAATGT CACCCAAACT TGACGAAAGT TGACGAAAGT TGACGAACATG GGAACCACAA GTATCAGTAT	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTCA CCACCCCAGA CCGTGGACACG AAGGAGAACTC CCACCCCCCC CAATTCGGCG CCACTTACTC ATCCGCTTGC ATCCGCTTGC ACCACTAATG ACCGCTTGTCCTG CACATTCTG CCACTCTCTCG	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT CAAGGAACAA ATACCATGTA ACTATCCTGA ACTATCCTGA AACTACATAC CGCAGAACGT ACAAAATGGC GGGATTCAAT CACAAATGGG CACCAGGGCT CAAAATGGAT CAAAATGAGT CAATCTATAA TATGAAGCCT CCATCCAATT ACTGTGATTA	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGCTATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA ATACAAAAAC GCACCAAAAA GCACCAAAAA GCACCAATTTAT GCAGCCAACA GCGACTGATGT GCGAATGCTTG GCGAATGACTT TTCACATGGA GAGAGAATT TTCACATGGA GCGACAAAACC CGGACAAAACC	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAAAGGACAA AAAAGGTACA ATGGATTTTT AAGCAGGAGA ATTTCCATGA AATTCCATGA AATTCCATGA AATTCCATGA CAATTACA GCCAATTACA GCCAATTACA GCCAATTACA GTCTTGCCTT CGAATAACA AGATCAAGAT AAGATGAAGA TACAATGTC TACAGCATT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATCA CGCTTGCTTC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACACTT AGCACAATTG GAAACAGCAG CAGCATTACT CGAGTGACAA CTCACTTACA GTGATGTCACA CTGTGTTAAATT
4001 4101 4201 4301 4401 4501 4501 5501 5501 5501 5501 55	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACATA AAAGGGGGTT TTATATACAT AAAGGGGGTT TTGACAGCTC TTTTATCATA CGAGACTACT TTGACAGCTC TTTGTCACG AGTCTATGT TTGTCACGG AGTCTATGTT	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTAA CTTCCATTTG ATAATTGAGT CTGCACATTT GAATGTATT GAATGTTATT GAATGTTATT GAATGATAACA ATAGTAATGATAA AGTTTCTAAT ATGTTCCATC CGCCGATGAC AACGGGAATG CAGAATACCC CTCCACCGAT GGATTTCTGG GAATTCCCAC GAACTCCCAC	ATATGATTTA TGGTAATCCA AAAACTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAATTATAA AGACAATCAA TTGCATCAAC ATGCATCAAC ATGCATCAAC ATGCATCAAC ACATTAGC CTATAACAAA CGCACGTGAAC	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCTA ATATACATAT CTACTTCTTA TCGGAGGGGT AGCTTAAAAA TACCCAGAAG GCAAGAGCCT CGGATTCCCG ATCATCCACT CAGGTCCAGA AACCGAGGA AGTGGCCTTT TATCACATTA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACGA GACGTTTTTT AAACCATAAA CGCCTCCAAG TCACAATCG ACTCAAATGT CGGATTTAAT CACCCAAATGT CGGATCTATG CACCCAAAGTG TGGGAACATG GACAACAGA GTATCAGTAT GATCAAGATG	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCAGA CCACCCAGA CCACCCAGA CCACTCAGCGC CCACTTACTC ATCCGTGGC ATCCGTGTCA GAGTTGTCA GAGTTGTCA GGCACCCAA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAAATC TTTTAAATC TTTTAACATT CAAGGAACAA ATACCATGTA ACTTCCTGA AACTACCATGTA ACTATCCTGA ACAAAATGGC GGGATTCAAT CACAAATGGGG CACCAGGGCT CAATACGAG CACTACAAT ATGAAGCCT CCATCCATT ACATCCATT ACATCTGTATA	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACTAAAAAAAAAA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAG TATTTCAAAA ATTTCCATGA ATATCCATGA ATATCCATGA ATATCCATGA ATATCAATACA GCCATTACA GCCATTACA GCCTTGCCTT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTCC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCT AATGCAGATA ACTTTTCTCT GGAAGCAGT GGAAGCAGC CAGCATTACA GTGATGTCACA GTGATGTCACA GTGATGTCACA GTGATGTCACA CTTTTAAAT ACACGTGGG
4001 4101 4201 4301 4401 4501 4501 4501 5501 5501 5501 55	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATGGCATAA TGATTTTTAC GTTCCGTTAT TCATTCTCGA AAATATCCCC GCATACATAT TATATAACATA AAAGGGGGTT TTATCTTAC GTACGCAAC GAGCTGTCAT TTGACAGCAC CCAGAAATG ACCTTTTATCAG ACCTTTTATCAG	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTGA ATAATTGAGT CTCCACTTTT CAGGAAAACA GAATGTTATT GAATTGTGCT TTGCTTTATTAG CAAATGATAATGG CAAATGATAAA AGTTTCCAATC CGGCGATGAC AACGGGAATG CAGAATACCC CTCCACGAAT GAGACACCACA GAATTCCCAC AATTAAAGCA GAATTCCCAC AACTACCACA	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA AAAATGGCGT TCAAATTATA AGACAATCAA CCTTCGCCA TTGCATCAAC ATCAAAGTCT GAATGTCACA CTATAACAAA CCACCTTGA CGACAGTCAC ATAAGCCCGG GAAAGAATCA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AATATATCTA ATATACATAT AAAAACATA TCGTGAGGGGG GCAAGAGCGT CGAGTCCAGA ATCATCAGAG CACCCAAGA CCACCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CACCCAAGT AAAGCGAGA AGTGGCCTTT TATCACATTA ACTCGGCCGT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACGA GCATATGTAT AAACGAAAAAA CGGCTCCAAG TCACAATCCA ACTCAAATGT CGGATTTAAT CACCCAAACT TGACGAACATG GGAAGCACAA GTATCAGAAG	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCAGA CGTGGACAG AAATGAACTC CAACTCGGC CCACTTACTC ATCCGCTGC ATCCGCTGC ATCCGCTGC CAGTTGTCAC GTGCACCCAA ACAATTGAGT	TTTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTTCGTGAT ATTCCATGTAA ATACCATGTAA ACTTTCCTGA AAAAATATC AACTACATAC CGCAGAACGT GGGATTCAAT CACAAATCGC GGGATTCAAT CACAAATGGC CACCAGGGCT CAAAATGGAT ATTCCACTT CAATCTATAA TATGAAGCCT CCATTCCATT	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA ATACCAAAAAC AGGGGAGAGG GTGCCAAAAA GCACATTAT GCGACGACAGTG GCCAAATGCT CAACAGTCT CAACAGTCT CAACAGTCT CAACAGTCT CGACTGAGGAGTT TTCACAGGG GCAAAAACAC GCGACAAAGTG GCCACAAAA GTAATTGATG	CAGACAAACA AGCTGTGCTT CCATAAATA AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCAAAA ATTTCCATAA ATTTCCATAA ATTTCCATAA AAAACAATACA GCCATTACG ATCTCCAACAT TAACAATACA GTCTTGCCTT CGAATAACAA AGATCAAGAT AAGATGACAA TTCAATGTTC TAACTACAGC TACTGCCATT CTAACTACAGC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTGCAGATA ATTGCAGATA GGACAATTG CAGCATTACT GGATGTCACA GTGATGTCACA GTGATGTCACA GTGATGTCACA CTTTTAAATT ACGACGTGGG CACAACACC
4001 4101 4201 4301 4401 4501 4501 4701 4701 4801 5501 5501 5501 5501 5501 5501 5501 5	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT TATATACCAT TTATCTTAC GTACAGCAAC GAGCTGTCAT TTGACAGCTC TTTTATCAA CCGAGAAATA CCGAGAAATGA CCAAAAATGA	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT GAATGTATTG GAATGTTATT GAATGTCAT TTGCTTCCAA ATAGTAATCG CAAATGATCAA AGTTTCTAAT CGGCGATGAC AACGGCATG CAGAATACCC CTCCACCGCAT GGACTTCCCAC AATTAAAGCA AATTAAAGCA AATTACCACC AAGTACCACC	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGACAT TCCAAACTTATA AGACAATCAA TCCTTCGCCA ATCAAAGTGT GGATGTGGACT AAATACTTAG CTATAACAAA CCACCTGAA AGACAACACA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATATGTTAT CGAAATTTGA AAATATCTA ATATACATAT CTCTTCTTA TCCGAGGGGGT AGCTTAAAAA TACCCAGAAG GCAAGAGCCT CGAGTTCCCG ATCATCCAGAGA AGTGGCCTTT TATCACATTA ACTCGTCGGGT CATGGTCTTC	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACGA GACGTTTTTT AAACCATAAA CGCTCCAAG ACTCAAATGT CGACAATCCA ACTCAAATGT CGCAACCAAATGT TGACGAAAGT TGACGAAAGT TGACGAACATG GGAAGCACAA GTATCAGTAT GATGAAGATGC CAACCAGTAT	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCCAGA CGTGGACACG AAATGAACTC ATGCAGCACCCAC ATCCGCTGC CAATTCGGCG CCACTTACTC CACGTTGTCC GCACTTACTG GTGCACCCAA ACAATTCAGG GTGCACCCAA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT CAAGGAACAT ATACCATGTA ACTTCCTGA ACTATCCTGA ACTATCCTGA ACTACATAC GGCAGACGT ACCAACATAC CGCAGAACGT CACAAATGGG CACCAGGGCT CAAAATGGAT CAAAATGGAT CAAAATGGAT CAATCTATAA TATGAAGCCT CCATCCTGTT GCAACCTGCT GCAACCCCCC	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACTAAAAA ACCCAAAAAG GTGCCAAAAA GCACATTTAT GCAGCCACAA GCGACTGCAGCATT TTCACATGGA GCGACAGTGT GCGACACAGTG GCGACAAAGG GCGCACAAAA GTAATTGATG AAAGACTTGC GCGCCACAAA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AACCAGGAGA TATTTCCATGA AATTCCATGA AAAACAAACA TCATCTAACA GCCAATTACG GCCAATTACG GCCAATTACA GCCAATTACA GTCTGCCAT TAACAATACAAGAT AAGATGAAGA TTCAATGTCT TACTGCCATT TCTATGGAAC CCTTAATCGAC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATCA GGCTGCTTC AATAAATCAA GGCTGCTTC AAAACAATAA AGAACTATAT ATTCCACACCT AATGCAGAATAG CAACACACCA CTGCACTACA CTGCACTACA CTGCACTACA CTGTTAAATT AGGACGTGGG CACAACAATC
4001 4101 4201 4301 4401 4501 4501 5001 5101 5301 5501 5501 5501 5501 6001 6201 6301 6401 6401 6501	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACATA AAAGGGGGTT TTATATCTTAC GTACAGCAAC GAGCTGTCAT TTGACAGCAC CCGAGAATGA CCGAGAATGG AGTCTATGT TTGTCACGG AGTCTATGT TTGTCACGG AGTCTATGT ACATTTAAGA CCAAAAATGA AACACAACAT	TCTTTTTTTA ATACAAAATG CTTCCATTTAA CTTCCATTTAA CTTCCATTTGA ATAATTGAGT CTGCACATTT GAATGTTATT GAATGTTATTG GAATGTTACTG TTGCTTTATAG CAAATGATAA AGTTTCCAAC CGGCGATTCCAA AACGGGAATG CAGAATACCC CTCACACGAT GGATTTCTGG AATTAAAGCA GAATTCCGACC AGTCCGGCACT	ATATGATTTA TGGTAATCCA AAAACTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCCAAATTATA AGACAATCAA TGCATCAAC ATGCATCAAC ATGCATCAAC ATGCATCAAC ACAACTTGA CAAAGTGTG CTATAACAAA CGCACTTGAA CGCACTGAA ATAACCTGA ATAACCCGG AAAGAATCA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCATAT AAATAACATAT CTACTTCTTA AGCTTAAAA TCCGAGGGGGT AGCTTAAAAA TCCCAGAGA GCAAGAGCCT CGAGTCCCG ATCATCCACT CGAGTCCCG ATCATCCACT CAGGCCTTT TATCACAGGA AGTGGCCTTT ACTCGGCGGT CATGGTCTTA ACTGGGGCGT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC AACTATGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACGA GTATATGAAGT GACGTTTTTT AAACCATAAA CGCCCCAAGT CGGATTTTAT CACCAAATCG CACCCAAATGT TGGGAACATG GGAAGCACAA GTATCAGTAT GATGAAGATG CAAAGATGCC GCAACGCC	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC TCAC TTAGTACTTA AGGACCAAAC CCACCCCAGA CCTGGACACG AAATGAACTC ATCCACCACA CCACTTACTC CCACTTACTC ATCCGTGGC CACTTACTC ATCCGCTGC CACGTTGTCA GAATTTAAAG CGTGGTCCAA CGTCGTCCAA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAAATC TTTTAAATC TTTTAACATT CAAGGAACAA ATACCATGTA ACTTCCTGA AACTACATGA ACTACATAC CGCAGAACGT ACAAAATGGC GGATTCAA ATGATAGGGG CACCAGGGCT CAATACGAA TCTTCCATT CAATCTATAA TATGAAGCCT CCATCCATTA TCCATCGATTA TCCATCGTGTTG CCAATCCCA	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACTACTTA TTTATTTGA ACGCGAGAGAG GTGCCAAAAA CTGGAGCATTAT GCACCCACAA CTGGAGCATT CAACAGCAGCA CGCAACAGTCT CAACAGCAGCA GCAAAACAC GCGACAATGC GCACACTTGATG GAAAAACAC GCGCACAAGT CGCCACAAAA CTGGAGATTGC CGCCACAAAA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCAAAA ATTTCCAAGA TCATCTATAT GCCAATTACA GTCTTGCCTT CGAATAACAA AGATCAAGAA TTCAATGTCT TAACAAGAA AGATCAAGAT TCCATGTCT TAACAAGACA AGATGACAAT TCCATGCCATT TCTATGGAAC CCTTAATCAG CAAGCTAACC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTCC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCT ATTGCAGATA ACTTTTCTCT GGAAGCAGATT CGACGATTACA GTGATGTCACA GTGATGTCACA GTGATGTCACA GTGATGTCACA GTGATGTCACA GTGATGTCACA GTGATGTCACA GTGATGTCACA CTTTTAAATG CAAATGCCT GGACGTGGG CACAACAATC ACCACGTCGG CACAACAATC
4001 4101 4201 4301 4401 4401 4501 4501 5001 5101 5201 5301 5501 5501 5501 5501 5601 6001 6201 6301 6401 6501	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACACAT TTGACAGCAC GAGCTGTCAT TTGACAGCAC CAGAGAATA CTAATTCTAG CAGAGACTGT ACATTTAAGA CCAAAAATG AACACAACAT ACAGCAGTCA	TCTTTTTTA ATACADAATG TTTCATTTAA CTTCCATTTGA ATAATTGAGT CTGCACATTTT CAGGAAAACA GAATGTTATT GAATTGTGCT TTGCTTTATTAG CAAATGATAATG CAAATGATAATG CAAATGATAATG CGGCGATGAC AACGGGAATG CAGGAATG CAGGAATG CAGGAATG CAGGAATG CAGGAATG CAGGACATG CAGCACCACA AATTAAAGCA GAACTCCCAC AAGTACGACAG GCCTGCTTCT ACTTCCACTC	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA ATTATTACAT AGACATCAA TCCAATCAAA TCCAATCAAA CTTCGCCA TTGCAATCAAA TCCACTCGCCA TTGCAATCAAA CCACCTTGAA CGGAAGTCAC ATAAGCCCG GAAAGAATCA ACCACACAAA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTCA AAATATCTA AAATACATAT AAAAAACATA CTTCTTCTTA TCGGAGGGGT CGAGTTCCAGAG GCAAGAGCCT CAGTACAGGA CCACCCAGAG ACATCCACTT AAAGCGAGGA CACCCAGTT TATCACATTA ACTCGCCGT CAGTGCTTTA ACTCGGCGT CATGGTCTC AACGGCCTT CATCGCCCAGA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAACGA GTATATGTAT AAACGAAAAA GTATATGTAT GACGTTTTT AAACCATAAA CGGCTCCAAG TCACAATCA ACTCAAATGT CGGATTTAAT CACCCAAACT TGGCAACATG GGAAGCACAA GTATCAGATAT GAGGACATA GTATCAGATAT GAAGAACATG CACACATCAC	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGG ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCCAGA CAGTGGACAG AAATGAACTC CAACTCGCTGC CCACTTACTC ATCCGCTGC ATCCGCTGC ATCCGCTGC CACGTGTCCG GTGCACCCAG CAATTTAAAG CGTTGTTCTG GGCACCTAA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTTCGTGAT ATACCATGTA ATACCATGTA ACTTTCCTGA AAAAATATC AACTTCCTGA AAAAATATC GGGATTCATA CGCAGAACGT CACAAATGGC CACAAGGGCT CAAAATGGAT ATTTCCACTT CAAATCGAT ATTACAAGCCT CCATCTCAT ACTGTGATTA ACCATCTGTT GCAGCCTTGC CAAAATCCAC	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA ATCCAAAAAC AGGGGGAGAG GTGCCAAAAA GCACATTAT GCGGCCAGCA TGCCAAGTCT CAACAGTCT CAACAGTCT CGACAGTCT GCGAATAGT GCGACAATGC GCCACAAA GTAATTGATG GCCACAATGC AAGGATTTGC GTCGAATATT CTCAATCCTCA	CAGACAAACA AGCTGTGCTT CCATAAATA AAGAAACCAA AAAACGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCAAAA ATTGCATGA ATTTCCATGA AAAACAATACA GCCATTACG ATCTCCAACAT TAACAATACA GTCTTGCCTT CGAATTACA GTCTTGCCTT CGAATCACA AGATCAAGAT AAGATGAAGA TTCAATGTCC TACTGCCATT CTAACGGACT CCTTAATCGAC CCTTAATCAG	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCATCA GGATGCTTC AATAAATCAA GGCTGCTTC AAAACAATAA AGAACTATAT ATTGCACACTT ATTGCAGATA CAGCATTACT CGATGTCACA CTGATGTCACA CTGTTTAAATT ATTACAATG CCAACTGCGG CAACAATC ACCAGCATTT GCACGTCGGG CAACAATC ACCAGCATTT GTGTCACGCG
4001 4101 4201 4301 4301 4501 4501 4501 5501 5501 5501 5501 55	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATACCAT TATATACCAT TATATACCAT TATATACCAT GAACGGGGTT TTGACAGCAAC GAGCTGTCAT TTGACAGCTC TTTTATCAA CCGAGAAATAG CAAGACTTGT ACACTATCTAG AACAACAACAT ACACCAACAT ACACCAACA	TCTTTTTTTA ATACAAAATG TTTCATTTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT GAATGTATTG GAATGTTATT GAATGTTATT GAATGTTATTG GAATGTCAT CGCGATGAC ATGTTCCATC CGGCGATGAC AACGGGATGAC AACGGCATG GAATTCCCATC GGACTTCCCAC GGACTCCCAC AAGTACGACA GGACTCCCACC AGTCCGCCAG AGTCCGCCAG AGTCCCGCCAG AGTCCCGCCAG AGTCCCGCCAG AGTCCCGCCAG	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAAATTATA AGACAATCAA TTCCTTCGCCA TTGCATCACC ATCAAAGTGT GAATGTGGCT AAATACTTAG CTATAACAAA CGGAAGTCAC ATAAGCCCGG GAAAGCACA CAGCCATTT	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATATGTTAT CGAAATTTGA AAATATTCTA ATATACATAT CTCTTCTTA TCGGAGGGGT AGCTTAAAAA TACCCAGAAG GCAAGAGCCT CGAGTCCCG ATCATCCACTA AAAGCGAGGA AGTGGCCTTT TATCACATTA ACTCGTGGGTTTA ACTCGTCGTT GTACTGCTCC	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACGA GTATATGAAGT GACGTTTTTT AAACCATAAA CGGCTCCAAG TCACAATCCA CACCCAAATGT CGGAATTAAT CACCCAAATGT CACCACAACCA TGACGAAAGT TGACGAAAGT TGACGAACATG GGAACCATAA GATGAAGAATGC GTATCAGTAT GACGAACATCA GCACCTCAA	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCAGA CGTGGACACG AAATGAACTC ATCCCCAG CCACTTACTC CACCCCAC ATCCTGTGCC ATCCGCTTGC ATCCGCTTGC CACTTACTG GTGCACCCAA GAATTTAAG GTGCACCCAA ACAATTGAGT GAGATAGTTC GAGATAGTTC GAGATAGTC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT CAAGGAACAA ATACCATGTAA ATACCATGTA ACTTTCCTGA AACTACATGTA ACTATCCTGA ACTACATAC CGCAGAACGT ACAAAATGGC GGGATTCAAT CACAAATGGG CACCAGGGCT CAAAATGGAT ATTTCCACTT CCATCTATAA TATGAAGCCT GCAGCCTTGC CAATACCA ATGATACGAT CCAATCCAAT	TAAAATAAGT TTCGAGGAAA AACTCTACGT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACTAATAAA ACGGGGAGAG GTGCCAAAAA GCACATTTAT GCAGCCACAA CTGGAGCCATT GCGATGTTGG GCCACAATTCT TTCACATGGA GAGAGAGTT TTCACATGGA GCGACAAGTG GCGACAAAGTG GCGACAAATG GTAATTGATG AAGGATTTGC GTCGAACATTCC AACAACAAGC	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAAACGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AACCAGGAGA TATTCCATGA AATTCCATGA AATTCCATGA AATTCCATGA AATTCCATGA CCCATTACA GCCAATTACA GCCAATACAA AGATCAAGAT AAGATGAAGA TTCAATGTCC TACTGCCATT TCTATGGAAC CTATGGCATT TCTATGGAAC CCTTAATCAG CAGCTAACC TCCCTCATC TCCCTCATC TCCCTCATC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATCA GCATGCTTC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCCT ATTGCACACATTG GAAACAACAATG GAAACAACAATG CAGCATTACA CTGATGTCACA ATTACAAATG CAGATGTCACA ATTACAAATG CAAATGCCT TGTGTTAAATT ACGACGTGGG CACAACAATC ACCAGCATTACA
4001 4101 4201 4301 4401 4501 4501 5001 5101 5301 5501 5501 5501 5501 6001 6201 6301 6301 6401 6501 6601 6701 6701 6701	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATGCCATATG GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCA GCATACATAT TATATACATA AAAGGGGGTT TTATATCTTAC GTACAGCAAC GAGCTGTCAT TTGACAGCATCA CCGAGAATTGT ACGAGAATGA ACCAAAATGA ACCAAACAT ACGCCAACGC	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTAA CTTCCAGTTTG ATAATTGAGT CTGCACATTT GAATGTTATT GAATGTTATT GAATGTTATA GAATGATAACG CAAATGATAA AGTTTCTAAT ATGTTCCATC CGGCGATCAC AACGGGAATG CAGAATACCC CTCACACGAC GAATTCCGGC GAATTCCGGC AATTAAAGCA GAATTCCGCC AGTCCGCCA GGCCTCCCACGAT ACTCCACCAC	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCCAAATTATA AGACAATCAA TGCATCAAC ATGCATCAAC ATGCATCAAC ATGCATCAAC CTATAACAAA CGCACTGGA ATTAACATAA CGCACGG AAAGAATCA CCAGCCAAAA CATACCATTT GCATTACAAA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCTA AAATATCTA ATATACATAT CTCTTCTTA AGCTTAAAA TCCGAGGGGGT CGAGTTCCG ATCATCCACA CAGCCAAGA CCACCCAGA AGTGGCCTTT TATCACATA ACTGGGGGGT GTACTGGTGGT CATGGTCTC AATGGACAA TTTCCTTATT	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC AACTATGCAT GAATGTCAAT GAAGTAAGTG AAACAAACGA GTATATGTAT AAACAAAACGA GTATATGAAG GACGTTTTTT AAACCATAAA CGCCCCAAG TCACAATCG CACCAAATCG CACCCAAATG TGGGAACATG GGAAGCACAA GTATCAGTAT GATCAGATG CAAAGATGCC GCAACCATCAAC	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCCAGA CCACCCCAGA CCACCCCAGA CCACTTACTC CCACCCCAG CCACTTACTC ATCCGCGCG CCACTTACTC ATCCGCTGCC CACGTTGTCA GACATTGAGT GACAATAGT CACACAACA CACTCAATG GCCACCCAA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTTCGTGAT CAAGGAACAA ATACCATGTA ACTTCCTGA AACATACATAC CGCAGAACGT ACAAAATAGC GGACTTCATA ACGAGAGCGT CACAAATGGAT CACAAATGGAT CAACATGGATTA ATGAAGCCT CCATCCATTA TCCACTGTTA TCCACTGTTA CCATCGTGTTC CAAAATACCA GCAGCACTTC CAAAATACCA GCAGCACCTTA	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACCAAAAAC AGGGGAGAG GTGCCAAAAA CTGGAGCATTAT GCACCCACAA CTGGAGCATTC CAACAGCTCA CGCATGTTGG GCCAAAAGC CGCACAGTG GCCAAAAGC CGCACAGTG GCAAAACAC CGCACAAGG CGCCACAAG GAAAAACAC CGCACAAGG CGCCACAAG GTAATTGATG AAGGATTTGC TCCAAATTT TCCAAATCCA AACACCAAGC	CAGACAAACA AGCTGTGCTT CCATAAATA AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAG TTTTCCATGA ATTTCCATGA ATATCCATGA ATATCCATGA ACATCTATAT GCCAATTACG ATCTCAACACA GTCTTGCCTT CGAATAACAA AGATCAAGAT TCCATGTCT TAACAAGAGA TTCAATGTC TACTACAGA TTCAATGTC TACTACAGA TCCATGCCATT CTATGGAAC CCTTAATCAG CACGCTATCG GACGAAGAGC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCTTCC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCT AGCACAATTG GAAACAACAG CAGCACTTACA GTGATGTCACA GTGATGTCACA GTGATGTCACA CTTTTAAATG CACAACAGTGGG CACAACAATC ACCAGCAGTGG GACAACAATC ACCAGCATTAC ACCAGCATTAC ACCAGCATTAC ACCAGCATTAC
4001 4101 4201 4301 4301 4501 4501 5501 5101 5201 5501 5501 5501 5501 5	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACAAT TATATACAAT TTGACAGCAAC GAGCTGTCAT TTGACAGCAC CAAGACTTGT ACATTTAAG CAAGACTTGT ACATTTAAG ACCAACAT ACCAACAT ACACCACCA CTCTCTGCAAC CGCCAAGGCT	TCTTTTTTA ATACAAAATG ATACAAAATG TTTCATTTAA CTTCTGTTG ATAATTGAGT CTGCACTTTT GAATGTATTG GAATGTTATT GAATGTTATT GAATGTCAT TGCTTTCCA ATGTTATAG CAAATGATAA AGTTTCTAAT AGTTTCTAAT AGTTTCTAAT AATGTCCACTC CGGCGATGAC GGATTTCTG GGATTTCTG GGATTTCTG GGATTCCCAC AATTAAAGCA AGTCCCACC AGTCCGCAG AGTCCCACC AGTCCGCAG AGTCCCACC AGTCCCACC AGTCCCGCAG AGTCCCACC AGTCCCGCAG AGTCCCACC ACTCCACCC ACTCCACCC ACTCCACCC ACTCCACCC ACTCCACCC ACTCCACCC ACTCCACCC	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA ATTATTACAT AGACATCAA TCCAATCAAC ACAATCAAC ACAACACAAC	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCTA ATATACATAT AAAAAACATA CTTCTTCTTA TCGGAGGGGT CGAGTTCCAGAG GCAAGAGCCTT CAGTACAGGA CACCCAAGA CCACCCAGAG ACATCCACT AAGCGAGGA CACCCAGGT CAGGTCCGG ACATCGCCTT TATCACATTA ACTCGGCGTT CATCGCCTC CATCGACAA CTCGCCCAG CACCGCCGT CATCGCCTC CATCGACAA CTCGCCTC CATCGACAA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAACGA GTATATGTAT AAACGAAAAA TGTATGAAGT GACGTTTTT AAACCACATAAA CGGCTCCAAG TCACCATAAA ACTCAAATGT GGGAACATGT GGGAACATGT GGGAACATAG GGAAGCACAA GTATCAGATAT GATGAAGATG CACCATCAT GAAGATGCC GTCATCGTCA CACCATCATC GGAACCTTCA TAAGCAAGAAC	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGG TTAGTACTTT AGGACCAAAC CCACCCAGA CAGTGGACAAG CAACTGGACACG CAACTTACGC CCACTTACTC ATCCGCTGC CCACTTACTC ATCCGCTGC CACGTGGTCCA GAATTGAAG CGTGGTCCTG GTGCACCCAA ACAATTGAGT CACGATAGTT AGCAACAACA TCTCCAAATG GACAAAAAG CACATAATG GACAAAAAG	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTTCGTGAT ATTCGTGATA ATACCATGTA ACTTTCCTGA AAAAATATC AACTTCCTGA AAAAATATC GGGATTCATA CGCAGAACGT CACAAATGGC CACAAGGGC CACAGGGCT CAAAATGGATA ATGATCCACT CCATCCAAT ACTGTGATTA ACCATCTGTT GCAGCCTTGC CAAATACCA GCAGCAGCTT TTGGCTGCCC CAAGCTTTTC	TAAAATAAGT TTCGAGGAAA CAAATTATAG AACTCTACTT TACCTGCGCG CGGTATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA ATACCAAAAAAC AGGGGGAGAG GTGCCAAAAA GCACATTTAT GCGACCAGCA CTGCAGGCATTG GCCAAATGCT CAACAGTCT CAACAGTCT CAACAGTCA GCACAATGCT CGACAATGCG CGCACAAAA GTAATTGATG AAGAATTTCCAACACAC AACAACAACA	CAGACAAACA AGCTGTGCTT CCATAAATA AAGAAACCAA AAAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGA TATTTCAAAA ATTGCATGA ATTTCCATGA AAAACAATACA GCCATTACG ATCTCAACAT TAACAATACA GTCTTGCCTT CGAATAACA AGACTGACAT CTAACTACAG TTCAATGTCC TACTGCCATT CTAACGACA CCTTAATGAC CCTTAATCAG CCTTAATCAG CCTCATCACAT CCCTCATCAG CCTCATCACAC TCCCCCCCCCC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAAATCATCA CCAAATCTTC AATAAATCAA GGCTTGCTTC AAAACAATAA AGGACTATAT ATTGCAGCAT ATTGCAGCAT GGAGACATACA CGCAGCATTACA GGAGACTACA CTCACTTACA GGAGGTCACA CTGTTTAAATT ATTACAAATG CCAACTGCC CAACACCT GGACGTCGGG CACAACAATC ACCAGCATTT GTCGTCAGCG GCCCAGTTAT ATGACAGCG
4001 4101 4201 4301 4301 4501 4501 4501 4701 4501 5501 5501 5501 5501 5501 5501 55	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATACCAT TATATACCAT TATATACCAT TATATACCAT GAACGGGGTT TTGACAGCAAC GAGCTGTCAT TTGTCACGG CAGACTTGT TTGTCACGG AGTCTATGTT ACAGCAACAT ACAGCAACAT ACAGCAACAT ACAGCAACAT TTGTGCACGGA ATTCTGCAACG	TCTTTTTTTA ATACAAAATG TTTCATTTTAA CTTCCTTTTG ATAATTGAGT CTGCACTTTT GAATGTTATT GAATGTTATT GAATGTTATT GAATGTTATTG GAATGTTCCAA ATGTTCCATC CGGCGATGAC AATGTCAATG CAGAATACCC CGGCGATGAC AATGTCCATC CTCCACCGCATG GAATTCCCAC AGTCCGGCAG GCTTCCCAC AGTCCGGCAG AGTCCGGCAG ACTCCCACTC AACAAGAAAA CCTGCGAACA ACAGCTATGG ACTCCGCATCT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAATTATA AGACAATCAA TCCTTCGCCA TTGCATCACA CAACTTGA CGAAGTCAC ATAAGCCCGG GAAAGAATCA CATACCATTT GCATTACAAA CATACCATTT GCATACAACA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATATGTTAT CGAAATTTGA AAATATTCTA ATATACATAT CTACTTCTTA TCGGAGGGGT AGCTTAAAAA TCCCAGAG GCAAGAGCCT CGAGTTCCCG ATCATCCACT AAAGCGAGGA AAGCGCAGGA CACCCAAGT AACCGGCGTT TATCACATTA ACTGGCGGGT GTACTGCTTATT ACTGGGACAA TTTCCTTATT ACTGGCACGA CACGTACTCA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACGA GTATATGAAGT GACGTTTTTT AAACCATAAA CGGCTCCAAG TCACAATCGT CGGATTTAAT CACCCAAATGT CGGAACATG CGCACCAACT TGACGAAAGT CACCCAAAGT CACCCAAATG CACCCAAATG CACCCAAATG CACCCAAACT TGACGAACATG CACAAGACATA CATCAACATC CACCACATCA CACCATCATC GGAACCTTAA TAAGCAAGAA ACTCAAAGAC	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTA AGGACCAAAC CCACGCCAGA CGTGGACACG CAATTGAGACTC ATCCGCCCC CAATTCGGCG CCACTTACTC CACCCCAC CACGTGTCC ATCCGCTGC CACGTGTCC GTGCACCCAA GATTTAAAG CGTGGTCTCG GTGCACCCAA ACCATTGAGT GACAATAGTC GACAACAACA TCTTCAAATG ACCAACAACA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT CAAGGAACAA ATACCATGTAA ACTTCCTGA AACACATGTA ACTTTCCTGA AACAACATGC GGCAGACGT CGCAGAACGT CGCAGAACGG CGCAGAGCT CAAAATGGG CACCAGGGCT CAAAATGGCT CCAATCCAAT	TAAAATAAGT TTCGAGGAAA AACTCTAAGTA AACTCTACTT TACCTGCGCG CGTAATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA ATCCAAAAAC GCGCCGACAAAAA GCACATTTAT GCAGCCACAA GCGACGGAGTT GCCAAAATCTT TTCACATGGA GAAAAAACAC GCGACACAAGTG AAGGACTAGTG AGGACTAGTG AGGACTAGTG AAGGATTTGC GTCCAAAAA GTAATTGATG AAGGATTTGC GTCCAAAAACAC GTAAACAAGC AAACAACAAGC AAACAACAAGC AAATGAACAT ATCTAAACCC	CAGACAAACA AGCTAGACTA AAGAAACCAA AAAACGGGAT TTGTAAAACA AAAAGGTACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAG TATTTCAAAA ATTTCCATGA TCATCTATAT GCCAATTACA GTCTTGCCTT AAGACAACAACA TTCAATGCCA GTCTTGCCAT TAACTACAGAT AAGATGAAGA TTCAATGTCC TAATGCCAT TCTATGGAAC CCTTAATGGAC CAGCTAACG TCCCCTCATC TGCTGTTTTG GACGAAGAGA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCACGA TGAAAATTAC CCAATCATCA CCAATCTTCC AATAAATCAA GGCTIGCTTC AAAACAATAA AGAACTATAT ATTCCACCT ATTGCAGATA ACTTTTCTCT GGAAGCAG GACACATTACT CGACGTCACA GTGATGTCACA GTGATGACAT CTTTTAAATT ACGACGTGGG CACAACAATC ACCAGCATTA CTTTTAACT ACGACGTGGG CACAACAATC ACCAGCATTA ACGACGTGGG GCCCAGCTTA ATGATGAAGA GAGGATGACA
4001 4101 4201 4301 4401 4501 4501 5001 5101 5301 5501 5501 5501 5501 6001 6201 6301 6301 6601 6601 6601 6601 6701 6601 6701 67	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT AAAGGGGGTT TTATATCATTAC GTACAGCAAC GAGCTGTCAT TTGACAGCCT CTATTTATCAA CCGAGAAATGA ACCAAAAATGA AACACAACAT ACGCAAGCATCA ATTCTGCAAC GGCCAAGGCA CGGCAAGGCA CAGCATGAGA	TCTTTTTTTA ATACAAAATG TTTCATTTTAA CTTCCATTTAA CTTCCATTTGAGT CTCCACTTTT GATATTGAGT CTGCACTTATT GAATGTTATTG GAATGTTATAG TGCTTTATAG CAAATGATAACG CACAATGATAA AGTTTCCAATC CGGCCATTCCA CACACACGACA GAATTCCAGCC TCGACGACACA GAATTCCGGCA GAATTCCGGCA GAATTCCGGCA GAATTCCACCC ACTCCCACCA ACTCCGCCA ACTCCCACCA ACTCCCACCA ACTCCCACCA ACTCCCACCA CCCCGCTTCT ACAAAGAAA ACGACTATGG CTCCTCTATG	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCCAAATTATA AGACAATCAA TTGCATCAAC ATGCATCAAC ATTATACAAA CTATAACAAAA CTATAACAAAA CAACCATTA GCATTACAAA ACAAGAACA AACAAGAACA CAACAGAACA CAACAGAACA CAACAGAACA CAACAAGAACA CAACACAAAA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCTA AAATATCTA ATATACATAT CTCTTCTTA AGCTTAAAA TCCCAGAGG GCAAGAGCCT GAGTTCCCG ATCATCCACA GCAACACAAGA CCACCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CACCCAAGA CATCGTCCTC AATCGAACAA TTCCTAATA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTCA ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAC GTATATGAAGT GACGTTTTT AAACCATAAA CGGATTTAT CACCAAATCG CACCAAACT CGGAATTAAT GGAAGACATG GGAAGCACAA GTATCAGTAT GATGAAGATGC GGAACCTTAA CACCATCATC GGAACCTTAA CACCATCATC GGAACCTTAA CACCATCATC GGAACCTTAA CACCATCATCA CACCATCATCA CACCATCATCA	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC TATTTCTGGC TCACCCAAG CCACCCAAGA CCTGGACACG AAATGAACTC ATCCACACAG CCACTTACTC AATCTGGGG CCACTTACTC ATCCGCTGC CACGTTGTCA GACATTAAAG CGTGGTCTG GTGCACCCAA ACAATTGAGT GAGATAGTTC AGCATAATG GACAACAACA TCTTCAAATG GCAAAAATG ACGACACAA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTTCGTGAT ATTCGTGAT ATACCATGTA ACATTCCTGA AACACATGTA ACCATGCATAC CGCAGAACGT CACAAATCGC GGGATTCAAT CACAAATCGA CACCAGGGCT CAAAATGGAT ATGTATAGGGG CACCAGGGCT CCATCCGAT CCATCCGAT CCATCTGTG CAACCTGTG CAACCTGTG CAACCTGTG CAACCTGTC CAACCTGTC CAACCTGTC CAACCTCTCC	TAAAATAAGT TTCGAGGAAA AACTCTACGT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACCAAAAAC AGGGGAGAG GTGCCAAAAA CTGGAGCATTAT GCACCCACA CTGGAGCATT CAACAGCA GCGATGTTGG GCCAAAAGC CGGACAGTG GCCAAAAGC GCGACAATGCT CAACACTGA GAAAAACAC GCGACAATGG GCGACAATGG GCGACAATGG GCGACAATGG GTAATTGATGG AAGGATTGC GTCGAAAATT TCCGAACAAGC AAACAACAACC CAGCCTCTCA	CAGACAAACA AGCTGTGCTT CCATAAATA AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAG TTTTCCATGA ATTTCCATGA ATTTCCATGA TCATCTATAT GCCATTACAT GCCATTACA GTCTTGCCTT CGAATAACAA AGATCAAGAA TTCAATGTC TACCAAGAT CTCAACGC TACTACACG CCTTAATCAG CCTTAATCAG CCTTAATCAG GACGAAGAGC TACATAGCA AGCTCAAGAA AGCTCAAGAA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATCA CCAATCATCA CGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCT ATTGCAGATA ACTTTTCTCT AGCACAATTG CAGACATTACA CTGATGTCACA CTGATGTCACA CTGATGTCACA CTGATGTCACA CACAATGG CACAACATT ACGACGTGGG CACAACATT ACGACGTGG GCCCAGTTAT ATGATGAAAG AGAGGATGACA ATTCCTCGA ACCAGCATTA
4001 4101 4201 4301 4301 4501 4501 4701 4701 4701 4701 5501 5501 5501 5501 5501 5501 5501 6001 6301 6401 6501 6601 6701 6801 7001 7001 7001	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACAAT TATATACCAT TATATACCAT GTACAGCAAC GTACAGCAGCAC CGAGAAATGA CCAAAAATGA ACCCAAGCATCA CCGCAAGGCT AGTCTATGTT ACATTTACAG AGTCTATGTT ACAATTTAGA ACCACAACAT ACACCAGCAG CGCCAAGGCT	TCTTTTTTA ATACAAAATG ATACAAAATG CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGCAAATCA GAATGTTATT GAATTGTCTT TTGCTTCCCA ATAGTAATCG TGCTTTATAG CAAATGATCA CAGAATGATCA AGTTCCAAT CGGCGATGAC AATTAAAGCA GGACTCCCAC AGTACCACCA CACCCGCAG CCTGCCCCCA ACCGCAACAA CCTCCACCA CACCACCACA	ATATGATTTA TGGTAATCCA AAAACTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAAA ATTATTACAT AAAATGGCGT TCAAATTATA AGACAATCAA TCCTTCGCCA TTCAAAGTGT GCATGTGACCA ATCAAAGTCAC CTATACAAA CGGAAGTCAC ATAAGCCCG GAAGAACAA AACAAGAACA AACAAGAACA CAACCATTG CCATCACACA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCTA ATATACATAT AAAAACATA CTTCTTCTTA TCGGAGGGGT GAGTTCAGAGA GCAAGAGCCT CAGTACAGGA CACCCAAGA CCACCAAGA CCACCAAGA CACCCAGAG ACATCCACTT TATCACATTA ACTCGCCGGT CAGGCTTTT ACTCGCCGTC GTACTGCCTC AATCGAACAA TTCCTTATT ACTCGCAAGA CCCAAGACCA CACCCAAGT CATCGAACAA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTGGCTT GAATGTCAAT GACATTGGCTT GAACAACGA GTATATGTAT AAACGAAAAA TGTATGAAGT GACGTTTTT AAACCACATAAA CGGCTCCAAG TCACCATAAA ACTCAAATGT GGGAACATGT GGGAACATGT GGGAACATG GGAAGCACAA GTATCAGAAG CACCATCAT CACCATCAT GAAGAAGATG CACCATCAT GAAGAACTTA CACCATCATCA CACCATCAT CACCATCATCA CACCATCAT CACCATCATCA CACCATCAT CACCATCATCATCATCA CACCATCATCA CACCATCATCATCATCA CACCATCATCATCATCA CACCATCATCATCATCATCATCATCATCATCATCATCATC	ACATAATAT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGG ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCAGA CAGTGGACAG AAATGAACTC CAACTCGCTGC CCACTTACTC ATCCACTAATG GTGCACCCAA CAATTGAGAT CACATAAG GTGTGTCTG GGGATAGTTC AGCAACAACA TCTCCACTAATG GAGATAGTTC AGCAACAACA TCTCCACAAT GAGAACAACA CACATTCAACAT TCGAGCACTA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTTCGTGAT ATTCGTGAT ATACCATGTA ACTTTCCTGA AAAAATATC AACTTCCTGA AAAAATATC GGCAGTACAT CGCAGAACGT CACAAATGGC GGGATTCAAT CACAAATGGC CACAGGGCT CAAAATGGAT ATTTCCACTT CAAATCTATAA TGGAGGCTTG CCAATCTGTG GCAGCCTGCC CAAATACCCA GCAGCAGCTT TTGGCTGCCC CAAGCATCTCT AACTTGATG	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGCTATCTTG ATAATATAAA AAAACCTCAA TTTTATTGA ATACAAAAAC GCACCAAAAAC GCACCAAAAAC GCACCAAATCTT GCGACGACTGTG GCCAAAACCTCC CGACTGCACAATCT TTCACATGGA GCGACAACTCT CGACGACAATCT CGACGACAACAGC AAAGACAAC GCAACAACAAC ACCAACAACA AATGCAAACAAGC AATGCAAAAACAG CACCAACAAC	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAAAGGGAT TTGTAAAACA ATGGATTTTT AAGCAGGAGA ATGGATTTTT AAGCAGGAGA ATTTCCATGA ATTTCCATGA ATTTCCATGA ATTTCCATGA ACATCTACAA TCATCTACAT GCCAATTACA GCCAATTACA GCCAATTACA GCCAATTACA GCCAATTACA GCCATCTACAT TCAATGTCC GCAATACAGAT AAGATGAAGA TTCAATGTCC TACTGCCATT TCATGGCATT TCATGGCATT TCTATGGAAC CCTTAATCAG GACGAAGAGC TTTAATTGAA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCACGA TGAAAATTAC CCAATCATCA CCAATCATCA AGAACTATAA AGAACTATAA AGAACTATAT ATTCCACT ATTGCAGATA ACTTTTCTCT AGCACAATTG GAACAGCAG CAGCATTACT CGAGTGTCACA CTGATGTCACA GTGATGTCATA ATTACAAATG CAAACAATT ACGACGTGGG CACCAGTATT ACGACGTGGG CACCAGCATTT ACGACGTGGG CACCAGCATT ATGGTCAGCG GCCCAGTTAT ATGATGAAG AGAGGATGAC AATTCCTGG CACAATCAG
4001 4101 4201 4301 4301 4501 4501 4501 4501 5501 5501 5501 55	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATACCCC GCATACATAT TATATACCAT TATATACCAT TATATACCAT GAGCTGTCAT TTGACAGCAC CGAGAAATA CCGAGAAATAG CCAGAAATGA AACTATTTAGA CCAAAAATGA AACACAACAT ACACCAACAT TTGTGACGGT ACACCAGGAG ATTCTGCAACG CACCATGAGA GTCCAGAAGA	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCCTTTTG ATAATTGAGT CTGCACTTTT GAATGTTATT GAATGTTATT GAATGTTATT GAATGTTATTG GAATGTTCCAA ATGGTCAATCA CGGCGATGAC AACGGGATGAC AACGGGATGAC CAGAATACCC CGGCGATGAC AACTGCCACC AGTCCGGCAG GAACTCCCAC AGTACGGCAA ACTCCCACCCA ACTCCGCACTA ACACAAGAAA CCGTGAAACA ACAGCTATG GAACTCCGCACT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAATTATA AGACAATCAA TTGCATCAAC ATCATCGCCA CTACAACTTGA CGATGTGACT AAATACTTAG CGAAGTCAC CTATAACAAA CGGAAGTCAC ATAACCTTGA ACAAGCCCGG GAAAGAACA CATACCATTT GCATTACAAA ACCATTCACGGC AAATACCATTT GCATTACAAA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATATCGTTAT CGAAATTTCA AAATATCTA ATATACATAT CTATACATAT CTCTCTTA TCGGAGGGGT AGCTTAAAAA TACCCAGAAG GCAAGAGCCT CGAGTCCCG ATCATCCACG ACCACAGA CCACCAAGA CCACCAAGT CACGCAGGA CAAGGACGAG AGTGGCCTTT TATCACATTA ACTCGGCGGCGT GTACTCCTTATT ATCGGCAGA CATCGCACAA CACCAAATGA CATCGACCAA CACCGACCA CACGGTCTTA CATGGTCTGC ATCGCACAA CACCCAATGA CACCCAATGA CACCGACCA CACGGTCCCC	TATCACAAAC CTTCAAGAAT TTTGAGTAGA ACTTTGGCTT GAATGTCAAT GACATTTGGCTT GAAGTAAGTG AAACAAACGA GTATATGTAT AAACGAAAAGA GACGTTTTTT AAACCATAAA CGCCTCCAAG TCACAATCGA CGCACTCAATGT CGGATTTAAT CACCAAATGT CGGAACATG CACCCAAATGT TGGGAACATG GAAGCACAAG GTATCAGTAT GATGAAGATG CACCATCATC GGAACCTTAA TAAGCAAGAA ACTCAAACTC CACCACACC	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC TATTTCTGGC TCACCCAAC CCACGCCAAC CCACGCCACA CCACGCCACA CCACGCCCAC CCACTTACTC CCACCCCCC CACTTCGGCG CCACTTACTC CACCCCCCC CACGTGTCC CACGTGTCC GGGCACCAA ACAATTGAGT GGGCACCAA ACAATTGAGT GGCACAACA TCTTCAAATG ACCACCACCA ACCACACACA CCTTCCAACA CACAATGAGT CCACACCACA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT CAAGGAACAA ATACCATGTAT ACTTCGTGAT ACTTCCTGA AACTACCATGTA ACTTCCTGA AACTACATAC CGCAGAACGT ACAAAATGGC GGATTCAAT CACAAATGGGG CACCAGGGCT CAAATAGGGG CACCAGGGCT CAATCTATAA TATGAAGCCT CCATTCCATT	TAAAATAAGT TTCGAGGAAA AACTCTAAGTA AACTCTACTT TACCTGCGCG CGTAATTATAA AAAACCTCAA TTTTATTTGA ATAATATAAA AGGGGAGAGA GTGCCAAAAA GCACATTTAT GCAGCCAGCAA GTGGCAGACAT CTGGAGGAGAG CTGGAGGAGTT GCGAATGTTG GCGACAGTCT GCGACAGTGA GTAATTGATG AAGACAAGCA ACAACAAGC AACAACAAGC AATCAAAGAG AATCAAAGAG ACAACCAACA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAG TATTTCAAAA ATTTCCATGA AAACAAACAA TCATCTATAT GCCAATTACA GTCTTGCCTT CGAATAACAA AGATCAAGAT TAACTACAGAT TAACTACAGAT TAACTACAGAC TACTGCCATT CTATGGAAC CCTTAATGTAC GCAAGCTAACC TCCCTCACCACC TGCCGTTTTG GACGAAGACCT CTAAATTCAA AGTCAAGACA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATCA GCATACATTAC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCTT ATTCCACCTT AGCACAATTG GAAACAGCAG GGACACTTACA GTGATGTCACA GTGATGTCACA GTGATGTCACA CTCTTAAATT ACGACGTGGG CACAACAATC ACGACGTGGG GACAACAATC ACGACGTGGG GCCCAGTTAT ATGATGAAAG AGAGGATGAC ATTCCTTGA
4001 4101 4201 4301 4301 4401 5001 5001 5101 5201 5501 5501 5501 55	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGC GCATACATAT TAATACCAT TAATACCAT AAAGGGGTT TTATCTTAC GTACAGCAAC GAGCTGTCAT TTGACAGCTC TTTTTATCAA CCGAGAAATA CCGAGAAATA CCAAGACTGT ACATTTAAGA CAAGACTGT ACATTTAAGA AACACAACAT ACACCAACAT ACACCAACAT TTGTGCACGG AGTCTATGGAA ATTCTGCAACG CGCCAAGGCT TTGTGCACGG CGCCAAGGCT TTGTGCACGG CGCCAAGACT CGCCAAGACAT CGCCAAGACAT	TCTTTTTTTA ATACAAAATG TTTCATTTTAA CTTCCATTTAA CTTCCATTTGA ATAATTGAGT CTGCACATTT GAATGTATG GAATGTTATT GAATGTTACCA TTGCTTTATAG CAAATGATAA AGTTTCCTAAT ATGTTATAG CACACACCACGAT CACGCCATCAC AACGGCATCAC GAATTCCACCC CTCCACCGAT GGACTATCCAGCC GGACTTCCTG AACTACGACC AGTCCGGCCA GCCCGCTTCT ACCACGACCA ACCACGACAA ACCACCTACG CTCCTCTTAT GAAGCCCTCGA	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCCAAATTATA AGACAATCAA TTGCATCAAC ATGATATACAA CTATAACAAA CTATAACAAA CTATAACAGAACA CAACCATGAA CGACCAGAA CAACAGAACA CAACAGAACA CAACAGAACA CAACAGAACA CAACAGAACA CAACAGAACA CAACAGAACA CAACAGAACA CATACCATTA GCATTACAAA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATATCGTTAT CGAAATTTGA AAATATCTA AAATATCTA ATATACATAT CTCTTCTTA AGCTTAAAA TCCCAGAGG GCAAGAGCCT GAGTTCCAC ATCATCCACA AGTGGCCTTT TATCACATA AGTGGCCTT CAGGACGA GTACTGGTCGT CATGGTCGT CATGGACAA TTCCTTATT ACTGGAAGA CTCAAATCA ACTCAATCA CACCAAGA CTCAAATCA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTCA AACTATTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAACG GACGTTTTTT AAACCATAAA CGCCCCAAG TCACCATAAA CGGATTTAAT CACCCAAACT CACCCAAACT TGGAAGACATG GGAAGCACAA GTATCAGTAT GATGAAGATGCC GCAACGTCATCA GGAACCATCAC CACACAACA CACACAACAG ACTCAAACAG ACTCAAACAG ACTCAAACAG ACTCAAACAG CACAGACAAACA CGCAACACAACA	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC TATTTCTGGC TCACCCAAC CCACCCAAG CCTGGACACG AAATGAACTC ATCCACACAC CAATTCAGGG CCACTTACTC ATCCGCCCA ATCCGCCCA ATCCGCTGCC CACGTTGTCA GACATTAAAG CGTGGTCCCA GCACTTACTC GCGCACCAACACA ACATTGAAGT GACAATAGTC ACCACAACACA TCTCAACAACA CACGACACACA CCCCCCA CCACTATCATCG CCCACACCCA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTTCGTGAT ATTCGTGAT ATACCATGTA ACATTCCTGA AAAAATATC CAAGGAACAT ACTACATAC CGCAGAACGT ACAAAATAGG GGGATTCAAT CACAAATCAC CACAGAGCT CAAAATGGAT ATGTAAGGG CACCAGGGCT CAAAATGAAT CCATCTGATA ATGGTGACAT CCATCTGATTG CCACTCTGC CAACTCTGT CAACTTGATG CCAGCACCTTC CAACTTCTC CAACTTGATG CAACTTGATG TGGGCTTCC AACTTGATG	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACCAAAAAC AGGGAGAGG GTGCCAAAAA CTGGAGAGAT CTGGACATTAT GCACCCACA CTGGACAATTC CAACAGTCTCA GACAATGCT CAACAGTGAGTT TCCACATGAG GCAAAAGCA GCGACAATGC CGGCCACAAA GCACATTGTGA GAAAAACAC GCGACAATGC GTCGAATATT TCCAACAGTG AACGACTCTCA AATCGAACAACA ATCGAAAAGCA ACCACCAACA	CAGACAAACA AGCTGTGCTT CCATAAATA AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAG TTTTCCAAGA ATTTCCAAGA TCATCTATAT GCCATTACA GTCTTGCCTT CGAATAACAA AGTCTAACAA GTCTTGCCTT CGAATACAAGAT ACTGCCATT TCTATGGAGA TCCATGCCATT TCTATGAGAA CCTTAATCAG CCTTAATCAG CCTTAATCAG CCTTAATCAG CCTTAATCAG CCTTAATCAG CAGACTAACGA AGCTAAGGAA AGTGAAACTC CCTCACTCC TGCTGCTCC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATCA CGATGCTTC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATATA ATTGCAGCAT ATTGCAGATA ACTTTTCTCT AGCACAATG CAGACATTACA CTGATGTCACA CTGATGTCACA CTGATGTCACA CTGATGTCACA CTGACTTACA CTGATGTCACA CACAGCATTA ACGACGTGGG GCCCAGTTAT ATGATGAAAG AGAGGATGC ACAATCTAA ATGACGACACA
4001 4101 4201 4301 4301 4501 4501 4501 5001 5101 5201 5501 5501 5501 5501 6001 6301 6301 6401 6501 6501 6701 6701 7001 7101 7201 7301 7401	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT TATATACCAT TATATACCAT GTACAGCAAC GAGCTGTCAT TTGACAGCCA CCGAGAAATGA CCGAGAAATGA ACACAAAATGA ACACAAAATGA ACACAAACAT ACAGCAGTCA ATTCTGCAAC CGCAGGAG ATTCTGCAGA CGACAGCAG CCAGAGACT CGCCAGAGAC CGCCAGGCA CGCCAGGCA CGCCAGGCA TGCAGAACA TGCAGAAGA	TCTTTTTTTA ATACAAAATG CTTCTGTTTG ATAATTGAGT CTGCACTTTT CAGCAAATGA GAATGTTATT GAATTGTGTT TTGCTTGCTT TTGCTTCCCA ATAGTAATCG TGCTTTATAG CAAATGATAA AGTTTCTAAT GGCTGCACCA CAGGAATACCC CGGCGATGAC AACGCGCATG GAATTCCCAC AGTCCGGCAG CACTCCCACCA AGTACGACAC ACTCCCACCTC AACTACGAACA CTCCCACCTC ACCACATAGG CTCCTCTTAT GAAGCCCTGA TCAAGGCCTG ACACCTCCGA	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAATTATA AGACAATCAA TCCTTCGCCA ATCAAAGTGT GCATGTGACTA ACACAGTACA CGGAAGTCAC CTATACAAA CCACCTTGA CGAAGTCAC ATAAGCCGG GAAAGAACA CACCATTA CACCATTA CACATACAAA ACCACGCCAAA ACCACGCCAAA ACCACGCCAAA CCACCCCGCC CTTAGCCCGAC TGATATATAT	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATATGTTA AAATATCATA ATATACATAT CGAAATTCTA ATATACATAT CTCCTCTTA TCCGAGGGGT AGCTAAGAGCCT CGAGTTCCCG ATCATCCACAGA CCACCCAGAG AGTGCCTTT AAAGCGAGGA AGTGGCCTTT TATCACATTA ACTCGTCGGT GTACTGCTCG CATGGTCTC ATCGCACAA TTCCTTATT ACTCGCACAA CTCCCACACA CTCCCACACA CATCGCTCC ATCCCACACA CATCGCACA CATCGTCCC ATCCACACA CATCACTCC	TATCACAAAC CTTCAAGAAT TTTGAGTAGA ACTTTGGCTT GAATGTCAAT GAATGTCAAT GTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAACG	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACCAAAC CCACCCAGA CCACCCAGA CCACCCCAGA CCACCCCAG CCACCCCAC CAATTCGGCG CCACCTAATG CCACCCCAC ATCCGCTTGC CACCTTACTC GAGATGATCTG GAGATAGTTC AGCAACACAA TCTTCAAATG ACCAACACAA TCTCCACATG ACCAACACA CACTATTCATG CACAACACA CACATTCATCA CACAACACA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT CAAGGAACAA ATACCATGTA ACTTCCTGA ACTATCCTGA ACTATCCTGA ACTACATAC CGCAGAACGT ACCACATAC CGCAGACCAG CACAAATGGC CACAAATGGG CACAAATGGG CACAAGGAT CAAATGGAG CACAGGGCT CAAATCCAAT CCATCTGATT GCAGCCTGGC CAAATACCA CCAGCCTGC CAAATCCAAT CCAATCCAAT	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACTAATAAA ATACCAAAAAC GCGCCACAAAAG GCGCACAATTGT GCGACGCACTG GCGACAGTGT GCGACGACTG GCGACAGCAG GCGACAAGTG GCGACAAGTG GCGACAAGTG GCGACAAGTG TTCTAAAACAC GCGCACAAAAG GTGACTACTC AACAACAAGA AATTGTAAAACC CAGCTCTCA ACAACTACTA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAAAGGTACA ATGTAAAACA ATGGATTTTT AAGCAGGAGA TATTTCAAAA ATTTCCATGA AATACCATACA TAATCCATGA CAATCTACAA TCATCTACAT CGAATAACA GCCAATTACA GCCAATTACA GCCAATTACA GCCAATTACA GCCAATACA GCCATCACAT TAACAATGCACA TACACACACT TACTGCACATT TCTATGGAAC CCTCAACGT TACTGCACATT TCTATGGAAC CCTCACCTCA	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCACGA TGAAAATTAC CCAATCATCA GCATGCTTC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCT ATTGCAGATA ACTTTTCTCT AGCACAATTG GAAACAGCAG CAGCATTACT CGAGTGACAA CTCACTTACA CTGATGTCACA ACGACGTGGG CACAACAATC ACGACGTGGG CACAACAATC ACGACGTGGG GCCCAGTTAT ATGATGAAAG GAGAGTGACA ATTCCTTGA AATTCCTTGA AATTCCTTGA CACAATCAACA CTCTTTTTGT
4001 4101 4201 4301 4301 4401 4501 4501 5001 5301 5501 5501 5501 5501 5501 5	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATACCCC GCATACATAT TATATACCAT TATATACCAT TATATACCAT CAGGGGGTT TTATCTTAC GTACAGCAAC GAGCTGTCAT TTGTCACG CAGAGACTTGT ACATTTAGA CCAAAAATGA AACACAACAT ATTCTGCAACG CAGCATGAAA CAGCATGAAA CAGCATGAAA CAGCATGAAA CAGCATGAAA CAGCATGAAA CAGCATGAAA	TCTTTTTTA ATACAAAATG TTTCATTTAA CTTCTGTTTG ATAATTGAGT CTGCACTTTT GAATTGTGTT GAATGTAATG GAATGTTATT GAATGTTATTG GAATGTTAATG TGCTTTATAG CAAATGATAA AGTTTCTAAT ATGTTCCATC CGCGATGAC AACGGGAATG CAGAATACCG GAATTCCCAC AGTCCGGCAG GCACTCCCAC ACTACCACCA CACAACACAC CTCCCCCTCAT CGCGAACAC CTCCTCTTAT GAACTCCGCAC CACACTACG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CACACTCCG CCACTTTTTTTTTT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAATTACA AGACAATCAA TTGCATCAAC ATCATCGCCA ATTACTTAG CTATAACAAA CGAAGTCAC CTATAACAAA CGAAGTCAC ATAAGCCGG GAAAGAATCA CATACCATTT GCATTACAAA ACCACCATTA CACCTTGAC CATACCATTT GCATTACAAA ACCACCATTT GCATTACAAA CGCACCACAAA ACCAGCCAAAA ACCACCACGC AAGTCCCCGA CCAACAGCAC CCAACAGCAC CCAACAGCAC CCAACAGCAC	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCATA ATATACATAT CTACTTCTTA ACTTCTTCTA AGCTTAAAA TCCGAGAGGGT CGAGTCCAGA GCAAGAGCCT CGAGTCCCG ATCATCCACT CAGCTCCCG ATCATCCACAGA CCACCCAAGA CCACCCAGAG CCACCCAGAG CCACCCAGAG CCACCCAGT TTCCCCGGGCGT TACCGCGGCGT TACTGCCAGA CATCGCACAGA CCACGTCCCA AATCGCACGAG CATCGCACAGA CCCCAATCA CATCGCACGAG CACCCAATCA CATCCAAATCA	TATCACAAAC CTTCAAGAAT TTTGAGTAGA ACTTTGGCTT GAATGTCAAT GACATTTGGCTT AAACAAACGA GTATATGTAT AAACAAACGA GTATATGAAT GACGTTTTT AAACCATAAA CGCTCCAAG TCACAATCA CGCATCTAT CGGATTTAAT CGGATCTAT GGGAACATG GAAGCACAA GTATCAGTAG GAAGCACAA GTATCAGTAG GAAGCACAT GGGAACCTTAA TAAGCAAGAG ACTCAAACG GCATCCGTCA CACCATCAAC CGCAACACAC CACACAACAC CACACAACAC CACACAACA	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC TTAGTACTTA AGGACAAAC CCACGCCAGA CCACGCCAGA CCACGCCAGA CCACGCCCAG CCACTTACTC CACCCCCAC CACTTACTC CACCCCCAC CACTTGTGCC CACGTTGTCA GACATTGAGA GGTGTTCTG GTGCACCAA ACAATTGAGT GGCACACAAT GACAACAATG CACGAGACAAT CCTCACAATG CCACATCATCA CCACAACCCA CACATATCAT CCACAACCCA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT CAAGGAACAA ATACCATGTAT ACTTCGTGAT ACTTCCTGA AACACATGTA ACTTCCTGA ACATACATAC CGCAGAACGT CACAAATGGAT CACAAATGGAT CACAAATGGAT CAATCTCATA ATGAAGCCT CAATCTCATA ACGAGCATTGT CCATCTGTT GCAGCCTTGC CAAGCATCTT TTGGCTGCCC CAAGCTTTAT AACTTCACT CAATCTCTCT AAACTGAGT CCACTGGATCA CCACTGGATCA CCACTGGATCA CCACTGCCC CAAGCTTTAT CCACTCTCTT AAACTGAGTCC CCCCCCCC CAGGCTATC CACTCTAAA AAAGATTGTA	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGTATATTATA AAACCTACATT TACCTGCGCG CGTATATTTAA ATACTACATT TACTGCGCG CGCACAACA GCGCCCACAAAAA GCACATTTAT GCACCCACCA GCGACGTTGG GCCAAAAGC GCGACAGTTGG GCCAAATGCT CACACGACAAG GCACACTGTG GCGCACAAGT CCGCACAAGT CTCCAATATTC CTCAATCCA AACAACAAGC AACACAACA ACAACTCTCA ACAACCAACA ACAACTCTCA ACAACCAAC	CAGACAAACA AGCTGTGCTT CCATAAATA AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAGA TCATCTATAT GCCAGTATACA GTCTTGCCTT CGAATAACA TCATCTACAT GCCATTACAG GTCTGCCTT CGAATAACA TCAATGACA TCAATGACA TCAATGACA TCAATGACA TCAATGACA TCAATGACA CCTTAATCAG CAAGCTAACGA CCTTAATCAG CAGCAAGAC CCTTAATCAG CAGCAAGAC TCAATGACC TCCTGCATTTG GACGAAGACC TCAATGAACCT TCAATGAACCT CAATGAACAT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATCA GCATACATT ATGAAATAAA AGAACTATAT ATGCACCT AATGCAGATAA ACTTTTCTCT AGCACAATTG GAAACAATG GAAACAACAG CAGCATTACA GTGATGTCACA GTGATGTCACA GTGATGTCACA CTTTTAAAT ACGACGTGGG GACAACAATC ACGACGATTA ACGACGTGGG GCCCAGTTAT ATGATGAAAG GGCCAGTTAT ATGATGAAAG CACATCTAA ATGCGGCGCC AATTCCTACA
4001 4101 4201 4301 4301 4401 5001 5001 5001 5501 5501 5501 55	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA GTCATCCCATAA GTTCCCGTTAT TAATAAAATG TCATTTTCCG AAATATCCCC GCATACATAT TATATACCAT TATATACCAT GTACAGCAAC GAGCTGTCAT TTGACAGCAC CGAGAAATA CTAATTCTACA CCGAGAAATA CCAAGACTGT ACATTTAAGA CCAAAAATG AACACAACAT ACCCAAGGC CCCAAGGCA CCCAAGGCA CTCTGTGACGA CTCAGAGACTG TTGTGCACGA CACCACAGCA TTGTGCACGA CACCACAGCA CTCCACAAGA CTCACAGACTG ATTTCCACAA TCCACAACAT CGAGACTGG ATTTCCACAA TCCACAACAC CACACACCAC CACACACCAC CACCACACAC CCCAAGCAC CACCACACAC CTCCCACAGCAC CACCACACAC CACCACACAC CACCACACAC CCCCACAGCAC CACCACACAC CACCACACAC CACCACACAC CACCAC	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCCATTTAA CTTCCATTTAA CTTCCACTTTG ATAATTGAGT CTGCACATTT GAATGTATG GAATGTTATG GAATGTTACA TTGCTTATAG CAAATGATAA AGTTTCCAATC CGGCCATCAC AACGGGAATG CAGAATACCC CTCCACGAT GGATTTCTG GAATTCCACC CTCCACGAC AGTCCGGCAG GCCCGCCTCT AACAAGAAA ACGCCAGAC ACCGCCATGG CTCCTCTTAT GAAGCCCTGA ACCACGACCA CTCCACTCT GAAGCCTCGA CTCCACTCT CAAATGTCATC CCGCAATATG CAAATGTCATC	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTAA TAGCGCGTGG AAACATAAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCCAAATTATA AGACAATCAA TTGCATCACCA TGCATCACCA ATTATACAAA CTATAACAAAA CTATAACAAAA CTATAACAAAA CAACCATTA CCACCCTGAA CAACAAGAACA ACAAGAACA CAACAGAACA CAACAGAACA CAACAGAACA CAACAGAACA CATACCATTA GCATTACAAA ACTACCACGC CTAACACACA CCACCCGC CTAACACACAC CCACCCGCC CTAACACACACAC	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AATATACATAT CTATACATAT CTATACATAT CTCTCTTGT AGCTTAAAA TCCGAGGGGG GCAAGACCC ATCATCCACT CAGTACAGGA CGACTACAGGA CGACTACAGGA CGACCCAGT AAAGCGAGGA CACCCAGGT AAAGCGAGGA CACCCAGGT CATGGTCGT CATGGTCGT CATGGCCTT TATCACATTA ACTCGGCGCT GTACTGCAGA TTCCTTATT ATCGCAAGA CTCAAATCA CTCCAGTGC ATCCACCA CACCAAGTCC ATCCACTAG CACGTACTCA CATCGTCCA CACGACAG CTCAAATCA TTCCGTGAGG CACCCAGTG CACCCAGTCC ATCCGTGCAG CACCCAAGTCC ATCCGTCGAG CACCCAAGTCC ATCCGTGCAG CTCAAATCAA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTCA ACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACGAAAAG GCATTTTT AAACCATAAA CGCCTCCAAG TCACAATGT CGGATCTTAT CACCCAAACT CGGAACATG GGAAGACACAA GTATCAGTAT GACGAACATG GGAAGACCAA GTATCAGTC GGAACCTCAC GGAACCTCAC GGAACCTCAC GGAACCTCAC GGAACCTCAC GGAACCTCAC CACCAAACG GTCCGCAAAG GTCCGCAAAG GTCCGCAAACA ACTCAAACAG ACACCATCAC CCCCATGCA ACACCATCAC	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC TATTTCTGGC TCACCCAAG CCACGCCAAG CCACGCCAAG CCACGCCACAG CAATGAACTC AATCGCGCG CACTTACTC ATCCGCTGCC CACGTTGTCA GTGCACCCAA GCACTTAATG GTGCACCCAA ACATTGAGT GCGCACCAA CACATTGAGT CACACAACACA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTACTCAC TTTTAACATT CAAGGAACAA ATACCATGTAA ACTTTCCTGA AACATACATAC CGCAGAACGT ACAAAATGGC GGGATTCAAT CACAAATCAA ATGATAGGG CACCAGGGCT CAAAATGGAT ATGTAGAGCCT CCATCTGTGATTA TCGAGCATCTG GCAGCCTTGC CAACTCTGTT TTGGCTGCCC CAACTTCTTC AACTTGATG TCGAGCATCT CAACTTGATG TCGAGCATCT CAACTTGATG TCGAGCACCT CAACTTGATA AGAGCTTTA ACTGGATCACCA TCGACACCTCC AACTCTTAAA AAAGATTGTA	TAAAATAAGT TTCGAGGAAA AACTCTACATT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACCAAAAAC AGGCGACAG GTGCCAAAAA CTGGAGAGAG GCGATGTTGG GCCAAATGCT CAACAGTGTGG GCCAAATGCT CAACAGTGAG GTACTATGA GAAAAACAC CGGACAATG GCGACAATGT TCCACATGA GAAAAACAC GCGACAATGT AACGACTCTCA AATCAAACAGC ACAACTGTGA AACAACTGTG AAACAACG AAATCAACAACA	CAGACAAACA AGCTGTGCTT CCATAAATA AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAG TTTTCCAAAA ATTTCCATGA ATTTCCATGA TCATCTATAT GCCATTACAT GCCATTACAA GTCTTGCCTT CGAATAACAA GTCTTGCCTT CGAATACAAGAT TCCATGTCATC TACAATGTC TACTACAGC TCCTTAATCAG CCTTAATCAG CCTTAATCAG CCTTAATCAG GACGAAGAGC TACTAAGAA AGTGAAACCT AACTACGACATC TGCTGATATC TGCATGACCTC TGCTGATATC TGCATGAACCTC CTGATGTATC TGCATGAACCTC CTGATGTATC TGCATGAACCTC CTAATAAACT	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATCA CCAATCATCA CCAATCATCA CGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCT AATGCAGATA ACTTTTCTCT AGCACAATG CACACATTACA CTGATGTCACA CTGATGTCACA CTGATGTCACA CTGATGTCACA CTGATGTCACA CACATCTACA CGACGTGGG GCCCAGTTAT ATGACGAGGG GCCCAGTTAT ATGATGAAAG AGAGGATGACA ATTCCTGA AGAGGATGC ACAATCTAA CACATCTAA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACATCTACA CACACACA
4001 4101 4201 4301 4301 4501 4501 4501 5001 5101 5201 5501 5501 5501 5501 5	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT TATATACCAT TATATACCAT TTATCTTAC GTACAGCAAC GAGCTGTCAT TTGACAGCTC TTTGTACAGCAC CCGAGAAATA CCGAGAAATA CCGAGAAATA CCGAGAACTTGT TTGTGCACG AGTCTATGTT ACAGCAGTCA ATTCTGCAAC CGCCAGGCA ATTCTGCAGAA GTCCAGAAGA TGAGAGATGA TGTGTGGGGA TATGAAAAA	TCTTTTTTTA ATACAAAATG CTTCTTTTTTTTTTTTTT	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGCGT TCAAATTATA AGACAATCAA TCCTTCGCCA ATCAAAGTGT GCATGTGACTA CACCTTGGC CAAAGTCAC CTATAACAAA CCACCTTGA CGAAGTCAC CATACCATT GCATTCACGG AAAGACCA CACCATTA CACCATTA CACCATTA CACCATTC CTAACCACA CCACCACGC CAACAGCAC CCACACGCC CTAACCCGTC CTAACCCGTC CGAACAACAC	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATATGTTA CGAAATTTGA AAATATCTA ATATACATAT CTCTTCTTA TCGGAGGGGT AGCTTAAAAA TCCCAGAG GCAAGAGCCT CGAGTTCCCG ATCATCCACAG CCACCAGAGC CCACCCAGG CCACCAGGTCCCG ATCGCCCAGT CATGGCCTTT AACTCGCAGAG CTCGCACACA ACTCGCACACA CTCCCCAATGA CTCCCAAATCA CTCCCAAATCA CTCCCAAATCA CTCCCGGAGA CACCGTGTCCC ATCCAAATCA	TATCACAAAC CTTCAAGAAT TTTGAGTAGA ACTTTGGCTT GAATGTCAAT GACATTTGGCTT GAACTATGACT AAACAAACGA GTATATGTAT AAACAAACGA GACGTTTTTT AAACCATAAA CGGCTCCAAG TCACAATCACA CGCACTCAATGT CGGATCTTAAT CACCAATCAC GGAAGCACAA GTATCAGTAT GATGAAGATG CACCATCATC GTATCAGTAT GATGAAGATG CACCATCATC GTATCAGTAT GAAGAACATC GTACCATCATC GGAACCTTAA TAACAAAAG ACTCCAATCAC GTACCATCATC GGAACCTTAA CACCATCATC CACCACTCA CACCATCATC CACACATCAC CACACATCAC CACACATCAC CACACATCAC CACTCATCATC CACACATCAC CACTCATCATC CACCACTTAA	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG GATATGACTT TATTTCTGGC ATCTGCTTCA TTAGTACTTT AGGACAAAC CCACGCCAGA CGTGGACACG AAATGAACAA CGTGGACACG CACTTACTC CACCCCAC ATCCTGTGC CACTTACTC GACATTCAG GTGCACCCAA GAATTTAAG GCGATGTTCTG GAGATAGTTC AGCAACAAAT CCTCCATATG ACCACACACA CCTTCCATCAT CCACACCAA CACTATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACATATCATCA CACAACAAA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTATTAAATC TTTTAACATT CAAGGAACAA ATACCATGTAA ACTTCCTGA ACTTCCTGA ACTATCCTGA ACTACATAC CGCAGAACGT ACCAACATAC CGCAGAACGT CACAAATGGC GGACTCAAT ATGATAGGG CACCAGGGCT CAAAATGGAT ATTCCACTT CCATCTGAT GCAGCCTTGC CAAAATCCT TTGGCTGCCC CAAACTGATC TTGGCTGCCC CAAGCTTTCT AACTTGAAG TTGGATGATCC TCCACGTCC ACCATCGATC ACTCTTAAA ATTCTTCAAT CACTCTTAAA ATTCTTCAAT CACTCTTAAA TTGAAGATCT	TAAAATAAGT TTCGAGGAAA AACTCTAGGGGG CGATATCTTG AAACTCTACTT TACCTGCGCG CGTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATCCTAAAAAG GCGCACAAAAA GCACATTTAT GCAGCCACCA CTGGACGCACTG CGACAGTGT GCGAATGTTG GCGACAAGTG GCGACAAGTG GCGACAAGTG GCGACAAGTG GTGGAACTACG GTGCACAAAAG GTAATTGAAGAT ACGACTACAAG AACACAACAA ATTGTAAACA ACACTTCTC ACACACTTCA ACACCACCA ACACTTCTCA ACACACTACA ACACTTCACAAAAGA ACACTTCACA AAAACACTGTC	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AACCAGGAGA TATTTCCATGA AAAACAAACA TCATCTATAT GCCAATTACG GCCAATTACG GCCAATTACA GTCTTGCCATG TACACAACAT CGAATAACA AGATCAAGAT AAGATGAAGAT CTAATGCACAT TCCATGTACG CAAGCTAACG CAAGCTAACG CAAGCTAACG TACAGCAACA TCCACTCTC CGCAGTTTTG GACGAAGACC TTAAATGAACCT AAGTCAACAT TGCATGAACCT CTGATGTATCC TGATGTATACC TGATGTATCC TGATGTATCC TGATGTATCC TGATGTATCC TACACCTCC TGATGTATCC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCACGA TGAAAATTAC CCAATCATCA CCAATCATCA CCAATCATCA CGCTTGCTTC AATAAACAATAA AGAACTATAT ATTCCACCTT ATTCCACCTT ATTGCAGAATAG CAACAATTG CAACACAATCG CAACACAATCG CAGCATTACA CTGATGTCACA CTGATGTCACA CTGTTAAATT ACGACGTGGG CACAACAATCT ACGACGTGGG CACCAGCATTAT ATGACGAGAGG GCCCAGTTAT ATGACGACGCC GCCCAGTTAT ATGACGACGCC AATTCCTTGA CACATCTAA ATCCGGCGCCC AATTCCTTGA CACAACAATCTAA CTCTTTTTGT TATATTTTGT AGTTTTTATA
4001 4101 4201 4301 4301 4301 4501 4501 5001 5301 5501 5501 5501 5501 5501 5	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA TGATTTTAC GTTCCGTTAT TAATAAAATG TCATTCTCGA AAATATACCCC GCATACATAT TATATACCAT TATATACCAT TATATACCAT CAGGGGGTT TTATCTTAC GTACAGCAAC GAGCTGTCAT TTGTCACAG CCGAGAAATA CCAAAATGA CCAAAAATGA ACACAACAT ACACCAACGT ACATTTAGAAACT CAGCAGGACA CCCAGAGCTCG ATTCTGCACAG CAGCATGAAA CAGCAGCACA CGCCAAGGCT TTGTGACGGCA CAGCATGAAA CAGCAGCAC CGCCAAGGCT TTGTGACGAC CGCCAAGGCT TTGTGACGAC CAGCATGAAAAAA CAAGAACTCG ATTTTCGAAC CGCCAAGGCT TTGTGACGGGA TTGTGGGGGA TATTTCCATAG TATTTCCATAG TATTTCATAG	TCTTTTTTTA ATACAAAATG TTTCATTTAA CTTCCTTTTG ATAATTGAGT CTGCACTTTTG ATAATTGAGT CTGCACTTTTG GAATGTTATTG GAATGTTATTG GAATGTTATAG CAAATGATAA ATGTTTCCAAT CGCCGATGAC ATGTCCACC CAGCGAATGCCG CAGAATACCC GGACTCCCACC AGTCCGGCAG GCTTCCCACC AGTCCGGCAG GCTCCCCCCTCT ACTACCACCA ACTACCACCA CACATACGACA ACTCCCACTC CAACAACAAA CCGTGAAACA CCGCGAACAC ACACCTCTG CAACATCCCG CAACTACCG CACCTCTG CAAATATGTG AAAATGTGAT TTAAAACATA TCAAGTTATA TTAAAACATA	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAATTATA AGACAATCAA TTGCATCAAC ATCATCGCCA ATGCATCAAC TGCATCACA CGAAGTGTGAC AATACTTGAC CATAACATAA	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCATA ATATACATAT CTGAAGGGGGT AGCTTAAAAA TCCGAGGGGGT CGAGTCCCG ATCATCCACA CGAGTCCCG ATCATCCACA CGACTACAGGA CCACCAAGA CCACCAAGA CCACCAGAG CACCCAGA AGTGGCCTTT TATCACATTA ACTGGGGCGT GTACTGCCAGA CACGTCCCA ATCCCACATCA AATCGACCA AATCGACCA CACGTCCCA CACGTCCCC ATCCCACAGA CCACAATCA CACCGTGCC CACCCAATCA CACCGTGCC CACCCAATCA CACCGAGCC CACCCAATCA CACCGTGCC CACCCAATCA CCCACAATCA CACCGTGCC CACCCAATCA CACCGTGCC CACCCAATCA CACCGTGCC CACCCAATCA CCCACCACCAC CCCACCCACC CCCCCACCCA	TATCACAAAC CTTCAAGAAT TTTGAGTAGA ACTTTTGTAC AACTTTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACAAACGA GTATATGAAG GACGTTTTTT AAACCATAAA CGCTCCAAG TCACAATCA CGGATTTAAT CACCAAATGT CGGATCTAAT CACCCAAATGT CGGAACACAG GAAGCACAA GTATCAGTAG GAAGCACAT CACCATCATC GGAACCTTAA TAAGCAAGAG ACTCAACTC CACCATCAC CACCATCAC CACCATCAC CACACAACAC ACTCAACTC CACCACACAC CACACACA	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC TTAGTACTTA AGGACAAAC CCACGCCAGA CCACGCCAGA CCACGCCAGA CCACGCCAGA CCACTTACTC CACCCCAG CCACTTACTC CACCCCAG CCACTTACTC CACGTTGTCC GACATATGAGT GGCACCAA ACAATTGAGT GGCACCAA CACTATCAG GGCACCAA CACTATCAG CACGAGACAAT CCACAACAAT CCACAACCAA CACTATCATCA CACAATCATCA CACAACCAA CACTATCATCA CACAACCAA CACTATCATCA CACAACCAA CACTATCACAA	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTACCATGTA TTTTAACATT CAAGGAACAA ATACCATGTA ACTTCCTGA AACTACATAC CGCAGAACGT ACATACATAC CGCAGAACGG GGATTCAAT CACAAATGGAT CAATACGAG CACCAGGGCT CAATACCAA ATGATAGAGCCT CCATTCCAAT ACCATCTGTT GCAGCCTTGC CAAATACCA GCAGCAGCTT TTGGCTGCCC CCAGCCTCC CAACTCTATA AGCAGCACCT CAATCTCTCT AACTTGAGC CCACTGCC CCAGCCTCC CAACTCTATA AGCAGCCCC CCAGCCTCC CAACTCTATA AGCACCCC CCCCCCCC CCCCCCCC CAGCTTCAAT TTCAACTCAAT TTTAAAAATC CAACTTAAA	TAAAATAAGT TTCGAGGAAA AACTCTACTT TACCTGCGCG CGCTATCTTG ATAATATAAA AAACCTCAA TTTTATTTGA ATACTACTTA AAAACCTCAA TTTTATTTGA ACCCACCA GGGGGGGGG GCCCAAAAA CTGGAGGCATT CGGAGGCATT CGCACACTTGA GCCAAATGCT CACACGCACA GGACACACG GCCACACTGA GAAAAACAC GCGACACTGC GTCCAATATTC CTCCAATATCCA AACAACAAGC AACAACAAGC AACAACAAGC ACAACCAAC	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAGA TCATCTATAT GCCATTACAG ATCTCAATACA GTCTTGCCTT CGAATAACAA AGATCAAGAC TCATCGCATT CTAATGACA CCTTAATGTC TCAATGACA CCTTAATCAG CAGACAACA CCTTAATCAG CAGACAACA CCTTAATCAG CAGACAACA CCTTAATCAG CAGACAACA CCTTAATCAG CAGACAACA CCTTAATCAG CAGCAAGAC TACACCTCC TGCTGTATTG AACTAACAACC TAATAAACTAAAC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCAGGA TGAAAATTAC CCAATCATCA CCAATCTTCC AATAAATCAA GGCTTGCTTC AAAACAATAA AGAACTATAT ATTCCACCT ATTGCAGCAT ACTTTTCTCT GGAAGCAGTG CAGCATTACA GTGATGTCACA GTGATGTCACA GTGATGTCACA GTGATGTCACA CTTTTAAAT CTACTTACAATC GCCCAGCATTA ACGACGTGGG GCCCAGCATT ATGATGAAG GGCCAGTATC ATGATGAAG CACATCTAA ATTCCTGG AGAGATGACA AATTCCTAA ATTCCTACA AATTCCTAA CTTTTTTAAT ATGGGCGCC AAGTGCACA
4001 4101 4201 4201 4301 4401 5001 5001 5001 5001 5501 5501 55	CAGAAATTAG AAATTTTGGA GTCATACTTA ACCTACATTG ATTGCCATAA GTCATCCCATAA GTACCGTTATG TGATTTTAC GTTCCCGTTAT TAATAAAATG TCATTCTCGA AAATATCCCC GCATACATAT TATATACCAT TATATACCAT GTACAGCAAC GAGCTGTCAT TTGACAGCAC CGAGAAATA CTAATTCTACA CCAAGACTGT ACATTTAAGA CCAAAAATG AACACAACAT ACCCAAGGC AGTCTATGTT ACAGCAGCAC ATTCTCGCAA CTCCAGAAG CTCAGAGACTG TTGTGACGGA CACCACAGC TTGTGGACGA CACCACAGC TTGTGGACGA CACCACAGC CTCAGAAAA CCACAACAT TGCAGAGCTG ATTTCCACAA CTCAGAAAAATGA AACACTACAC	TCTTTTTTTA ATACAAAATG TTTCATTTTAA CTTCCATTTAA CTTCCATTTGA ATAATTGAGT CTGCACATTT GAATGTTATT GAATGTTATTG GAATGTTATTG GAATGTTATAG CAAATGATAA AGTTTCTAAT ATGTTCCATC CGGCGATCAC AACGGGAATG CACACACGAC GAATTCCAGCC TGGAGACACG GAATTCCAGCC AGTCCGGCAG GGATTTCCG GAATACCACC ACTCCACCAT CCGCCATCTT ACTTCCACTC GAACACCACGAC ACCACGACCA CTCCTCTTAT TCAAGGCCT GAACACTTCG CAAATATGGAAT CTCGTGTTTG TTAAATACAA TAAAACATA TAAAAACATA TAAAAACATA TAAAAACATA	ATATGATTTA TGGTAATCCA AAAACTTTAT CTGTAAGCCT AGAAAAAACA TCCTGCCTTT ATTTAAATAA TCCGATACTA TAGCGCGTGG AAACATAAA CTTTTCAGTA ATTATTACAT AAAATGGCGT TCAAATTACAT AGACAATCA TGCATCACCA TGCATCACCA TGCATCACA ATTATACATA CAACATTACA CTATAACAAA CTATAACATAA CCACCTTGAA CGACATCAC ATTACCATTA GCATTACAAA CACACCAGC AAAGAACAT CCACCCGC GAAAGAACA CATACCATTT GCATTACAAA ATCCACGCC CTAGCCCAAA ACCACCGCC CTAGCCCAAA ACCACCCGC GGAGATATAT GCCCCCCGC GGAGATATTA GCACCCGTCG GGAGATATTC CTATAAAAAAT AATGTTGTTT TACAAAACAT	GGTAGCTATA CTATAGAAGC ATTTTAAGA AATTCGTTAT GTGATCCTTT AAATAAGAGT CGAAATTTGA AAATATCTCA ATATACATAT CTGAGGGGGT CGAGTCCCG ATCATCCACAGAG CCACCAAGA CCACCCAGAG CCACCCAGT CGGGTCCCG ATCATCCACT CAGTACAGGA AGTGGCCTTT TATCACATA ACTCGGCGGGT GTACTGCTCCA ATCGCAGGA CTCATCCACAGA CACGCTACTCC ATCGCAGGG TTCCTTATT ACTCGCAGAG CTCAAATCAA CTCCAAATCA CTCCAAATCC CACGTGCTCC CATCGTCCG CACTCCCC CACCCAAGA CCCAACAGACA CCCAACACA CTCCACAGACA CTCCACACA CTCCACACA CTCCACTCC CATCGTCCC CATCGTCCC CATCGTCCC CATCGTCCC CATCGTCCC CATCGTCCC CTTTTTCCTTT ACAAATCAACC CCAATATATA	TATCACAAAC CTTCAAGAAT TTTGAGTAGG ACTTTTGTAC AACTATTGGCTT GAATGTCAAT TGTCAAAGTG AAACAAACGA GTATATGTAT AAACAAACGA GTATATGAAG GACGTTTTT AAACCATAAA CGGCTCCAAG TCACAATCA CGGATTTAAT CACCAAACTG CACCAAACTG GGAAGCACAA GTATCAGTAT GGAAGACACAA GTATCAGTCA GGAACCTTAA CACCATCACC GGAACCTTAA CACCATCACC GGAACCTTAA CACCATCACC GGAACCTTAA CACCATCACC GGAACCTTAA CACCACAACA ACTCAAACAG GGTCCCTTA CACGACAACA ACTCAACTC CACACTCACC CACACTCACC CACACTCACC GGAACCTTAA	ACATAATATT ATATTAAAGT TATGTATCTA ACACATAAAG CCACAGAATA GATATGACTT TATTTCTGGC TATTTCTGGC TCACCCAGA CCTGGACACAG CCACCCAGA CCTGGACACG CAATTGAGCG CCACTTACTC ATCCGCCCAC CCACTTACTC ATCCGTGTGC CACGTTGTCTG GTGCACCCAA GAATATGAGT GAGATAGTTC AGCATAGTTC AGCACAACACA TCTCCAATGG CCCCCA ACATTGAGC CACGTTGTCTC GCACACACACA CACATTGAGC CCCCCCA CACATAGCTC CCACACCCA CACATGCTC CCCCAACCCCA CCCCCCA CCCCCCA CCCCCCA CACATGCTC CCCCACCCA CACATGCTC TCCGCCCCA CCCCCCA CCCCCCCA CCCCCCCA CCCCCCCA CCCCCC	TTTTTTAAG TGAGCTATTT AGTACCTCAC TTTTTAACATT AGTACCACGTAT TTTTAACATT CAAGGAACAA ATACCATGTA ACTTCCTGA AACATTCCTGA AACAAATACC CGCAGAACGT ACAAAATAGC GGATTCATA CGCAGAGCT CAAAATGGAT ATGATAGGGG CACCAGGGCT CAAAATGGAT TTTGCACTTC CAATCTGTG CCATCTGTG CCATCTGTG CCATCTGTG CCATCTGTG CCACCTTGC CAACATCTCT TTGGCCCC CAACATCTCT AACTTGATC TCGATCTCTC AACTTGATC CACCTCTCAAA AAAGATTGTA ACTCTTTAAA AAAGATTGTA	TAAAATAAGT TTCGAGGAAA AACTCTACGT TACCTGCGCG CGCATATCTTG ATAATATAAA AAAACCTCAA TTTTATTTGA ATACTACTTA ATACCAAAAAC AGGGGAGAG GTGCCAAAAA CTGGAGCATTAT GCACCTCACA CTGGAGCATT CAACAGCAACA CGGACATTGT GCCAAAAGC GCGACATTGT GACAATGCT CAACAGCACA GGAAAACAC GCGACAATGCT CTACATGAT GAAAAAACAC GCGACAATGT TCCAAATGTT AACAACAACA ATCAAAAGAG ACAACCAACA ATCAAAAGAG ACAACCAACA ACAACTGTG AACAGTTCAC AAATAATATTA AAAATATTTA AAAATATTTA AAAATATTTA AAAATATTAA	CAGACAAACA AGCTGTGCTT CCATAAATAT AAGAAACCAA AAACCGGGAT TTGTAAAACA AAAAGGTACA ATGGATTTTT AAGCAGGAGAG TTTTCCATGA ATTTCCATGA ATTTCCATGA TCATCTATAT GCCATTACA GTCTTGCCTT CGAATAACAA AGTCTAGAAGA TTCAATGTCT CGAATACAAG TTCAATGTC TACATGCCAT TCTATGGAAC CCTTAATCAG CAGACTAACAG CCTTAATCAG GACGAAGAGC TACTAAGAAC TGCTGTTTTG GACGAAGAGC TACTAAGGAA AGTGAAACCT AACTAACACTC TGATGTATCC TGATGTATCC TGATGTATCC TGATGTATCC TGATGTATCC	CCTGATTAAA CCAGCAATCA TTTTCTCATT CCAATCACGA TGAAAATTAC CCAATCATCA CCAATCATCA CCAATCATCA CGCTTGCTTC AAAACAATAA AGAACTATATA ATTCCACCT ATTGCAGATA ACTTTTCTCT AGCACAATTG GAAACAACAG CGACGATTACT CGATGTCACA CTCACTTACA CTCACTTACA CTCACTTACA CTCACTTACA CTCACTTACA CTCACTTACA CTCACTTACA CTCACTTACA CACAACAATC ACCAGCATTT GTGTCACCG GCCCAGTTAT ATGATGACAG AGAGGATGACA CACAATCTAA CACAATCTACA CACATCACACACA

Source: Two different PCR products (1..5352 and 4568..6207), both amplified from independent genomic DNA templates, and phage Eba-hb ph10 (5433..8200). Alignment with cDNA sequences: Positions 1..66 correspond to parts of the first exon of the *Episyrphus hunchback* P1 transcript (SEQ04), positions 3568..3606 correspond to the first exon of the P2 transcript (SEQ05), and positions 4561..4685 correspond to the first exon of the P3 transcript (SEQ06). Positions 5313..7625 are presumably common to the second exon of all three *Episyrphus hunchback* transcripts. Three putative polyadenylation signals (Birnstiel *et al.*, 1985) were identified in the genomic sequence (7582..7587, 7744..7749, 7768..7773), and two putative *nanos* response element (NRE) sequences (Wharton and Struhl, 1991) were identified (7547..7578, 7626..7657).

### SEQ 08 Megaselia hunchback, cDNA, partial P2 transcript.

1	AAAGGTTGTA	GAACCAAGTC	AGTTGAAGCA	GAGAAATCGA	AGAGATAGAT	АТАСААСААА	AATCAAAATG	CAGAATTGGG	AATCATTACA	ACAAACAGCT
101	TCGTATGAGC	ATAATTGGTA	CGGAAATATG	TTTCCAGCCA	CACAAATCAA	AACAGAGCCT	CTGGAGCCAT	CCAGTCAACC	ATCGCAATTG	GAACAGTATC
201	TCACATCGAT	GAAACAACAA	CAGCAACACA	CCAACGAAAT	GAATTCAATG	ACTCCATCAC	CAAGAGGTGA	GAACGAAACA	CAAAGTTTCT	TCGGAAACGG
301	TAGCACTCAG	TTGGGCTTCA	ATCCTTTAAC	CCCACCTGGT	CTACCCAGTG	CAGTCTTACC	ACCAATTTCA	CATTTCCATC	ACGCTATGCA	AAGTCAATTG
401	GCAGCCTCCG	CCAATAACAC	ACCCACTCCA	ACTAGTACTC	CTCCTATGGA	TGTTACCCCA	CCGAAGTCCC	CAAGTTTCCT	GATGGACACC	TCTGCTAAAG
501	ACTCAAACAC	CGATCACGAA	ATGATGTCAA	ATTCAAGTGA	AGATGGTAAG	GATCTCTTAG	AAAGTGAAGA	CGATGAAGCA	ATCAACATGC	CAATCTACAA
601	CTCTCATGGT	AAAATGAAGA	ATTACAAGTG	CAAAAGCTGT	GGTTTCACTG	CTATTACAAA	AGTGTCTTTC	TGGACCCATA	TGCGATCTCA	CATKAAACCA
701	RAAAAGGTGC	TCCAATGCCC	AAAATGCCCA	TTTGTCRCCG	AACTAAAACA	CCATTTGGAA	TATCACATTC	GCAAACACAA	GAACATCAAA	CCTTTCCAAT
801	GTGATAAGTG	CAACTATAGT	TGTGTAAACA	AGTCCATGCT	GAACTCACAC	AGGAAATCTC	ATTCCTCTGT	ATA		

Source: 5' RACE product, amplified from an embryonic cDNA template. At position 422, a cytidine (C) has been added to the sequence. The sequence of this particular 5' RACE product lacks this cytidine, resulting in a frameshift and a premature stop of the ORF compared to the sequences of three independently cloned, though slightly shorter putative zygotic 5' RACE products. The additional cytidine was also found in the genomic clone (see SEQ09, 2300). Therefore, the cytidine has been included in this sequence.

#### SEQ09 Megaselia hunchback, genomic.

1	AAATTATGAA	ATTTGAAAGA	ATTCAGTTAT	ТААААСТААА	GAAACTACGT	TTTTATAATG	AACTAAGAAA	AATTGTTATG	ACATAAATTG	TAAATTTTTA
101	AAATAAAAT	GTTGGCTGTA	AAATATTTCT	TAGTTTAATT	AAATTTTGTT	ТААСТАААТА	TTATTAATGA	TAATAATTTG	CTCAAATTTA	TTTGGTTTTG
201	ATAATGCTGT	CTTAACTATG	TTCATATATT	TGTCAAATTC	TTCAAGTCTT	TATAAAAGTT	TACGGATGGA	CAAAATGTTG	TATTGTAAAA	TAATGAAGAT
301	TCCCAACAAA	ACCATCTTTC	ACTTTTTGTT	TTTTCACTTT	CAATAAATAA	TGTTGCGTTA	AAAGAAGAAC	CCCTTAAGGT	CATTCAAATT	GACCGAGAAG
401	ACCATGTTGA	TGGCCGAACA	TAACTAGAGG	TCAAGGCGAA	GAAAACAAAT	CCATCGTCTT	TAATGAAAAC	AAGAAACCGA	AACCGAAGGA	AATATAAATT
501	CGTGTGAATT	ATTCAAATCA	TCTTGCAAAT	ATAAACAAAT	AGTGATTAAA	TGCAATAAAA	ACCAACACTA	CTACAAAACT	ACCTACCCTA	TCTACCGACC
601	CTCCACCACA	TCAATGTATG	TGTGTCACAA	GGACCACAAA	TCTCTCTCTC	TCTCTATTTT	GTCTAGCCGA	AGATTGCTTT	GCCTGCCTCT	CCTACAGCCA
701	TGTAATGAGA	GAATCTTTCA	ATGTATCTGT	GTAAAAACTA	TATGAGAAAA	AAATTGGAAA	AATGTATATA	ATTTTTTGTA	AAGTGATTTT	TTTGTACAAA
801	ATCAAATCAG	TCGTGCTCGC	AACTTAACAC	GAGAAAGGCA	AAGGATGTTA	GCTGCTCGTG	CCGTGTAGAT	CACGAAATTA	TTACAATATT	TTTTACAAAA
901	AGAAAAAAAC	AGTGACAAAA	ATTTTAAGAA	TTCAATAAAC	ААААААТААА	AAGAGAAAAA	AACAAATAAA	GTGACAAAAA	ATCAACCGTT	ATTTCCAAAA
1001	AAATTGGATT	TAAATTTACG	AACAATTTTC	GCTAAAAAGG	TAAGAATGGA	TTAACTTTTA	AGCCAAATGG	<u>ATT</u> AGTTATT	ACTGTTTCAT	ATAAGTAGTT
1101	AGGAAAGGTT	GTAGAACCAA	GTCAGTTGAA	GCAGAGAAAT	CGAAGAGATA	GGTAAGCGAA	TAAGTCAGTC	CGTCTGTCTG	TCTAATCAGC	TGATGCACAG
1201	ATAAAGTGTT	GAAATTGATA	GACCGTCATC	CTTCAAAAAG	ACAAAGAAGA	ATTTTTTCA	AGAATAAAAA	TTATTTCAAA	GTCAAAGAAC	ACGCAAAAAA
1301	TAACAAAACA	AAAATATTAT	TTAGAAAAAC	TTGGCAAAAC	AGTTTTTCTT	TTTTCATTTT	TCAATTTTCA	TTGCATTTGA	CTATGCAAAA	GAATATGACG
1401	AACAGTTTTC	TTTCTTTCTC	ATTTTCATTT	TTTTTTGTTT	ААААААТААА	TTTCTAATAA	AAATAGAAAA	ATGTATATTT	TTCTTTGCCG	TTTTTTCTTTT
1501	TTGTTCGCCA	ATTTTCTTTA	TATGCACTTT	ATAGTGCATT	TTGCTGCTGC	ATCTGTTTAT	CTCGGCATTC	AGGCAAACAA	CAATTTTTAT	TTTATTTTCT
1601	TTTTTTTTG	TTTCTTGCTT	TGCACATTTT	TATGAAAATT	TATAGTTTTT	ATTTATAGTA	GTCGGGCATG	GTGTGTATTG	TATATACGTA	TAGTTTTTGA
1701	ATAGTCCCAA	GAACTTGACC	GGAAAGCGGG	CGGGCGTCTA	GATATTACGC	AGTGCATTTC	AATTTTTTCA	TAGACTTTTT	GTTTTTGATA	ACCTTGAATA
1801	GTTTATTAGA	AATTCTAAAT	AAAAAGGATA	AGGGATGTTT	GAGTTTACGG	ACACGCGTAC	AACTTGTGCC	TGAAGAAAAC	AGAAATCAAA	ATGGCGTTTT
1901	ACTAAAAATC	TCTTTCTTCC	TTACAGATAT	ACAACAAAAA	TCAAAATGCA	GAATTGGGAA	TCATTACAAC	AAACAGCTTC	GTATGAGCAT	AATTGGTACG
2001	GAAATATGTT	TCCAGCCACA	CAAATCAAAA	CAGAGCCTCT	GGAGCCATCC	AGTCAACCAT	CGCAATTGGA	ACAGTATCTC	ACATCGATGA	AACAACAACA
2101	GCAACACACC	AACGAAATGA	ATTCAATGAC	TCCATCACCA	AGAGGTGAGA	ACGAAACACA	AAGTTTCTTC	GGAAACGGTA	GCACTCAGTT	GGGCTTCAAT
2201	CCTTTAACCC	CACCTGGTCT	ACCCAGTGCA	GTCTTACCAC	CAATTTCACA	TTTCCATCAC	GCTATGCAAA	GTCAATTGGC	AGCCTCCGCC	AATAACACAC
2301	CCACTCCAAC	TAGTACTCCT	CCTATGGATG	TTACCCCACC	GAAGTCCCCA	AGTTTCCTGA	TGGACACCTC	TGCTAAAGAC	TCAAACACCG	ATCACGAAAT
2401	GATGTCAAAT	TCAAGTGAAG	ATGGTAAGGA	TCTCTTAGAA	AGTGAAGACG	ATGAAGCAAT	CAACATGCCA	ATCTACAACT	CTCATGGTAA	AATGAAGAAT
2501	TACAAGTGCA	AAAGCTGTGG	TTTCACTGCT	ATTACAAAAG	TGTCTTTCTG	GACCCATATG	CGATCTCACA	TGAAACCAGA	AAAGGTGCTC	CAATGCCCAA
2601	AATGCCCATT	TGTCACCGAA	CTAAAACACC	ATTTGGAATA	TCACATTCGC	AAACACAAGA	ACATCAAACC	TTTCCAATGT	GATAAGTGCA	ACTATAGTTG
2701	TGTAAACAAG	TCCATGCTGA	ACTCACACAG	GAAATCTCAT	TCCTCTGTAT	ACCAATACCG	TTGTGCTGAT	TGTGATTACG	CCACTAAGTA	TTGCCATTCA
2801	TTCAAATTGC	ATCTCAGAAA	GTATGACCAC	AAACCAGGCA	TGGTTTTAGA	TGAAGAGGGT	ATCCCAAACC	CATCAATCGT	TATTGATGTT	TACGGAACCC
2901	GCCGTGGCCC	AAAGATGAAG	GGAGGTATAA	GCACACCATC	AGTTTCCCAT	AGGAGAATTG	TGCCTGATCA	AAAACCAAGT	TTGTCCGATT	TGAAAATACC
3001	CTTCTCACAT	TTGCCAACCT	CCCCCGCTAA	AAGTACAACT	TCATCCAACT	CGGAATACAA	CACCCCACCA	AGTTCCGCCA	ATCAAATGAA	GCCCAATGGA
3101	CAAATCTCAA	ACCTCCTCCC	ACCTTTGGTT	CAGAGTATGC	TTCAGCAACA	ACAACAAATG	AGCGGTTTCT	TCCCCTACTG	GAACTTGAAC	CTCCAAATGC
3201	TTGCTGCCCA	ACAACAACTA	GCTCAATTGT	CGCCAAGTAT	GAGAGAAAGT	CTTCAACATC	AACAACAACG	TTTTGACAAG	GATTCCAGCC	GCGAATTCGA
3301	TGTTTACGAA	GACGAGGAGG	AAGAAGACGA	ACACGACCAG	AAAGAAGAGC	ATGTCGCTGC	TGCCATCGAT	TTATCCGCCC	AAGCTTCTAC	ACCTATCAAA
3401	GACGAAGAAG	AAGCTAAGGA	GGAAGAAACC	AGCAGTAACA	CTCCCACCGT	CAGCACAACC	TTCATTTCAA	GAAGGAAAGG	ACGTGTCCTC	AAACTAGATA
3501	CTACAACCAA	CACTCAAAGT	CAAGTTGATG	AAGACTCTGA	CCGTCAATCC	CCATCCTCTT	CCTTTGAAGA	ACCAAAAGAA	ACAGCTGCCA	CGTCGACCCC
3601	TAGTCCAGCC	CCAGCCCCAG	CATCTCCACC	CAGTTCAAAC	CTCTTCGAGT	GCAAATATTG	TGATATTTTC	TTCAAGGACG	CTGTCCTCTA	CACCATTCAC
3701	ATGGGCTACC	ACAGTTGTGA	TGATGTATTC	AAGTGCAACA	TGTGCGGGGA	GAAGTGCGAA	GGACCCGTCG	GACTATTTGT	GCATATGGCA	AGAAATGCTC
3801	ACTCGTAAAC	TTCTCTTAAA	GATAAGTTTT	CTGAAAACGT	TAATTTTGTA	ААААААААА	АСААААААА	ATAGTTTGTA	AGGCCAAAAA	CTGTTGTCAA
3901	AGATTGTACA	TATGAATTTA	GACTGTAAGT	GTTTTTTTCTT	TTAATTTAAG	CTATAGTATT	TTTTAAGAAA	CGCAAAACAA	ACAAACCAAA	AATTGTCTTA
4001	ATTTTAATTA	AAAAACTTTT	TTAATATTGT	AAAAATTACA	AAATATTTTT	TCCTTATTTC	TAAAAATAT	AATTTTTCCT	TGCGTGTAAA	ATAATTTATT

4101	TCGTTTTGAA	AAGAAATTTT	CTAGTCAATT	TTTAGTTTTT	TAGAATAAAT	TTTTATTTAG	TTTAAAGAAA	ATATACTCAA	CAAATTCGAC	ААТАСАТААА
4201	TGGTACAAAA	CTGATGAAAT	GTACAGTAGA	ATTAAAGAAG	AAATGGGTGG	TCGAGGCTTT	TCGAGCGTCA	CGAATCAAAA	TGAATACTTT	TTATTTTAA
4301	GCTAATTTTA	TTTTTCATTT	TTTACTAAGA	ATAAAATGTT	ATTTTTCAAA	ATTATTTCAA	TTGTTTTTTG	GGGGGTATAT	ATTTTTTTTT	AGGTTTTTAG
4401	GGGTTTTTTT	TAAATATTGT	TATATTTCGA	GAAAAATTT	TGAGATATAT	TTTTAAAATT	TTAAAATTTC	ATTAAAATTA	GGTTTTGATT	ААААААТССА
4501	САААААААА	ATGGATTTAA	TAGATACAAC	TAATTTTTAA	AATTTAAAGC	TTGAAAAAAA	AATATTTGTA	AAAAAGTTTT	GAAATTGAAT	CTTGATTAAA
4601	GTTATAGATT	TAAAGTTTAT	TTATTTATAA	CCTTGAATAT	GTTATTAATT	TTTAGGAATT	TTTCTAAAAA	ATATATATTT	TTTTTATAAA	AAAAACAGAT
4701	ATAGAACAGA	ТААААТАААТ	TCGAACTTGA	TTAAAGTTGC	AGGATAGGTT	CAATGAAGTC	TGGAATAGCA	GATAGGCTTT	TTTCGTAATG	TTACGTTACG
4801	GTTAATGTTA	CGGTTTTGAT	GAAATTACGT	AACGTAGTTA	AGATTTTCAT	AGACTTCTAA	GTTTTTTCCA	AGTAATTCTG	GGGTTGAAGG	CCAGGGAAA

Source: Genomic DNA of phage Mab-hb ph2a, which was partially sequenced in a subcloned *SpeI* (1..2309) and an *XbaI* (1738..4899) fragment. Alignment with cDNA sequences: Positions 823..1039 correspond to the first exon of the *Megaselia hunchback* P1 transcript (Stauber *et al.*, 2000), and positions 1104..1151 correspond to the first exon of the P2 transcript (SEQ07). Positions 1927.. 3861 are presumably common to the second exon of both *Megaselia hunchback* transcripts. Three putative polyadenylation signals were identified in the genomic sequence (4144..4149, 4330..4336, 4714..4719) and one putative NRE sequence (3891..3912). Bicoid-binding sites identified by DNaseI footprinting are underlined (Philip Shaw, supplemental Figure S3; 1041..1044, 1047..1059, 1063..1073).

#### SEQ10 Platypeza hunchback, cDNA, P1 transcript.

1	AGTTTAGTTC	AGATTTGTGT	TTCTGTGAAT	ATTTATTTAT	TCGTTTTGTT	TGTGCGGTTG	GACAAGTTTT	CTTTTTCAAA	GATCAAAATG	CGATTTGTTT
101	TTGCTTTAGA	TCTCTAACAA	AATGCAAAAT	TGGGACGCAC	TTCAGCCAGC	TAGCTACGAG	CACAATTGGT	ACAGCAACAT	GTTCCAAAAT	ATTAAACAAG
201	AGCCGCAGAG	CCAACCCACC	TCCCAACTGG	AGCAATATCT	TACAATGAAA	TCGCAACAGC	ATCAGCAACA	CCAGCAACAA	CAACAACAGC	AACAGCATCA
301	TCATCATCAT	CAACAACAGC	AATCGCAAAA	CGTTGATATG	AATTCGCTAA	CACCTTCGCC	GAGAGCGGAT	AACACAGATG	GACAAAGTTT	CTTCGATCAT
401	ATGCCACATC	CGCTAAGCGG	TTTCAATCCG	CTAACCCCAC	CGGGTCTGCC	GAATGCCGTC	TTGCCGTCCA	TGTCGCACTT	CCTTGCCACC	AGTCCGCCAG
501	AGAATTTGAA	TGCGCAAACG	CAATCGCTGA	CGCCACGCAA	CACACCGCCA	ATGGATGTGA	CGCCGCCAAA	ATCGCCGAAA	CCGGAGTTTT	CCATGTTTAT
601	GGATAAAGAG	CAAGATTTGA	TTTCCAACTC	CAGCGATGAT	ACGAAATTTT	TGGAAAGCGA	AGACGACGAG	AACATTCGGA	TGCCGATTTA	CAATTCGCAT
701	GGCAAAATGA	AGAGCTACAA	ATGCAAGAGC	TGCGGATTAA	CAGCTATTAC	GAAAATCGGC	TTCTGGCAGC	ACGCTCGCAC	TCATATGAAG	CCCGAGAAGA
801	TCTTGCAGTG	CCCCAAGTGC	CCATTTGTCA	CCGAGTTGAA	GCACCACTTG	GAGTACCACA	TCCGCAAGCA	CAAGAACTTG	AAGCCATTCC	AATGCGACAA
901	GTGCAGTTAC	AGCTGCGTCA	ACAAATCGAT	GCTCAACTCG	CATCGCAAGT	CCCATTCATC	GGTCTACCAG	TACCGTTGCT	CCGATTGCGA	TTACGCAACG
1001	AAATACTGCC	ATTCGTTCAA	GTTGCATTTG	CGCAAGTATG	ACCACAAGCC	TGGCATGGTT	TTGGATGAGG	AGGGTCTGCC	CAATCCCTCG	ATTGTCATCG
1101	ATGTGTACGG	TACACGTCGT	GGCCCGAAAG	TGAAGAACGC	CAACAACAAG	GCAAACAAAG	CGGCTGCCAT	CAAGTCTGAA	ATGAAAATTC	CACAACATCC
1201	CAACCATCAT	CAGCTGCCAG	CCTCGCCTGC	CAAGAGCACC	ACATCTTCAT	CGTCTGACCA	CCCCAACCAA	CAACAAATGT	CGCAACAAAC	GCCCCAAATG
1301	GCATTGGCAT	CGATCCTCCA	ACAAGGCCAC	AACATGCCCT	CATTCTTCCC	CTACTGGAAT	CTCAATCTGC	AAATGTTGGC	TGCCCAGCAA	CAAGTGTTGG
1401	CGCAAATGTC	GCCACGTATG	CGGGAAGCCA	CCCTCCAAAA	TTTGCATGGC	GGACAAAGCA	ACGACGACAA	TAACGCTGAA	GACAACCACA	GCTTCGAGGA
1501	TGAGGACAAC	TTTGACCAAA	AATCGGAAGG	CAGCGCAATG	GACTTGTCCC	AAGGCTCCCC	ATTGAAAAAC	GAATCGTCAC	CCCCTGTCCT	ACCATACAAC
1601	CTATTGAAAA	TGACCGAAGA	GGAAGTCAAC	ACCCCCACAA	TCAGTTCGTC	AAGCAGCTCC	CGGCGAAAGG	GCCGTGTCCT	CAAGCTAGAC	ACATCCACCC
1701	AACATCTCCC	AGTCGCCGAG	GAAATCGCCG	TGCCCGAGCC	AATCCGCAGC	ACTGAGTCCC	CCTCGTCGTC	GTCTTTCGAA	GAGCCAAAAA	TGGTGCAATC
1801	CCCCGCACCG	GTCGCAGCAC	CAGTTGCTGT	TGCCCCAATC	ACCACACCCT	CCCCGCCCGC	CGTAGTCCCG	TCGAACAGTA	ACATCTTCGA	GTGCAAGTAC
1901	TGTGATATTT	ACTTCAAAGA	CGCGGTCCTC	TACACCATCC	ACATGGGCTA	CCACAGTTGC	GACGATGTGT	TCAAGTGCAA	CATGTGCGGG	GAGAAATGCG
2001	ACAGCCCTGT	CGGACTCTTC	GTCCACATGG	CCCGCAATCC	ACACTCATAA	TTCTATTACC	CGTTATTTGT	TTTATTATTA	TTTGTTGTCA	AAGATTGTAC
2101	ATACGA									

Source: 5' RACE product (1..893), 3' RACE product (1457..2103) and an additional PCR product (864..1704), all amplified from a larval cDNA template. At positions 301..303, an additional TCA repeat has been added to the sequence. The sequence of this particular 5' RACE product lacks this repeat compared with two independent, though shorter 5' RACE products. The additional TCA repeat was also found in the genomic clone (see SEQ11, 6594..6596). Therefore, the TCA repeat has been included in this sequence.

### SEQ11 Platypeza hunchback, cDNA, partial P2 transcript.

1 AGTCGTGTCT GTAACCGTAC AAGAAGAAGG ATGTGTACTG CCGTCGTGTA GATCACGTGA ACAAAAAAAG AAAATAAATA TTATAATTT TAATTGAAAT 101 TTTAAAAAT CAAAAAAAA AGAAAAGGA TTTTGTGTGT TTTTAATAAC TGTTATGTGC TAAACGAGG AAATATAAA ACAAATATTA TTATAAATTT 201 TTTCCCACCA AAACTAACGG ATTATTGAT AGAGCAATTT ATTGGTGTT CCAGGGAAGG AGTTAATTGG GATCTCTAAC AAAATGCAAA ATTGGGACGC 301 ACTTCAGCCA GATACTGG AGCACAATTG GTACAGCAAC ATGTTCCAAA ATATTAAAAA AGAGCGCAG AGCCAACCCA CCTCCCAACA GGAGCAATAT 401 CTTACAAATGA AATCGCAACA GTATCAGCAA CACCAGCAAC AACAACAACA GCAACCACA CATCATCAACA ACAACAACA

Source: 5' RACE product, amplified from a larval cDNA template.

#### SEQ12 Platypeza hunchback, genomic.

1	CTGATTCGAG	TTATGACATC	CAAATCAGTT	CCACCGGGTG	CGGCGATTTC	GCTGCCTAGG	TATGTGAACC	TTTCCACCGT	CACAATTTTG	TTGTCCAGCG
101	ATTCGGGTTT	CCTTCTAGGC	TTAAAACCTT	AGTCTTTTTA	AAGTTGATCT	TTAACCGGAA	GACTTCTTCT	CCTCAAACTC	TAGCGATTGG	GCCACGTTGG
201	ATAAATCGAA	TATCCGGTGT	GCTATGGAGG	CATATGTCGT	CGGCATAATC	CAGGTGTTGC	ATCAGGAAGC	TACATATAGT	TTATATAAAT	AATAATTTTT
301	GTAATAAAAG	ACGAAATTTC	AGTTAGCTTC	TCAGATATTT	AAATTAAAAT	TTTATTTAAA	AAAAAGAAAG	AATAAAGTTA	AAGTAAGCAA	ACTGTTTAGT
401	TAATCTCAGT	TAATAAGCAT	CGATATAATG	AGTTCAGGTG	CTTGATTGAT	ATTCAATTTG	AAATGTTATT	CCGGGATTTG	TTCAAATTGG	CTTATTTCTG
501	TTTTCTCTTT	TTTCCACTCT	AAGACTCTCT	TGGTGTCCAT	ATTAAAATAA	ATGATTGTGA	AATGTTAATC	ATTTGTTCTA	ATTTTGTATG	AGTTACGATT
601	TATAATTTTT	TTCCCAAATG	TTACATCTTT	AAAAATGTA	TATCATAGTT	ATTACAAATA	TGCATTAACA	ATGTAAATTA	ACGAGTTGCT	GAAAAGAGAA
701	CTTTTTAATT	ACAATCCTAT	GTATATACGA	AATAAAGGAA	GTCTCAAGTA	TTACAAGAAT	GATTGATAAA	AGAAGATAAA	AACAATTCAA	CCAAAATATC
801	TAAAACGGAA	TCTTCCTCAT	TGACTTCTCA	TTACACAAGA	AAAAAAAAAC	GTTTCCTTTC	TAGAACAGTA	CCTAAATGTG	TCCTTAACTC	TAATCCGTAC
901	ATCTTAACTA	CTTTAATCAC	CACACAAATG	TCCAGTCCGT	CCAGTCGCAT	GCAGACCTCA	TCAACCATCA	CAAAAACAGA	CACAATGGAG	TTCCAACCTC

TOOT	TTTAACACTT	TTGTTCTGTT	GTCTGTTGTC	TGCTGGTCTT	TTCCATATCT	CTGCTAATCG	TCAATGAATA	TTCGTTTCTG	GGCTTATTAC	AAACACATTA
1101	CCATGACTAT	GCAAATTTGT	TGTGTCGAAG	GACCTTCTTT	TGGGTGTGTG	TACATTCTCT	AACCGCATAT	GTCGGACGAA	TGTGCGATTT	TGATTGAAGA
1201	AACCCACAGG	AATTTTTGAT	TTTTGTAAAT	ATTTACCGAA	GCGAACGGCA	GAAGGCGCAA	TCGGTAACCG	CTACTGGATT	TGCGATTTTC	TTATCTGGGT
1301	CTATTGTTGT	TTGAAAGGAG	AAAGGAGGTA	GATGTAAGGA	CATGAGGTCA	ACTGATGTGA	TGACGTTGCT	ATTATTACGA	TGGGTCAAGG	GGGACGATCC
1401	CCTTACGATG	AAAGATTTTA	TATATTATGT	AAATTAATTA	AAAGGAAATT	TCTTTAAATG	GACATTATTG	GGTCAAGGAG	TGTTTTGAAA	ACAATTGACG
1501	TGCGCCAAGG	TCAACTGACT	ATTGGGCAAT	TCTTGAAAAT	AAATAAATAT	TTGTCGATTT	ATTAATTGTA	CATTATCGCA	TATCAATAGA	AAACAGAAAA
1701	ATACCGCACA	TGTGAGGCCC	CATGCGCATT	CACAGCAGAA	AAACAAATCT	TTCATATCTT	TTTTTTAAAAT	CTCAATTGCG	ATTAAACGAT	TTGCGATTTA
1901	GGCTATGCAA	AAAAA	CCUTCANC	TCAATCATTC	ACTITAAAAT	TATCTTTAAT	CGCGATTAAG	TCCTTTTTGT	GTTAGCTTAT	ACATAATTAA
1901	CCTACAAAAA	ПОССССААС ТО	TATTA	CCTTCACTT	CCTTATIAG	AATACCTACC	CIIGIGIGGI	ACACTCCTCA	CACACCATCA	AAAATCACAA
2001	AGAAAAGTGT	AATAAAAATG		TTTTTAACCTT	ΔΔGΔΔΤΔΔΔΤ	АПАССІАСС	ТСССААТСАТ	AUGATCCACA	TATAAAACGC	ттаааатат
2101	ТАААААСТАА	TATCGGCATT	TTTACTAACA	AATTATAAAG	TTATGGGGGA	AAGACATTTA	АААТАТАТТТ	тттаттсата	ТАССТААААТ	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
2201	TTCATTTTTT	AGCAAACATC	AGAGTAATAT	TGGAAATTGT	ATATTACCAT	TTTTTTTAAA	ATATACCACG	AACCATATAA	CGAACCAAAA	AATATAAGAT
2301	TAGATTTAAA	ACTAACTGCA	TCAGTGATTA	AATGGCTAAT	TATTATATAG	TATTAACTAC	AGCCCAGGAC	CCCGATTAGA	TAGAGGAACT	CGTAAGACGA
2401	ΑΤΤΤΑΑΑΑΑΑ	AAGAAACTAA	GCACATGTCC	TTCTATTAGT	AGCTTTGGGA	CAAAATAGAT	GTTTGATCCG	AATCGGCGGA	AAAGCAGATT	TATGTATAAT
2501	TTAAAATTGG	ATAGCTCTTA	TATTTAAATT	ТАААААААА	TCCGCTTATT	АТАААТАААТ	TATTAGATTT	TTTTTTAAAT	TTAAATATAA	GAGCTATGTT
2601	TAACTAAATA	TATATTTTAA	TAAAGACGTA	ATTGACTGTT	TGTTGTGTTA	TGCCAAAATA	ТАТААААТАА	GTAAAAGAGT	CTGTGAGAAA	ACGACAGCAG
2701	TCGATATTCC	AGCTCATATG	AGTTCCTGGT	ACAAACGAGA	TTCTTTTTCC	TTCATTTTTT	TTCTTCAATA	ACAATCAGTC	AGTAAAACGA	TTCCTTGAAA
2801	TAAAAATGTA	TAAAAAACAT	ACATCGAAGG	ATGTGTTTAC	ATCACATTGA	TGGAGCAGAC	TAGACGGCTG	TGTGTGGATT	AATGGTATGG	GGAGATAAAT
2901	GAAAGCAATA	AGCTTCGCCA	TTTTTTTTTA	TTAATCGAGG	GGTTTTATTT	AAAATAAACG	ATTTGAAATT	CTTCGAAGTC	AACATATGAA	CGTAGGCAAA
3001	ATCGTGAAAA	ATATGACAAT	TTACCGTTAT	ATAAATATTT	TATAGTAATG	GTTTTTTATA	TTTTAGAATT	TAGAAGAACA	AAAAAACATT	TAAAAATGTT
3101	AAATGTTAAA	AAATATATAT	TTTATATAAT	AGTTTTTTT	TGTTGAGATA	ATAGAATAAA	CAAAATTTTT	ATAAGAACAA	AAATAAATAA	AATATTTTAA
3201	AAATCAAGGA	AAATATTCAA		TATAAAAAAT	ATATCCCTAC	AATAAATTTT		AATAATACAA	TGAAAAAATA	TTACTATTTT
3401	ACCONTRA	TTTTTTTCCAT TTTTTCCAT	CATTITITITAA	CACCTCACCA	TATAIGCAGT	ATATATAT	AATTAAGAAT	MTAGTTTTCG	CTATTAACAGCA	AATATATATT
3501	TATTCTTACT	AAATAACATC		TTATATAT	TGACIGGIAI	СТАЗАТАТА			CATTALAACT	TTCTATTCT
3601	ATCCCTATT	TTAATACAIC	AAIAIIIAAA	CTAAAACATA	CAAAATATAA	A ATTCT A ACT	AAACAGAAAA	ACCCATAAAA	адтатасста	TAATAAAG
3701	ΔͲͲΔΔΔΔͲͲϚ	AAAAATATGG	AAAGGTCTGT	TGACGGATTG	GCAGACAGAT	ΔΔΠΔΓΔΔΠΠΠ	TAATAACCCC	ATCGACTACC	ͲͲͲͲͲϾΔΔΔΔ	THEATCACC
3801	ATAGTTAGAA	ACCTGCCGAT	ТАТТТААТСА	GAGCAATATT	CCTAATCTGG	GTACAAATAT	AAATATTAGA	TTACAAACAG	TTTAGAAGGC	AGTCAAATCG
3901	CATTCAGGCA	GAAAGGTCAA	CGGATAAAAA	TAAAATATAC	AATTTGACAG	AAACAGAGGA	AAGGATAAAA	ATATATCCCC	TTTTTTTGTTT	ATATTATTT
4001	CGATGATGTC	TACCTATCAC	TATATATTTG	CTAGCCAGCA	TCGTTAAGGA	TATGTAAAAA	GGATATATGT	AGGTATAACG	TCATTCATTC	ААААААТСАА
4101	CAAAAATCCC	AAAAGAAAAT	AAACCCAAAT	CGATAACCGA	AATGAATTCC	GTTCATCTAT	ACACAAACAC	ATGGAAAACC	CACAAAAAAA	АТАТААТААА
4201	АТААТАААТ	AAAAAGAAAC	AGAAAGTACA	ATTTCCTTGT	AACAACGAAC	GGGCATCCCT	TTCTGCAGCA	TACAAACCGG	CAACTGGCCA	TCTCTGTCCC
4301	AACCGTCTGC	AGCCGATAAT	GTACGGGAAA	ACTGTCATGT	AATTTTTCGA	CGAGATGAGA	ATTTTCGGTT	TTTGAAATCA	GTCGTGTCTG	TAACCGTACA
4401	AGAAGAAGGA	TGTGTACTGC	TGTCGTGTAG	ATCACGTGAA	CAAAAAAAGA	AAATAAATAT	TATAATTTTT	AATTGAAATT	TTTAAAATTC	ААААААААА
4501	GAAAAGCGAT	TTTGTGTGTT	TTTAATAACT	GTTATGTGCT	AAACGAGTGA	AATATACAAA	CAAATATTAT	TATAAATTTT	TTCCCACACA	AACTAACGGA
4601	TTATTTGATA	GAGCAATTTA	TTTGCTTTTC	CAGAGAACGA	GTTAATTGAG	GTAAGCATAT	ATTTTTAATT	TTTTTTTTAA	TTCCAATACG	ATCAGGTCAC
4701	CCTGCTTATT	TTTCCTTCCT	TACCTTTACT	TAAACAAATC	CTTAACATCA	AACCTGCAAA	TATCCCTAAA	ACGTACCTGC	CCGTCCTAAA	ACGAGTTGAT
4801	TAAAAATCAA	CAAAAACAAT	TTCGAAAAAA	TCGTTGGATT	GATTGACTTA	GTCTGAGATA	ACCCCCGGCC	ATCAGACATG	TCCTCATTTT	GTTGCAAATC
4901	CGCGTATCAA	TTCGGGATTA	GGCGCTGGGA	CAGATACCTG	CAMMINAAGTGC	TATATATICC	GCCGATATTT	TGATTTTGAT	ATCAGTCGTG	AAATAGTATT
5101	TTCTTTCCATAGI		TIGCAGACAA	AATCCAACCC	CALITAATT		CTCAATCCC	TITITGGIGA	ATATTATT	AAIAAIIAGI
5201	CCAGATGCAC	ALILITAATA	AAAAC	AAICCAAGGC	TTCCCTCCAT	DADTATCADA	AATACACCCA	CCCCCCCCCCC	TATALITACAL	ACAIAIAAAAA
5301	ТАТАТАСААА	AATATACCCC	CTTTGTCCAA	ACCGTACAAA	CCGTCCAAAC	GGTATTCGA	GCACACACGC	СССССССАСТ	ТІПАСАТІІС	ΔΔΠΔΔΔΔΔΔΔ
5401	TGCATAAATA		CGCGAAAAAG	CAAAAACGAA	GATCGTTCCG	GTCAAGTTCT	TGGTTTTTTT	CCTCACTCGT	1111 110111001	
5501	maaaamaama	AAACIAIIII							TTATTTAATT	TTTTTGTTGT
5601	IGCACICATA	TGAGCATCAG	ATATTTTGTG	CATTAGACAC	ATTATTTTAT	TTATATACAT	ATTCATTTTT	CGATATTTTT	TTATTTAATT AAAAAATGAT	TTTTTGTTGT GCATGTGCAA
5701	AATAGTGCAA	TGAGCATCAG TCTCAGTGCA	ATATTTTGTG ACAACAATGC	CATTAGACAC ACACAAAAAT	ATTATTTTAT ATCAACAACA	TTATATACAT CAAAAAAAAA	ATTCATTTTT ATATTTAGAA	CGATATTTTT AACATTTGCC	TTATTTAATT AAAAAATGAT AAGTTTTTTA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA
0,01	AATAGTGCAA ATAATTGGCC	TGAGCATCAG TCTCAGTGCA TGTGTCTATA	ATATTTTGTG ACAACAATGC TTTGTAATTT	CATTAGACAC ACACAAAAAT TCAAGTATTT	ATTATTTTAT ATCAACAACA TAAAAATTTA	TTATATACAT CAAAAAAAAA AAAAAAAATT	ATTCATTTTT ATATTTAGAA AACAAAAAAA	CGATATTTTT AACATTTGCC GAAAGGATTT	TTATTTAATT AAAAAATGAT AAGTTTTTTA GCGATAGTAA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA
5801	AATAGTGCAA ATAATTGGCC GGTATATATG	TGAGCATCAG TCTCAGTGCA TGTGTCTATA CAAATTATGA	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA	CATTAGACAC ACACAAAAAT TCAAGTATTT AAGGTTGTTT	ATTATTTTAT ATCAACAACA TAAAAATTTA TTGCGAAATT	TTATATACAT CAAAAAAAAA AAAAAAAATT TGGGTAATAA	ATTCATTTT ATATTTAGAA AACAAAAAAA AAGAATTACT	CGATATTTTT AACATTTGCC GAAAGGATTT CCAGATTTGG	TTATTTAATT AAAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTT	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAAAAGGC
5801 5901	AATAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT	TGAGCATCAG TCTCAGTGCA TGTGTCTATA CAAATTATGA CATTTATAAC	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC	CATTAGACAC ACACAAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA	ATTATTTTAT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGTA	TTATATACAT CAAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT	АТТСАТТТТТ АТАТТТАGAA ААСААААААА ААGAATTACT ТАССТААААА	CGATATTTTT AACATTTGCC GAAAGGATTT CCAGATTTGG CAGGCACAAC	TTATTTAATT AAAAAATGAT AAGTTTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTCT	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAAAAGGC TTATATACAC
5801 5901 6001	ATAATAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTTGAAC	TGAGCATCAG TCTCAGTGCA TGTGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA	CATTAGACAC ACACAAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC	ATTATTTTAT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGTA TCTTGGATAT	TTATATACAT CAAAAAAAAAA AAAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA	ATTCATTTTT ATATTTAGAA AACAAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT	CGATATTTTT AACATTTGCC GAAAGGATTT CCAGATTTGG CAGGCACAAC GAGTTGAAGT	TTATTTAATT AAAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTCT TGAGATTTAA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAAT
5801 5901 6001 6101	AATAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CGCCTAAAAA	TGAGCATCAG TCTCAGTGCA TGTGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA TACGCTTATG	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC	CATTAGACAC ACACAAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA	ATTATTTTAT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGTA TCTTGGATAT AAGGGATGGT	TTATATACAT CAAAAAAAAAA AAAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATA	ATTCATTTT ATATTTAGAA AACAAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG	CGATATTTTT AACATTTGCC GAAAGGATTT CCAGATTTGG CAGGCACAAC GAGTTGAAGT TGTTGCTAAT	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTCT TGAGATTTAA GTAAAAATAC	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAAT GTAAAAATAC
5801 5901 6001 6101 6201	AATAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CGCCTAAAAA CCTTAGGGGA	TGAGCATCAG TCTCAGTGCA TGTGTCTATA CAAATTATGA CAATTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTATG	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATTA	CATTAGACAC ACACAAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT	ATTATTTAT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGTA TCTTGGATAT AAGGGATGGT TAGGGGTTGG TAGGGGTTGG	TTATATACAT CAAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA ATATAATGGAAA	ATTCATTTT ATATTTAGAA AACAAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT	CGATATTTTT AACATTTGCC GAAAGGATTT CCAGATTTGG CAGGCACAAC GAGTTGAAGT TGTTGCTAAT TGGTGCACAGC	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTCT TGAGATTTAA GTAAAAATAC CATTGCTTGT	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAT GTAAAAATAC GCCAATTTTT
5801 5901 6001 6101 6201 6301	AATAGTGCAA ATAATTGGCA GGTATATGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CGCCTAAAAA CCTTAGGGA TATTCGTTTT	TGAGCATCAG TCTCAGTGCA TCTCAGTGCA TGTGTCTATA CAAATTATGA CAATTATAGA CATTTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT	АТАТТТТСТС АСААСААТСС ТТТСТААТТТ ААСТААТТА СССТТАСТС АСТАСТАААА ТАСТСТАААС ТТАТСТАТААС ТТАТСТАТАА ТТАТТТТТТАТ	CATTAGACAC ACACAAAAAT TCAAGTATTT AAGGTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTTA	ATTATTTAAT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGAA TCTTGGATAAT AAGGGATGGT TAAGGGTTGG TTAAGTATAATA	ТТАТАТАСАТ СААААААААА ААААААААТТ ТGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATA ATATAATGGA AAACGTTTTA	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGGT CACCCCGATG GGAAATTACT AAATGGCGGA	CGATATTTTT AACATTTGCC GAAAGGATTT CCAGATTTGG CAGGCACAAC GAGTTGCAAAC TGTTGCTAAT TGGGCACGCG CTTTATTTA	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTCT TGAGATTTAA GTAAAAATAC CATTGCTTGT TAAAATTTTTT	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCAAGTTTGA TCCAAAAGGC TTATATACAC ATCCAAAAAT GTAAAAATAC GCCAATTTTT TTTAATTTCC
5801 5901 6001 6101 6201 6301 6401	AATAGTGCAA ATAATTGGCCA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CCCTTAAGGGA TATTCGTTTT AGATCTCTAA GAACCATCTAA	TGAGCATCAG TGCAGTGCA TGTGTCTATA CAAATTATGA ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATTA TTTTTTTTTT	CATTAGACAC ACACAAAAAT TCAAGTATTT AAGGTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTTA CACTTCAGCC TTCTTCACC	ATTATTTAAT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGAA TCTTGGATAAT AAGGGATGGT TAGGGATGGT TAGGGATAGT AAGTAGTAATAA	ТТАТАТАСАТ СААААААААА ААААААААТТ ТGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATA ATATAATGGA AAACGTTTTA GAGCACAATT	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT GGAAATTACT AAATGGCGGA GGTACAGCAA	CGATATTTTT AACATTTGCC GAAAGGATTT CCAGGATTTGG CAGGCACAAC GAGTTGCAAC TGTTGCTAAT TGGGCACGCG CTTTTATTTA CAGGTTCCAAC	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTTT TGAGATTTAA GTAAAATAC CATTGCTTGT TAAATTTTTT AATATTAAAC	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCATAGTTCA TCCAAAAGGC TTATATACAC ATCCAAAATAC GCCAATTTTT TTTAAATTCC AAGAGCCGCA
5801 5901 6001 6101 6201 6301 6401 6501 6601	AATAGTGCAA ATAAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CGCCTAAAAA CCTTAGGGGA TATTCGTTTT AGATCTCTAA GAGCCAACC CATCAACAAC	TGAGCATCAG TGAGCATCAG TCTCAGTGCA TGTGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA ACCTCCCAA	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTAAGC TTATGTATTA TATTGGACG TGGAGCAATA AAACGTGATG	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTTA CACTTCAGCC TCTTACAATG	ATTATTTATT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGTA TCTTGGATAT AAGGATGGT TAGGGGTTGG TTAGTATATA AGCTAGCTAC AAATCGCAAC	TTATATACAT CAAAAAAAAA AAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA ATATAATGGA AAACGTTTTA GAGCACAATT AGCACCACAT	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT GGAAATTACT GAAATGGCGGA GGTACAGCAA ACACCAGCAA	CGATATTTT AACATTTGCC GAAAGGATTT CCAGATTTG CAGGCACAAC GAGTTGAAGT TGTTGCTAAT TGGGCACGCG CTTTTATTTA CATGTTCCAA CACAACAAC	TTATTTAATT AAAAATGAT AAGTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTCT TGAGATTTAA GTAAAAATAC CATTGCTTGT TAAATTTTTT AATATTAAAC AGCAACAGCA TTTCCTTCCG2	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAAT GTAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCCCA CATCATCATC
5801 5901 6001 6101 6201 6301 6401 6501 6601 6701	AATAGTGCAA AATAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CGCCTAAAAA CCTTAGGGGA TATTCGTTTT AGATCTCTAA GAGCCAACC CATCAACAAC	TGAGCATCAG TGAGCATCAG TCTCAGTGCA TGTGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA ACCTCCCAAC AGCAATCGCA	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATAA TATTGTATTAT AATTGGGACG TGGAGCAATA AAACGTTGAT	CATTAGACAC ACACAAAAAT TCAAGTATTT AAGGTATTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTTA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCGGGCTCT	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGAA ATCTTGGATAT AAGGGATGGT TAGGGGTTGG TTAGTACATGC AAATCGCAAC TAACACCTTC GCCGAATGCC	TTATATACAT CAAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATA ATATAATGGA AAACGTTTTA GAGCACAATT AGCACACAGCG GCCGGGAGGCG GTCTTGCCCCT	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA ACACCAGCAA GATAACACGC CCATGTCGCA	CGATATTTTT AACATTTGCC GAAAGGATTT CCAGGATTTG CAGGCACAAC GAGTTGAAGT TGTGCACAGC CTTTTGCTAAT CATGTTCCAA CATGTTCCAA ATGGACAAAG CTTCCTTGCC	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTTT GCCATTTTT GCCATTTTT GCCATTTTT ATAATATTAAC AGTAACAGCA ATTCTTCGGT ACCAGTCCCC	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAAT GTAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCAT CATATGCCAC CAGAGA ATTT
5801 5901 6001 6101 6201 6301 6401 6501 6601 6701 6801	AATAGTGAA AATAGTGAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CGCCTAAAAA CCTTAGGGGA TATTCGTTTT AGATCTCTAA GAGCCAACCC CATCAACAAC ATCCGCTAAG GAATGCGCAA	TGAGCATCAG TGTGCATCAG TCTCAGTGCA TGTGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA ACCTCCCAAC AGCAATCGCA CGGTTTCAAT	ATATTTTGTG ACAACAATGC TTTGTAATTC AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATAA TTTTTTTTAT AATTGGGACG TGGAGCAATA AAACGTTGAT CCGCTAACCC TGAGGCACG	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTTA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCGCGTCT CACACACCCG	ATTATTTATT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGATA TCTTGGATAT AAGGGATGGT TAAGGATGGT TAAGGATGGT AACTAGCTAC AAATCGCAAC TAACACCTTC GCCGAATGGATG	TTATATACAT CAAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATA ATATAATGGA AAACGTTTTA GAGCACAATT AGCATCAGCA GCCGAGAGCG GTCTTGCCGT TGACGCCGCC	ATTCATTTT ATATTTAGAA AACAATAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT GGAAATTACT GGTACAGCAA GATAACACAG CATGTCGCC AAATGCCCG	CGATATTTT AACATTTGCC GAAAGGATTT CCAGGATTGG CAGGCACAAC GAGTTGAAGT TGTGCTAAT TGGGCACGGC CTTTTATTTA CATGTTCCAA CAACAACAAC ATGGACAAAG CTTCCTTGCC AAACCGGAGT	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTTT GCCATTTTT GCCATTGCTTGT TAAAATATC CATTGCTTGT TAAATTTAAAC AGCAACAGCA TTTCTTCGAT ACCAGTCCGC TTTCCCATGTT	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAATAC GCCAAATTTTT TTTAAATTTCC AAGAGCGCA TCATCATCAT CATCATCAT CAGAGAATTT TATGGATAAA
5801 5901 6001 6101 6201 6301 6401 6501 6601 6701 6801 6901	AATAGTGCAA AATAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CCCTTAGGGA TATTCGTTTT AGATCTCTAA GAGCCAACCAC CATCAACAAC ATCCGCTAAG GAATGCGCAA GAACAACAT	TGAGCATCAG TGCAGTGCA TGTGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT CAAAATGCAA ACCTCCCAAC AGCAATCGCA CGGTTTCAAT ACGCAATCGCA	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTAAGA TTTTTTTTAT AATTGGGACG TGGAGCAATA AAACGTTGAT CCGCTAACCC TGAGCCACG GTCCAGCGAT	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTAA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACCG GATACGAAAT	ATTATTTATT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGAA TCTTGGATAT AAGGGATGGT TAGGGATGG TTAGGGATGG TAAGTACTAC AAATCGCAAC TAACACCTTC GCCGAATGGC CCAATGGAAG TTTTGGAAAG	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA ATATAATGGA AAACGTTTTA GAGCACAATT AGCATCAGCA GCCGAGAGCG GTCTTGCCGT TGACGCCGCC CGAAGACGAC	ATTCATTTT ATATTTAGAA AACAAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GATAACACAG CATGTCGCCA AAAATCGCCG GAGAACATTC	CGATATTTT AACATTTGCC GAAAGGATTT CCAGACTTGG CAGGCACAAC GAGTTGAAGT TGTGCTAAT TGGGCACGAG CTTTTATTAT CATGTTCCAA CAACAACAAC ATGGACAAAG CTTCCTTGCC AAACCGGAGT GGATGCCGAT	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTTT GCCATTTTTT GCCATTTTTT TGAGATTTAA GTAAAAATAC CATTGCTTGT TAAAATTTTTTT AATATTAAAC AGCAACAGCA TTTCTTCGAT ACCAGTCCGC TTTCCATGTT TTACAATTCG	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAAT GTAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCAT CATATGCCAC CAGAGAATTT TATGGATAAA CATGGCAAAA
5801 5901 6001 6101 6201 6301 6401 6501 6601 6701 6801 6901 7001	AATAGTGCAA AATAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CCCTTAGGGGA TATTCGTTTT AGATCTCTAA GAGCCAACAC ATCCGCTAAG GAATGCGCAA GAGCAAGATT TGAAGAACTA	TGAGCATCAG TGCAGTGCA TGTGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA ACCTCCCAAC AGCAATCGCA CGGTTTCAAT ACGCAATCGCA CAAATGCAA CAAATGCAAG	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATTA TTTTTTTTTAT AATGGACG TGGAGCAATA AAACGTTGAT CCGCTAACCC TGACGCAGAT AGCTGCGGAT	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCC GTTATTAATCA TTGTACATTT TTGATTTTA CACTTCAACC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACCG GATACGAAAT TAACAGCTAT	ATTATTTATT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGTA TCTTGGATAT AAGGGATGGT TAGGGGTTGG TTAGTATATA AGCTAGCTAC AAATCGCAAC CCAATGGATG TTAGAAAG TACGAAAATC	TTATATACAT CAAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA CTTATGTATA ATATAATGGA AAACGTTTTA GAGCACAATT AGCATCAGCA GCCGAGAGCGG GTCTTGCCGT TGACGCCGCC CGAAGACGAC GGCTTCTGGC	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GATAACACAG CATGTCGCA AAAATCGCCG GAGAACATTC AACACGCTCG	CGATATTTT AACATTTGCC GAAAGGATTT CCAGATTTGG CAGGCACAAC GAGTTGAAGT TGTTGCTAAT TGGCCACGCG CTTTTATTTA CAACAACAACA ATGGACAACA ATGGACAACA GGATGCCGAT GGATGCCGAT CACTCATATG	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTTG GTCAATTTTCT GCCATTTTCT GTAAAAATAC CATTGCTTGT TAAATTTTAAC AGCAACAGCA TTTCTTCGAT ACCAGTCCGC TTTCCATGTT TTACAATTCG AAGCCCCGAGA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAATAC GTCAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA CATGGCAAAA CATGGCAAAA AGATCTTGCA
5801 5901 6001 6201 6301 6401 6501 6601 6701 6801 6901 7001 7101	AATAGTGCAA ATAATTGGCC GGTATATATG AAGACTGGT ATTTTTGAAC CCCTTAGGGGA TATTCGTTTT AGATCTCTAA GAGCCAACCC CATCAACAAC ATCCGCTAAG GAATGCGCAA GAGCAAGATT TGAAGAACTA GTGCCCCAAG	TGAGCATTCAG TGAGCATCAG TCTCAGTGCA TCTCAGTGCA CAAATTATGA CAAATTATGA CAATTATAAC ACGAGTCAAA TACGCTTATG GATGTCCAAA ACCAATCGCA AGCAATCGCA CGGTTTCCAA CAAATGCAAG TGCCCATTTG	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAATAA TACTGTAAGC TTATGTAAGC TTTATGTAAGC TGAGGCAATA AAACGTTGAT CCGCTAACCC TGACGCACG CTCCAGCGAT AGCTGCGGGT	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCC GTTATTAAAA TTGTACATTT TTGATTTTA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACCG GATACGAAAT TAACAGCTAT GAAGCACCAC	ATTATTTATT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGTA TCTTGGATAT AAGGGATGGT TAGGGGTTGG TTAGTATATA AGCTAGCTAC AAATCGCAAC TAACACCTTC GCCGAATGCC CCAATGGAAG TATGGAAAG TACGAAATC TTGGAGTACC	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA CTTATGTATA ATATAATGGA AAACGTTTTA GGCCACAATT AGCATCAGCA GCCCAGCAGCG GTCTTGCCGT TGACGCCGCC CGAAGACGA GGCTTCTGGC GACTCCGCA	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT GGAAATTACT AAATGGCGGA GGTACAGCAA GGTACAGCAA GATAACACAG GATAACACAG GAGAACATTC AGCACGCTCG GCACAAGAAC	CGATATTTT AACATTTGCC GAAAGATTT CCAGATTTG CAGGCACAAC GAGTTGAAGT TGTTGCTAAT TGGCCACGCG CTTTTATTTA CATGTTCCAA CATGACAACAA CATCCTTGCC AATGGACAAAG CTTCCTTGCC AAACCGGAGT CGATGCCGAT CACTCATATG	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTT GCCATTTTTT TGAGATTTAA GTAAAAATAC CATTGCTTGT TAAATTTTAAC AGCAACAGCA TTCCTTCGAT ACCAGTCCGC TTTCCATGT TTACAATTCG AAGCCCGAGA TCCAATGCGA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCATAGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAT GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCAT CATCATCATC CAGAGAATTT TATGGATAAA CATGGCAAAA CAAGTGCAGT
5801 5901 6001 6201 6301 6401 6501 6601 6701 6801 6901 7001 7101 7201	AATAGTGAA AATAGTGAA AATAGTGAA AAGGACTGGT AAGGACTGGT AATTTTGAAC CGCCTAAAAA CCTTAGGGGA TATTCGTTTT AGACCTCTCAA GAGCCAACCC CATCAACAAC ATCCGCTAAG GAATGCGCAA GAGCAAGATT TGAAGAACTT GTACCCCCAAG TACAGCTGCG	TGAGCATTA TGAGCATCAG TCTCAGTGCA TCTCAGTGCA CAAATTATGA CATATATGA CAATTATAAC ACGAGTCAAA TACGCTTATG GATGTCCAA ACCAATCGCA AGCAATCGCA CAAATGCAA CAAATGCAAG TGCCCATTG TCAACAAATC	ATATTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTAAGC TTATGTAAGC TGAGGCAATA AAACGTTGAT CGGCTAACCC TGACGCACG CTCCAGCGAT AGCTGCGGAT GATGCTCAAC	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCC GTTATTAACA TTGTACATTT TTGATTTTTA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACCG GATACGAAAT TAACAGCTAT GAAGCACCAC TCGCATCGCA	ATTATTTATT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGTA TCTTGGATAT AAGGATGGT TAGGGGTTGG TTAGTATATA AGCTAGCTAC AAATCGCAAC TAACACCTTC GCCGAATGCC CCAATGGATG TTTGGAAAG TACGAAATCC TAGGAAATCC CAATGGAAATCC CAATGGAAATCC AGTCCCATTC	TTATATACAT CAAAAAAAA AAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA CTTATGTATA ATATAATGGA AAACGTTTTA GAGCACAATT AGCACAACAG GCCTAGCGG GCCTTGCCGC GGCTTCTGCC GGCTTCTGCC GGCTTCTGCC GGCTTCTGCC ACATCCGCAA	ATTCATTTT ATATTTAGAA AACAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GATAACACGC CCATGTCGCA AAAATCGCCG GGAACATTC AGCACGCTCG GCACAAGAAC	CGATATTTT AACATTTGCC GAAAGGATTTG CCAGATTTGG CAGGCACAAC GAGTTGAAGT TGTGCACAGG CTTTTATTTA CATGTTCCAA CATGTTCCAA ATGGACAAAG CTTCCTTGCC AAACCGGAGT GGATGCCGAT TGGAGCCAT GCTCCGATTG	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTTT GCCATTTTT GCCATTTTT GCCATTTTT GAGATTTAA GTAAAAATAC CATTGCTTGGT TAAAATTAAAC AGCAACAGCA ATTCTTCGAT ACCAGTCCGC TTTCCATGTT TTACAATTCG AAGCCCGAGA TCCAATGCGA CGATTACGCA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCAATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAAT GTAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCAT CATGGCAAAA AGATCTTGCA ACAAGGCAGT CAAGAGCAGT
5801 5901 6001 6201 6301 6401 6501 6701 6801 6901 7001 7101 7201 7301	AATAGTGCAA AATAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CGCCTAAAAA CCTTAGGGGA TATTCGTTTT AGATCTCTAA GAGCCAACCC CATCACAACAC ATCCGCTAAG GAATGCGCAA GAGCAAGATT TGAAGAACTA GTGCCCCAAG TACCACCGCG GCCATTCGTT	TGAGCATCAG TGAGCATCAG TGTCAGTGCA CATGTCTATA CAAATTATGA CATTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA CGGTTTCCAA CAGCTACCAATCGC TGACTACCAATGCAA TGCCCATTGCAA CGAGTTGCAAT CAAAAACCAATCC	ATATTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATAG TTTTTTTTAT AATTGGACGG TGGAGCAATA AAACGTTGAT CCGCTAACCC TGACGCCACG CTCCAGCGAT AGCTGCGGAT TCACCGAGTT GATGCTCAAC TTGCGCAAGT	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTAA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCACACG GATACGACAT TAACAGCACA TGACCACAA	ATTATTTATT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGAA TCTTGGATAT AAGGGATGGT TAGGGATGG TTAGGGATGG TAAGCACCTTC GCCGAATGCC CCAATGGAAG TATTGGAAAG TATGGAAAATC TTGGAGTACC TTGGAGTACC GCCTGGCATG	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATA ATATAATGGA AAACGTTTTA GAGCACAATT AGCATCAGCA GTCTTGCCGT TGACGCCGCC CGAAGACGAC GCCTTCTGCGC ACATCCGCA ATCCGCCAA ATCGGTCTAC GTTTTGGATG	ATTCATTTT ATATTTAGAA AACAATTACT TACCTAAAAA CACATAGGGT CACCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GATACACAG CATGCGGCA AAAATCGCCG GAGAACATTC AGCACGCTCG GCACACACT AGGAGGGTCT	CGATATTTT AACATTTGCC GAAAGGATTT CCAGGATTGG CAGGATTGAAGT TGTGCACAAC GAGTTGAAGT TGGGCACGAG CTTTTATTTA CATGTTCCAA CATGACAACAAC ATGGACAAAG CTTCCTTGCC AAACCGGAGT GGATGCCGAT CACTCATATG GCTCCGATTG GCCCCATCCC	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTTT GCCATTTTT GCCATTTTT GCCATTTTT ATAATATTCT GTAAAAATAC AGCACAGCAC	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCAATGTTCA TCCAATGTTCA TCCAAAAAGGC TTATATACAC ATCCAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCGCA CAGAGAATTT TATGGATAAA CATGGCAAAA AGATCTTGCA CAGAGTGCAGT TCGATGTGTA
5801 5901 6001 6201 6301 6401 6501 6601 6701 6801 6901 7101 7201 7301 7401	AATAGTGCAA AATAGTGCAA AATAGTGGCA GGTATATATG GGTATATATG AAGGACTGGT ATTTTTGAAC CCCTTAGGGGA TATTCGTTTT AGATCTCTAA GAGCCAACACC CATCAACAAC ATCCGCTAAG GAATGCGCAA GACAAGATT TGAAGAACTA GTGCCCCAAG TACAGCTGCG GCCATTCGTT CGGTACACGT	TGAGCATCAG TGTGAGCATCAG TGTGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA ACCTCCCAAC AGCAATCGCA CGAATTCCAA CAAATGCAAG TGACTATCCAA CAAATGCAAG TCAACAAATC CCAGTTGCAT CGTGGCCCGA	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATTA TTTTTTTTTAT AATTGGACG TGGACCAATA AAACGTTGAT CCGCTAACCC TGACGCCACG CTCCAGCGAT TCACCGCAGT GATGCTCAAC TTGCGCAAGA	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTAA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACG GATACGAAAT TAACAGCTAT GAAGCACCAC ATGACCACAA CGCCAACAAC	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGAT TCTTGGATAT AAGGGATGGT TAGGGATGG TTAGGGATGG TAAGTACTAC AAATCGCAAC TAACACCTTC GCCGAATGCC CCAATGGAAG TATTGGAATAC TTGGAGTACC AGTCCCATTC GCCTGGCATG AAGGCAAACA	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATA ATATAATGGA AAACGTTTTA GAGCACAATT AGCATCAGCA GCCGAGAGCG GTCTTGCCGT GACGCCGCC CGAAGACGAC GGCTTCTGGC ACATCCGCAA ATCGGTCTAC ATTGGATG AAGCGCCTGC	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTAACACAA GATAACACAG CATGTCGCA AAAATCGCCA GAGAACATTC AGCACGCTCC GCACACAGAC CAGTACCGTT AGGAGGGTCT CATCAAGTCT	CGATATTTT AACATTTGCC GAAAGGATTTG CCAGATTTGG CAGGCACAAC GAGTGAAGT TGTTGCTAAT TGGGCACGCG CTTTTATTTA CATGATCCAAC AACGAACAAC AACGAACAAC AACGAACAAC AACGACAAC GGATGCCGAT GCCCAATCCC GAAATGAAAA	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTTT GCCATTTTT GCCATTTTT TGAGATTTAA GTAAAAATAC CATTGCTTGT TAAAATTTTTT AATATTAAAC AGCAACAGCA TTTCTTCGAT ACCAGTCCGC TTTCCAGTCGT TTACAATTCG AAGCCCGAGA CGATTACGA TCCACAGCACA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCAATGTTCA TCCAATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAAATA GTAAAAATAC GCCAATTTTT TTTAATTTTCC AAGAGCCGCA TCATCATCATC CATGAGCAACA CAGGCAAAAA AGATCTTGCA CAAGTGCAACA TCGATGTGTA TCCCAACCAT
5801 5901 6001 6101 6201 6301 6401 6501 6601 6701 6801 6901 7101 7201 7301 7401 7501	AATAGTGCAA AATAGTGCAA ATAATTGGCC GGTATATATG AAGACTGGT ATTTTTGAAC CCCTAAAAA CCTTAGGGGA TATTCGTTTA AGATCTCTAA GAGCCAACCC CATCAACAAC ATCGGCTAAG GAACAACAAC ATCGCCAAGATT TGAAGAACTA GTGCCCCAAG TACAGCTGCG GCCATTCGTT CGGTACACGT CATCAGCTGC	TAAGCATTAT TGAGCATCAG TCTCAGTGCA TGTGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT CAAAATGCAA ACCTCCCAAC AGCAATCGCA CGATTCCAA CAAATGCAAG TGCCCATTG CAAACAAATC CAAGTGCAT CGAGCCCGA CAGCCCCGAC	ATATTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATTA TTTTTTTTAT AATGGAGCAATA AAACGTTGAT CCGCTAACCC TGACGCAAG CTCCAGCGAT GACGCCAGGAT GATGCTCAAC TTGCCCAAGA TGCCAAGAGA	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAATCA TTGTACATTT TTGATTTATA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACCG GATACGAAAT TAACAGCTAT GAAGCACCAC ATGACCACAA ATGACCACAC	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGAA TCTTGGATAT AAGGGATGGT TAGGGATGGT TAGGGATGGT AAATCGCAAC TAACACCTTC GCCGAATGCAC TTTGGAAAATC TTGGAAAATC TTGGAGTACC AGTCCCATCC GCCTGGCAACA CATCGTCTGA	TTATATACAT CAAAAAAAAA AAAAAAAAT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATA ATATAATGGA AAACGTTTTA GAGCACAATT AGCATCAGCA GCCGAGAGCGG GCCTTGCCGT TGACGCCGCC CGAAGACGAC GGCTTCTGGC ACATCCGCAA ATCGGTCTAC GTTTGGATG AAGCGGCTGC CCACCCCAAC	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTAACACAA GGTACAGCAA GATAACACAG GACAACAACA GGACAATTC AGCACGCTCG GCACAAGAAC CAGTACCGTT AGGAGGGTCT CATCAAGTCT CAACAACAAA	CGATATTTT AACATTTGCC GAAAGATTTG CCAGATTTG GCAGCACAAC GAGTTGAAGT TGTTGCTAAT TGGCCACGCG CTTTTATTTA CAACAACAACA ATGGACAAAC AATGGACAAAG GGATGCCGAT GGATGCCGATG GCCCAATCC GCCCCAATCC GAAAGAAAA TGTCCCCAACA	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTCT TGAGATTTTCT GCAATTTCT GTAAAAATAC CATTGCTTGT TAAAATTTTTT AATATTAAAC AGCAACAGCA TTTCTTCGAT ACCAGTCCGC TTCCATGGTA TCCAATGCGA CGATTACCA ACGATTGCCA AATGCCCCAA AATGCCCCAA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCAT CATATGCCAC CAGAGAATTT TATGGCAAAA AGATCTTGCA CAAGTGCAGT ACGAAATACT TCGAATGTGA ACGACCAT ATGGCATGG
5801 5901 6001 6101 6201 6301 6401 6501 6601 6701 6801 6901 7101 7201 7301 7401 7501	AATAGTGCAA AATAGTGCAA ATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CCCTTAGGGGA TATTCGTTTT AGATCCTAA GAGCCAACCC CATCAACAAC ATCCGCTAAG GAATGCGCAA GAGCAAGATT TGAAGAACTA GTGCCCCAAG TACAGCTGCG GCCATTCGTT CATCAGCTGC CATCAGTCCC	TAGAGCATTCAG TGAGCATCAG TCTCAGTGCA TCTCAGTGCA CAAATTATGA CAAATTATGA CAATTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA AGCAATCGCA CGGTTTCCAA CAAATGCAAT CGCAATTGCAT TCCACAATCGC TCCACAAATC CAAGTTGCAT CGTGGCCCGA CAGCCTCGCC CCCACAAGGC	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAC TACTGTAAGC TTATGTAAGC TTATGTATA TTTTTTTTTAT AATCGCAACAT CCGCTAACCC TGACGCCACG CTCCACCGAT AGCTGCCGAGTT GATGCTCAAC TTCGCCAAGAT AAGTGAAGAA TGCCAAGAAC	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCATAATCA CAGTTCCCC GTTATTAATCA CAGTTCCCC GTTATTAAAA TTGTACATTT TTGATTTTTA CACTTCAGCC TCTTACAATG CACCGGGTCT CAACACACCG GATACGAAAT TGACCACAC ATGAACACAC CGCCACCACA ATGACCACACA CGCCACACACTCTT CCTCATTCTT	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGTA ATGTTGGATAT AAGGATGGT TAGGGGTTGG TTAGTATATA AGCTAGCTAC AAATCGCAAC TAACACCTTC GCCGAATGCC CCAATGGATACC AGTCCCATTC GCCTGGCATC GCCTGGCATC AGTCCCATCG AGTCCCATCG CATCGTCTGA	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA CTTATGTATA ATATAATGGA AAACGTTTTA GGCACAATT AGCATCAGCA GTCTTGCCGT TGACGCCGCC CGAAGACGAC GCCTTCTGGC ACATCCGCAA ATCGGTCTAC GTTTTGCATC AAGCGGCTGC CCACCCCAAC AATCTCAATC	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GGTACAGCAA GATAACACAG GATAACACAG GAGAACATTC GGCACAAGAAC CAGTACCGTT AGGAGGGTCT CATCAACTAA TCAACAAGT	CGATATTTT AACATTTGCC GAAAGATTTG CCAGATTTG CCAGATTTG GAGCACAAC GAGTTGAAAT TGTTGCTAAT TGGGCACGCG CTTTTATTTA CAGCACAAC ATGGACAACA CAACGGAGAT GGATGCCGAT GGATGCCGATG GCTCCGATTG GCTCCGATG GCTCCGATG GCTCCGATG GCTCCCATCC GAAATGAAAA TGTCGCAACA GCTGCCAG	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTT GCCATTTTT GCCATTTTT GCCATTTTT GCCATTTTT ATATTTAA GTAAAATAC AGCACCACACA ACTTCTCGAT ACCAGTCCGC TTTCCATGTT TTACAATTCG AAGCCCGAGA TCCAATGCCA TCCACAACA ATTCCACAACA AATGCCCCAA CAACAAGTGT	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCATGTTCA TCCAAAAGGC TTATATACAC GTAAAATAC GTCAAAATAC GTCAAAATTTTT TTTAATTTCC AAGAGCCGCA CATGGCAAAA CATGGCAAAA AGATCTTGCA CAAGTGCAAA AGATCTTGCA CAAGTGCAAT TCGAAATACT TCGCAACCAT ATGGCATGTGTA TCCCAACCAT
5801 5901 6001 6201 6301 6401 6501 6601 6701 6801 7001 7101 7201 7401 7501 7401 7501	ATAGTACTARIA AATAGTGCA AAGGACTGGT AAGGACTGGT AAGGACTGGT ATTTTGAAC CGCCTAAAAA CCTTAGGGGA TATTCGTTTT GAGACCACCC CATCACAACAC ATCCGCTAAG GAATGCCCAAG GACGCAAGATT TGAAGAACTA GTGCCCCAAG TACAGCTGCG GCCATTCGTT CGGTACACGT CATCAGCTGC CATCAGCTGC CATCAGCTGC	TAGAGCATCAG TGAGCATCAG TCTCAGTGCA TCTCAGTGCA CAAATTATGA CATATATGA CATTATAAC ACGAGTCAAA TACGCTTATG GATGTCTAT CAAAATGCAA ACCTCCCAAC AGCAATCGCA CGGTTTCCAA CAAATGCAAG TGCCCATTG TCAACAATCC CAAGTTGCAT CAAGTGCACAA CAACCTCCCAC CCAACAAGCC CCAACAAGCC	ATATTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTAAGC TTATGTATAA TATTGGAACA TGAGGCAATA AGATGCGAGTT GATGCCAACG TTCACCGAGTT GATGCCAACA TGCCAACATGC CACAACATGC CACACCTCCA	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCC GTTATTAATCA CAGTTCCCC GTTATTAAAA TTGTACATTT TTGATTTTTA CACTTCAGCC TCTTACAATG CACCGGGTCT CAACACACCG GATACGAAAT TAACAGCTAT GAAGCACCAC TCGCATCGCA	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGTA TCTTGGATATT AAGGGATGGT TAGGGGTTGG TTAGTATATA AGCTAGCTAC AAATCGCAAC TTACGAAAGCTAC CCAATGGAAG TACGAAAGCTAC AGTCCCATTC GCCTGGCATG AAGGCAAACA CATCGTCTGA CCCCTACTGG GGCGGACAAA	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA CTTATGTATA AAACGTTTTA GACCACAATT AGCACAACG GTCTTGCCGT TGACGCCGCC CGAAGACGAC GGCTTCTGGC GGCTTCTGCCA ACATCCGCAA ATCGGTCTAC GTATTGGATG AAGCGCCTGC CCAACCCCAAC AATCTCAATC GCAACGACGAC	ATTCATTTT ATATTTAGAA AACAATTACT TACCTAAAAA CACATAGGGT CACCCCGAT GGAAATTACT AAATGGCGGA GGTACAGCAA GATAACACG CATGTCGCG GAGAACATTC CAGTACCGTT AGGAGGGTCT CACTAAGTC CACTACAGTC CACTACAAT CAATAACGCT	CGATATTTT AACATTTGCC GAAAGATTTG CCAGATTTG CCAGGTCAAC GAGTTGAAGT TGTTGCTAAT TGGGCACGG CTTTTATTTA CATGTTCCAA CATGACAACA CATCATAG GATGCCGAT GGATGCCGAT GCTCCGATG GCTCCGATG GCCCCAATCCC GAAAGACAACA	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTTT GCCATTTTT GCCATTTTT GCCATTTTT GCCATTTTT TAAAATTAAC AGTACAGCC AGCACAGCAC	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCAATGTTCA TCCAATGTTCA TCCAATGTTCA TCCAATGTCA GTCAAAAATG GTCAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCAT CATGGCAAAA CAAGGCCAAAA CAAGGCCAAAA CAAGGCCAAAA CAAGGCCAAAA CAAGGCCAAAA CAAGGCCAAAA CAAGGCCAACAT TCCGATGTGTA TCCCAACCAT ATGGCAATGG GGATGAGGAC
5801 5901 6001 6201 6301 6401 6501 6601 6701 6801 6701 7001 7101 7201 7301 7401 7501 7601 7701	AATAGTCATA AATAGTCAA AATAGTCAA ATAATTGGCC GGTATATATG CGCCTAAAAA CCTTAGGGGA TATTCGTTTT GAGACCACCC CATCACAACAC ATCCGCTAAG GAATGCCAAGATT TGAAGAACTA GTGCCCCAAG TACAGCTGCG GCCATTCGTT CATCAGCTGC GTCGCCACGT CATCAGCTGC GTCGCCACGT AACTTTGACC	TAAGCATTATA TGAGCATCAG TCTCAGTGCA TCTCAGTGCA CATATATAGA CATATATAGA CATATATAGA TACGCTTATG GATGTGCTAT CAAATGCAA ACCTCCCAAC AGCATCGCA CGGTTTCAAT ACCCAATCGCA TGCCCATTGC TGACTATCCAA CAAGCTGCAT CAACAAATCCAA CAGCTCGCA CAACTGCA CAAGTGCAT CCAACAAGCC ACCCCCAC ACACCCCCAC	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATAA TATTGTATTA AATGGGACG TGAGGCAATA AAACGTTGAT CCGCTAACCC TGACGCCACG CTCCAGCGAT GATGCTCAAC TTGCGCAAGAT AGCCAAGAG AGCCACAGAG CACAACATGC CACACCTCCA AGGCAGCCA	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTAA CACTTCAGAC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACCG GATACGAAAT TGACACACAC TCGCATCGCA	ATTATTTATT ATCAACAACA TAAAAATTTA TTGCGAAATT GTGCTTGGAT TCTTGGATAT AAGGGATGGT TAGGGATGG TTAGGGATGG TAAGCACCTTC GCCGAATGCAC CCAATGGAAG TATTGGAAGAAC TTTGGAGTACC CCCATCGGCATG AGCCCACTCG GCCTGGCATG AGGCCAACA CATCGTCTGG GGCGGACAAA CCCCAACGCCC	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA CATCATAGA CTTATGTATA AAACGTTTTA GAGCACAATT AGCATCAGCA GCCGAGAGACGAC GGCTTCTGGCG CCAACGACGACGAC AATCGCCTAC CCAACGACGACGA CCCACTGAAC	ATTCATTTT ATATTTAGAA AACAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GATACAGCAA GATAACACAA GATAACACGA GAGAACATTC AGCAGCGCCG GACAAGACC CAGTACCGTT AGGAGGGTCT CACCAACAA CACACAACAA CACACAACA	CGATATTTT AACATTTGCC GAAAGGATTT CCAGGATTTGG CAGGCACAAC GAGTTGAAGT TGTGCAAGT TGGGCACGCG CTTTTATTTA CATGTTCCAA CATGTTCCAA CATGACAAAG CTTCCTTGCC AAACCGGAGT GGATGCCGAT CGCTCCGATTG GCCCCATCC GAAATGAAA TGTCGCAAC GGCTGCCCAG GAAGACAACC CACCCCCTGT	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTTT GCCATTTTT GCCATTTTT TGAGATTTAA GTAAAAATAC CATTGCTTGT TAAATTTATAAC AGCACAGCAC	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCAATGTTCA TCCAATGTTCA TCCAATGTTCA TCCAATAAGC TTATATACAC ATCCAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCGCA TCATCATCAT CATGAGAATAA CATGGCAATAA CATGGCAATAA CAGAGTGCAGT TCGATGTGTA TCGATGTGTA TCGATGTGTA TGGCCATGTG AAGACTATTGA AACCTATTGA
5801 5901 6001 6201 6301 6401 6501 6701 6701 6801 6701 7101 7301 7401 7501 7601 7701 7801	AATAGTCATA AATAGTGCA AATAATTGGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CCCTTAGGGA TATTCGTTTT AGATCTCTAA GAGCCAACCC CATCAACAAC ATCCGCTAAG GAATGCGCAA GACAAGATT TGAAGAACTA GTGCCCCAAG TACAGCTGGG GCCATTCGTT CGGTACACGT CATCGATCCT AACTTTGACC AAATTGCCCA	TAGAGCATCAG TCTCAGTGCA TCTCAGTGCA TCTCAGTGCA CATGTCTATA CAAATTATGA CATTTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA CGGTTTCCAA CAGATCGCA TGACCAATCGC CAAATGCAAG CGGTTTCCAA CAGATGCAAT CGAGCAATCGCA CAGCTACGCA CAGCCACGG CAACAAGC ATGCCGGAAG AAAAATCGGA AGAGGAAGTC	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATTA TTTTTTTTTAT AATCGGACAATA AAACGTTGAT CCGCTAACCC TGACGCCACG CTCCAGCGAT TCACCGCAAGT CACCCACGAGT AAGTGAAGAA TGCCCAAGAG CCACACATGC CCACCCCCA AAGCCCCCA	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TTGATTTTAA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACCG GATACGAAAT TAACAGCTAT GAAGCACCAC ATGACCACCA CGCCAACAAC ACGCCAACAAC ACGCCAACAAC ACGCCAACAAC ACGCCAACAAC ACGCCACATCTT CAATCAGTTCT CAACACTTGT CAACACCTGT	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGAA TCTTGGATAT AAGGGATGGT TAGGGATGG TTAGGGATGG TAGGTAGCTAC AAATCGCAAC CCAATGGAAG TACGAAAATC TTGGAGTACC AGTCCCATTC GCCTGGCATG AAGCCAACA CACCCTACTGG GCCGGCACAAA CCCCAAGGCTC GTCAAGCAGC	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAATT TACTCATAGA CTTATGTATAA CATCATAGA AAACGTTTTA ACATCAGCA GCCGAGAGCG GTCTTGCCGT GACGCCGCC CGAAGACGACG ACTCTCAGCA ATCGGTCTAC GTTTTGGATG AAGCGCCGCCAC CCACCCAAATC GCCACGCCAAC CCCCCGCCAAC	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GGTACAGCAA GATAACACAG GACAACACTC GAGAACACTC GAGAACACTC CAGTCAGGTCT CATCAAGTCT CAATAACGCT AACGAATCGT AACGAATCGT AACGAATCGT AACGAATCGT AACGAATCGT	CGATATTTT AACATTTGCC GAAAGGATTTG CCAGATTTGG CAGGCACAAC GAGTTGAAGT TGTTGCTAAT TGGGCACGCG CTTTTATTTA CATGGTCCAAG ATGGACAACA ATGGACAACA CATCCTTGCC AAACCGGAGT GCACTCCTTGCC GAAATGAAAA TGCGCCAATCCC GAAATGAAAA GGCTGCCCAG GAAGACAACC CACCCCCTGT CCCCAACCTA	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTA TTTGACTTTTT GCCATTTTTT GCCATTTTTT GCCATTTTT TGAGATTTAA GTAAAAATAC CATGCTTGT TAAATTTTTT AATATTAAC AGCAACAGCA TTTCTTCGAT ACCAGTCCGAC TTCCATGGT TTACAATTCG AAGCCCGAGA CGATTACGCA CGATTGCCA ATTCCACAACA CAACAAGTGT ACACCTCCA CCTACCATAC GACACATCCA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCAATGTTCA TCCAATGTTCA TCCAAAAGGC TTATATACAC ATCCAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCGCA TCATCATCAT CATGAGCAACA CAGGAAATAT ACGACACATTGCA CAGGAAATAC TCCAACCAT TCGATGTGTA TCCCAACATTGG GGATGAGGAC AACCTATTGA CCCAACATCTG
5801 5901 6001 6101 6201 6301 6401 6601 6701 6801 6901 7101 7201 7301 7401 7501 7701 7801 7801 8001	AATAGTGCAA AATAGTGCAA AATAGTGCC GGTATATATG GCC GGTATATATG CCCTAGAT ATTTTTGAAC CCCTAGGGA TATTCGTTTT AGATCTCTAA GAGCCAACCC CATCAACAAC ATCCGCTAAG GAATGCCAA GACAAGATTG TGAAGAACTA GTGCCCCAAG CCATTCGTC CGGTACACGT GTCGCCACGT GTCGCCACGT GTCGCCACGA CCCAGTCGCACC	TGAGCATCAG TGAGCATCAG TCTCAGTGCA TCTCAGTGCA TCTCAGTGCA CAAATTATGA CAAATTATGA CAATTATAAC ACGAGTCAAA TACGCTTATG GATGTGCTAT TTTATTGGTT CAAAATGCAA AGCAATCGCA CGGTTTCCAAT CAACAATCGCA TGCCCATTG TCAACAATCGCA CAGCCTCGCC CAACAAGCC ATGCGGAAG AGAGGAAGTC GAGGAAATCG CAACGATTCCA	ATATTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATTA TTTTTTTTAT AATGGAGCA TGGAGCAATA AAACGTTGAT CCGCTAACCC TGACGCAACA CTCACGAGTT GATGCCAAGAGT GACGCCAAGAGC CACACAGCGC CACACATGC CACACCCCA AGCCCCCA ACACCCCCA CCGTGCCCCA	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCCC GTTATTAAAA TTGTACATTT TGAATTTTTA CACTTCAGCAT TTGTACATG ATGAATTCGC CACCGGGTCT CAACACACCG GATACGAAAT TGAACACCCC ATGACCACAA ATGACCACAA ACGACAACAAC ACCACATCTT CCTCATTCTT AAATTTGCAT GAACAGTTCG CAATCAGTTC GCCAATCCGC	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGAA TCTTGGATAT AAGGGATGGT TAGGGGTGG TTAGGGATGG TAGGTAGCTAC AAATCGCAAC TAACACCTTC GCCGGAATGCC CCAATGGAAG TTTTGGAATAGC AGTCCCATTC GCCTGGCAAG CATCGTCTGA CATCGTCTGA GGCGGACAACA CCCAAGCTC GTCAAGCAGC	TTATATACAT CAAAAAAAA AAAAAAAAT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGGTATA ATATAATGGA AAACGTTTTA ACACACAGA GCCGAGAGCGG GTCTTGCCGT TGACGCCGCC GGATGCCGCA ATCGGTCTCCGCA ATCGGCCGAC GTTTTGGATG AAGCGGCTGC CCACCCCAAC AACCGGCGAA CCCCCCCGAA TCCCGGCGAA	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GGTACAGCAA GATAACACAG GACAACAAG CCATGTCGCG GAGAACATTC AGCACGCTCG GCACAAGAAC CATCAAGTCT CATCAAGTCT CAACAACAAA TGCAAATCGTT AACGAATCGT AACGAACTCGT AACGACTCTTCC CACCACCAACAA	CGATATTTT AACATTTGCC GAAAGATTTGC CCAGATTTGG CCAGGTTGAAG TGTTGCTAAT TGGTCAAG CATGTTCCAA CAACAACAAC ATGGACAACA ATGGACAACA GGATGCCGAT GGATGCCGATG GCCCAATCCC GAAAGAACAA GGCTGCCAAG GAAGACAACA CACCCCTGT CCTCAAGCTA GAAGACCAACC	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTTA GCCATTTTTA GTAAAAATAC CATTGCTTGT TAAAATTTTA ATAATAAAC AGCAACAGCA TTTCTTCGAT ACCAGTCCGA CAGTCCGAC AGCAATGCGA CGATTACCA AATGCCCAA AATGCCCAAC AAAGGTTCGA CAACAACGACA CAACACCAACCA AAATGCTCCA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCATGTTCA TCCATGTTCA TCCATGTTCA TCCATGTTCA GCCAATTTTC GTAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCATCAT CATGGCAAAA AGATCTTGCA CAGGGCAAAT ACGAGGCAGT ACGAATATG TCCAACCAT GGATGGCGCAAAT GGATGAGGAC AACCTATTGA CCCAACATCT ATCCCCCCCA
5801 5901 6001 6201 6301 6401 6401 6401 6401 6401 7001 7301 7301 7401 7501 7401 7501 7601 7701 8001 8001 8101	AATAGTGCAA AATAGTGCAA AATAGTGCC GGTATATATG AAGGACTGGT ATTTTTGAAC CCCTAGAGAG CCTTAGGGGA TATTCGTTTA AGACCAACAC ATCCGCTAAAA GAGCAACAAC ATCCGCTAGA GACAACAAC ATCCGCAACG GCATTCGT TGAAGAACTA GTGCCCCAAG GCCATCGT CGTACACGT CGTCACACT CATCGCCACGT GCCACTCG AAATGACCGA CCCAGTCGCA CCCAGTCGCA	TAAGCATTA TGAGCATCAG TCTCAGTGCA TCTCAGTGCA CAAATTATGA CAAATTATGA CAATTATAAC ACGAGTCAAA TACGCTTATG GATGTCCTAT CAAAATGCAA ACCACTCCCAAC AGCAATCGCA CGGTTTCCAA CAAATGCAA CAAATGCAA CAAATGCAA CAAATGCAA CAAATGCAA CAAATGCAA CAACTTCCA CAGCTTGCC CCACCAATCGC CACCAGTGCC CAACAAGC AGAGGAAGTC GAGCACATCG CACCAGTTGC	ATATTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTAAGC TTATGTATAT TTTTTTTTAT AATTGGGACG TGGAGCAATA AAACGTGAT GATGCTCAAC TGCCAGCGAT TCACCGAGTT GATGCTCAAC CCACCACAGAG CCACAACATGC CACACCACA AGGCAGCGCA AGGCACCCCA CGTGCCCGA TTTTGCCCCA	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCATAATCA CAGTTCCCC GTTATTAATCA TTGTACATTT TTGATTTTTA CACTTCAGCC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACG GATACGAAT TGACACACA ATGACACACA ATGACCACA ATGACCACA ATGACCACA ACCACATCTT CACCACACAC CCCATCGC CAACCACCG ATCACCACAC	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGTA TAGGGGTGGT TAGGGGTGG TTAGTATATA AGGGATGGT TAGTACTAC AAATCGCAAC CCAACCTTC GCCGAATGCC CCAATGGATACC AGTCCCATCC GCCTGGCAAG CCCTGGCAACA CATCGTCTGA CCCCTACTGG GGCGGACAACA CCCAAGCACC AGCACTGACGT CCTCCCCGCC	TTATATACAT CAAAAAAAA AAAAAAAAT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA CTTATGTATA AACGTTATGA AAACGTTTTA GAGCACAATT AGCATCAGCA GCCGAGAGCGG GCCTTGCCGT GAAGCCGCC CGAAGACGACG GCTTTCTGGC AATCCGCAAC AATCCCAAC AATCTCAATC GCCACCCCAAC AATCCCCAAC CCCCCGGCGAA CCCCCCGCGCA	ATTCATTTT ATATTTAGAA AACAAAAAA AAGAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GGTACAGCAA GATAACACAG GATAACACAG GACAACAAGAAC CATGTCGCA GCACAAGAAC CAGTACCGTT CACAACAAA TGCAAATGTT CAATAACGCT AAGGACCGTTT AAGGACCGTTT AGGACCGTTT CACAACAAA	CGATATTTT AACATTTGCC GAAAGATTTG CCAGATTTGG CAGGCACAAC GAGTTGAAG TGTTGCTAAT TGGCCACGCG CTTTTATTTA CAGGCACGCG CTTCCTTGCC AAACGGACAAC ATGGACCAATG CACTCCATATG GCTCCGATTG GCTCCCAATCCC GAAAGACAACA GGCTGCCCAG GAAAGACAACA CACCCCCTGT CCTCAAGCTA GAAGACCAA GCACCCCTGT CCTCAAGCTA	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTT GCCATTTTT GCCATTTTT GCCATTTTT GCCATTTTT AATATTAAC GTAAAAATAC CATGCTTGT TAAATTTAAAC ACCAGCCCGA TTTCCATGTT TTACAATTCG AAGCCCGAGA TCCAATGCGA CGATTACGCA CGACTCGAA CCAACAACGA CAACAAGTGTAA CCAACCATCCA AAATGGTGCA CGACTGCAA CGACGCCAA	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCAATGTTCA TCCAATGTTCA TCCAATGTTCA TCCAATGTCA TCCAAAAATG GTAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCAT CATCATCATCAT CATCATCATCAT CATGGCAAAA AGATCTTGCA ACGAATACT TCGAGAGATTG TGGCGCAAAT GGATGAGGAC AACCTATTGG TGGCGCAAAT GCCACCCCCCA TATCCCCCGCA
5801 5901 6001 6201 6301 6401 6501 6601 6701 6801 7701 7701 7701 7301 7301 7301 7801 8001 8101 8201	ATAGTCATA AATAGTGCA AATAATTGGCC GGTATATATG AAGGACTGGT AAGGACTGGT ATTTTTGAAC CGCCTAAAAA CCTTAGGGGA TATTCGTTTA GAGCCAACCC CATCACAACAC ATCCGCTAAG GAATGCCCAAG TACAGCTGCG GCCATTCGTT CATCAGCTGCG GCCATTCGTT CATCAGCTGCC AAATGACCAGT CATCGGCCCCGT AACTTTGACC AAATGACCGA CCCAGTCGCC CCCGGTCGCAG TTTACTTCAA	TAAGCATTA TGAGCATCAG TCTCAGTGCA TCTCAGTGCA CAAATTATGA CATATATGA CAATTATAAC ACGAGTCAAA TACGCTTATG GATGTCTAT CAAAATGCAA ACCACTCCCAAC CGGTTTCCAAA CAAATGCAA TCCACAATCGC TCAACAAATCCAAG TCCACCACTCGC CCAACAAGCC ATGCGGGAAG AAAATCGGA CAACCTCCCAC CCACAGTGCC CCACCAGTGC CAACAGTCC	ATATTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAATAA TACTGTAAGC TTATGTAAGC TTATGTATAT TTTTTTTTTAT AATTGGGACGAT CGCTAACCC TGACGCCACG CTCCACGAGTT GATGCTCAACGTT AAGTGCAGGAT TGCCCACGAGT AAGTGAAGAC CCACACTCCA AGGCCCCCA AGCGCCCCA CCCTACCCCA	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCC GTTATTAAAA TTGTACATTT TTGATTTTA CACTTCACATC ACCTCAGCC TCTTACAATG ACACTCAGCC CACACACCCG GATACGAAT TAACAGCTAT GAAGCACACA ACGACACACA ACGACACACA ACGACACACA ACGACATCTT CAAATCAGTTC CACACACACC CCCAATCCGC ATCACCACAC CCCACATCGC	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGAA TCTTGGATAT AAGGGATGGT TAAGGGATGGT TAAGGATGGT CAAGCTAC AAATCGCAAC TAACACCTTC GCCGAATGGCA CCAATGGATA TTTGGAAAA TTTGGAATAC CCCATGGCATG AGCCCCATTC GCCTGGCATG AGGCCAACAA CATCGTCTGG GGCGGACAAA CATCGTCTGG GGCGGACAAA CACCACAGGCTC GTCAAGCACC	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA CTTATGTATAA GACCACAATT AGCACAATT AGCACAACGC GTCTTGCCGT TGACGCCGCC GGCTTCTGCGC ACATCCGCAA ATCGGCCGAC GACTCCCCAAC AATCTCAATC GCAACGACGA CCCACCCCAAC ACCCCTCGTC TGCGACGATG	ATTCATTTT ATATTTAGAA AACAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GGTACAGCAA GATAACACG GACAACAGCA GACAACACG GACAACATCC CAGTACCGTC AGGACCGTT CACCAAGAAC CAGTACCGTT CACCAACAAA TGCAAATGTT CAACAACAAA TGCAAATGTT CAACAACGAT CACCAACAAA TGCACTTCC CGCGCGTCTTC CCGTCGTCATTC	CGATATTTT AACATTTGCC GAAAGATTTG CCAGATTTG CCAGATTGG CAGGTGAAGT TGTTGCTAAT TGTGCAAGC CTTTTATTTA CATGTTCCAA CATGTCCAA ATGGACAAGG CTTCCTTGCC AAACCGGAGT GGATGCCGAT GCTCCGATG GCTCCGATG GCTCCCATCG GAAATGAAAA TGTCGCAACA GGCTGCCCAG GAAGACAACA CAACCCCTGT CCTCAAGCTA GAAGAGCCAA GTACATCTTC	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTA TTTGACTTTTT GCCATTTTTT GCCATTTTT GCCATTTTT GCCATTTTT GCCATTTTT TAAAATAAC GTAAAAATAC CATTGCTTGGA TTACAATTCGA AGCACACAGCA TTTCCACGAT TTCCACAGTCCA AGCCCCAAA TCCAATGCCA CGATTGCCA CGATTGCA CAACAACTCGA CAACAACTCGA GCGATTCGA GACCACTCCA AAATGGTGCA GGGGGGAAATTCC	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCAATGTTCA TCCAATGTTCA TCCAATGTCA TCCAATAAGCC TTATATACAC ATCCAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCAT CATGGCAAAA CAAGGCGCAAAT CATGGCAAAA CAAGGCGCAAAT TCGATGTGTA TCGATGTGTA TCGATGTGTA TGGCCGCAATT GGATGAGGAC AACCTATTGA AACCTATTGA ACCAACCATCT ATCGCCCGCA
5801 5901 6001 6201 6301 6401 6401 6401 7001 7101 7001 7101 7301 7401 7501 7501 7501 7801 8001 8101 8301 8301	AATAGTGGAA AATAGTGGAA AATAGTGGAA AATAGTGGAA ATATTGGCC GGTATATATG CGCCTAAAAA CCTTAGGGGA TATTCGTTTT GAGACCACCC CATCACAACAC ATCCGCTAAG GAATGCCCAAG TACAGCTGCG GCCATTCGTT CAGCACACGT CATCAGCTGCG GCCATTCGTT CATCAGCTGCC AATGACCACGT CACCAGTCGCC CAGGTCGCACG CCCGGTCGCAG TTTACTTCAA TGTCGGACCCC	TAAGCATTATA TGAGCATCAG TCTCAGTGCA TCTCAGTGCA CAAATTATGA CATATATAAC ACGAGTCAAA TACGCTTATAAC ACGAGTCAAA TACGCTTATG GATGTCCAA ACCTCCCAAC AGCAATCGCA ACCAATCGCA TGACTTCCAA CAGATTGCAT CAACAATGCAA CAGCTACCAA CAGCTCCCAA CAAGCTGCA CAAGCAATCGC ACCAATCGC ACCACAAGGCA AGAGGAAGTC GAGGAAATCG CAACGAGTGC CACCAGTTGC TCGCCACAA	ATATTTTGTG ACAACAATGC TTTGTAATTT AAGTAAATAA GGCTTAGTCC ACTACTAAAA TACTGTAAGC TTATGTATAA TATTGTAAGC TGAGGACAATA AAACGTTGAT CGGCTAACCC TGACGCCACG CTCCAGCGAT AGCTGCGGAT TGACGCAGGAT AGCTGCGAGAT AGCTGCAGAGA CACAACATGC CCACCCCA AGCCGCCGA AACACCCCA AGCGCCCCA TGTGCCCCA TGTGCCCCA	CATTAGACAC ACACAAAAT TCAAGTATTT AAGGTTGTTT TTCTTAATCA CAGTTCCCC GTTATTAAAA TTGTACATTT TTGATTTTA CACTTCAGAC TCTTACAATG ATGAATTCGC CACCGGGTCT CAACACACCG GATACGAAAT TGACACACAC TCGCAACACAC ACCACATCTT CAACACACACAC ACCACATCTT CAATCAGTTC CACACACAC CCCACATCGC ATCACCACAC CCCACATCGC ATCACCACAC TCCACATCGC ATCACCACAC TCCACATCGC ATCACCACAC TCCACATCGC TCCACATCGT	ATTATTTATT ATCAACAACA TTAAAAATTTA TTGCGAAATT GTGCTTGGAT TCTTGGATAT AAGGGATGGT TAGGGATGG TTAGGGATGG TAAGCACCTTC GCCGAATGCAC CCAATGGAAG TATCGAAAA TTTGGAGTACC CCCATCG GCCTGGCATC GCCTGGCATCG GCCTGGCATCG GCCGGACAAA CATCGTCTGG GGCGGACAAA CATCGTCTGG GGCGGACAAA CATCGACTGAG GCCAAGCAC GTCAAGCAGCT CTCCCCCGCC CTACCACGC TAATTCTATT	TTATATACAT CAAAAAAAA AAAAAAAATT TGGGTAATAA AGGTAGAAAT TACTCATAGA CTTATGTATAA CATCATAGA CTTATGTATA AAACGTTTTA GAGCACAATT AGCATCAGCA GCCGAGAGACGAC GGCTTCTGGCG CGAAGACGACG ACTCCCCCAAC AATCCCCAAC AATCCCCAAC AATCCCCACCAAC AATCCCCACCAAC AATCCCCACCAAC CCCCCTCGTC CGCCGCGTAGTC TGCCGCCGTAGTC TCGCGCCGACTC CGCCGCTAGTC	ATTCATTTT ATATTTAGAA AACAATTACT TACCTAAAAA CACATAGGGT CACCCCGATG GGAAATTACT AAATGGCGGA GGTACAGCAA GATAACACAG CCATGTCGCA AAAATCGCCG GAGAACATTC CACAAGAAC CAGTACCGTC AGGAGGGTCT CATCAAGTCT AGGAACATCT CAACAACATCT CACAACAAC TGCAATCATT CAACAACTT CACCAACTT CACCAACATCT CACCAACATCT CACCACACTT CACCAACTTC CCGTCGAACA	CGATATTTT AACATTTGCC GAAAGATTTG CCAGATTTG CCAGATTGG CAGGTGAAGT TGTTGCTAAT TGTGCCACAG CTTTTATTTA CATGTTCCAA CATGTTCCAA CATGTCCAAAG CTTCTTGCC AAACCGGAGT GGATGCCGAT GGCTCCGATTG GCCCAATCCC GAAAGACAACC GAAGACAACC GAAGACAACC CACCCCTGT CTCCAAGCAA GTACCCCCAG GAAGACCAAC CACCCCCTGT CTCCAAGCAA GTACATCTT CACACATCTC CACACATCTC CTCAAGCAAT	TTATTTAATT AAAAATGAT AAGTTTTTTA GCGATAGTAA TTTGACTTTTT GCCATTTTTT GCCATTTTT GCCATTTTT TGAGATTTAA GTAAAAATAC CATTGCTTGT TAAATTTAAAC AGCACAGCAC	TTTTTGTTGT GCATGTGCAA GTGACTTTGA TCCAATGTTCA TCCAATGTTCA TCCAATGTTCA TCCAATAAGCC TTATATACAC ATCCAAAAATAC GCCAATTTTT TTTAATTTCC AAGAGCCGCA TCATCATCAT CATGACAATA CAGGCAAAA CAAGTGCAGA CAAGTGCAAAA AGAGTGCAGT TCGATGTGTA TCGATGTGTA ACGAACATCT GGATGAGGAC ACCTATTGA ACCCACATATGA TACTGTGATA

Source: Genomic DNA, phage Pco-hb ph1. Alignment with cDNA sequences: The first exon of the *Platypeza hunchback* P1 transcript (SEQ10) could not be aligned with the genomic sequence, positions 4380..4650 correspond to the first exon of the P2 transcript (SEQ11). Positions 6403..8399 are presumably common to the second

exon of both *Platypeza hunchback* transcripts. In the genomic sequence, two NRE sequences were identified (8375..8396, 8466..8487).

# SEQ13 Lonchoptera hunchback, cDNA, P1 transcript.

1	AGTAATAGTT	GTTCTCTTGC	AGCAGCGAGA	GGTCCTGTGT	TGTTATTATT	TCGTGCCATA	TATTTTATTT	TATTTTTTA	AATTTGTTAT	AATTCGATTA
101	CAATTCAGAC	GTTTCATTTT	ААААТАААА	САААААААСА	ATTAACAAAT	TTACTATTAA	CCAAAAGTAT	ATTGTTTTCA	ATTGAAGTGT	ACAAAAACAA
201	TTCATTAAGT	TTAAATTAAA	ATCACTTTAT	CCTTTCACAT	ΑΑΑΑΑΤΑΤΑ	AAATAGCAAA	AAGTGTAATT	AAGCATAAAG	TAATTAATAT	TATGAACATC
301	AACAAATTTT	AAAAAAATT	AATCAAATTT	GAGGAAAAGA	CAATGATGTT	GTTATAATAT	TTTCGATTAT	AGTTACATAT	ATTGCTATTA	TTTGCTAAAA
401	ATGTTTGTGT	AAAATTAAAT	AATTGTTAAA	CAGTTTTGCT	AATATTATAT	GTTAAATGAG	ТТАСАААААА	AAGTGGCCGA	CGCGTTCCGA	TTAACGGATA
501	ТААТААСТАА	TACTTATGCT	GCTGATATTA	ACAACATTTT	AGAGACGGAT	TAACTAGATC	GCTCTCCATC	GCAGTTAAAT	TTGAAGTGAA	ТСААААААА
601	TTTATCAAAA	CAGTTGCTAT	TGAAACCAAA	AACAAATACA	AAATGCAAAA	TTGGGATACA	TTACAACCGA	CAGCCAGTTA	TGAACATAAC	TGGTACGGTA
701	ACATGTTTCC	AACAATTAAA	ACAGAATCGA	TGACTCATTC	GCCAACTACA	TCACAATTAG	AACAATATTT	AACAATGAAA	CAACAAGAAA	TAAATTCGTT
801	AACACCATCA	CCGCGTGCTG	ATATAAACGG	AACTTCAGAC	ATTCAAAATT	TCTTTGATAG	CCACAATCTG	CAAAATGGCA	GTTTGCATCA	TAATCATCCA
901	TTAGGATTTA	ATCCATTAAC	ACCACCTGGT	CTACCAAATG	CTGTTCTACC	AGCAGTTTCA	CATTTCCATC	ATAACATTAC	AGGAATCCAT	GCAAATAATC
1001	CAACGAGTCC	ACTGGCACAG	CAATCCGAGG	GAAATACTCA	ATCATTAACA	CCACGCAATA	CGCCACCTAT	GGATATTACT	CCGCCAAAGT	CACCCAAATT
1101	GTACTCCGAA	TATGTTGAAA	AGGATCACGA	TATGATATCA	AATTCAAGTG	ACGACACAAA	ATTCCTTGAA	AGCGATGATG	ACTCATCGAT	TCGTACACCA
1201	ATTTATAATT	CGCATGGCAA	GATGAAAAAT	TACAAATGTA	AAAGTTGTGG	ATATATGGCT	GTAACAAAAG	TAGCATTTTG	GGAGCATGCA	AGTTCTCATA
1301	TGAAACCCGA	AAAAATTCTG	CAGTGTCCAA	AGTGTCCATT	TGTTACCGAA	TTAAAACATC	ATTTGGAGTA	CCACATTCGC	AAACACAAAA	ATTTGAAGCC
1401	ATTCCAATGC	GACAAATGCA	ATTATAGCTG	TGTCAATAAA	TCAATGCTGA	ATTCACATCG	TAAATCCCAT	TCTTCTGTGT	ATCAGTATCG	TTGTGCCGAT
1501	TGTGATTATG	СААСАААТА	TTGTCATTCT	TTTAAATTAC	ATTTACGAAA	ATATGACCAT	AAGCCCGGTA	TGGTCTTAGA	TGAAGATGGT	TGCCCGAACC
1601	CTTCGATAAT	TATAGATGTT	TATGGAACAC	GTCGTGGCCC	CAAAATGAAG	TCGCAATCCG	GCGGTGGTGG	TGGTAAAATA	AGTTCGGGAA	CAAAAAATT
1701	ATCCGCAATT	AAAGCTGAAT	TAAAAGTTCC	ATGTGGTGGC	TCTCAACTAT	CAGCGGCCTT	ACAAGGCCAG	CTCCATTTTC	CAGCATCTCC	AGCTAAAAGT
1801	AGTAATTCAT	CATCTTCGGA	ATACCCGGCG	GTCTCTTCAT	CCTCACTCTC	ACTAAGTCAG	CAAGTTTATA	ACCAACAACA	AAATCAACAA	CAACAGCAAC
1901	AATACCAACA	ACAGCAACAG	CAACAGCAGC	AGCAGCAACA	CCAACAACAA	CAGCAGCAGA	AACAACAACA	GTCATTACCA	CAAATATCCA	ATTTGTTACC
2001	ACCATTAGCC	TCAATTCTGC	AACAAGGAAG	AAATATGTCA	TTCTTTCCAT	ACTGGAATAT	TAATCTACAA	ATGCTTGCAG	CTCAGCAGCA	AGTTCTTGCT
2101	CAAATGTCCC	CAAGTATGCG	TGAAACTACC	ATTCAAAATC	TACAGAATGG	ACAGTCTCAA	GTTATTGAAA	ATGATAAAGA	TTCCGTCCAG	GATTTCGAAT
2201	GCGAGACAGA	CGATGAATTT	AATCGTCGTT	CAAATGGAAG	CGCTATAGAC	CTGTCACAAT	CAAACGGAAC	ACCCACCAAA	ATCACAAATT	TCAATCAAGA
2301	TACATTTCAA	ATGTCAACCA	ATGTATCAAA	TGTGTTAGCC	GACAGTGGCA	TGCAGCAAAT	TAAAAGTAAA	GAAGAAATCA	ATACTCCAAC	AATAAGTTCA
2401	TCATGCAGTT	CACGGCGCAA	GGGACGTGTT	CTTAAACTCG	ATATAAATGC	AATCAATCGA	AATTTAGTAA	AAAGTCCCGA	AATTCATTCC	CATCAATCGC
2501	GATCGACAGA	ATCGCCATCG	TCTTCATTTT	TTGAAGAACC	AAAACTGCAA	GAAAAATCGC	CTCCTCATAA	CAGCATGGAT	ACCAACAGCA	TTTCCATGCT
2601	GAATTCTCCA	GAGTCATTAG	ATCTTTCATC	AACTGCCGTT	TCAGCTCCTG	TTCCATCAGC	ATTTACAACA	ACAACCATTA	GCAACACTAA	TAATAATTCA
2701	ACAACTCCAA	ACACTACGAG	CACAACTAGT	ACCACAAGTA	GTAGTAACAT	AAGTAATGTT	AGTAGTAGTA	ATAGAATTAG	CAACAATCCA	ACCTCCTCTC
2801	AAGGCAACAT	TTATGAGTGC	AAATATTGCG	ATATCTTTTT	TAAGGATGCT	GTTCTATACA	CAATTCATAT	GGGCTATCAT	AGTTGTGATG	ATGTTTTCAA
2901	ATGCAACATG	TGTGGTGAAA	AATGCGATGG	ACCTGTTGGA	CTTTTTGTTC	ATATGGCCAG	AAATCCACAT	TCGTAGGCGT	TTAATTTTTT	TCGCGAATTG
3001	ТААСТАТАТА	TATGACTCTA	TTCCTAGAAA	AGGAAAATAT	TTATGGTAAA	TGGTTAAAAT	AATTTTAAGA	GATTAACTCA	TCAAATTCTA	TTTATTTAAA
3101	ААААААААА	AAATGTTTA								

Source: 5' RACE product (1..737), 3' RACE product (580..1308), and an additional PCR product (504..3119), all amplified from cDNA templates of adult females.

# SEQ14. Lonchoptera hunchback, genomic.

1	ACTAATACTT	ATGCTGCTGA	TATTAACAAC	ATTTTAGAGA	CGGATTAACT	AGGTATGTGT	ATTGCAACAA	ATAACTGCAA	CTTTCGAATA	ATTTCAAAAA
101	TAAAAATAT	TTTTCTTGTT	TTTTCATAAA	TTATCAAACT	AATGTAAAGT	TAAATCGATT	TTTATGAATT	TATTTTATGT	CAAACATTAA	CAATAAGGAT
201	AAAGGACAAA	AATACTCTTT	CTTATTAGAA	TTAAGTAATT	TAATTTAATA	AATTACGGTT	AATGACGTCA	TGCTGTAACA	CTTCATCTAA	AATAATATAT
301	CCCTGAATTA	TTGCATTTCA	CTAAACCAAC	GTCGAATAAT	GTATTTGTAA	ACTTTTTCTA	TCTCCTTTTG	ACGAGGTGAT	TTATTTGCTG	GAAACGGTTT
401	AAAGGATAAA	ATGATATCAC	AATCAAAATT	TTGCTAAATC	AACAAATCCC	TAATAATAGG	TGATTTTTAA	TTGTTTGTTC	GTACCTGACG	GCTACTGAGT
501	GTACTCTTAA	CTATTTCCAT	CTCATAATTA	ATTATTTTTT	ACCTGAAATC	ATAAATCGTC	ATCCATGTAA	ATTATCTATT	CGGTTAAACA	AACGTATGAC
601	CGACCATGGC	CATGAAATTC	TGAATCAAGT	GAAAATCTCA	GTTGTGTCTT	GAAACCCTTT	CGAAAAGCAT	CAAATTATTC	AAAAGTTTCG	GTTTAGTTTT
701	TGTTTTTTTT	TTGGTTTTTT	TTTTGTGCAT	TTTCAAGTCT	AAATCTGCTT	TTAATTGTGT	CAGAATTTTA	ATTCAAATTC	AAAAGTGAAG	GTAATTAATT
801	TAGTTTTCCT	TTTTATTTAT	TTTGATATTG	TTTTTCGGGC	ATAAAAAAT	TAACAATTTG	TGACAACATT	CGAAAATCTA	AGAAATATTT	TAAAATAGCC
901	AAACCAGAGA	ACTGAAAGTC	ACATTAAAAA	ACTTGCCAAA	TCGCATCTGC	ATCATCAGCT	CATAATAAAT	TTTGTATTGT	ATATCCATCC	ATCCCACAAA
1001	TGAAATTTTG	TTATTCTTAT	ATTTTGTTAT	AAAAATTGTG	TTTAACTACT	AACATACACG	AATAAAAATA	TAAACAAATT	ATTTCAAAAG	GAATTTTCCT
1101	ААТААААСТА	AAGTAAATAA	ATAATCGCCG	TAAAATTTTA	AATTTACTAC	TTTAAAGTTT	GTATTTTATA	TATATATATA	TTTAAATATA	TATATATTCA
1201	АТАААААТА	AAGTGCAGTA	CAGAAAATAA	TCTCGTGTAC	CCCTATCCTA	TCCAATTGCT	TTATTCAAAA	TGATTAACGT	CATTCGGTGT	ACGAAAAAAA
1301	TCAAACGAAA	TTTAAAAAAA	AAAAAATAT	ААТААААТА	AACTTTAAAA	AATATATAAT	TAATAAAACA	TCTTGTGTGC	ATCATCTACC	AAGAACTTGA
1401	CCTGCACATT	TGGCGATCGC	ATTTTAAAAT	GCATTTACAC	АААААСАААА	ATCTAAATCA	CACGATTTTA	TTAATGTAAA	TTATTTTGT	TTAAACAAAA
1501	CCAAAAAATT	GCATATTTTT	AGAGAAATAA	ATAAAAATT	CTTTATTTTC	TTGTTTTGCA	TCAATTTTTG	CTGCTTAAAG	TTTACCCAAT	GCAGATCTTT
1601	CTCGTCTATA	CTATCTTAAA	TCGTACTCGT	ATCAACCTTA	ACCGTCCGCT	CTATACATGA	AAATGCAATT	ATTTACTAAA	AAGAAGAAGA	AGAAGAAAAA
1701	AAAAACTTTA	TTTTGGTTTT	TAAACTTTTA	ATAGAACTGT	GTTGGGCACA	AATGAAAAGA	AAACTTAGAA	AATTTTAAAA	TTTTTCCAAA	AAATTATTTT
1801	TTTAACCCCA	CACCATAACA	AAATCATTTG	CTCCGTTTTT	ATTTGATGGT	GGTGAGTGTA	CTCTGTAATT	AATCAATCTT	AAATATTAAA	TCGAAATTTT
1901	TTTCTAACTC	TAGGTTTTCA	АТААААААА	ААААААААА	TAATTGATGC	TTATTTTCAT	CCCTGCATGT	GATGCACTTA	ACTAACAGAC	AGAACAAAAC
2001	TAATCTAACC	ACCTTAACGG	CCAATGCATT	TCATTCAGTA	AACCATCTGC	TTTTTTTCAC	CAAAACAAAT	ACTACATGAT	GATGAAAATT	TCAGTTTAGT
2101	GCTCCAAAGA	AGCATTTCGC	ATTTCTGCCA	TTTGAAAATC	AGATATTGGA	AAATCTTCTC	CTTTTTTTTT	ACATTTTTAT	TCAATCATTA	GGTTAGTAGA
2201	GCCAACTTAA	TGTACTGTTA	CGGCTAAAAA	GACTCTAATA	TATATTACAT	ATATTCGCAA	CTTTTGTGTA	TAGGAGGATA	GTTTTTCTAA	TTAATATGTG
2301	TTTTGCTATA	ATTTTTTATT	TATTTAATTT	TTATTTTTTT	GTTTTTAATT	TTTAGTTTAT	CTACACTTGG	CCTCTAAAAT	AATTTCGATG	TGTGGATTTT
2401	TTATTAGATT	TGTTACTAAA	TTCTCATCTT	CCTCTAAATT	TTTGAATATC	TTGGTTGCAT	ATATGCGATG	CAGTCAGAAT	TCATGTAATG	TATGTATGTA
2501	CTTGGTTGTG	GCATGAGGAT	GGGCGATTAT	ATATCAAAAT	TCATGCATTC	ATGCATCGTT	ATTTCCGAAA	ATAGAATTTT	TGTTTGTTTT	TATCGATTAT
2601	TTCATTTTTG	CAAGAGTCCT	CAAATCGCCT	TTCCTAATGG	TGTTTAATAA	AGGTGCCGTT	TTTTTAAATT	TTTTTTTATT	TTTTTTGGAA	AAAACCAAAA
2701	AAAATTTCAC	TATCGTTTTA	TTTAAGATCC	TACAGGACTT	TGTTTAAAAT	AGTTTCACTT	TAATCCTATC	GACTTGTAAA	TGATTAACAG	TTCTTTTACA
2801	TTTGCGAATT	CAATTTGACA	AATACAAGTT	AAAAGTGAAA	GTAGATTAAG	ACAATACTAC	CTGGGTAGTA	ATTAAGTTTC	CCATCATGAC	AAGCCATTAA
2901	CCCATTTAAG	GAAAAAAAT	ААААААААА	AATTATTAAA	CAGTTGAATA	CGAGTTTTAT	CACTGTGAAT	GTTTGTTTTT	ATAACATTGC	ATTTGGAATA
3001	TGTAAATATT	TCTTTTGTTC	TTCTGAAAAT	TCTAAAGTAA	CAAATAACAA	TCCGAAGAAG	GAAAATTAAC	GGTTGCTTGT	TTATAAATTA	CCTGGTCCTT
3101	TTGTTTCTGT	ATATATCTAT	AATATTTTTA	CAGTAATACA	ACAGTGAACA	GTCAAAGTGG	TTTGACTTTT	CTGATGTTTG	ACCATCCCGA	CTCCCTCAAG
3201	TTTTATTAGT	TATTTCAATT	TTTGCAAATT	TCTTTTTTAT	ACCTATATTA	TTGTCTTTTA	AATCTGTCCA	TGTACAGTAC	AGGGCGTTAG	ATTAGTGTTT
3301	TAGTTTTAAT	CAGTTTCTTT	TTGGAAATCG	GATATTTTTC	TTTAACATGA	GACAAATAAG	TTTTATAAAT	TCAAATCTTT	TTGTATCATT	AGTTTTTGTT

3401	TGTTTTACTA	TTTATTGATT	TGAGTTGCAT	ATACATTAGA	AGATTAGTGA	TTCTTTGAAA	TGAATGTCAC	AACTAAAATG	CAAATGAATA	TTTTTTAACG
3501	GGATTTGTTT	TCTTTTCTTT	TTTTTTTCTAA	AGATATGAGT	ACATGACATT	ТТАТААААА	AGGTAAAATC	ACATAACAAA	AATCAAGCCA	CTCCCCGTTA
3601	AAAAAAATT	GTGACCAAGA	CGCCATTTTT	TCGGTATATT	TTTTACAATC	ААТАААААА	TTGGATGGTA	AAAAACTCTC	CACACAGCAC	AGTTTACGCC
3701	AATAAAAAAC	CTGGTAAACA	AAAAAAGCCC	TGACGGTAGC	AGGCACACGA	ACTGCGCGTG	CCCCTATTTA	AAGCAACCCC	TAAAAATTTG	AAAATTGTAT
3801	GAATACCAAG	ATAAAGAATA	TAATTCCACA	CCACGCTACC	ATACCCATAA	CTGAGTACCA	TAAAAACTTC	TCCCCTTCAT	TTATTGTGTA	TTGTTTTTTA
3901	TTTTTATTCCA	ACAATGTCTA	TCAAAACGCG	CCATGGTAAT	AATGGTTCAC	CAGGGTGTAC	ATAAGACCCT	CCTATATTTA	AAGTACACAT	CCGACTAGTA
4001	GTATATAATC	ACAATGAGAA	AAGATATAAA	АААААССТАА	ATGGAATACA	GAGGGCGAGG	GTATAGCACG	CGGTGGTTTT	TTTTTTTAAC	ATTTTTTCGT
4101	TTTGTTTTAT	TTTTATACGT	TTTTTTTTTC	GATAACTTTG	GGTTTTGAAT	TGTTTTGGTT	CACTTTTATA	CGAAGATTCT	AAATCGGATT	TGGTAAATTG
4201	AGAAATGAAT	TCTCAATCAC	CCTACGTAGA	AACATTAATA	CGTGAAACCA	TTTTTGTTTT	TTCATTTTGT	TTTTGATTTT	GCTTTTGTAA	TATTTTATAG
4301	TTGTAAAAAC	CCTTGGAGTA	ATAACATTTC	GCTCTCCACT	GTGTTTCCCC	TGGTTTTACG	GCTTTGTGTA	TATATACAGT	ATTTATTTGG	CTCGTGGCTT
4401	TAATATGGAA	GAATGAATTA	AAGATTTTTA	TTTGAGGTAA	TTCCTCTTGA	CTGAAAGTGG	ATTTATTTTA	TAGCTAATGT	ATTTTTGTCA	TATGGATTTA
4501	TATTAATTAA	AATCTTTATT	TTTTTTCAGA	TCGCTCTCCA	TCGCAGTTAA	ATTTGAAGTG	AATCAAAAAA	AATTTATCAA	AACAGTTGCT	ATTGAAACCA
4601	ААААСАААТА	CAAAATGCAA	AATTGGGATA	CATTACAACC	GACAGCCAGT	TATGAACATA	ACTGGTACGG	TAATATGTTT	CCAACAATTA	AAACAGAATC
4701	GATGACTCAT	TCGCCAACTA	CATCACAATT	AGAACAATAT	TTAACAATGA	AACAACAAGA	AATGAATTCG	TTAACACCAT	CACCGCGTGC	TGATATAAAC
4801	GGAACTTCAG	ACATTCAAAA	TTTCTTTGAT	AGTCACAATC	TGCAAAATGG	CAGTTTGCAT	CATAATCATC	CATTAGGATT	TAATCCATTA	ACACCACCTG
4901	GTCTACCAAA	TGCTGTTCTA	CCAGCAGTTT	CACATTTCCA	TCATAACATT	ACAGGAATCC	ATGCAAATAA	TCCAGCGAGT	CCACTGGCAC	AGCAATCCGA
5001	GGGAAATACT	CAATCATTAA	CACCACGCAA	TACGCCACCT	ATGGATATTA	CTCCGCCAAA	GTCACCCAAA	TTGTACTCCG	AATATGTTGA	AAAGGATCAC
5101	GATATGATAT	CAAATTCAAG	TGACGACACA	AAATTCCTTG	AAAGCGATGA	TGACTCATCG	ATTCGTACAC	CAATTTATAA	TTCGCATGGC	AAGATGAAAA
5201	ATTACAAATG	TAAAAGTTGT	GGATATATGG	CTGTAACAAA	AGTAGCATTT	TGGGAGCATG	CAAGTTCTCA	TATGAAACCC	GAAAAATTC	TTCAGTGTCC
5301	AAAGTGTCCA	TTTGTTACGG	AATTAAAACA	TCATTTGGAG	TACCACATTC	GCAAACACAA	AAATTTGAAG	CCATTCCAAT	GCGACAAATG	CAATTATAGC
5401	TGTGTCAATA	AATCAATGCT	GAATTCACAT	CGTAAATCCC	ATTCTTCTGT	GTATCAGTAT	CGTTGTGCCG	ATTGTGATTA	TGCAACAAAA	TATTGTCATT
5501	CTTTTAAATT	ACATTTACGA	AAATATGACC	ATAAGCCCGG	TATGGTCTTA	GATGAAGATG	GTTGCCCGAA	CCCTTCGATA	ATTATAGATG	TTTATGGAAC
5601	ACGTCGTGGC	CCCAAAATGA	AGTCGCAATC	CGGCGGTGGT	GGTGGTAAAA	TAAGTTCGGG	ААСААААААА	TTATCCGCAA	TTAAAGCTGA	ATTAAAAGTT
5701	CCATGTGGTG	GCTCTCAACT	ATCGGCGGCC	TTACAAGGCC	AGCTCCATTT	TCCAGCATCT	CCAGCTAAAA	GTAGTAATTC	ATCATCTTCG	GAATACCCGG
5801	CGGTCTCTTC	ATCCTCACTC	TCACTAAGTC	AGCAAGTTTA	TAACCAACAA	CAAAATCAAC	AACAACAGCA	ACAATACCAA	CAACAGCAAC	AGCAGCAGCA
5901	GCAACACCAA	CAACAACAGC	AGCAGAAACA	ACAACAGTCA	TTACCACAAA	TATCCAATTT	GTTACCACCA	TTAGCCTCAA	TTCTGCAACA	AGGAAGAAAT
6001	ATGTCATTCT	TTCCATACTG	GAATATTAAT	CTACAAATGC	TTGCAGCTCA	GCAGCAAGTT	CTTGCTCAAA	TGTCCCCAAG	TATGCGTGAA	ACTACCATTC
6101	AAAATCTACA	GAATGGACAG	TCTCAAGTTA	TTGAAAATGA	TAAAGATTCC	GTCCAGGATT	TCGAATGCGA	GACAGACGAT	GAATTTAATC	GTCGTTCAAA
6201	TGGAAGCGCT	ATAGACCTGT	CACAATCAAA	CGGAACGCCC	ACCAAAATCA	CAAATTTCAA	TCAAAATACA	TTTCAAATGT	CAACCAATGT	ATCAAATGTG
6301	TTAGCCGACA	GTGGCATGCA	GCAAATTAAA	AGTAAAGAAG	AAATCAATAC	TCCAACAATA	AGTTCATCCT	GCAGTTCACG	GCGCAAGGGA	CGTGTTCTTA
6401	AACTCGATAT	AAATGCAATC	AATCGAAATT	TAGTAAAAAG	TCCCGAAATT	CATTCCCATC	AACCGCGATC	GACAGAATCG	CCATCGTCTT	CATTTTTGA
6501	AGAACCAAAA	CTGCAAGAAA	AATCGCCTCC	TCATAACAGC	ATGGATACCA	ACAGCATTTC	CATGCTAAAT	TCTCCAGAGT	CATTAGATCT	TTCATCAACT
6601	GCCGTTTCAG	CTCCTGTTCC	ATCAGCATTT	ACAACAACAA	CCATTAGCAA	CACTAATAAT	AATTCAACAA	CTCCAAACAC	TACGAGCACA	ACTAATACCA
6701	CAAGTAGTAG	TAACATAAGT	AATGTTAGTA	GTAGTAACAG	AATTAGCAAC	AATCCGACCT	CTTCTCAAGG	CAACATTTAT	GAGTGCAAAT	ATTGCGATAT
6801	CTTTTTTAAG	GATGCTGTTC	TATACACAAT	TCATATGGGC	TATCATAGTT	GTGATGATGT	TTTCAAATGC	AACATGTGTG	GTGAAAAATG	CGATGGACCT
6901	GTTGGACTTT	TTGTTCATAT	GGCCAGAAAT	CCACATTCGT	AGGCGTTTAA	TTTTTTCGCG	AATTGTAACT	ATATATATGA	CTCTATTCCT	AGAAAAGGAA
7001	AATATTTATG	GTAAATGGTT	AAAATAATTT	TAAGAGATTA	ACTCATCGAA	TTCTATTTAT	TTAAAAAGAA	ATAAAAATG	TTTATTTTCT	GCCTTGTCAT
7101	CTTTTGTACA	TTTTTTCTAAA	TACTGAAAAT	TTAACTTAAC	TGATTTGTAT	TTTTTATCAG	TTTAACAAAG	TATTTTAGAC	TTTTAGAAGT	TGTAATGAAT
7201	TTTTTTAATG	AAATACATTA	ACTAAATGAG	TTATAACAAA	TAAGCACAAT	GTTGTCATAA	TTGTACATAA	AAATGTTTTA	TTTTAAGTTA	ATACATTAAC
7301	ATTTAGTTTT	AAAGTTTATT	TTATTATTTT	CGTTTTTTCT	TAAAAGTTTT	TTTTTTTTTA	TTTAAATAAT	TAATTTAGTT	TTAATTAACC	AAATTATGGC
7401	AATTAGTTTT	ATCCTTATTT	АААААТАААА	CCATAAATAA	AGGACCATAA	ATGGTGGTAA	AAAACCGTAT	AAAATTAACC	AAAGCCATTT	TAGCCATTAT
7501	TATTATCCTA	AACCATAGCC	ATGGCCTTTA	AAGGAACCAT	AGTCAAGGAT	ATATATATCC	TTATATAACC	AAAGGAAATA	TTTTTTTTTT	AAGGTAGGTA
7601	AAGTAAATCC	CACAAGAAAA	AAGGTGTTCC	ATGGTAAAAG	AATTGGACAT	TTATTTGGTT	AAATTTTAAT	AATTTTAATA	ATTTTAAAAA	AAACCGAACT
7701	TATATTGGTG	GTGAACTATT	AAA							

Source: Two different PCR products (1..4557, 2726..5470) amplified from independent genomic DNA templates, and phage Llu-hb ph2 (2726..7723). Alignment with cDNA sequences: Positions 1..50 correspond to parts of the first exon of the *Lonchoptera hunchback* P1 transcript (SEQ13), and positions 4528..5280 correspond to parts of the second exon of the *Lonchoptera hunchback* transcript. Two putative polyadenylation signals were identified in the genomic sequence (7424..7429, 7436..7441), and one NRE sequence (7249..7269).

### SEQ15 Empis hunchback, cDNA, partial P1 transcript.

1	AGTCATGTCT	GATGTCTTAT	AATGAGAAGG	ACGTGTGTGC	АСТАААААА	AAAAAAGGT	GAAAACAAAA	AACATTTAAT	TTACACCGAA	ΑΤΑΑΑΤΑΤΤΑ
101	TTTTTTAATT	ΑΑΑΑΤΤΑΑΤΑ	AAAAAATGAA	TTAAAAGAGT	TTAATTTTGA	ATATCTTTTT	TTTTTTTTTT	GGATATTTAA	TAATTTATAA	GAATTTTTAG
201	TTTTAAAAAC	GATTTATTGG	AATAAAAATG	TGTTAACGGA	ATAAATTAAA	AGTGTTTTTA	TTAAATAAAA	САААААТАА	AATTAAATTA	AAATTAAAA
301	AAAATTGATT	GAACAATAAA	ΑΑΤΑΑΤΑΤΤΑ	AAAATTAGCG	CAAGAATTTT	TCAAGAGATA	ATTAAAAATT	TGTGGATTAT	AACGTTTTCA	GTCGTTCAAG
401	TACGAAATTT	AAAAGAGTTT	TTCAATATTT	TTGAGATCCA	TAATGCAAAA	TTGGGATTCA	TCATTACAGC	CAGCTAATTA	TGAAAATAAT	TGGTATGGTA
501	ATATATTTCC	АСАААТТААА	ATTGAACCAC	AAACAGCAAT	TAATAGTAAT	GATAATGGTC	CACCATCAGC	ATCATCATCA	TCACTACCAC	TATCAACACA
601	ACAACCACTA	CCACCTCCAA	CATCAACATC	AATATCCGTG	ATAGATCATT	ATTTCAATAT	TAAAAATCAA	CATTTACAAA	ATCAACAAAA	AGAAATTGAT
701	ACGAGTTCAG	TATTAATACA	ATCACCACGC	AATGATAGTA	ATGATGAACA	GCAATTTTTT	GATAATAATA	ATACTAAAAC	TAATCATATA	AATATTAATA
801	ATAATGGTAA	TAATAATTTA	TTGCATCAAC	AATTGCATCA	TCATCCATTA	GGTTTCAATC	CATTAACACC	ACCTGGCTTA	CCAAATGCAA	TTTTACCACA
901	AATAAATTAT	AATAATAATG	ATTTGTCATT	AGAATCAATA	CAACAACGTC	CATCATCATC	GATAGGACAA	TTAAATCATA	ATCAACAAGA	ТАТАААТАТА
1001	ACAAAATTAG	ΑΤΑΑΤΑΑΤΑΑ	TTCATTAACA	CCTCGAAATA	CACCACCAAT	AGATATAACA	CCACCAAAAT	CACCAAAAAA	TGGACCAACA	AATATGTATA
1101	ATAATAGTGG	CAATAGTGAA	TATTTAATAA	TTGATAAAGA	TCAAGAAATT	TTATCAAATT	CATCAGATTT	AGTTGATAGT	GATGATGATG	AAATAATACG
1201	TATGCCAATA	TATAATTCAC	ATGGTAAAAT	GAAAAATTAT	AAATGTAAAA	GTTGTGGTTA	TATGGCAATA	ACAAAAATTG	CATTTTGGCA	ACATACACGA
1301	TCACATATGA	AACCAGAAAA	AATTTTACAA	TGTTCAAAAT	GTCCATTTGT	TACAGAACTA	AAACATCATT	TAGAATATCA	TATTAGGAAA	CATAAAAATT
1401	TAAAACCATT	TCAATGTAAT	AAATGTAATT	ATACTTGTGT	ТААТАААТСА	ATGTTAAACT	CACATTTAAA	ATCACAT		

Source: 5' RACE product, amplified from an embryonic cDNA template. The sequence of an independent 5' RACE product, which was amplified from a cDNA template prepared from ovaries, aligns exactly with the embryonic cDNA but lacks the first 47 bp of the embryonic cDNA.

# SEQ16 Empis hunchback, genomic.

1	ATGAGAAGGA	CGTGTGTGCA	СТААААААА	AAAAAGGGTG		ልሮልጥጥጥልልጥጥ	тасасссааа	ጥልልጥልጥጥልጥ	<u> </u>	ልልጥጥልልጥልል
101	АААААТСААТ	TAAAAGAGTT	ТААТТТСАА	ТАТСТТТТТТ	ጥጥጥጥጥጥጥጥር	GATATTTAAT	AATTTATAAG	AATTTTTAGT	TTTAAAAACG	ATTTATTGGA
201	ATAAAAATGT	GTTAACGGAA	ТАААТТАААА	GTGTTTTTAT	ТАААТААААС	ΑΑΑΑΑΑΤΑΑ	ΑΑΤΤΑΑΑΤΤΑ	ΑΑΑΤΤΑΑΑΑΑ	AAAATTGATT	GAACAATAAA
301	ААТААТАТТА	AAAATTAGCG	CAAGAATTTT	TCAAGAGATA	АТТААААТТ	TGTGGATTAT	AACGTTTTCA	GTCGTTCAAG	TACGAAATTT	AAAAGAGTTT
401	TTCAATATTT	TTGGTAAGTT	TTTTTTAAAT	TTTAATTATT	GTTTTTAATA	ААААААААТ	GTTTTTAATT	TTAAGTGAAT	AACAAGTCAC	AAGATATAAT
501	TTCTACATAA	AAACTAAAGC	ААСААААААА	AAGATGAATT	TTCAAAATTA	ATTATGAAAA	TTCATCTTTT	TTATTTATTT	TTGCTAATTT	TAATTTTAAT
601	TTAATTTAAT	TTAATTACAA	AAAATGTTAA	GGAATATTTT	ATATTTTGTC	TTGCGCATTG	ATTGAAAACA	ATCTACGACA	TTGTTGATAT	TTTTTAATAT
701	TTATACTTAG	TGTCACACAT	ААААААТААС	AAGTTTTGTC	ATTAAGCTCA	TCATTAAATT	AACCTCTTCT	AAAAAGGATA	ACAATATAGA	AAAAAAAGT
801	ССААААААТА	TATTCACTGA	AAGTACTATG	TTGTCAGATA	GTGAAAGAAA	TATGATGTAT	ACATAATTTA	ATTTATTCCT	TTAACAAATT	CATTTCTGTA
901	AAATTATCAT	TTTTCAAAAA	AAAATTAGCG	ATTACCTCTC	TATTGTAAAA	TGTAATATTC	GTTTCAAGTT	TTATACAAGA	CTATTTGCCG	GTGTAGACTA
1001	TGACAATAAT	ATATAAGTTT	AAATAAATTA	ATAGAACTTT	CATTAACTTA	TTTAAGTTGC	АААААААААА	ATTACTATTA	AATTAGGTTT	GCAATATCAT
1101	TATTTTGACA	CTTCTTTTTC	CAATTCTAAA	GTAAATCCAA	AGCACGTTAA	TAGTACAAAA	TGTAATTTAT	CAATTGCGTG	GGTAATCATC	TGTTAATGAC
1201	ATGAAATGTA	TTTTTTAGCT	GAGTTTTATT	TTAAGGATTT	AATTTTATAT	TATATTAATA	TATGTATTAT	ATTTTTTTTC	TACTGGAAGT	AGAGGAATTT
1301	ACTGGAAAAA	GTAGATGTTC	ATTACGTTTA	GTTTATTATT	CAAGAATTAG	AGATTTAAAT	TGTAATTCTT	TAAATATCCA	AGGATTAGGT	ATTTATTACC
1401	TAAGCCTGGT	ATATTTAAAG	AATTACAATT	TAAATTTTTT	TCATTTTTT	GCATTGTTAT	AGCTTATTAT	GCCCAACAAA	TTAGTTTCAA	GCCAATTATA
1501	ATTAATGATT	TTCTATGGTT	TAGTAAGAAA	ATGGTAATAA	TTTAATATTA	AGTTCTATTT	TAATAAAAAT	TACTAATTTA	ATAACAATTT	ACCACTTACG
1701	COMMONAGIT	1 TAGGTAATA						AGTATATTAA	AAAAAAAATAC	
1801				THE	CHAIAAAAA	TTTCACCATC	CCACAAAAAA	1GACICAAIA		
1901	CCTTCATTA	TGCATITICT	AACACCATAT	TIGITI IGCA	AGAAAAAATT	ATAATAATAT		CACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		TTTTAACIIG
2001	TTGTTTTTAT	AAAATTAGAA	AAGAAAATCA	TTTAAAAACA	ТСАСТААААА	СТСТАТААТА	AAAATTCGCG	AAATTTAGCA	ATGCGCATCT	CCAACATCAA
2101	AAAAATGCAC	ACACATTTAA	AATTAATTTT	ААААТААТАА	TTCAAAGTCA	GTAAAAAAGT	TGGCAAAAAT	TAAATGCATC	TGAAAAAAAA	ATTTATATTT
2201	TTTATAAAAA	TATCTCACAA	AAATGAAAAT	GAAGAAAAAC	TTGACCGGAA	AATATTAAAA	ATCACTAATA	CATCCAAAGA	TGATTTTATA	TTTAAATTAA
2301	TTTTTTTTTT	TTTGTTGCCA	TAAATTTTAA	AGCTAAATTC	ATTTATAAAA	TTAGTTTTTG	GAAAAATTTA	TTGTTTATTA	GTTTTAAGTT	ATTATTGCTT
2401	TGTTTTTGTC	AGCCCTTGAC	TTTTTTCTTAC	TTAACTACAG	GTTGATTTAA	TGGCTGTATA	TGACGACTTT	TAAATGAGTT	TTAATATTCT	AAAGATATCG
2501	ТАААААААА	CAATGAAATT	TAAGCGTTTC	AAGAAAAGAA	AAGTCAATCT	AACCTCAATG	TTATTAAATT	TGGGCTCCTA	CCGTCTGGGA	TCTACACCTT
2601	GAAAATCTGT	CAAAGAAAAA	CCGTACTTGA	AACATTGTCT	TAAAGTACAT	ATCTTTAAGT	AATTTTTTAC	TGATTTAACT	GACAAACAAA	TAATTTGTTT
2701	GATATTTTCT	TTGACTAATT	TTCAAGTTGT	ATACTGAAGT	AGAGCCCAAA	CGTGATACAT	TAGATTTTTA	TTGTATTGAA	ATTAAAATCA	TAGTTTATAT
2801	TTTCATTCAC	GAAATAATTT	AATCTACTGA	AATATTTTTA	ATTAAATACG	TGCTAATTAT	TTTCAGTAAT	AGATTGAATT	GTTGAGTAAA	ATTTTCTGGT
2901	TATTTTCTGG	TAATAGTTTA	AATTTGTGCA	AAATTTTCTC	ATTTGGTTCA	ACCGAATTTT	TTGCCATAAT	TGGTTAAAAT	ATTTTAGTTT	ATTGGATTTT
3001	TCAAATTAAT	ATAAT TAGA	TCTTCTAAGA	ATATTAGCCC	ACATTGTATT	TAAATTATAT	GCCACCAAAT	TATTGTTACG	ATATTTATT	TATTICTTAA
3101	ATTTATAAAT	ACCCTATEC	ACOMMONIA	TGTATGATCA	TGGACCCCAC	AAAAAATTAT	ATTTAAGTTT	TTGAAAAAGA	TCTGAATTTA	AATTATAAAA
2201				MAGCIIIIIA	CCCAMCCAM		AACTITCATC	GAAAAAAAII	ANNANAMAC	
3401		CCTTTAGAAT		TITIGCAGII	GGGAIGCAII TATTCTTTTT	TACTTTTCTA	ACAICGCAAI	TTTTA A ATA A A	AMAMAMAIAC	ATTATAAATT
3501	ТАТСТТАААА	АТАGААТТТА	САСААААААА	AAACGAATTT	GAAAGAAAAT	TGTGTTGGCG	ATGCATTCAG	TTTTGCGTTA	TAATATGTTT	ттааатааас
3601	TTTAAAAAGC	GTATTAGAGA	GATTTTCTGA	AAATTTATTT	TTCTAAGTTT	TTTTCAAAAT	TTTATGGGAA	CAATGGTCAC	TTTTAGGATG	GTTTCTGCGA
3701	AAATATTTAT	TATGGTCACT	TTTCATTTAA	GGAAAAAACC	CTTCAAATAT	TTTTATAGAT	GAAAAAGAAA	ATGTTTTTTT	TTTCCAAAAA	ААААААСААА
3801	AAAATTGATT	AACTTTATGT	TAATTGCGGG	ТААТААААА	AAAAGAACCG	GACTTGGAGT	TAAATATAAT	TAAAAGGTAT	ACATACACCA	GAAATCATTT
3901	TTTGAAAATC	ATAATATAAT	ATGTAAAAAT	AATATTATTG	TTAGTTTTTT	CCCTTTTACT	TTATTACCTG	ATCTTTAAGG	TAAGATGACA	CTTATTTTGT
4001	CTGTATATCA	GTTAAAAGGT	CTTCACTTAA	AATCATTTTT	AATTTACATT	TATATAGTAA	AATTTTTACG	ATATGGCTTT	TTGTCTATGA	GAAAGAAAAA
4101	TATAGAAATT	TATTTGATTT	TAATCTATTT	AAGATGGTGG	TTTTTACAGA	GATATTTAAA	AGTTTGCAAA	ТТААААААА	ACTTGTCTTT	TTTTTTATGA
4201	TATATCATAA	AATTTTGTTT	CTATCAGATT	TAGTTTGGTT	TTACTGAACA	TTGTAAAGGA	TTGTTTAAAT	AGTGAACCTA	AAACAAACTG	AAATAAATAA
4301	AATTTAAATA	TGGCAGGGTG	ATTTATTATA	TTTTGTGTAA	GTGTAATAAA	TTTAAAATGC	AGAAATATAA	CTTTTCATTA	AATTTTTAAG	TTCATGAAGG
4401	TCACTGCCAA	ATTTGTATGT	AAAGCACTCA	СТСААТАААА	TTTTAAATTT	CAAAACAACA	ATACAAAATG	ACAATTTCTT	AATAGCCTTT	AAATTATATC
4501	TGCTATCTCA	TAAATAATGC	TAAATTTATG	TTAATTCAAA	ATGGGTCAAT	TATATACAAA	TTCATCTTCT	GGTACAGAAT	TTTGGGTATA	ACTAAATTAA
4601	CTTAAATTGT	AATATTTAAA	TCTAAATCGA	TTAATTATAC	AATAACGACA	AATATACGTT	TGAGTTTAAA	ACTACTTAGA	TTTCAGCACT	TTTTCAGATT
4701	CLIGCAGCII	ACCA MUMA MU		ADDATES	CUURAGAGAAAT			AATATTGAA	AAGAAACTAA	AGGTGCATAG
4001		CGACGACATT	CTTTCACACA	ACCOTOTT	TATCTCAAAA	TICCACAAAA		TTTTCICAL TTTTC ATTTTA	ATTAICCAC	TTTTATTTCIAA
5001	GAAATAATTT	TATTGTATTT	CAAATCGTGG	ΑΑΑΤΑΑΑΤΤΤ	TGAAATAAGT	GAAAAAAAA	CAGTGTGAAT	TGGTGGATGC	TTCTTACTT	TTATCCTCCC
5101	ACCAAAAAAT	GTTTTATGAC	AAAACTACAA	CCGCGAACCA	ACAAAAAAAA	TTATAATTTT	TTGTGAGGTT	TAAGAATTCC	AAAAAATATA	TACGAGGGGT
5201	GAGTTAAACA	CATTACAAAA	TTGTTGCCAG	AGGGGTCGTG	TAGTCAATAT	GAGCTTATAT	ACTTTAATGT	TGAGTATAAA	AAACAAAGAG	GATAAACTCA
5301	TATATATGTC	TTTATATAAA	TTTAAGTAGA	TCCTAGGGTC	GATTTATAAA	TATTACATTA	TTATGATTAT	TTTTTAGTTT	TTTCAGTATT	CAGGAAGGGG
5401	GAGAATTACA	TTATTATTAT	TATTACATTT	TTTTTTCAAC	ATTGTGATTT	TAATGTTGTT	GATGTTACTT	GAATTAGGTA	TGAGATGGTT	TTTTTATGTG
5501	ТАТААТАТАА	ATTTATATAT	ATACATATAG	ТАААААТАТА	AAATTTGGTT	GTGAGACACG	CGCAGTTCTT	ATGCCAACAT	TACAGAATTT	ACCCCCTCCT
5601	ATATATTATG	TATACAAAAT	ATATTTTTA	TTCATGTCTA	CCCCGAAGGA	TATGAGTTTT	TTTTTTGATT	TTTATAAAAA	AATACTATTT	AGATTAATTT
5701	TTTATTCAAT	TCTGGTTGGT	TGGAAATTAT	CAACGTTGTT	TTTACGATTT	TTTTTTTAAA	TGGCGACAAG	CATAATGGCG	TATATAATAA	TTGTATTTAA
5801	AAGAAAAATA	AAAAGTATTC	TAATATTAAT	AAATTAAATT	TTTTAACTTT	CAGAGATCCA	TAATGCAAAA	TTGGGATTCA	TCATTACAGC	CAGCTAATTA
5901	TGAAAATAAT	TGGTATGGTA	ATATATTTCC	ACAAATAAAA	AT					

Source: PCR product, amplified from genomic DNA. Alignment with cDNA sequences: Positions 1..413 correspond to parts of the first exon of the *Empis hunchback* P1 transcript (SEQ15), and positions 5854..5942 correspond to parts of the second exon of the *Empis hunchback* transcript.

# SEQ17 Haematopota hunchback, cDNA, P1 transcript.

1	GTGGAATTAA	GACAAAAATT	TTTTCGTAAA	TGAGTTTTTT	TTTTTGGGAA	AATTAATATT	TTTTTCGCAA	ATGAATAAAA	AAAAAAAGT	AAATAAATGA
101	CAGTTTTGAA	AAAAGAAATT	GTGAATTTTT	GTGTTAATAA	ATTTATGTTT	TCGAAAAAAC	TTATGTGAAG	ATAAAAACAT	TTAAAGTGAT	TTTGTGAAAA
201	TTATGAAATA	ATTTGGACGC	AATTTACGTG	TTTGTTTAAA	AAAAAGGAT	ACCAAATTTG	TTGAACAATA	TTCACCGGAA	TAGGATAGAA	AGAGACGTAT
301	CTCATCATGC	ATGGTTGGGA	ATCATTGCCG	CAAGCAACAT	ATGATCATAA	CTGGTGTGGA	AATATGCTAC	CAATTAAAAC	AGAACCACAA	ACTACAACAT
401	ACCCATCAAT	GGAACATCAT	CATATGCATC	ATATGAATCA	AAAAACAAGT	TCATTGGGTG	GCGGAAGTTC	ACCACTTTCT	ACCCCCAGCA	TGGATGGAAT
501	GGAAACACAA	AACTTTTTCT	ATGATCAATT	TAGCAGTTTA	CATCGACCAC	TTGGTTTTAA	TCCACTCACA	CCACCTGGAT	ATCCAAATGC	TATGATACCA
601	CAATCGAGTT	TGCATCAAGA	TTCTATGTTA	AACGTAGACC	AATCACCGTT	ACAGCAACTT	AATCATGGTA	GTATATCACA	ATTTGCTGCT	TTTGCCAAAA
701	ACGATGGTAG	TAATCCATCG	TTAACACCAA	GCCATACCCC	GCCAATGGAT	GTTACACCAC	CAAAATCACC	AAAATTTCCC	GTTGATATAC	CAACACCGGA
801	AAAGGATAAT	GATTTAAATT	CAAATTACAA	TGATTCAGAA	GATACACGAT	CATTGGAAAG	TGATAATGAT	GATGAATCTA	TACGTACACC	AAAGATAAAT
901	TCACATGGAA	AAGTTAAAAA	ATTCAAATGT	AAACAATGTA	ATTTCATAGC	TGTAACTAAA	CTAAGCTTTT	GGGAACATAC	TAAAGGTCAT	ATAAAACCGG

1001	АСААААТСТТ	AAAATGTCCA	ΔΔΔͲΩͲϹϹΔͲ	ттстастса	СТАСАААСАТ	САТТТССААТ	<b>ΑΤΓΑΤΤΤΑΓ</b>	ааатсатсат	GGCTCGAAAC	CATTCCAATG
1101									almama ama	
1101	TAACAAATGT	AGCTATAGTT	GIGITAATAA	GTCAATGCTA	AATTCACATT	TGAAATCACA	TTCGAATATA	TATCAGTATA	GATGTGCTGA	TIGIAGTIAT
1201	GCTACTAAAT	ATTGTCATTC	ACTAAAGCTG	CATCTACGAA	AATACGGTCA	TAAACCAGCC	ATGGTCCTGA	ATGAAGATGG	AACACCGAAT	CCACTGCCAA
1301	TCATCGATGT	CTACGGTACA	CGAAGAGGCC	CGAAGATGAA	GTCATCGAAG	AAACGTGATC	CACCATCACA	GCAACTCTTC	AAACAGGAAA	CACAGAATGG
1401	ACCATCATCC	CTCGCACAAA	ACTCTAAAAC	TCCACCCCTT	CAGCAACAGC	AACAATCGCA	GCAATTTCAA	GCGACTGTGC	CAACCCCATC	ACCGGCTAAT
1501	CTAATGTCCA	ATTTCCTGCC	TACAACATTA	GCGAGTATGT	TGCAACAGAG	CGGCAACACG	ATGCCATTCT	TCCCGTACCT	GAACCTAAAC	CTTCACATGC
1601	TAGCGGCACA	ACAGCAAGCG	GCTCTCGCCC	AAATGTCACC	AAACATGCGA	GATGAGACAA	CGAACATGAG	TAAATGTGAG	AGCGATGAAG	AAGATGCTAT
1701	GAGTGACTAC	GAAACTGATG	AACGATGTGA	GAGTCGAATC	GATAACGATG	CCATGGATCT	GTCGCAAACA	ACCCCAACGA	AAAATGTTGC	CAACCAGAGC
1801	GACCCCATTG	AGCCACCCAA	AGAAATACCA	ACAACACCCT	CAACAGTTAC	ATCAACATGG	CGGAATCATA	GGAGGAAAGG	TCGCGCGTTT	AAATTAGACT
1901	CGTCAGTAAC	ACCCGCAGAA	AATGAGAATC	TAACAATGGA	CACCACCCTT	CTCAAGAAGC	AACCAACGGA	GGTTATTGAA	ATGGATAACT	CAAGTCGGTT
2001	GGAAATGTCG	GGGGATGAAG	ATGTTCCAAC	ATCATCATCA	TCGGTGGTGT	TGGAGAACAA	AGACGATGCT	AGTGATGAAA	CAAATAAGAA	ACCAGAGAGC
2101	AGCACAACAC	CTTCCCTGGA	AGTAGAGAAT	AAAGAAACAT	CAAAGAGTAC	CTCACCGAAC	AATGCTAGTA	ACTGCACGCA	AGAGAACTAC	GAATGCAAAT
2201	TTTGCGGCAT	CTCTTTCAAG	GATGCTGTCC	TGTACACTAT	CCACATGGGC	TACCATGGAT	ACAACGACGT	CTTCAAATGC	AACATGTGCG	GCGAAAAGTG
2301	TGAAGATCGT	ATATCATTCT	TCCTACACAT	TGCCAGGAAC	CCACACTCAT	AAAGTTCTGC	ATCTATAAAG	TCGAATTTTA	CACTACGAGC	GAGAAATTTT
2401	ACGACTTGAG	CTTTCCAACG	AATGATTTTC	ATTCGTCAAT	TTTGTCCGTT	A				

Source: 5' RACE product (1..1091), 3' RACE product (1065..1433), and an additional PCR product (221..2451), all amplified from a cDNA template prepared from ovaries.

### SEQ18 Haematopota hunchback, genomic.

		_	-							
1	AGCAAATATA	ATATTACACA	TGAGCGCCCC	AGAGTACAGC	CATTTGTGGG	ATCCCTGCGG	AACAAGAGCA	AAATTTCCTT	CATATTCTTA	TGCTTCGGTA
101	GATTGACGAG	GAGAAGACTT	GGGTCTGTGG	TCTTAAAACC	AAGTGAATTG	TAACATATGT	ACATGTCCCG	TTGGATACCT	GCTAAAGTTC	TCCGAGAAAA
201	TAGGCTTGCT	GAACGAGGAA	TAAAGTGGAG	AAGGGAAAAA	CAGACCACTG	GATGGAAGTG	GCAAGGGGCC	AAAACAGAAC	TGCCCCACCT	TTCTAATGAA
301	ACAAAATGGG	GTGGAAGCAT	TCATTTTACT	AAACTTTTAA	CTAATTTGAT	AAAACTTTTT	TGGCAAATTT	TTTTTTATTTA	TTGTTTTTTG	TTATTATATT
401	TAATAAATAC	GCCAAATAAC	ACATTTTCCG	TTGGTCAAAA	GTGGAAAACC	TTAAAAACAC	AAATCAGCTG	AGAGAAATTT	TTCAAAATGT	CATCAAATGA
501	GGGGAAAACA	AACATTTTCA	GCTATAAAAA	GTCCTTGGCT	GAATTCCCAT	TTGTCAGCGC	TAATGCTGCG	TTTACACTCG	AAACTCACAC	ACACCGCACA
601	TGCAAACGTG	AATTTTCCCC	TCCAGGACAC	TAATTTCCTC	TCACATCAAA	TACTCGTCTC	TGATTGGTTC	AAAATCGGGA	AAACATTCTA	AACGATGAAG
701	AGAGACAGCA	ATACGAAACA	AAAAATGCAC	AGAAAAAAAA	TTTCTCAAAT	ACCATCGCAA	TCCGGCATTT	TTCCGTAATA	ATTTTTCTTC	AAAAAGTGAA
801	AAAAAACGA	AGATGGGAAA	AACTTTTTGA	AAATGCGTTT	TCGGAAAAAC	GAAGGGGCAG	GGGTTGGTTA	ACGAACGGAA	CGTGTCAGGT	CGATAAAAAT
901	CTTAAAATTT	TTCAGTACGA	ATCGGGATGT	CGTGGAATTA	AGACAAAAAT	TTTTTCGTAA	ATGAGTTTTT	TTTTTGGGAA	AATTTATATT	TTTTTCGCAA
1001	ATGAATAAAA	AAAAAGTAA	ATAAATGACA	GTTTTGGAAA	AAGAAATTGT	GAATTTTTGT	GTTAATAAAT	TTATGTTTTC	GAAAAAACTT	ATGTGTGAAG
1101	ATAAAAACAT	TTAAAGTGAT	TTTGTGAAAA	TTATGAAATA	ATTTGGACGC	AATTTACGTG	TTTGTTTCAA	AAAAAAGGA	TACCAAATTT	GTTGAACAAT
1201	ATTCACCGGA	ATAGGATAGA	AAGGTAAAAC	CAAAATTAAT	ТААААААТА	AATTCAAATT	CAAAAATTTC	TAAAAATTTA	ATGACACATA	ACGGGACTTT
1301	GGCTAGCGAG	CCGTGAGGCT	TAATTTAGTG	GAATTTTCAA	ААААААААА	ATATTTTATT	TTTATCAATA	ATTCATTAAA	AAACTTGAAT	TGTTTTAATG
1401	CATCTGACCC	САААААААА	CACAAAAAAT	CCCTGCAGGC	TATTGAACGA	ATATTCCCCA	AAAATTCCAA	AAACGAACTT	GACAGTCGTC	GAGGTTGTTT
1501	CCATTCCGTT	GAGGTTAGAA	AAAAAAACAT	TTTTTTTTAT	TACTCGATTT	TGGATTTTCA	ATTTCCAATG	ACCTTCCGGA	ATGCAGGCAA	AATTTGGTCA
1601	AAAATTCACA	AAAATTTTGT	TTACTTTAAG	AAGCAGTAAA	AACTAATGAA	CGGATTAGAA	AAAATTTCGG	GGAGGGCAAA	AAGTCAATTA	CATCCATAAA
1701	CTGCACTGCA	TCCCTTGCAT	TGTAACGGCT	TTAAAATGGG	GGTGGATACG	TGTATAGGGT	CTCATAAAAT	TTTATTGAAA	TATTTGCTTA	AAAATACCCT
1801	TGCGAACAAA	CATTTTTATA	ACGACGGAGA	GAGACACCAT	TTTGTTTGAC	AATTTATTTT	TAAAAATCAA	TTTTTGGGGT	TGGTTCATAG	CAGGAGGGGT
1901	AAATATAACG	AATCAAAAAC	TAAAAAAACG	AAAGACGTCC	CACTTCTTTA	CACTCCTCAT	TGGTGTTGAA	ATAAGTACAT	АААААААААС	ATTTGTAGGC
2001	АТАТААААА	TACTAAGAGG	GGTTGCTCAA	CACAAAAAAA	AAATTTCATC	CCCTAAATAA	AATTTGCTAA	AAAAGTAAGA	ATCGCGTACG	TTTGCCACGA
2101	GAACTGCGCG	TGCCCCTTCA	ATATAAACAT	CATCCGGGGG	TTGTGTTGTA	TGGTAGGGTG	GTGTTGAATA	CATTTTTGTT	GTTGTTAAAT	CACTCAAAAA
2201	GGAAAAAAA	AAGAAGCAGG	GTAACTACCC	TCGTGTTGTC	ATACACATTT	CTATCTCATT	TTATCCAATA	CACATTCGTG	AGTAAAGTCA	CCATTTTTCA
2301	AGTATATAAG	CGTATTTTTA	TACTTAGAAT	TCTTTAGTTT	TCTCCCTTCT	GTACCAAAAA	CCATCCCTTC	AAGCTTAATG	ATCCTAGTAT	GGGGTGGTAA
2401	AATATAGTAA	AAAAAATGTC	GAGGGGAGAA	TTTTTTTTTCT	TCTGTTCACA	ATTCTGGTAT	TTTTTTCCAT	TTTGGAAATC	ACATTTTTTT	TTATTCTGTT
2501	CCCTGTGATG	CTGCCCTTCC	TTTGATTCAC	TTAAACCGAT	TGGAAATCAA	GTAGAGTCCT	TGAGAGGGAT	GAGGTTGTGA	GGGAGACTCT	AGGAGGATAA
2501 2601	CCCTGTGATG AGAAAATATC	CTGCCCTTCC AACCCCCGCG	TTTGATTCAC TCTGGTGGAA	TTAAACCGAT CTTATTCATC	TGGAAATCAA CATAGGCACT	GTAGAGTCCT GAAAGGCGAC	TGAGAGGGAT AAAAACGACA	GAGGTTGTGA TTAAGGTACC	GGGAGACTCT CCAAAAAGGG	AGGAGGATAA AAGCTTTTAG
2501 2601 2701	CCCTGTGATG AGAAAATATC AAACTTGTAG	CTGCCCTTCC AACCCCCGCG ACCAGATAAC	TTTGATTCAC TCTGGTGGAA GCACAATGAT	TTAAACCGAT CTTATTCATC GTAGAAAAGA	TGGAAATCAA CATAGGCACT TGTAAATGAA	GTAGAGTCCT GAAAGGCGAC TACTAAAATC	TGAGAGGGAT AAAAACGACA GTTAGTCTGA	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT	GGGAGACTCT CCAAAAAGGG CATTAGTTAT	AGGAGGATAA AAGCTTTTAG TAGAATCTAA
2501 2601 2701 2801	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG	CTGCCCTTCC AACCCCCGCG ACCAGATAAC AATTCATTGA	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT	TGAGAGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG	GGGAGACTCT CCAAAAAGGG CATTAGTTAT GAATCATTGC	AGGAGGATAA AAGCTTTTAG TAGAATCTAA CGCAAGCAAC
2501 2601 2701 2801 2901	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT	CTGCCCTTCC AACCCCCGCG ACCAGATAAC AATTCATTGA AACTGGTGTG	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCAATTAAA	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC	TGAGAGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC	GGGAGACTCT CCAAAAAGGG CATTAGTTAT GAATCATTGC ATCATATGCA	AGGAGGATAA AAGCTTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT
2501 2601 2701 2801 2901 3001	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA	CTGCCCTTCC AACCCCCGCG ACCAGATAAC AATTCATTGA AACTGGTGTG GTTCATTGGG	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT TGGCGGAAGT	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCAATTAAA TCACCACTTT	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCCAG	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA	TGAGAGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT	GGGAGACTCT CCAAAAAGGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA	AGGAGGATAA AAGCTTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT
2501 2601 2701 2801 2901 3001 3101	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC	CTGCCCTTCC AACCCCCGCG ACCAGATAAC AATTCATTGA AACTGGTGTG GTTCATTGGG ACTTGGTTTC	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT TGGCGGAAGT AATCCACTCA	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCAATTAAA TCACCACTTT CACCACCTGG	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCCAG ATATCCAAAT	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATAC	TGAGAGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT TTTGCATCAA	GGGAGACTCT CCAAAAAGGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT	AGGAGGATAA AAGCTTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT TAAACGTAGA
2501 2601 2701 2801 2901 3001 3101 3201	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCAATCACCG	CTGCCCTTCC AACCCCCGCG ACCAGATAAC AATTCATTGA AACTGGTGTG GTTCATTGGG ACTTGGTTTC TTACAGCAAC	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT TGGCGGAAGT AATCCACTCA TTAATCATGG	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCAATTAAA TCACCACTTT CACCACCTGG TAGTATATCA	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATAC CTTTTGCCAA	TGAGAGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT	GGGAGACTCT CCAAAAAGGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT CGTTAACACC	AGGAGGATAA AAGCTTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT TAAACGTAGA AAGCCATACC
2501 2601 2701 2801 2901 3001 3101 3201 3301	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAACAA TACATCGACC CCCAATCACCG CCGCCAATGG	CTGCCCTTCC AACCCCCCGCG ACCAGATAAC AATTCATTGA AACTGGTGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATGTTACACC	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA	TTAAACCGAT CTTAATTCATC GTAGAAAAGA ACTTTTATTT ACCAATTAAA TCACCACTTT CACCACCTGG TAGTATATCA CCAAAATTTC	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCCAG ATATCCAAAT CAATTTGCTG CCGTTGATAT	GTAGAGTCCT GAAAGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATACA CTTTTGCCAA ACCAACACCG	TGAGAGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAAAAGGATA	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTAGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT ATGATTTAAA	GGGAGACTCT CCAAAAAGGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT CGTTAACACC TTCAAATTAC	AGGAGGATAA AAGCTTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT TTAACGTAGA AAGCCATACC AATGATTCAG
2501 2601 2701 2801 2901 3001 3101 3201 3301 3401	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCCAATCACCG CCGCCAATGG AAGATACACG	CTGCCCTTCC AACCCCCCGCG ACCAGATAAC AATTCATTGA GTTCATTGGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATGTTACACC ATCATTGGAA	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTATTT ACCAATTAAA TCACCACTTT CACCACCTGG TAGTATATCA CCAAAATTC ATGATGAATC	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCCAG ATATCCAAAT CAATTTGCTG CCGTTGATAT TATACGTACA	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATAC CTTTTGCCAA ACCAACACCG CCAAAGATAA	TGAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAAAAGGATA ATTCACATGG	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT AAGAGTTAAA	GGGAGACTCT CCAAAAAGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT CGTTAACACC TTCAAATTCAAAT	AGGAGGATAA AAGCTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT TAAACGTAGA AAGCCATACC AATGATTCAG GTAAACAATG
2501 2601 2701 2801 2901 3001 3101 3201 3301 3401 3501	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCAATCACCG CCGCCAATGG AAGATACACG TAATTTCATA	CTECCTTCC AACCCCCEGG AACTAGATAAC AATTCATTGA GTTCATTGGG GTTCATTGGGTGTG TTACAGCAAC ATGTTACACC ATGTTACACC ATGTTACACA	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT TGGCGGAAGT ATTCACATCA ATTAATCATCG ACCAAAATCA AGTGATAATG AACTAAGCTT	TTAAACCGAT CTTATTCATC GTAGAAAGA ACTTTATTT ACCAATTAAA TCACCACTTT CACCACCTGG TAGTATATCA CCAAAATTTC ATGATGAATC TTGGGAACAT	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCCAG ATATCCAAAT CAATTTGCTG CCGTTGATAT TATACGTACA ACTAAAGGTC	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATAC CTATGATAC CCTATGACAC ACCAACACCG CCAAAGATAA ATATAAAACC	TGAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAAAAGGATA ATTCACATGG GGAGAAAATG	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT ATGATTTAAA AAAAGTTAAA TTAAAATGTC	GGGAGACTCT CCAAAAAGGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT CGTTAACAC TTCAAATTCC AAATTCCAAAT CAAAATGTCC	AGGAGGATAA AAGCTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT TAAACGTAGA AAGGCATACC AATGATTCAG GTAAACATG ATTTGTAACT
2501 2601 2701 2801 3001 3101 3201 3301 3401 3501 3601	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCCAATCGACC CCCAATCACCG AAGATACACA GAGTACAAAC	CTGCCTTCC AACCAGATAAC AACTGGATAAC AATTCATTGA GTTCATTGGG GTTCATTGGGT ACTTGGTTTC TTACAGCAAC ATGTTACACC ATCATTGGAA GCTGTAACTA ATCATTTGGA	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATTTA	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCACTTATT CACCACTTT CACCACCTGG TAGTATATCA CCAAAATTTC ATGAGAACAT TTGGGAACAT CGAAATCATG	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CGGTGATAT TATACGTACA ACTAAAGGTC ATGGCTCGAA	GTAGAGTCCT GAAAGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATAC CTTTTGCCAA ACCAACACCG CCAAAGATAA ATATAAAACC ACCATTCCAA	TGAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAAAT	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT ATGATTTAAA TTAAAAGTTAAA TTAAAATGTC GTAGCTATAG	GGAGACTCT CCAAAAAGG CATTAGTTAT GAATCATTGC ATCATATGC ATCATATGC ATTGATCAA GATTCTATGT CGTTAACACC TTCAAATTAC AAATTCAAAT CAAAATGTCC TTGTGTTAAT	AGGAGGATTAA AAGCTTTTAG TAGAATCTTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT TTAACGATGAA AAGCCATACC AATGATTCAG GTAAACAATG ATTTGTAACT AAGTCAATGC
2501 2601 2701 2801 3001 3101 3201 3301 3401 3501 3601 3701	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCCAATCGC CCCAATCG CCGCCAATGG AAGATACACG GAGTACAACA TAAATTCACA	CTECCETEC AACCCCGCG AACCAGATAAC AATTCATTGA AATTCATTGA GTTCATTGG GTTCATTGG ACTTGGTTTC TTACAGCAAC ATGTTACACC ATCATTGGAA GCTGTAACTA ATCATTTGGA TTTGAAATCA	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT TGGCGGAAGT AATCCACTCA TTAATCATGG AACTAAAGTG AACTAAAGTG AATTACATTTA CATTCGAATA	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCAATTAAA TCACCACTTT CACCACCTGG TAGTATATCA CCAAAATTTC ATGATGATCA CTTGGGAACAT CGAAATCATG TATATCAGTA	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAA ATATCCAAAT CAATTTGCTG CCGTTGATAT TATACGTACA ACTAAAGGTCA ATGGCTCGAA TAGATGTGCT	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATACACC CTTTTGCCAA ACCAACACCG CCAAAGATAA ATATAAAACC ACCATTCCAA GATTGTAGTT	TGAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATGCCATCAT ATGCCATCA CACAATCGAG GAAACGATGGT GAAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAAAT ATGCTACTAA	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT ATGATTTAAA AAAGTTAAA TTAAAATGTC GTAGCTATAG ATATTGTCAT	GGGAGACTCT CCAATAAGGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT GGTTAACACC TTCAAATTAC AAATTCAAAT CAAAATGTCC TTGTGTTAAT	AGGAGGATAA AAGCTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTAGCAGTAT TAAACGTAGA AAGCCATACC AATGATTCAG GTAAACAATG ATTTGTAACT AAGTCAATGC TGCATCTACG
2501 2601 2701 2801 3001 3101 3201 3301 3401 3501 3601 3701 3801	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCAATCACCG CCGACATCG AAGATACACG TAAATTCACA AAAATACGGT	CTECCETTCC AACCCCGCG AACCAGATAAC AACTGGTGTG GTTCATTGGG ACTTGGTGTG TTACAGCAAC ATGTTACACC ATGTTACACC ATCATTGGAA GCTGTAACTA ATCATTGGA CATCAATCAC CATAAACCAG	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT TGGCGGAAGT AATCCACTCA TTAATCATCAG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATTTA CATTCGAATA CCATGGTCCT	TTAAACCGAT CTTATTCATC GTAGAAAGA ACTTTTATTT ACCAATTAAA TCACCACTTG TAGTATATCA CCAAAATTTC ATGATGAATC TTGGGAACAT CGAAATCATG GAAATCAGTA GAATGAAGAT	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CCATTTGCTG CCGTTGATAT TATACGTACA ACTAAAGGTC ATGGCTCGAA TAGATGTGCT GGAACACCGA	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATACAC CTTTTGCCAA ACCAACACCG CCAAAGATAA ATATAAAACC ACCATTCCAA GATTGTAGTT ATCCACTGCC	TGAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAAAGGATAG ATTCACATGG GGAGAAAATG TGTAACAAAT AATCATCGAT	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT ATGATTTAAA ATAAGTTAAA TTAAAATGTC GTAGCTATAG GTACTACGGTA	GGGAGACTCT CCAAAAAGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATATGCA CATAGATCAA GATTCTATGT CGTTAACACC TTCAAATTCAAAT CAAAATGTCC TTGTGTTAAT TCACTAAAGC CACGAAGAGG	AGGAGGATTAA AAGCTTTTAG TAGAATCTTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT TAAACGTAGA AAGCCATACC AATGATTCAG GTAAACAATG ATTTGTAACT AAGTCAATGC TGCATCTACG CCCGAAGATG
2501 2601 2701 2801 2901 3001 3101 3201 3301 3401 3501 3601 3701 3801 3901	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCAATCGACC CCGCAATGG AAGATACAG TAATTTCATA GAGTACAAAC TAAATTCACA AAAATACGGT AAGTCATCGA	CTECCETEC AACCCCGCG AACCAGATAAC AATTCATTGA AATTCATTGG GTTCATTGGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATCATTGGAA ATCATTGGAA TTGAAATCA ATCATTTGGA AGAAACCAG AGAAACCAG	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAAGCTT ATATCATTTA CAATGGATAA CCATGGTCCT TCCACCATCA	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCACTTT CACCACTTG CACACCCGG TAGTATATCA CCAAAATTTC ATGATGAACAT CGAAATCATG TATATCAGTA GAATGAAGAT CAACAACTCT	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CCGTTGATAT TATACGTACA ACTAAAGGTC ATGGCTCGAA TAGATGTGCT GGAACACGA	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATAC CTTTTGCCAA ACCAACACCG CCAAAGATAA ATATAAAACC ACCATTCCAA GATTGTAGTT ATCCACTGCC AACACAGAAT	TGAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATGCAATCAT ATGGAAACAC CACAATCGAG GAAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAAAT ATGCTACTAA AATCATCGAT GGACCATCAT	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGATGG ATGGAACATC AAAACTTTT TTTGCATCAA AGTAATCCAT ATGATTTAAA ATAAATGTC GTAGCTATAG ATATTGTCAT GTCTACGGTA	GGGAGACTCT CCAAAAAGG CATAGTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT CGTTAACACC TTCAAATTCA AAATTCAAAT CAAAATGTCC TTGTGTTAAT CCACGAAGAG AAACTCTAAA	AGGAGGATAA AAGCTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTA TTTAGCAGTA AAGCCATACC AATGATTCAG GTAAACAATG GTAAACAATG TGCAACTACG CCCGAAGATG ACTCCACCCC
2501 2601 2701 2801 2901 3001 3101 3201 3301 3401 3501 3601 3701 3801 3901	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCAATCACCG CCGCAATGG AAGATACAG TAATTCACA AAAATACGGA AAGTCATCGA TTCAGCAACA	CTECCCTCC AACCCCGCG AACCAGATAAC AACTGGTTGG GTTCATTGGG GTTCATTGGG GTTCATTGGA ACTTGGTTTC TTACAGCAAC ATCATTGAA ATCATTGGA TTTGAAATCA CATAAACCAG GCAACAATCG GCACAATCG	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATTTA CATTCGAATA CCATGGTCCT TCCACCATCA	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCACTTT CACCACTTG TAGTATATCA CACACCTGG TAGTATATCA CAAAATTTC ATGAGAAATCATG TAGGAACAT CGAAATCATG TATATCAGTA GAATGAAGAT CAACAACTCT AAGCGACTGT	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CGGTGAAAT TATACGTACA ACTAAAGGTC ATGGCTCGAA TAGATGTGCT GGAACACCGA CCAACCCA	GTAGAGTCCT GAAAGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGGATGGA GCTATGGATGC CCAAGACACCG CCAAAGATAA ACCAATCCAA GATTGTAGTT ATCCACTGCC AACAACAGAT TCACCGGCTA	TCAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAATA ATGCTACTAA AATCATCGAT GGACCATCAT GGACCATCAT	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGATGG ATGGAACATC AAAACTTTTT TTGCATCAA AGTAATCCAT ATGATTTAAA ATAAAGTTAAA TTAAAATGTC GTAGCTATAG ATATTGTCAT GTCTACGGTA CCCTCGGCACA	GGAGACTCT CCAAAAAGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT CGTTAACACC TTCAAATTAC AAATTCAAAT CAAAATGTCC TTGTGTTAAT TCACTAAAGG AAACTCTAA	AGGAGGATTAA AAGCTTTTAG TAGAATCTTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTA TTTAGCAGTA AAGCCATACC AATGATTCAG GTAAACAATGG GTAAACAATGG ATTTGTAACT AAGTCAATGC TGCATCTACG CCCGAAGATG
2501 2601 2701 2801 2901 3101 3201 3301 3401 3501 3601 3701 3801 3901 4001 4101	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCCAATCGACC CCCAATCGC CCGCCAATGG AAGATACACG TAATTTCATA GAGTACAAACA TTCAGCAACA GCTGCAACAG	CTECCETEC AACCCCGCG AACCAGATAAC AATTCATTGA AATTCATTGG GTTCATTGGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATGTTACACC ATCATTTGAA GCTGTAACTA GCTGTAACTA CATAAACCAG AGAACATCG AGTGCGAACA	TTGGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATAATGAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATTTA CATTCGAATA CCATGGTCCT TCCACCATCA CAGCAATTC CGATGCCATT	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCAATTAAA TCACCACTTG TAGTATATCA CACACCTGG TAGTATATCA CCAAAATTTC ATGATGAATC TTGGGAACAT GAATGAAGAT CAACAACTCT AAGCGACTGT CTTCCCGTAC	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CCGTTGATAT TATACGTACA ACTAAAGGTC ATGGCTCGAA TAGATGTGCT GGAACACCGA GCCAACCCCA CTGAACCTAA	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATACAAC CATGGATGA ACCAACACCG CCAAAGATAA ATATAAAACC ACCATTCCAA GATTGTAGTT ATCCACTGCC AACAACAGAAT TCACCGGCTA	TGAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACA CACAATCGAG GAAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAAAT ATGCTACTAA AATCATCGAT GGACCATCAA ATCTAATGTC CCTAGCGGCA	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT ATGATTTAAA ATAAAGTTAAA TTAAAATGTC GTAGCTATAG ATATTGTCAT GTCCTCGCACA CAATTTCCTG CAACAGCAAG	GGGAGACTCT CCAAAAAGG CATAGTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT CGTTAACACC TTCAAATTCAAAT CAAAATGCAAAT CAAAATGTCA TTGACTAAAGC CACGAAGAGG AAACTCTAAA CCTACAACAT CGGCTCTCGC	AGGAGGATTAA TAGAATCTTAG TGCAAGCAAC TCATATGAAT TTAGCAATCTAA AGCCATACGAGAT TAAACGTAGA AAGCCATACC AATGATTCAG GTAAACAATG CACGAAGATG CCCGAAGATG CCCGAAGATG CCCAATGTCA
2501 2601 2701 2901 3001 3101 3201 3301 3401 3501 3501 3601 3701 3801 3901 4001 4201	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATGA ATATGATCAT CAAAAAACAA TACATCGACC CCCAATCGC CCCAATCGC CCGCCAATGG AAGATACACG TAAATTCATA AGAGTACAACA TCAGCAACA GCTGCAACAACG CCAAACATGC	CTECCETTEC AACCCCGCG AACCAGATAAC AACTGGTGTG GTTCATTGG ACTGGTGTG GTTCATTGGA ACTGGTTCC TTACAGCAAC ATGTTACACC ATCATTGGA GCTGTAACTA ATCATTTGGA ATCATTTGGA CATAAACCAG GGAACAATCG GCAACAATCG GCAACAATCG GGAACAACG	TTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATGCT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AATGATAAGTA AATGATAAGTT ATATCATTTA CATTCGAATA CCATGGTCCT TCCACCATCA CAGCAATTCC CGATGCCATT AACGAACATG	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCAATTAAA TCACCACTTT CACCACCTGG TAGTATATCA CCAAAATTTC ATGATGATCA TGGAAACATC TTGGGAACAT GAATGAAGAT CAACAACTCT AAGCGACTGT CTTCCCGTAC AGTAAATGTG	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAA ATATCCAAAT CAATTTGCTG CCGTTGATAT TATACGTACA ATGACTGAA ATGGCTCGAA TGGAACACGA GCCAACCCGA CTGAACCTAA AGAGCGATGA	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATACAAC CTTTTGCCAA ACCAACACCG CCAAAGATAA ACCAACACGC AACCATCCAA ACCATCCACAGAT TCACCGGCTA ACCATCCACT	TGAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATGCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAAAT ATGCTACTAA AATCATCGAT GGACATCAT ATCTAATGTC GCTAGCGGCA	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT ATGATTTAAA AAAGTTAAA TTAAAATGTC GTAGCTATAG GTAGCTATAG GTCTACGGTA CCATCGCACA CAATTTCCTG CAACAGCAAG ATGAACTGA	GGGAGACTCT CCAATAAGGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGC GGTTAACACC TTCAAATTAC AAATTCAAAT CAAAATGTCC TTGTGTTAAT TCACTAAAGC CACGAAGAGG AAACTCTAAA CCTACAACAT TGGCTCTCGC TGAACGATGT	AGGAGGATTAA AAGCTTTTAG TAGAATCTTAA CGCAAGCAAC TCATATGAAT TTTAGCAATT TAAACGTAGA AAGCCATACC AATGATTCAG GTAAACAATG GTAAACAATG TGCATCTACG CCCGAAGATG ACCCGAGAGTAT CCAAATGCAA
2501 2601 2701 2901 3001 3101 3201 3301 3401 3501 3501 3601 3701 3901 4001 4101 4201 4301	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCAATCACCG CCGCAATGG AAGATACAAC TAATTTCATA GAGTACAACA AAAATCAGGT AAGTCATCGA TTCAGCAACA GCCAACAAG CCAAACAAGC	CTECCCTCC AACCCCGCG AACCCCGTCA ACCAGATAAC AATTCATTGA AATTCATTGG GTTCATTGGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATCATTGGAA ATCATTTGGA TTTGAAATCA AGAAACCAG AGAACGTGA CACACAATCG AGGAGCAACA TGCCATGGAT	TTIGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATTA CAATGGTCATTA CCATGGTCCT ACGACATTC CAGCAATTC CAGCAATTC CAGCAATTC CAGCAATTC CCATGCCATA	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCACTTT CACCACTTG CACACCTGG TAGTATATCA CAGAGACATTTC ATGATGAACAT TTGGGACATG TATATCAGTA GAATGAAGAT CAACAACTCT AAGCGACTGT CTTCCCGTAC AGTAAATGTG CAACCCCAAC	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CCGTTGATAT TATACGTACA ACTAAAGGTC ATGGCTCGAA TGGACGACGAT GCCAACCCCA CTGAACCTAA AGAGGGATGA GAAAAATGTT	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGGATGGA GCTATGGATAC CCATGCAACACG CCAAAGATAA ATATAAAACC ACCATTCCAA GATTGTAGTT ATCCACTGCC AACACAGAAT TCACCGGCTA ACCACTCACAT AGAAGATGT	TCAGGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG GAAAAGGATGA ATTCACATGG GGAGAAAATG TGTAACAAATG TGTAACAAATG GGACCATCAT ATCTAATGGAT GGACCATCAT ATCTAACGGCA ATCAGTGACA	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGTTGG ATGGAACATC AAAACTTTT TTTGCATCAA AGTAATCCAT ATGATTTAA ATAAATGTC GTAGCTATAG TTAACGTATAG CCTCGCACA CAATTTCCTG CAACAGCAAG ATGAACTGAA TGAGCCACCC	GGGAGACTCT CCAAAAAGG CATAGTAT GAATCATTGC ATCATAGTA CTATGATCAA GATTCTATGT CGTTAACAC TTCAAATTAC AAATTCAAAT CAAAATGTCC TTGTGTTAAT CACATAAAG CACGAAGAGG AAACTCTAAA CCTACAACAT GGCTCTCGC TGAACGATGT AAAGAAATAC	AGGAGGATTAA AAGCTTTTAG TAGAATCTTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTA TTTAGCAGTA AAGCCATACC AATGATTCAG GTAAACAATGG GTAAACAATGG ATTTGTAACT AAGTCAATGC CCGAAGAATG ACTCCAACCCC TAGCGAGTAT CCAAATGTCAA CAACAACACC
2501 2601 2701 2801 3001 3101 3201 3301 3401 3501 3601 3701 3801 3901 4001 4101 4201 4301	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCCAATCGACC CCCAATCGC CCGCAATGG AAGATACAG TAATTTCATA GAGTACAAAC TAAATACGGA TTCAGCAACAA GCTGCAACAG TCGATAACAGT	CTECCCTTCC AACCCCGCG AACCAGATAAC ACCAGATAAC AATTCATTGA GTTCATTGGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATCATTGGA ATCATTGGA TTTGAAATCA GCAACAATCG AGAACCATCG AGTGGCAACA GCACTGAACAAC GAGAACGACA	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATTTA CATTCGAATA CCATGGTCCT TCCACCATCA CAGCAATTTC CAGTGCCATT AACGAACATG CTGTCGCAAA GGCGGAATCA	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCACTTT CACCACTTG CACCACCTGG TAGTATATCA CCAAAATTTC ATGAGAACAT TTGGGAACAT CGAAATCATG TATATCAGTA GAATGAAGAT CAACAACTCT AAGCGACTGT CTTCCCGTAC AGTAAATGTC CAACACCCAC	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CGTTGATAT ATAAGGTACA ACTAAAGGTC ATGGCTCGAA TGGATCGCGA GCAACCCCA CTGAACCCAA GCCAACCCAA GAAAAATGTT GGACGCGTG	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATAC CTTTTGCCAA ACCAACACCG CCAAAGATAA ATATAAAACC ACCATTCCAA GATTGTAGTT ATCCACTGCC AACACAGAAT TCACCGGCTA ACCATCACAT GCCAACACAG GCCAACACAGA	TCAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG GAAGAAATGG GGAGAAAATG TGTAACAAAT ATCCACATGG GGACAAAAT ATCTACATGGA GGACCATCAT ATCTAATGTC GCTAGCGGCA ATGAGTGACT TCGACCCAT CTCGTCCCATA	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGATGG ATGGAACATC AAAACTTTTT TTGGATCAA AGTAATCCAT ATGATTTAA ATAAAGTTAAA TTAAAATGTC GTAGCTATAG ATATTGTCAT GTCTACGGTA CCCTCGCACA CAATTTCCTG CAACAGCAAG ATGAAACTGA TGAGCCACCC ACACCCGCAG	GGAGACTCT CCAATAAGGG CATTAGTTAT GAATCATTGC ATCATATGCA CTATGATCAA GATTCTATGT CGTTAACACC TTCAAATTCC AAATTCAAAT CCAAAATGTCC TTGTGTTAAT TCACTAAAG CCACGAAGAG AAACTCTAAA CCTACAACAT CGGCTCTCGC TGAACGATGT AAAGAAATGACAA	AGGAGGATTAA AAGCTTTTAG TAGAATCTTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTA TTTAGCAGTA AAGCCATACC AATGATTCAG GTAAACAATG GTAAACAATG GCACAATGC TGCACTCAACG CCCGAAGATG ACTCCAACGCC TAGCGAGTCCAA CAACAACACC
2501 2601 2701 2801 3001 3101 3201 3301 3501 3601 3701 3801 3901 4001 4101 4201 4301 4501	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCACCG CCCAATCACCG CCGCCAATGG AAGATACACG TAATTCATA GAGTACAAAC TAAATTCGA AAATACGGT AAGTCATCGA TTCAGCAACA GCTGCAACAG CCCAACAGTC GACACCACCC	CTECCETEC AACCCCGCG AACCAGATAAC AATTCATTGA AATTGGTGTG GTTCATTGGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATGTTACACC ATGTTACACC ATGTTACACC ATGTATTGGA TTTGAAATCA GGACAATCG AGGAGGCAACA GAGATGGAAC GCACAATGG AGGAGGAACA GAGATGGAACA TGCCAAGAAT TTCTCAAGAA	TTIGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATTTA CATTCGAATA CCATGGTCCT TCCACCATCA CAGCAACTTC CGATGCCATT AACGAACAAC GCCACAACACG	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCACTTTT CACCACTTG TAGTATATCA CACACCTGG TAGTATATCA CCAAAATTTC ATGAGAACAT CGAAATCATG GAATGAAGAT CAACAACTCT AAGCGACTGT CTTCCCGTAC AGTAAATGTG CAACCCCAAC GAAGTATTG GAAGGTATTG	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CCGTTGATAT TATACGTACA ACTAAAGGTC ATGGCTCGAA TCAAACAGGA GCCAACCCCA CTGAACCCAA GAACACCCA GAACACCCA CAGACCCCA CAAACAGGATGA GAAAAATGGATAA	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAAGACGT AAACTACAAC CATGGATGGA GCTATGGATGGA GCTATGGATGGA ACCAACACCG CCAAAGATAA ATATAAAACC AACAATTCCAA GATTGTAGTT ATCCACTGCC AACAACAGAT ACCATTCACAT ACACTCACAT ACAACAGTCGG GCCAACCAGA TTAAATTAGA CTCAAGTCGG	TCAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAAAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAAAT ATGCTACTAA AATCATCGAT GGACCATCAT GGACCATCAT TCGAACGGCA ATGGAGTGACT TCGGACCCAT TTGGAAATGT	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGATGG ATGGAACATC AAAACTTTT TTTGCATCAA AGTAATCCAT ATGATTTAA AAAAGTTAAA TTAAAATGTC GTAGCTATAG CTACGCTATAG CAATTGTCAT GTCTCACGGTA CCCTCGCACA CAATTTCCTG CAACAGCAAG ATGAAACTGA TGAGCCACCG CGGGGGATGA	GGGAGACTCT CCAAAAAGG CATAGTAG ACCATAGTAG ATCATAGTA CTATGATCAA GATTCTATGT CGTTAACACC TTCAAATTAC AAAATCAAAT CAAAATGAC CACGAAGAGG AAACTCTAAA CCTACAACAT CGGCTCTCGC TGAACGATGT AAAGAAATA AGATGTTCCA	AGGAGGATTAA AAGCTTTTAG TAGAATCTTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTA AAGCCATACC AATGATTCAG GTAAACAATG ACTCCAATGC TGCATCTACG ACTCCACCCC TAGCGAGTAT CCAAATGTCA GAGAGTCCAA GAGAGTCCAA GAGAGTCCAA GAGAGTCCAA GAGAGTCCAA GAGAGTCCAA GAGAGTCCAA GAGAGTCCAA GAGAGTCCAA
2501 2601 2701 2801 3001 3101 3201 3301 3401 3501 3601 3701 3901 4001 4101 4201 4401 4501 4601	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCCAATCGC CCCAATCGC CCCAATCACCG CCGCCAATGG AGATACAACA TAAATTCACA AAATTCACA AAATACGGT AGGTCATCGA TTCAGCAACAG CCCAACAGT CCGATAACGA CTCGATAACGA	CTECCCTCC AACCCCCCCGC AACCACATTAC AACTCGTTAC ACCACATTACA AATTCATTGG GTTCATTGG GTTCATTGGA ACTGGTTACCA ATGTTACACC ATCATTGGAA TTTGAAATCA CATTAAACCAG AGAACATGG AGAACATGG AGAACAATCG GCACAACAT GACATGACAAC GCACAACAT GCACATGACAAC TTCCCAAGAA GTTGCAAGAAC	TTEGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATAATCAT GAAATAATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG AACTAAAGCT AACTAAGCTT CAATCGATATA CAATGGACATTC CAATGGCATTC CAATGGCAATTC CAGGAACATG CTTCCGCAACA GCCAGCAACGACGACGACG	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCAATTAAA TCACCACTTT CACCACCTGG TAGTATATCA CCAAAATTTC ATGATGATCA TTGGGAACAT GAATGAAGAT CAACAACTCT AAGCGACTGA CATCCGTAC AGTAAATGG CAACCCCAAC TAGGAGGAAA GAGGTTATTG CTACGGATGA	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAA ATATCCAAAT CAATTTGCTG CCGTTGATAT TATACGTACA ATGGCTCGAA TGGACGCGA GGAACACCGA GCCAACCCCA GCCAACCCCA GCAACCCCA GGAACACCTA AGAGCGATGA GAAAAATGAT AACAAATAAG	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATACAC CTTTTGCCAA ACCAACACCG CCAATGAACA ACCATCCAA GATTGTAGTT ATCCACTGCC AACACAGAAT TCACCGGCTA ACCTTCCACAT GCCAACCAGA TTAAATTAGA CTCAAGTCGG AAACCAGAGA	TCAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGCAAACAC CACAATCGAG AAACGATGGT GAAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAAAT ATGCTACTAA AATCATCGAT GGACCATCAT ATGCAACTGAC GCTAGCGGCA ATGAGTGACT TCGACCCCAT CTCGTCAGTA	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGACATC AAAACTTTT TTTGCATCAA AGTAATCCAT ATGATTTAA ATGATTTAA ATGATTTAA TTAAAATGTC GTAGCTATAG ATATTGTCAT GTCTACGGTA CCCTCGCACA CAACTCCACA CAACAGCAAG ATGAAACTGA CGGGGGATGA ACCTTCCCTG	GGGAGACTCT CCANAAAGG CATTAGTTAT GAATCATTGC ATCATATGTAT GATTCTATGT CTTTGATCAA GATTCTATGT CTTCAAATTAC AAATTCAAAT CAAAATGTCC TTGGTTAAT CACTAAAGC CACGAAGAGG AAACTCTAAA CCTACAACAT CGGCTCTCGC TGAACGATGT AAAGAAATAC AAATGAGAA AGATGTTCCA GAGTAGAGA	AGGAGGATAA AAGCTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT TAAACGTAGA AAGCCATACC AATGATTCAG GTAAACAATG TGCATCTACG ACCGAAGATG CCCGAAGATG CCCAAATGTCA GAGAGTCGAA GAGAGTCGAA CAACAACACC TCTAACAATG ACATCATCAT
2501 2601 2701 2901 3001 3101 3201 3301 3501 3701 3701 3801 4001 4101 4201 4401 4501 4401 4501 4701	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCAATCGACC CCCAATCGC CCGCAATGG AAGTACAACAG TAATTTCATA GAGTACAACAG TTCAGCAACAG CCAACACGGT AAGTCATCGA CCGATAACGAT GCCAACAGTT GACACCACCC CATCGGTGGT	CTECCCTCC AACCCCGCG AACCCCGTA ACCAGATAAC AATTCATTGA AATTCATTGG GTTCATTGGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATCATTTGGA ATCATTTGGA ATCATTTGGA ATTCATTGGA CATAAACCAG AGAACCAG GCACAGACG AGGATGGAACA TGCCATGGAT ACATCAACAT TTCCCAGGA ACTCCACCAG	TTIGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATTA CAATGGAATA CCATGGTCCT TCCACCATCA CAGCAATTTC CGATGCCATT AACGAACATG CTGTCGCAAA GCGGGAATCA GCAACCAACG AAAGACGATG	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCACTTT CACCACTTG CACACCCGG TAGTATATCA CCAAAATTTC ATGATGAATCA TTGGGAACAT CGAAATCATG GAATGAAGAT CACAACTCT AAGCGACTGT CTCCCGTAC CACCAACTCT AAGCGACGAA AGTAAATGTG CAACCCCAAC TAGGAGGAAA GAGGTTATTG CTAGTGAAGAT	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CGTTGATAT ATAAGTCCAA ACTAAAGGTC ATGGCTCGAA CTGAACCAGA GCCAACCCA CTGAACCAGA GCCAACCCA CTGAACCAGA AGAGCGATGA AACAAATGTT GGTCGCGCGT AAATGGATAA ACAAATAAC	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGGATGGA GCTATGGATAC CTTTTGCCAA ACCAACACCG CCAAAGATAA ATATAAAACC ACCATTCCAA GATTGTAGTT ATCCACTGCC AACACAGAAT TCACCGGCTA ACAACAGAAT TCAAGTCG AACACAGAA TTAAATTAGA CTCAAGTCGG AAACCAGAA	TCAGAGAGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG GAAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAAAT ATGCTACTAGA GGACCATCAT ATCTAATGGTC GGACCATCAT ATCTAATGGTC GCTCAGCGCA ATGAGTGACT TCGGACACAC	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGATGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT ATGATTTAA ATAAATGTC GTAGCTATAG ATAATGTCAT GTCTACGGTA CCCTCGCACA CAATTTCCTG CAACAGCAAG ATGAAACTGA GCGACGCACG CGGGGGATGA ACCTTCCCTG ATCTTTTCA	GGGAGACTCT CCAAAAAGG CATAGTAGTAT GAATCATTGC ATCATAGTA CTATGATCAA GATTCTATGT CGTTAACACC TTCAAATTAC AAATTCAAAT CAAAATGTCC TTGTGTTAATC CACGAAGAG AAACTCTAAA CCTACAACAT CGGCTCTCGC TGAACGATGT AAAGAAATAC AAAATGAGAA AGAATGCTGT	AGGAGGATTAA AAGCTTTAG TTAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTA TTTAGCAGTA AAGCCATACC AATGATTCAG GTAAACAATG GTAAACAATG AAGTCAATGC TGCATCTACG CCCGAAGATG CCCGAAGATG CCCGAAGATG CAAATGTCAA GAGAGTCGAA CAACAACACC TCTAACAATG ACAACACCC TCTAACAATG
25011 26011 27011 29011 30011 31011 32011 33011 35011 36011 37011 39011 40011 41011 42011 44011 44011 44011 44011	CCCTGTGATG AGAAAATATC AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCAATCACCG CCGCAATGG CCGCAATGG AAGATACAG TAATTTCATA GAGTACAAAC TAAATTCACA AAAATACGGT TTCAGCAACAG CCAAACATGC CCGATAACAG TCGATAACAG CTCAACAGTT GACACCACCC CATCGGTGGT ATCAACAGTG	CTECCCTTCC AACCCCGCG AACCACATTAC ACCACATTACA AACTGGTTGG GTTCATTGGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATCATTGGAA ATCATTGGA ATCATTGGA GCACATACTA GGAACAATCG AGAACCATCG GGAACAATCG AGTAGCAACA GTGGCAACA ACTCCACAGA ACTCCACAGA ACTCCACCGA GCTACCACCGA	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATTTA CATTCGAATA CCATGGTCCT TCCACCATCA CAGCAATTTC CAGCAATTTC CAGCGAATCA GGCGGAATCA GCCACCAACG AACAACGAC	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCACTTT CACCACTTG CACACCCGG TAGTATATCA CCAAAATTTC ATGAGAAATTTC CGAAATCATG TAGTGAAGAT CACAAACTCT AAGCGACTGT CTTCCCGTAC AGGGAGAAA GAGGTTATTG CTACGAGGAAA GAGGTTATTG CTACGCAG CTACTCAAAT	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CGTTGATAT ATAAGGTACA ACTAAAGGTC ATGGCTCGAA TGGACGCGA GGCAACCCCA CTGAACCCAA GGCAACCCAA GGCAACCCAA GGAACATGT AAATGGATAA ACAAATAG CAACAATGTG GCAACATGTG CGCAACCTG	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGGATGGA GCTATGATAC CTTTTGCCAA ACCAACACCG CCAAAGATAA ATATAAAACC ACCATTCCAA GATTGTAGTT ATCCACTGCC AACACAGAAT TCACCGGCTA ACCATCACAG TCAAGTCGG AAACCAGACA ACCAAGTCGG AACCAGACA ACGACGAAAAG	TGAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG GAAGAATGGT GAACAATGG GGAGAAAATG TGTAACAAAT ATCCACATGA GGACGACATCAT ATCTAATGTC GCTAGCGGCA ATGAGTGACT TCGACCCAT CTCGTCAGTA TTGGAAAATGT GCAGCACAAC ATTTGGAAAATGT GCAGCACAAC	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGATGG ATGGAACATC AAAACTTTTT TTTGCATCAA AGTAATCCAT ATGATTTAA ATAAAGTTAAA TTAAAATGTC GTAGCTATAG ATATTGTCATG CACTGCACA CAATTTCCTG CAACAGCAAG ATGAAACTG ACCCCCGCAG CGGGGGATGA ACCTTCCTTCA GTATATCATT	GGGAGACTCT CCAAAAAGG CATAAGTA GAATCATTGC ATCATATGC ATCATATGC ATCATATGC GATTCTATGT CGTTAACACC TTCAAATTCC AAATTCAAAT CCACTAAAAT CCACTAAAAT CCACGAAGAG GAACGATCTCAA CCTACAACAT CGGCTCTCGC CAGGACGATGT CAACGATGACAA AGATGTTCCA GAAGTAGCAA AGGATGCTGT CTTCCTACAC	AGGAGGATTAA AAGCTTTTAG TAGAATCTTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTA TTTAGCAGTA AAGCCATACC AATGCATACC AATGATTCAG GTAAACAATG GTAAACAATG CGCGAAGATG CCCGAAGATG CACACCACCC TAGCGAGTCAA CAACAACACC TCTAACAATG CACATCATCAT ATTACAACAATG CATCATCATCAT ATTACCAGGA
2501 2601 2701 2801 2901 3101 3301 3301 3301 3501 3701 3801 3701 4101 4201 4301 4401 4501 4401 4401 4401	CCCTGTGATA AAACTTGTATA AAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCACCG CCCAATCACCG CCGCCAATGG AAGATACACG TAATTCACA AAATACGAT AAGTCATCGA TTCAGCAACA GCTGCAACAG CCCAACAGT GACACCACCC CATCGGTGGT ATCCAACAGT ACCCACACTG ACCCACACTG	CTECCCTCC AACCCCCCGCG AACCACGATAAC ACCACGATAAC AATTCATTGG GTTCATTGGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATGTTACACC ATGTTACACC ATGTTACACC ATCATTTGGA ATCATTTGGA ATCATTTGGA GACACGTGA GCACCATGG AGTGGCAACA GACATCGACAC GCACCAGCAT TTCTCAAGAA GTTCGAGAAC ACTTCCACGA ACTCCACCGA ACTCCACCGA	TTIGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATCA AGTGATAATCA TATATCATTTA CATTCGAATA CCATGGTCCT TCCACCATCA CGATGCCATT AACGAACATC CGATGCCATC AGCGACAACG CTGTCGCAAA GCCACCAACG AAAAGCCAACG ACAACCAACG ATTCTACACT	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATT ACCACATTAT CACCACTTG TAGTATATCA CCACACTGG TAGTATATCA CCAAAATTCC ATGAGAACAT CGAAATCATG GAATCATG CAACAACTCT AAGCGACTGT CTTCCCGTAC AGCACACT CTACCGTAC AGCGCACAAC TAGGAGGAAA GAGGTTATTG CTACTGAAG GTCTTCCAACT ACGAGCGACA	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CGTTGATAT TATACGTACA ACTAAAGGTC GAACACCGA CTGAACCTGA GCAACCCCA CTGAACCCAA GAAAAATGTT GGTCGCGCGT AAATGGATAA AACAAATAG CAAGAGAACT	GTAGAGTCCT GAAAGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATACAC CTTTTGCCAA ACCAACACCG CCAAAGATAA ATATAAACC ACCATTCCAA GATTGTAGTT ATCCACTGCC AACACAGAAT CACCACCAGA GCTTCACAT GCCAACCAGA CTCAAGTCGG AAACCAGAGA CTCAAGTCGG CAAATGCAG CGCGCAAAAG CTTGAGCTTT	TCAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAACAATCGAG GGAGAAAATG TGTAACAAAT ATTCACATGG GGACGACAAT GGACCATCAT GGACCATCAT TCGAACGACA ATGGAGTGACT TCGGACACAAC ATTGGAAAATGT GCAGCACAAC ATTTGCGGAC	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGATGG ATGGAACATC AAAACTTTT TTTGCATCAA AGTAATCCAT ATGATTTAAA TTAAAAGTTAAA TTAAAAGTTAAA CTATCGCACA CCCTCGCACA CCATCGCACA CAACTCCCGCAG ATGACCACCG ACACCCCCCAG CGGGGGATGA ACCTTCCTTCA ATTTCCATTC	GGGAGACTCT CCAAAAAGG CATAGTAG GATCATTGC ATCATAGTA CTATGATCAA GATCTTATGC CGTTAACACC TTCAAATTAC AAATTCAAAT CAAAATGTCC TTGTGTTAAT TCACTAAAGC CACGAAGAGG AAACTCTAAA CCTACAACAT CGGCTCTCGC TGAACGATGCT AAAGAATAC AAAATGAGAA AGATGTTCCA GAAGTAGAGA AGATGTTCCA GTCCAACTTTG	AGGAGGATTAA AAGCTTTTAG TAGAATCTTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTA TTTAGCAGTA AAGCCATACC AATGATTCAG GTAAACAATG CATCTACG CCGAAGATG ACTCCACCCC TAGCGAGATT CCAAATGTCA GAGAGTCCAA CAACAACACC CCTGTACAAT ATTACAATAC ATCCCCGTACAG TCCGTTAAGT
2501 2601 2701 2801 3001 3101 3301 3501 3701 3701 3701 3701 4001 4101 4201 4301 4401 4501 4501 4701 4801 4901 5001	CCCTGTGTATA AGAAAATATC AAAACTTGTAG TCATGTATAG ATATGATCAT CAAAAAACAA TACATCGACC CCAATCACCG CCGCAATGG AAGATACACA TAATTTCATA GAGTACAAAC TAAATTCACA AAAATACAGA TTCAGCAACAG CCCAACAGT GACACACG ACCACACACA ACCACACACA	CTECCCTCC AACCCCGCG AACCCCGCG ACCAGATAAC AATTGATTGA GTTCATTGG GTTCATTGGG ACTTGGTTTC TTACAGCAAC ATGTTACACC ATCATTGGAA GCTGTAACTA GCTGTAACTA GCAATAGAACG AGAATGGAACA GCAACAATCG AGAGTGGAACA TTCCCAAGGAT ACACCAACAT GCTGCACGA ACCTCACCGA GCTACCATGG ATAAAGTCGA	TTTGATTCAC TCTGGTGGAA GCACAATGAT CAATAATCAA GAAATATCAT TGGCGGAAGT AATCCACTCA TTAATCATGG ACCAAAATCA AGTGATAATG AACTAAGCTT ATATCATAA CATTCGAATA CCATGGTCCT TCCACCATCA CAGCAATTCC CGATGCCATT AACGAACATG CTGTCGCAAA GCAACCAACG AAGACCAACG AAGACCAACG AAGACCAACG ATTTACAACGAC ATTTACAACGATG	TTAAACCGAT CTTATTCATC GTAGAAAAGA ACTTTTATTT ACCACTTT CACCACTTG TAGTATATCA CCACACTGG TAGTATATCA CCAAAATTTCA CCAAAATTCATC TTGGGAACAT CATGATGAAGAT CAACACTCT AAGCGACTGT CTTCCCGTAC AGTAAATGTG CAACCCCAAC TAGGAGGAAA GAGGTATTG GTACTGCACG GTCTTCAAAT ACGACGAGA AGAATTCAAC	TGGAAATCAA CATAGGCACT TGTAAATGAA ATGCTTCATT ACAGAACCAC CTACCCCAG ATATCCAAAT CAATTTGCTG CCGTGGTAAT TATACGTCAA ACTAAAGGTC ATGGCTCGAA TGGAACACCGA TCAAACAGGA GCCAACCCCA CTGAACACGA GGAACACGA GGAACACGA GGAACACGA GGAACATGTG GCAACCCTA AAATGGATAA AACAAATAAG CAAGAGAACT GCAACATGTG AATTTTACGA	GTAGAGTCCT GAAAGGCGAC TACTAAAATC TCAGAGACGT AAACTACAAC CATGGATGGA GCTATGATAC CTTTGCCAA ACCAACACCG CCAAAGATAA ATATAAAACC ACCATTCCAA GATTGTAGTT TACCAGGCTA ACCACAGGAT GCCAACCAGA TTAAATTAGA CTCAAGTCGG AACCAGAGA CTCAAGTCGA CTCGAGCTTT TTCCAAATAT	TCAGAGGGGAT AAAAACGACA GTTAGTCTGA ATCTCATCAT ATACCCATCA ATGGAAACAC CACAATCGAG AAACGATGGT GAAAAGGATA ATTCACATGG GGAGAAAATG TGTAACAAATG GGACCATCAT ATGCTACTGAT GGCACCACAT CTCGACGCAATG TTGGAAGTGACT TTGGAAGTGACT TTGGAAATGTT CCCAACGAATG GTAAATGTTT	GAGGTTGTGA TTAAGGTACC ACTTCCTTAT GCATGGACATC AAAACTTTT TTTGCATCAA AGTAATCCAT ATGATTTAA ATGATTTAA TTAAAATGTC GTAGCTATAG ATATTGTCAT GTCTACGGTA CCCTCGCACA CAACTCCACG CAACAGCAAG ATGAAACTGA TGAGCCACCC ACACCCCAG ACGGGGGATGA ACCTTCCCTG ATCTCTTCCA GTATATCATTC ATTTCATTC	GGGAGACTCT CCAAAAAGG CATTAGTTAT GAATCATTGC ATCATAGTTAT GAATCATTGC CTTCAAATTGC TTCAAATTCCAAT CAAAATGTCC TTCAAATTCCAAT CAAAATGTCC TTGGTTAAT CACTAAAGC CACGAAGAGG AAACTCTAAA CCTACAACAT CGGCTCTCGC TGAACGATGT AAAGAAATAC AAATGAGAA AGATGTTCCA GACATTTTG GAATATTTAT	AGGAGGATAA AAGCTTTAG TAGAATCTAA CGCAAGCAAC TCATATGAAT TTTAGCAGTT TTAACAGTAG AAGCCATACC AATGATTCAG GTAAACAATG ATTTGTAACT AAGTCAATGC CCGAAGATG CCGAAGATG ACTCCACCCC TAGCAGCAGA CAACAACACC CCTAACAATG ACATCATCAT ATTAGCAAGA CCTGTACATCA ATTGCCAGGA TCGGTAAGTT

Source: Genomic DNA, phage Hpl-hb phB. Alignment with cDNA sequences: Positions 932..1223 correspond to the first exon of the *Haematopota hunchback* P1 transcript (SEQ17), and positions 2855..4998 correspond to the second exon of the *Haematopota hunchback* transcript. In the genomic sequence, one putative NRE sequence (5022..5043) was identified.

# SEQ19 Clogmia hunchback, cDNA, P1 transcript.

CGGACCTTAC	GTTCTGTATT	CAGTCTCTGA	AAACTAAGAA	CGTATTTTAA	GIGI"IGI"IIA	TTAGTTAGTT	AAAAATCTGC	TCTCCCCAAA	AACGAT"I'GI'G
CTTATCTCCG	TTCTGTGAAT	ATTATCTATA	'I'I'AGAAG'I'GA	TTATTATATT	CGCAAAACTA	ACGACTATGT	GAGATGTAAT	TGTGATAAAG	GATACTCTAT
TTGTTGTGCC	CCATCTATCC	TGGATTACTC	GTATATTTCG	ATTATTGTGC	TAAAAAACTT	ATTGTGATAG	TGAAATCCCG	GCAGTGTCCA	GTTCTTTTTT
GCACACTTTT	CCAACTACAA	AAAGTCAAAA	TGCATAGCTG	GGACGTGATT	CCTCAGACCA	ACTACGAGAA	CAATTGGTAT	AACAACAATT	ACCAGATGAA
AACAGAGCCC	CACGATGGGT	TCAACGGGCA	ACAGCCCAAT	TCCCCGCAGA	GCATGGACAG	CATTCACCCT	GAAACACATC	ACAGTTCTCC	AGTTCAACAA
CAACATATGA	TGTTCGATTC	GTCCAATATT	ATAAACACCA	TGACCCAACT	ACACAACGTT	CAAATGCAGA	GACAGACCCA	CTTCAATCCC	CTTACTCCTC
CGGGTTATCC	AGGCGCTATG	ATCCATCCCC	AAAACTCTCA	GGCAAACTCA	ACACCATTTA	GAAGCTTTAC	AAAGGGACTG	GACTCGATTC	CTTTTGGAAA
TAATGTATCC	AACTTAACAC	CGAGTCACAC	TCCTCCAATG	GACATAACTC	CGCCAAAGTC	ACCAAAGTTC	AACGGCAAGG	AAACCCCTGA	AAAGGATTCT
CTAAAGCAGG	ACCAAAGTCA	ACTTCTAAAA	ACCCCAATCC	AGACGAATGG	AAATGGAAAC	CAGCAATCGA	CGTTCGACTC	TGGCGAAGAC	AGCCACTCAA
TTCCCGATAG	CGATCTCCTT	GAACCGGTAA	TCACCGACGG	TGCGGACGTA	GATGACGAAA	ACGATGCTGA	AGAGGACGAT	GACATTCGCA	CTCCGAAAAT
CAATTCGCAC	GGTAAAATGA	AGACGTACAA	GTGCAAACAG	TGTGACTTTA	TCGCAGTTAC	AAAACTGTCC	TTCTGGGAGC	ACAATAGGAT	CCACATCAAA
CCTGAGAAGA	TGCTCAAGTG	CCAAAAGTGT	CCCTTTATCA	CCGAATACAA	GCACCATCTT	GAGTACCACC	TGCGAAATCA	CAACGGATCA	AAGCCCTTCC
AATGTAAACA	GTGTAACTAC	TCCTGCGTGA	ACAAATCCAT	GCTCAACTCA	CACATGAAGT	CGCACAGTAA	TATCTACCAG	TACCGGTGCA	AAGACTGCAA
CTATGCAACC	AAGTATTGCC	ACTCCTTGAA	ACTGCATCTC	CGCAAATATT	CGCACAATCC	AGCCATGGTG	TTGAACTCTG	ATGGAACGCC	AAACCCACTG
CCGATTATTG	ATGTCTATGG	AACGAGAAGA	GGACCAAAAG	TAAAGTCTCA	TAAGGATGAA	GGAGGGCATA	ATTTACTTAA	СТСАААСАТА	AATACCAGTA
GAAGGAGCAA	GTCAGGGAAA	CGGGACAGTT	TTCCGAATTT	CGAACAGTCT	CAACATGTTT	CCAATAATCC	ATCAAGTCAG	GCTTTGGCAA	TGTTGCCCAA
TTTGGCCAAC	ATCTTCCAGC	AGAGTCCCAG	TATGCCCCTG	TTCCCCTACC	TGAACCTCAA	CTTTCACCAC	ATTCTGGCCC	AGCAAAAGGC	AGCCCTTTCA
CAAATCTCCC	CATCAATAAA	TGGGTGGCAA	AATGAGGAAA	ACTGCAACGA	GGAAGAGACT	CCAGAAAAGG	AGGAAGACCC	CAAACGAATG	TCTGCCCTTG
ATCTCAGCAG	CAATCCTAGC	ACCCCATCAA	CAATGAGCCA	AGTTAAACAT	AAGCGCAAAG	GCAGGGCATT	CAAATTAGAG	CTGATGAAGG	AGAGTTCCGA
TGACGAAGAA	GGCCAAACAA	TTCGGACTTT	AGGGGAGATC	AGGAGCGAGC	TGGAGACACC	AAAACCAGTT	CAACTTCAGT	TACCCACCTC	GAGCACCACC
ACACCTCTAA	AGACTACCTC	TGAAGATGAT	TCCACATCGG	TGGAACCTTT	GCAGAATTTG	TACGAGTGCA	AATTTTGTGA	TATCTCATTC	AAGCACGCCG
TTCTCTACAC	AATCCACATG	GGCTATCACG	GTTACAACGA	CGTCTTCAAG	TGTAATGCGT	GTGGCAAAAA	GTGCGAGGAT	CGAGTTGCGT	TCTTTTTGCA
CATTGCTCGG	GACGCACATG	CGTAAGTCAT	GTCACAAAAT	CACAAAACCT	ТААААТАААТ	CCATAAAACG	TTTGTACAGT	CCAGGAAAGT	GTGATATTAC
GACAATAATT	ССТААТТТТА	AGTATTTTAT	ACGTATTTTT	ATGTAATTAT	CTTAAGTATT	ТАААААТАА	CAAAATCGAG	AGAGGAAGAG	GCTAGTTCCA
AAAGTTAAGT	TATGTACACC	AAAAAGAGG	AGGGTAAAAT	CCGATCCGAA	AACCGTATAG	TCATTATGGG	GTGAATTTTT	TAGTTTTAAC	ATATTTTTTA
TTTGTGTTTTT	TCATTCCCTT	CGAATTAATA	CGGACAACTC	CTCTTCCCAA	GCCAGGGCTA	GTGAGGGTTG	TTGTCCTATT	TGGTCGTTGT	ATAGATTGTT
GTTATATTAC	CAAAATGTCC	CTAGGAGTCT	TTTTTGGCTT	ACAAGTGTAT	AAGAGACACT	CCTTTCTGAG	CGAACCAAAT	CAAAATGTCA	CCAAGGATAT
CAAATTTAGT	TTGTAAATTT	ATTTTAGAGA	TAAATTGTAT	АТАТТААТАТ	TTCAATCATG	TAGTTTATAT	ΑΤΤΤΑΤΑΤΑΑ	TGGTTGATAA	ТТАААСААТА
AA									
	CGGACCTTAC CTTATCTCCG GCACACTTTT AACAGAGCCC CAACATATGA CGGGTTATCC TAATGTATCC CTAAGCAGG TTCCCGATAG CCTCAGAAGA CCTCAGAAGAA CTATGCAACC CCGGATTATTG GAAGGAGCAA TTTGGCCAAC CACATCTCACAC ACCTCTAGAGA TAGCGAAGAA ACACCTCTAA TTCTCTCACAC GACAATAATT AAAGTTAAGT TTGTGTTTT GTTATATTAC CAAATTTAGT AA	CGGACCTTAC GTTCTGTATT CTTATCTCG TTCTGTGAAT TTGTTGTGCC CCATCTATCC GCACACTTT CCAACTACAA AACAGAGCCC CACGATGGT CAACATATGA TGTTCGATTC CGGGTTATCC AGCGCTATG TAATGTATCC AGCGCTATG TAATGTATCC AGCGCTATG TAATGTACC GGTAAAATGA CTTAGCAAGAG TGCTCAAGTG AATGTAAACA GTGTAAATGA CCTGAGAAGA GTGTAAATGC CCGAGAGAGA GTGTAACTAC CTATGCACC AGTATTGCC CGATTATTG ATGTCTATGG GAAGGACCAC ATCTTCCAGC CAAATCTCCC CATCAGTGC CAAGAGAGA GCCAACAA ATCTCAGCAG GAACCAACTC TCTCTCTACAC AATCTACC TCTCTCTACAC AATCTACC TCTCTCTACAC AATCCACTG GACAATATT CCTAATTTG GACAATATT CCTAATTTA AAAGTTAAGT TATGTACACC TTTGTGTTTT TCATTCCTT GTATATTAC CAAAATGTC CAAAATTTAC CAAAATGTC CAAAATTTAC TTGTAAATTT AA	CGGACCTTACGTTCTGTATCAGTCTCTGACTTATCTCCTTCTGTGAATATTATCTATATTGTTGTGCCCATCTATCCTGGATACTAAACAGAGCCCACGATGGGTTCAACGGCACAACATATGATGTTCGATCGTCCAATATCGGGTTATCCAGCGCTATGATCCAATATCGGGTTATCCAGCGCTAGAACACGAAAAACAGACAAAGCCTAACACCGACTATATCGGGTTATCCAGCGCTAGAACTCTAAAATCCCGATACGGACTCCTTGAACGGTAATTCCGATAGGGACTCCTTGAACGGTAACTGAAAATCACCGAGTACACCCCAAAGTGAAATGTAAACAGTGTAACTACTCCTGCGTGACTATGCAACCAAGTCTCAGGACGCACAGTTTTTGGACAACATCTTCAGCACGGGACAGTTTTTGGCCAACATCTTCAGCACGGGACAGTTTTTGGCCAACCATCTCCAGCAGAGTGCCAACAACTCACAAAGCCCAACAATTCGGACTATAACTCACACAGCCAAACAATTCGACAGTTAACCTCAACAAGACTACCCCAAAATACCACCAATATATCCTAATTATAGACTACTCAGACAATAATCCTAATTTATAAAAAGAGGTTTGTGTTTTTCATCACCACAAAAAAGAGGTTTGTGTTTTTCATCACCACCAAATAAAGAGGTTTGTGTTTTTCATCACCACCAAAAAAGAGGTTTGTGTTTTTCATCACCACCAAAAAAGAGGTTTGTGTTTTTCATCCCTCCGAATTAATAGTTATATACCAAAATGCCCTAGGAGTCTCAAATTAACCAAAATGACCCTAGGAGTCTCAAATTAACCAAAATGACCCTAGGAGTCTCAAATTAATCCAAAATGACCCAAATAATACCAAATTAAT	CGGACCTTACGTTCTGTATTCAGTCTCTGAANACTAAGAACTTATCTCCGTTCTGTGAATATTATCTATATTAGAAGTGATTGTTGTGCCCCATCTATCCTGGATTACTCGTATATTTCGGCACACTTTCCAACTACAAAAAGTCAAAATGCATAGCGGAACAGAGCCCACGATGGGTTCAACGGCAACAGCCAATCAACATATGATGTTCGATTCGTCCAATATTATAAACACCACGGGTTATCCAGCGCTATGATCCATCCCAAAACTCCACTAAGCAGGACCCAAGTCAACTCCCCAATGTCCCCCAATGCTAAGCAGGCCCAAAGTCAACTCCCCAATGCCCCCCACGGCCTGACAAGAGGTCTCCTTGAACGGTAAAGTGCAAACAGCCTGACAAGAGTGTAACTACTCCTGCGGGAACCAACAGTCACCTGACAACAGTGTAACTGCCCAAAGTCCACCTGCACAACAGCCTGACAACAGTGTAACTGCCCAAAGTCCATCCCAAAGTCCATCCCTAGGAACAGTGTACTGGACCACACTCTACAGCACATCCCTATGCAACAATCTCCAGGGAACGGCACAACAAGGGCAACACACTTCGACACACTTCCAGCAACGGCCAACATCCCAAAGGAAAATCTCCACACAACCCCCATCAAAATGAGGAAAATCTCACACAACTCACACAGAGGCCAACACATCCACACAGACAACCTCTAAAGACTACCTTGCACAACAATCCACACAGGCACACTCTAAAGACTACCTTGCACACAGAGTCACAACAACACCTCTAAAGACTACCTTCGACATCAGGTCACAACAACACCTCTAAAAGACTACCTTCGCACTCAGAAATGAGAAACTCTCACACAAGCCAACTCGTACAACAAATCCACACAGACACACTCTAAAAGACTACCCCTGAAAAAGAGGAGGGTAAAATCTCTCAC	CGGACCTTACGTTCTGTATTCAGTCTCTGAAAACTAAGAACGTATTTTAACTTATCTCCGTTCTGTGAATATTATCTATATTAGAAGTGATTATTATATTTTGTITGTGCCCCATCTATCCTGGATTACTCGTATATTTTCGATTATTATGTGCGCACACTTTCCAACTACAAAAAGTCAAAATGCATAGCGGGGACGACGATTAACAGAGCCCACGATGGGTTCAACGGCAACAGCCCAATTCCCCCGCAGACAACATATGATGTTCGATTCGTCCAATATTATAAACACCATGACCCAACTCGGGTTATCCAGCGGTATGGTCCAATATCAAAACTCAATGGGACAACTCCCAACATATGAACCTAAAGTCAACTCTAACACGGCGAACTCCCTAAGCAGACGACCAAGTCAACTCTAAAAAGCCCCAATGCGGCGAACGATACAATCGACAGGCAAAATGAACTCCACAGGTGCGGACGTACAAAGTTAAACAGGTCAAGTACCCAAAGTCAAGGCGAACATACATGCAAACAGTGTAACTACTCCTGCGAGGGCCAAACTCACATAGCAACAGTGTAACTACTCCTGCGAACAAGTCAACTCACATAGCAACAGTGTAACTACTCCTGCGAACAAGTCAACTCACAAAGTAAACAGTGCAACAACATTCCCCACTCCGAAACAATCCGAACAACATCTCCAGCACACACACACTCGCAAACAATCCGACAAACATCCGGGCAAAACGCCAACACATTCCCCACGGCAAACTACCCCAACATACACGGGCCAACACATCCCCACGGCTACAAGAAACCCCAATGGAACACACTAAAGACACCTCAAACTACCCCAACACACACGGCAAACCATTCCACAACAACTCCCACACAGGCCAACCACGTCACAACAATCCACACACGCTACAAGAAAACCCCAATGGGAAACACTAAACTACACAGCAA	CGGACCTTACGTTCTGTATCAGTCTCTGAAAACTAAGAACGTATTTTAAGTGTTGTTATCTTATCTCCGTTCTGTGAATATTATCTATATTAGAAGTAATTATATATTTCGCAAACTATTGTGTGCCCCATCTATCCTGGATTACTCGTATATTTTCGATTATTATGTCTAAAAACTTGCACACTTTCCAACTACAAAAGCCAAAATGCATAGCGGGACGTGATTCCTCAGACCAAACAGAGCCCACGATGGTTCAACGGCAACACCCCAATTCCCCCCAAAGCACACATTCGGGTTATCCAGGCGCATGGTCCATATTATAAACACCATGACCACCAACATTTCGGGTTATCCAGGCGCATGATTCTTCAAAAACCCCAATCGGCAAACTCAAACCACAACGTTCGGACAATATGAGGCGCATGATTCTTCAAAAACCCCAATCCAGACGAATGCAAACCACAACGTTCAACAAAGTCAAGCGTACAACTCCCCAAAGTCAAACGGAAAGAAACGGAAAGAATGGAAACACTCCCGAAAGGGATCTCCTTGAACGGTAAACCCCAATCCCAGACGAATAACGACCAAAGTCCCTGACAAGACGATCACCTTGAACGGTAAAGTGCAAACAGGTCAACTACCACACATCTTCATGCAAACAGTGTAACTACTCTGCGCTACAGCACAATCACCACACATCTTCATGCAAACAGTGTAACTACTCTGCGCAAAGTCCAATCACCACAATCACCATGCAAACAGTGTAACTACTCTGCGCAACAGCACAATCCCGCACAATCCCCATGCAAACAGTGTAACTACTCTGCGCTACCGACAATCCCCGACAATCCCCATGCAAACAGTGTAACTACTCTGCGCTACCGACAATCCCCGACAATCCCCTATGCAACAGTGTAACTACCTCTGCGCAAACGCCAAAGTCCGCAAAGTCCCTATGCAACAGTGCAACAATTCCCCCACACGCACAACAT <th>CGGACCTTACGTTCTGTATTCAGTCTCTGAAAACTAAGAACGTATTTTAAGTGTTGTTATTTAGTAGTTCTTATCTCGTTCTGTGAATATTATCTCATATTAGAAGGGATTATTATATTCGCAAAACTAACGACTATGTTTGTGTGCCCCATCTATCCTGGATTACTCGTATATTTGGATTATTGTGCTCAAAAACTAAATGCAAAAGCACACTTTCCAACTACAAAAAGCAAAATGCATACGGGGACGGGATTCCTCAGACCAACTACGAGAAAACAGAGCCCACACTACATGTCCAATATTATAAACACATGACCGCAAACCACACACCTCACACACGTCACACACCTCAACTATGATGTTCGATTCGTCCAATATTATAAACACCATGACCACACAACACAAACGTCAAATGCAGACGGGTTATCCAGGCGCTATGGTCCATCCCAAAACTCAAGGGACAAACAAACACAAAGTCACACAAAGTCCAAATTTACCAGCGCAAAGTCAACTTCCCAAAGTACCCCAAAGTAACACAAAGTCCCACAAAGTCAACACAAAGTCACAAATGCACAGGACTTACACCGACAAGTAAACACGATACAAGCACAAGTAACAGCATCGAAAAGCACAAACAAATGCACAGTGTAAATGAACACGGTAAAACACGATACAAGCACAATTACGACAACTCCACAAAGCACAACACAATGCAAAACAAATGTCACGAACGATAACCCCAAGTAAACACACACACACAAAAGCACACACAAAGCAAACAAACAACACGACACACATGCAAAACAGTGTAACTACTCCTGGCAAAAACAAACATTACGCAAAGTAACGACAACTACAGCACAACTACAAAAGCACACCACCATGTAAAAACAGTGTAACTACTCCTGGCAAAAACAGGTGAAACTACAGCACAACTACAGCACAACTACAAAATGAACACACACACACACACACACACACACACACACAC</th> <th>CGGACCTTAC GTTCTGTATT CAGTCTCTGA AAACTAAGAA CGTATTTTAA GTGTTGTTTA TTAGTTAGTT AAAAATCTGC CTTATCTCCG TTCTGTGAAT ATTATCTATA TTAGAAGTGA TTATTATATT GGCAAAACTA ACGACTATGT GAGATGTAAT TTGTTGTGCC CCACTATCC TGGATTACTC GTATATTTGG ATTATTGTGC TAAAAAACTT ATTGTGATG GGAGATGTAAT TTGTTGTGCC CCACCTACCA CAAGTCAAAA TGCATAGCTG GGACGTGATT CCTCAGACAA ACTACGAGAA CAATTGGTAT AACAGAGGCC CACCGATGGT TCAACGGCA ACAGCCCAAT TCCCCGCAGA GCATGGACGA CATACGAGAA CAATTGGTAT AACAGAGGCC CACCGATGGT TCAACGGCA ACAACCCCA TTACAACACT ACACACGTT CAAAGCAGAA GAACACAC GGCGTATCC AGGGCTATG ATCCATCCC AAAACTCCC GGCAAACTCA ACACACGTT CAAAGCAGA GACAGACCCA CGGGTTATCC AGGGCTATG ATCCATCCCC AAAACTCCA GGCAAACTCA ACACACGTT CAAAGCAAAC CAAAGGCC TTAAGCAGG ACCAAAGTCA ACTTCTAAAA ACCCCAATCC AGACGAATGG AAATGGAAAC CAGAAGTCC ACCGACAGG CTAAAGCAGG ACCAAAGTCA ACTTCTAAAA ACCCCAATCC AGAGGAATGG AAATGGAAAC CAGCAATGCA AGAGGACGAT CAATGTAACCA GGATCTCCT GAACGGTAA TCACCGGACGG TGCGGACGTA GATGGCGAAA ACGATGCTCA AGAGGACGAT CAATGTAAACA GTGTAACTAC TCCTGGGTAA CCACGGACGG TGCGGACGTA GATGGCGAAC CACAAGGTCC TGCGGAAATGA AATGTAAACA GTGTAACTAC TCCTGGGTGA ACAAATCCAT GCTGAACTCA CACACATCTT GAGTACCAC CGCGAAGTG CCTGAGAAGA TGCTCAAGG CCAAAGGTG TCCCTTTATC ACGAATGCA CACCATCTT GAGTACCAC CTGGGAAATCA AATGTAAACA GTGTAACTAC TCCTGGGTGA ACAAATCCAT GCTCAACTCA CACATGAAGT GGCACAGTA TATCTACCAG CTATGCAACC AGTGTATGC ACTCCTTGAA ACTGCATCT CGCAAATTT GGCACAATCC AGCACGGGG CATA ATTACTACCAG CTAGGAACAA GTCAGGAAAA TGGGAGAGA GGCCAAAGG TAAAGTCTCA TAAGGTGTA ACTATCACTG CCGAATATTG ATGTCTATGG AGGGACAGAG GACCCAAAG TAAGAGTCT CAACGATCC ACAATAACA TTGGCCAAC ATCTTCCAGC AGGGCACAGA AATGAGGAAA ACTGCACGA GAAGGACCT CAATAAGAG TTTGGCCAAC ATCTTCCAGC AGGGCGCAA AATGAGGAAA ACTGCAACGA TGGAACACC ATACTACGAG CACGGACATT CAATATGAG GACGAAGAA GCCCAACAA TTGGGACGACA TGCAAGAGA ACTGCACAAAAGG CCAGGGCATA CAATATGAG AAACTACTCC CAAAAAAGAGG GGCTAACAG GTGGAACCTT GAAACACC TTAAAGAG CAAGAGGCAT CAATATAGG GGCAAAAAA GGCCAACAA TTCGAACAG TTCACACGA GGCACACCA AAACCACCA AAACCACCA AAACCAGTT CAATATAGG GACAAAAAG AGCCAACCA TTGGAAGAGA GGCCAAAAACCT TGAAACACT TAAAAAAAGA GTGCGAGAAT CATTGCTACA AGACTACCC</th> <th>CGGACCTTAC GTTCTGTATT CAGTCTCTGA AAACTAAGAA CGTATTTTA GTGTTGTTTA TTAGTTAGTT AAAAACTGC TCTCCCCAAA CTTATCTCCG TTCTGTGAAT ATTATCTTATA TTAGAAGTGA TTATTATATT CGCAAAACTA ACGGACTATGT GAGATGTAAT TGTGATAAAG TTGTGTGTGCC CACTCATCC TGGATTACTC GTATATTCG ATTATTATCT CGCAAAACTA ACGGACTATGT GAGATGTAAT TGTGATAAAG GCCACTTTT CCAACTACCA AAAGTCAAA TGCATAGTG GGACGTGATT CCTCAGACCA ACTACGAGAA CAATTGGTAT AACAACAAT AACGACACCT CACGAGAGGT TCAACGGCCA ACACCCCAAT TCCCCGCAGA GCAGGACCA ACTACGAGA GACAACCAC ACCACGTC CAACATATGA TGTCGATTC GTCCAATAT ATAAACACCA TGCCCGAGA GCACGACGA CAATGCAGA GACGACCA CTTCAATCCC CGGGTTATC AGCGGCTAG ATCCATCCC AAAACTCCA GGCAAACTA ACACACGT CAAAAGCAG GACGACCA CTTCAATCCC CGGGTATACA CGCACAGCA CTCCTCAAAA ACCCCAATG GACGACACGA CACCAACGT CAAGGGAG GACGACGA GACGACCG CTAAGACACAC CGACGCAACA CTCCTAAAA CCCCCAATG GACGAGAGCA CACCAAGGT CACCAAGGA GACGACGA GACGACCG CTAAGGACG CCAAAGTCA ACTCTTAAA ACCCCAATC GGCGAAGCA GACGAAACGA CAGCAACGA CGTCGGACC TGCGGAAGCA CTAAGGAGG GCAAAAATGA AGCGTACAA CTCCCCAATG GCGGAGCGT GAGAGCAAC CAGCAATGCA CAGCGACGA CGACTGCGA CAATTGCAAC GGTAAAATGA AGCGTACAA CTCCTATAC CCCGAAAGT CACAACTGC CTCGGGAGAC CACATAGGAT CTGCGAAGG GCAAAATGA AGCGTACAA CTCCTTTAC CCGAAATCA GACCACTCT GGACAACTCC TGCGGAAGCA CACATAGGAT CTAGGAAGA GGCTAAATGC ACTCCTTGAA ACGGCACATA TGCGAAACAG GCACCAATTAGCAC TACGAGGAC CAATTGCAAC AGTAATTGC CACCCTTTAA ACCGATACA GCACACTCT AGGAGACC TGCGAAACCA CACCGGGACA ACATTGCAAC AGTAACTAC CCCCTGCTGA AAATGCAT GCACAATCC ACACGAGGA ACTACTCC TGCGAAACAC ACATAGGAT CTAGGAAAC GTAGAACTAC TCCCGCTGA AAAGTCAC GCACAATAC AGGACAATC CACCAGGAA ATTTACTTAA CTCAAACTA GAAGGACACC AAACAATTGCC CACCCATGA AATGGAGAA CTCCCACCT GGAAACAC ATTTGCGAA ACGACTAC AATGGAAC CTATGCAACA ATTCCAACG AAAGGACA ATTGCGACTG AGCACACTT GGAACCACT CAACTGC CACAAAAGG AGGACACAT TTGGCAACA ATTCTCAACA AAGGGCAAAAA CTCCCACAC ACGACAGAC CTTACACG AGGAAGACC CAAATAGAA AGCGTACAA TTCGGACTA AATGGAGAA ACGCACTAA ACGACAGC CAAAAAGGA AGCAACACT TGCGACACA ATTGGGAAA ACGCACTAC AAAGGAAA ACGACGACA ATTGTGGACAA ATTGCGAC AAAGGAAA CTCTACCAC AAACAACT TCCGAACAA TCCGAACAG TGCAAAAAACAACAA CGAGAACAC CAAAAAGAAA CCAACA</th>	CGGACCTTACGTTCTGTATTCAGTCTCTGAAAACTAAGAACGTATTTTAAGTGTTGTTATTTAGTAGTTCTTATCTCGTTCTGTGAATATTATCTCATATTAGAAGGGATTATTATATTCGCAAAACTAACGACTATGTTTGTGTGCCCCATCTATCCTGGATTACTCGTATATTTGGATTATTGTGCTCAAAAACTAAATGCAAAAGCACACTTTCCAACTACAAAAAGCAAAATGCATACGGGGACGGGATTCCTCAGACCAACTACGAGAAAACAGAGCCCACACTACATGTCCAATATTATAAACACATGACCGCAAACCACACACCTCACACACGTCACACACCTCAACTATGATGTTCGATTCGTCCAATATTATAAACACCATGACCACACAACACAAACGTCAAATGCAGACGGGTTATCCAGGCGCTATGGTCCATCCCAAAACTCAAGGGACAAACAAACACAAAGTCACACAAAGTCCAAATTTACCAGCGCAAAGTCAACTTCCCAAAGTACCCCAAAGTAACACAAAGTCCCACAAAGTCAACACAAAGTCACAAATGCACAGGACTTACACCGACAAGTAAACACGATACAAGCACAAGTAACAGCATCGAAAAGCACAAACAAATGCACAGTGTAAATGAACACGGTAAAACACGATACAAGCACAATTACGACAACTCCACAAAGCACAACACAATGCAAAACAAATGTCACGAACGATAACCCCAAGTAAACACACACACACAAAAGCACACACAAAGCAAACAAACAACACGACACACATGCAAAACAGTGTAACTACTCCTGGCAAAAACAAACATTACGCAAAGTAACGACAACTACAGCACAACTACAAAAGCACACCACCATGTAAAAACAGTGTAACTACTCCTGGCAAAAACAGGTGAAACTACAGCACAACTACAGCACAACTACAAAATGAACACACACACACACACACACACACACACACACAC	CGGACCTTAC GTTCTGTATT CAGTCTCTGA AAACTAAGAA CGTATTTTAA GTGTTGTTTA TTAGTTAGTT AAAAATCTGC CTTATCTCCG TTCTGTGAAT ATTATCTATA TTAGAAGTGA TTATTATATT GGCAAAACTA ACGACTATGT GAGATGTAAT TTGTTGTGCC CCACTATCC TGGATTACTC GTATATTTGG ATTATTGTGC TAAAAAACTT ATTGTGATG GGAGATGTAAT TTGTTGTGCC CCACCTACCA CAAGTCAAAA TGCATAGCTG GGACGTGATT CCTCAGACAA ACTACGAGAA CAATTGGTAT AACAGAGGCC CACCGATGGT TCAACGGCA ACAGCCCAAT TCCCCGCAGA GCATGGACGA CATACGAGAA CAATTGGTAT AACAGAGGCC CACCGATGGT TCAACGGCA ACAACCCCA TTACAACACT ACACACGTT CAAAGCAGAA GAACACAC GGCGTATCC AGGGCTATG ATCCATCCC AAAACTCCC GGCAAACTCA ACACACGTT CAAAGCAGA GACAGACCCA CGGGTTATCC AGGGCTATG ATCCATCCCC AAAACTCCA GGCAAACTCA ACACACGTT CAAAGCAAAC CAAAGGCC TTAAGCAGG ACCAAAGTCA ACTTCTAAAA ACCCCAATCC AGACGAATGG AAATGGAAAC CAGAAGTCC ACCGACAGG CTAAAGCAGG ACCAAAGTCA ACTTCTAAAA ACCCCAATCC AGAGGAATGG AAATGGAAAC CAGCAATGCA AGAGGACGAT CAATGTAACCA GGATCTCCT GAACGGTAA TCACCGGACGG TGCGGACGTA GATGGCGAAA ACGATGCTCA AGAGGACGAT CAATGTAAACA GTGTAACTAC TCCTGGGTAA CCACGGACGG TGCGGACGTA GATGGCGAAC CACAAGGTCC TGCGGAAATGA AATGTAAACA GTGTAACTAC TCCTGGGTGA ACAAATCCAT GCTGAACTCA CACACATCTT GAGTACCAC CGCGAAGTG CCTGAGAAGA TGCTCAAGG CCAAAGGTG TCCCTTTATC ACGAATGCA CACCATCTT GAGTACCAC CTGGGAAATCA AATGTAAACA GTGTAACTAC TCCTGGGTGA ACAAATCCAT GCTCAACTCA CACATGAAGT GGCACAGTA TATCTACCAG CTATGCAACC AGTGTATGC ACTCCTTGAA ACTGCATCT CGCAAATTT GGCACAATCC AGCACGGGG CATA ATTACTACCAG CTAGGAACAA GTCAGGAAAA TGGGAGAGA GGCCAAAGG TAAAGTCTCA TAAGGTGTA ACTATCACTG CCGAATATTG ATGTCTATGG AGGGACAGAG GACCCAAAG TAAGAGTCT CAACGATCC ACAATAACA TTGGCCAAC ATCTTCCAGC AGGGCACAGA AATGAGGAAA ACTGCACGA GAAGGACCT CAATAAGAG TTTGGCCAAC ATCTTCCAGC AGGGCGCAA AATGAGGAAA ACTGCAACGA TGGAACACC ATACTACGAG CACGGACATT CAATATGAG GACGAAGAA GCCCAACAA TTGGGACGACA TGCAAGAGA ACTGCACAAAAGG CCAGGGCATA CAATATGAG AAACTACTCC CAAAAAAGAGG GGCTAACAG GTGGAACCTT GAAACACC TTAAAGAG CAAGAGGCAT CAATATAGG GGCAAAAAA GGCCAACAA TTCGAACAG TTCACACGA GGCACACCA AAACCACCA AAACCACCA AAACCAGTT CAATATAGG GACAAAAAG AGCCAACCA TTGGAAGAGA GGCCAAAAACCT TGAAACACT TAAAAAAAGA GTGCGAGAAT CATTGCTACA AGACTACCC	CGGACCTTAC GTTCTGTATT CAGTCTCTGA AAACTAAGAA CGTATTTTA GTGTTGTTTA TTAGTTAGTT AAAAACTGC TCTCCCCAAA CTTATCTCCG TTCTGTGAAT ATTATCTTATA TTAGAAGTGA TTATTATATT CGCAAAACTA ACGGACTATGT GAGATGTAAT TGTGATAAAG TTGTGTGTGCC CACTCATCC TGGATTACTC GTATATTCG ATTATTATCT CGCAAAACTA ACGGACTATGT GAGATGTAAT TGTGATAAAG GCCACTTTT CCAACTACCA AAAGTCAAA TGCATAGTG GGACGTGATT CCTCAGACCA ACTACGAGAA CAATTGGTAT AACAACAAT AACGACACCT CACGAGAGGT TCAACGGCCA ACACCCCAAT TCCCCGCAGA GCAGGACCA ACTACGAGA GACAACCAC ACCACGTC CAACATATGA TGTCGATTC GTCCAATAT ATAAACACCA TGCCCGAGA GCACGACGA CAATGCAGA GACGACCA CTTCAATCCC CGGGTTATC AGCGGCTAG ATCCATCCC AAAACTCCA GGCAAACTA ACACACGT CAAAAGCAG GACGACCA CTTCAATCCC CGGGTATACA CGCACAGCA CTCCTCAAAA ACCCCAATG GACGACACGA CACCAACGT CAAGGGAG GACGACGA GACGACCG CTAAGACACAC CGACGCAACA CTCCTAAAA CCCCCAATG GACGAGAGCA CACCAAGGT CACCAAGGA GACGACGA GACGACCG CTAAGGACG CCAAAGTCA ACTCTTAAA ACCCCAATC GGCGAAGCA GACGAAACGA CAGCAACGA CGTCGGACC TGCGGAAGCA CTAAGGAGG GCAAAAATGA AGCGTACAA CTCCCCAATG GCGGAGCGT GAGAGCAAC CAGCAATGCA CAGCGACGA CGACTGCGA CAATTGCAAC GGTAAAATGA AGCGTACAA CTCCTATAC CCCGAAAGT CACAACTGC CTCGGGAGAC CACATAGGAT CTGCGAAGG GCAAAATGA AGCGTACAA CTCCTTTAC CCGAAATCA GACCACTCT GGACAACTCC TGCGGAAGCA CACATAGGAT CTAGGAAGA GGCTAAATGC ACTCCTTGAA ACGGCACATA TGCGAAACAG GCACCAATTAGCAC TACGAGGAC CAATTGCAAC AGTAATTGC CACCCTTTAA ACCGATACA GCACACTCT AGGAGACC TGCGAAACCA CACCGGGACA ACATTGCAAC AGTAACTAC CCCCTGCTGA AAATGCAT GCACAATCC ACACGAGGA ACTACTCC TGCGAAACAC ACATAGGAT CTAGGAAAC GTAGAACTAC TCCCGCTGA AAAGTCAC GCACAATAC AGGACAATC CACCAGGAA ATTTACTTAA CTCAAACTA GAAGGACACC AAACAATTGCC CACCCATGA AATGGAGAA CTCCCACCT GGAAACAC ATTTGCGAA ACGACTAC AATGGAAC CTATGCAACA ATTCCAACG AAAGGACA ATTGCGACTG AGCACACTT GGAACCACT CAACTGC CACAAAAGG AGGACACAT TTGGCAACA ATTCTCAACA AAGGGCAAAAA CTCCCACAC ACGACAGAC CTTACACG AGGAAGACC CAAATAGAA AGCGTACAA TTCGGACTA AATGGAGAA ACGCACTAA ACGACAGC CAAAAAGGA AGCAACACT TGCGACACA ATTGGGAAA ACGCACTAC AAAGGAAA ACGACGACA ATTGTGGACAA ATTGCGAC AAAGGAAA CTCTACCAC AAACAACT TCCGAACAA TCCGAACAG TGCAAAAAACAACAA CGAGAACAC CAAAAAGAAA CCAACA

Source: Lambda ZAP clone, isolated from a maternal cDNA library (Schmidt-Ott, unpublished). One putative polyadenylation signal was identified in the in the putative 3' untranslated region (2797..2802), and one putative NRE sequence (2584..2604).

# SEQ20 Clogmia hunchback, genomic.

1	TCTTCAAATT	TACACACAGT	ATATCTCTCT	ATATTATCTG	CCTCTGTGTC	CTAGTGGTTA	GTACCCTGGA	CATCCTCCTC	GAGGTCGTGG	GATCAAATCC
101	CGGGCGGATA	TTTATCTAGC	GCGAAAAAGT	TTCCATTTGT	GTGTATATCA	ACAACATTGC	GAAATAATAT	TCATAGCGAA	TTATAAATTT	GTGATGGCGC
201	АААСАААТАА	ATAAAATGGG	CGGATTGACT	GGATCCGTTT	ACAACCGGTC	TAGGTGCTAT	TAACTGACTA	CTAATACACT	GAAAAAACAA	ACAGTCGTGT
301	ACAGTCTAAC	ACCCGTTGGT	TGAGGTTGAG	ACTACCCCTA	GTTCCGTCGT	GAATGTCGTG	GGAATGATTA	TTATTATTAT	TATATCTCTT	TACAGATTAC
401	AACTTTAGGT	CAAACCAGAT	TTGCAAATAC	TTGAATCAAT	TTTGCGAAAT	TCAGAAAAAA	ACGAAAAACC	GCAAATGCCG	AATGGGTCAG	CGTTATAGCC
501	ACACCCCCT	TTCCTCAGCT	ATGGGGCATT	ATTTCAACAA	TCTATTGATG	GGGGTCGTTA	GACAATCCGA	AATATATTCA	GAACTGCGAA	AATTTGGGTC
601	GCGAAGTCAT	AGCAGGCCGG	CAGTAGCTCA	CATTGAAAGT	CTAAAAACCC	TGATTTCGAA	AAATGAGGAA	ATGGTATTAA	TAATTTTCGA	TTTTAAAAAA
701	AGTAATAAAG	ACAATCGAAA	ATGGGGGTTA	TTATGCATTA	TTAGTGATTA	AATGGGCACA	ATAGGACAGC	TAATAAAGTC	ACAGTTACGA	GTAACATTTA
801	CTGACAAAGC	GGGTTTTTAG	GCTACTACTA	GACGTAGCTA	AAAAACGGCT	CGATGGATTT	AACCCAAACC	ATATTAAAAT	AAAGATTATT	AAATTCCCTA
901	CAAACCGTCC	ATACATAGTT	TTAGAGTTGG	ACTGCTGGAT	TCTAGTCATT	TTGGCACTAA	AGAAAATGTT	TAATTTGCTC	AATTTGGATG	CATTTTAATT
1001	CCATATAGTA	ATTCGGCCAA	TATTTTTGAG	TTATTTCATA	TTCAAAATGA	TTCGAAATTG	ATTTAAGTAT	TTGCAAATCT	GCTTTGACCT	AAAGTTGTAA
1101	TCTGATTAGA	GATATACTGT	GTGTAAATTT	GAAGATCCAG	CTTCTTTTGG	CTTCAGAGAT	AACAAGGATA	AACGGGTGCG	GCTATAACCT	CAATTCACAC
1201	TGTATCTGCA	CAAAGGCAAT	AAACTTGGCA	AAAAATGGGT	TTTTTTCTTGA	TCATCTTTAT	TTTCATCGCC	GCCGATGTCG	TTGTTTTCCC	CATTCCTAAT
1301	TGTCTTCTTT	TAATAGTGTC	CTCCCCGTT	TTGTGTGTCC	TTTCCACGAG	TTTCAGTCAC	TGGCAGAAGC	CGTGGCATTG	GTGGACAATG	GGTCGAATTT
1401	ATCTATTGGG	CATAATAATC	ATTTATTTGC	TCAAAATTGC	CATTAGACAC	CTCTCACACT	TTGGCTCCTC	TTCATTTGTT	GTCTCTTTCC	AATAAATTGC
1501	AATTTTTCAT	AGACCAAAAC	TCTCCCTCGG	TCCATGTCAA	TGGAACAGCA	GTGAAGGGAT	TAGCACAGCT	GAGGATAAAA	TGCGAAAAAA	CCGACTAGGG
1601	GAAATTGGAT	CTGGCTTGGT	CCACAGATTT	CTTCTCCTCT	CCTTGCTGAT	GATGTCGAGG	GGTGGGTGGT	GTGCATTTTG	GTGAAAGTAG	GTGGCAGAGG
1701	GTTGTGTTTT	CTGTCAAAAA	GTATAAGAAT	TGAGGGGAGC	CGCAAAAAAC	CGTCTACGTG	CAAAAAGAAT	TCCCATGGCA	ACCAGAAAAA	TAAATGGGCA
1801	TTTCTCTTAT	CGGAGACAAC	GGAACAATCA	TGGATTATTA	TGGAATGGCA	ATATTATTTC	GCTTTGCAGA	AATAGAAAAG	TCTGCTTTTG	ATGATGTGAA
1901	TTAAAACAAA	TGGCAAAATT	ATAATTCTTC	AATTTATATT	AAATGCATTT	TATTCCAGAA	GATCAATCTC	TAAAATATTA	ACTAAAATTT	ATAATCAAAA
2001	AATAACTTTG	AAATGCTATC	TTTACATCGT	ATCATAATAA	TTTAAAAGAT	AAAAATACTT	TCATTTTTTA	TATCAAACCC	AATTATCATC	CCTTTAAATT
2101	CCCCAGTTCA	TTCGCATACA	TTCCCATCAT	CATCCCCTCT	TCAATTTCCA	CCAATAGTAT	TCCAGCAAAC	ACATTCAACA	AAAAAATTC	AATACAATCA
2201	CATCCCCTAA	ATGTTACACT	CGGCGACAGT	GGAATAGGAC	CTATTTAGGC	GCATGCTTTA	GATTTGTTTG	AATGTAGTGT	TAATAATTTT	TTCTACAAAT
2301	GTATTTTAAA	AGATAAATTC	TTATATGAAT	ATTAACTAAT	TGAATTTGCC	TTAAACAAAA	ATTTAGATAT	TTCAAAACAG	CCGTTTAGTT	TTTTTGAATA
2401	TATTTTAATA	TATTTCAAAT	ATTTTTGTTA	CATAAAATTG	ATATTACGGT	TTTAACATTA	CATTTTAAAA	ATGATTAGTT	TTTTTTCTTTA	ATTTTAAAGA
2501	AACGATTTTC	CTCAAATTAA	TTAAATTCGG	TATATGTTCT	TGACATTCTG	GAAATTTACT	TAAATAAAAA	TAGAAAGCTT	TTTAAACATA	TTATTATTAT
2601	TGGCTTAGAT	ATAGTCATTA	CCTAGTCAAT	TGATTTCTAG	TAATTCATAT	CATGCATTGC	AAAATTGTCG	AATGATAAAA	TGTATTATTT	ATATTTGTAT
2701	AATGGAATAT	CAGTACAGGG	AAAAAACCCT	CACAACATGT	TCTTATGATT	TAGTCCACAA	GGTGATATTC	ATACTAATGG	TCTATATTAT	TAACATCCTA
2801	TCCTCCAAAG	TGTACAGACC	TATACTACTC	TAGAGTGCGG	AGCCGAATGG	CGGACCTATT	AAACACTTTG	ACCTAACCCA	TTTTTAATAA	GTCATTTGCT
2901	ATTATCTTGA	AATACTATAT	ATAGCAACAA	ATTAGTTAGG	GCCAGCAGTT	GTGATAGGGA	GGGAGAGCTA	CTGTCCATTA	TACACGCGGT	GATACGTTAT
3001	CGTACATCGT	TACTATGATC	TGATTCATAT	ATTTATATAT	GAATAAGAAC	TATAATCCCC	AAACGGGTTC	AAAGCCTAAC	CTACCATACT	AATTAAATTC
3101	GTATAGAACT	GAGCTTCGTT	GTGAATAGTT	TGCTGCATTT	CATGAAATGT	ACATATCGGA	TCTCTATTCT	TAAATGAATA	ACGATTATTT	GAGAGAGGAC
3201	ATTTCCCAAG	CCCAAAAAAG	TGATATAAAA	AATATCAAGT	ATACACGAAA	CAAGAATTGC	AAACTGTTTA	AAATTAATAT	TATATTTGTA	TTAATTTACA
3301	ACACCAACAA	TATCCCCAGA	TTTTGGCATT	AATTTGTTTA	AAAGATATAA	ATTTTTGTAC	CGAATACGAT	CTATAAATTC	GATTAATAAT	AGTTATTTGT
3401	AAAACTAGCT	CAATAAAATG	AAGCATTATT	AAACAGAATC	AAATGTCTTC	TATAAAACAC	TCGCCAATCT	CGCAACTTCA	AAGTATTTTA	GTGTTTTTAT
3501	TATTATTTTA	TAAATTCAAA	TATTCCAATA	ACTAATAGTT	AATTTATGAA	ATGTATAAAG	GTAATAATTC	TATCATTTCC	AACTAAATAT	TATAACCTAT
3601	TTCACTGTCC	TAAAGTGTAC	TTCAGTCAAA	TTGTTACTTT	CTCATAGAAA	GGGGCAATAC	GAAAAAGGGA	AATTTTGATT	TTGAAACTTA	CACACACTAG
3701	TACAGCAATA	TTTTTTTTCTA	CACAAATCCT	GCCCTGAGCA	AATTTTGTTT	TGAATTCTAT	CAATTAATGG	CAATGATATT	TTTTCAAAAT	ATTTTTTAGT
3801	TTCACACCGG	AACGTCGTGT	AGGTCGGACC	TTACGTTCTG	TATTCAGTCT	CTGAAAACTA	AGAACGTATT	TTAAGTGTTG	TTTATTAGTT	AGTTAAAAAT
3901	CTGCTCTTAC	CAAAAACGAT	TGTGCTTATC	TCCTGTGAAT	ATTATCTATA	TTAGAAGTGA	TTATTATATT	CGCAAGACCA	CCGATAATGT	GAGATGTGAT
4001	TGTGATAAAG	GATACTCTAT	TTGTTGTACC	CCAACTATCC	TGGATTACTC	GTATATTTCG	ATTATTGTGC	TTAAAAACTT	ATTGTGATAG	TGAAATCCCG

4101	ACAGTGTCCA	GTTCTTTTTT	GCACACTTTT	CCAACTACAA	AAAAGGTTAG	TGTTTTATTC	ATATTTTTA	АСААТАААТА	TTTTTCATTT	GATCATATCA
4201	GAAATTATTT	AAAGAACCTG	TTTTTAATTT	TATTTAATAA	TGTCATCGTT	TATTTTTTTA	AAATTATTTA	CCAAGAATAA	AAATTGAATA	ATACAGTTTC
4301	ATCTTGAATA	AGTTTTTCAT	AATGTAAAAT	TTATTTTGTA	TTCATTACTT	TTGTGCTTGA	AAACAAGTTT	AGCAAAAACA	ATATAAACCT	AAATTTTAGA
4401	ATCAGACGAG	AATTGCAATG	AAATATACCG	AACCGGTTTT	TAAATAAATA	ACCTAAATGT	TGAATATCAA	TCATATTTGA	TTTATTTTAT	CTCAATTTAT
4501	AGGAATCATT	TCTTTCTAAA	ATTAAATGTG	ААТСАТААТА	TTACCTATTT	TATTATAAAA	CTGATTCATT	GCAATTTAAT	AAATTAATTA	TATTTCATAG
4601	TTATATATTC	AAATTTTAGG	ACACACTCTT	TTCTAGAATA	AACATGTTTC	TAGAGCATGT	CGTATTTATA	GACACACTTG	GACTATTTGG	ATTATTAGCC
4701	TTTTTGCAGA	AATGTGTCCC	ААСАААСААТ	TGCATTCTGC	CATTCGACAC	GCTCCGCCAC	CCTTTGTCCG	AACTGCTCTA	AATAGGATTC	TTTCTTGGAA
4801	TTGCTGCAAA	TCCTTTGAAA	CGGGTCCTAA	AACAAACAAC	GCAGGCACAT	AAAACTCGGT	CCGAATTTCC	ATTATGGGAT	CGATCGGGCC	TCTGGTAGCC
4901	TATCCGGACA	AAAGAAAGAA	GGCGTATGCA	AATAAAGTGC	GAATTTATGG	GCGAAAGTCG	GTGAAAAGTT	AATTTCTAGC	CGTAAAATCG	AAAGGATAAT
5001	AATGGGACGA	GTACCTAGAA	TTTTACAACA	CAGTCCCAAT	AAAGCAATAA	ACAAAAGTCA	ATTGCGAATT	ACTGCAAAAG	TCAATCTTAA	TCATTTTGGC
5101	AGATCATCAT	GATTTACCCA	AGGCCATAAA	AAATCATTTA	CCTACTGTCC	ACTATTGGCT	CCGCAACTTG	GTTCGTAATG	ATTTCCCAGA	ATTTAGACTC
5201	ATCCTCAATC	GGTTGGGTTT	TAAACCAGTT	TCCCACAGGC	AATGGATTTT	TAATGCTAAT	TATGTCTAAA	AATTTCAAAA	GGTGTCTTTG	CATTTTCTCG
5301	TTGTAAATTA	GTATTTTCGA	AGAAAAAAAG	AACTATCGTA	TTTTAAACTG	TATATGAGTT	AATTTTGATC	ACAGAAGATA	AGTATTTGGC	AATAGTTTTA
5401	ATGCTAATTT	CTTCTTAAAA	TTCCAAATAA	ATCCATTTCC	AAAATTCTTA	GCTTTTTTTA	AACCAGACAT	TTCTTATTTC	ATTGATAGAC	AAAAGATTAT
5501	TTTTAATTAA	TTTCGTAAAG	АААААТААТА	TTACAAAACT	TGTAATTCAA	GAAGCGTTGA	ATATGGAAAG	CCTTTGGGGG	TGAGAGAGAT	GATTTCGGAA
5601	GGCGCTATTA	ATAACATTAC	AAATTATTAC	CTAATAAATT	CAAATGAGCC	CATAAAAACT	AGACTTTATC	TGCTTCATTG	GCTTTTGTGC	TGGATGCTCG
5701	GATGATAAGA	AGGGAGAGAG	CGAAACAACT	TTTTATGGCC	GGATTGGCCA	CAAATACAAA	AGCTTATTGA	CAGGGAAAAA	CTGTTGACAA	GATAAAGAAG
5801	AGCAAATCAT	TGAATTGGAG	CAATCAATAA	AAATTTAATG	CGTCACGATT	GCAATAAATC	AAATCCAAAC	TGTTGGCAAT	TTTACGAATT	TTATTAGACC
5901	GTGAAGGAAA	TAAAAGTTGG	AAAGATCTTC	TTCTCCAATT	GATCACTCGT	TTATAATTAA	AAAAGGAACT	CTCGAATACA	AACAGTTTTA	TATTTGTTTT
6001	CAACTTTCCT	CCTCTAGTAT	TCCTTTGTAA	AGGATTCTTC	TTTATGGCGC	ATAGTAAAAT	TGGCAATCGT	AAAAACTTTG	CAAAACAGTA	ACATTCATTT
6101	TAATTCGCAA	GTTTAGATTC	CAGTCTCCAT	TTTATATATT	TCGTTTTTTA	TATACAAGTA	TTTTATGGCG	TATAAAATTA	AGGGAGTGGT	GCTTTTTTAT
6201	GTCCCCTAAA	AAAGCGAGCC	GTTTTTTCTCA	CCCCACTTTT	ATATGCGCAA	ΑΤΑΑΑΤΑΤΑΑ	GGGCAATTTG	GCCGTCCCTG	TGTTTGTACT	CGTATATTTT
6301	GGGGGATTGT	ATTTCCAATG	CGCGTGCCAC	ATATAAAATT	TCCACCCCAT	TTATAGCACA	GCGTATTTTT	TGCATTATAT	TCCCTTCTGA	CGATGTCGCC
6401	TACTTTTCTC	TCCTCGTCGA	CATCCTCGTC	CTCATCCTCC	CATTTTATTA	TATTAATGTT	TGTACAGCGT	TTCCATTTTT	ATGGACATTT	TTTGTCCATA
6501	TTTTATTTCT	TATACCCATA	TACGAGTATA	CATTTTACAC	AACTACTTGT	GCCGTTTTGG	TGATAAATTT	TATTCGCACA	CAGATATGGC	AAAAATGGA
6601	ATTAGAGTCC	AGTCCAAAGG	GGGCACAAAA	AACTATGTAC	TTTGTGCGAT	GCGTCACCAA	ACTCCTCAGA	GGCTTTCCAT	TTTTATCAAC	CGTTTTCCTT
6701	GTCGTTTTAA	TCTAGTTTCT	CTATGCGAAC	AGTCTTTTGA	GTTAGTTTTA	AGCCAGGAAT	TAGTAGAATC	AAGAAATCAA	ACTCTCCCTT	TACTGCCAGA
6801	TTTCCAATTC	TCGGACGTTA	ATTCTAAACA	AAATGCTAAT	TGGAACCCCT	TTTTTTCTATT	CTTTCAGTCA	AAATGCATAG	CTGGGACGTG	ATTCCTCAAA
6901	GCAACTACGA	CAACAATTGG	TATAACAACA	ATTACCAGAT	GAAAACAGAG	CCCCACGATG	GGTTCAACGG	GCAACAGCCC	AATTCTCCGC	AGAGCATGGA
7001	CAGCATTCAT	CCTGAAACAC	ATCACAGTCC	TCCAGTTCAA	CAGCAACATA	TGATGTTCGA	CTCGTCAAAT	CTTATAAACA	CCATGACCCA	ACTACACAAC
7101	GTTCAAATGC	AGAGACAGAC	CCACTTCAAT	CCCCTTACTC	CTCCGGGCTA	TCCAGGCGCT	ATGACCCTTC	CCCAAAACTC	TCAGGCAAAT	TCAACACCAT
7201	TTAGAAGCTT	CACAAAGGGA	CTGGACTCGA	TTCCTTTTGG	AAATAATGTA	TCCAACTTAA	CACCAAGTCA	CACACCTCCA	ATGGACATAA	CTCCGCCAAA
7301	GTCACCAAAG	TTCAACGGCA	AGGAAACCCC	TGAAAAGGAC	TCTCTAAAGC	AGGACCAAAA	TCAACTTCTC	AAAACCCCAA	TCCAGACGAA	TGGAAACCAG
7401	CAATCGACAT	TCGACTCTGG	CGAAGACAGC	CACTCAATGC	CCGATAGCGA	TCTGCTTGAA	CCGGTAATCA	CCGACGGTGC	GGACGTAGAT	GACGAAAACG
7501	ATGCTGAAGA	GGACGATGAC	ATTCGCACTC	CGAAAATCAA	TTCACACGGT	AAAATGAAGA	CGTACAAGTG	CAAACAGTGT	GACTTTATCG	CAGTTACAAA
7601	ACTGTCCTTC	TGGG								

Source: Two PCR products, both amplified from phage Cal-hb ph1. Alignment with cDNA sequences: Positions 3825..4143 correspond to the first exon of the *Clogmia hunchback* P1 transcript (SEQ19), and positions 6865..7614 correspond to parts of the second exon of the *Clogmia hunchback* transcript.

# SEQ21 Anopheles hunchback, cDNA, partial P1 transcript.

1 TCAGCAGCAG GACATCGTCG AACCTTTTGG TCTACATTTT CTTCGCCACA TTTGCATCGT GCAACCATTC AGAATGCA

Source: 5' RACE product, amplified from a cDNA template of adult females.

### SEQ22 Anopheles hunchback, genomic.

1	COCAGOCAA	ma ca ca amoo		man naan naa	A CA MCCMMMM	C 3 3 C 3 3 C 3 3 M	<b>NGG N N G N M N M</b>		mammamma	
1	GCGACCGGAA	TAGAGAATGG	AAACGTGTTA	TGAAGGAAGC	ACATGGTTTTT	GAACCAGAAT	ACGAAGATAT	GTTATAATTT	TATGTTGTTC	ACTGATAATT
101	ATTAAATTGA	ATCATTTCAA	GGCATTATTT	AAGACTGTTG	GTTAAAAAAA	TCGGTGATAA	TAGTCGTTGA	TCTAGAGGTT	AGTTTGAATG	TTAATAATAA
201	AATAGTTTAG	TAAAATGTCG	TAGCACAAAT	ATACAAGCCA	ATTGCAAATT	GGCAATAAAT	AAAAATAAAT	AAATAAATAA	ATAAATAAAT	AAATAAATAA
301	ATAAATACAT	AAATAAATTT	TTTCTTTCCT	ATTAAACACA	AATCTCACAT	TATCACATAA	AATTACACTT	GATTCTTGCA	TTAAATTCTA	CCTAAAGCCA
401	AAGAAAACTG	CTCTGTTCCC	CATGCCTTCG	GTTATTCGCC	CAGCCTAACA	ATGGGGGAGG	AAAAAAAACC	GCTTCCTTAC	GGGTCGTTTA	CTCCGTCAAC
501	GACTATCCCC	AGAATTTTGC	TGATTGCTCC	GGCTTGGTTA	GACGATTTCC	TTTCTGGATT	TTCTCTGCTT	CTTTTGCTAC	CACTAGCGGG	GTATGAGTTT
601	TCCCTTGACG	CAAGTGTGCA	ATGTGCGAAG	ААААТАААА	GCCCCGAGGG	AGTATTCTGT	GGCGGCGGTT	ACGAGTGGCG	AGGCAAATGT	CTATCCACTC
701	TTCTTTACAC	GGGTACAGCA	CACAAAAACA	AGGAAGAAAA	AGGCCCACTG	GGTGGTTAGT	GGGATGGGGG	CGCACTCTGA	TGCTGTTTTT	TTTTGTTTCG
801	TGCAGCCAAA	ATATTTTCTC	CATTTTTCCG	TTGATCATAG	CGATATGTAT	TAAAGCACGT	GTGTTAGTGT	ATGTGAAAGT	GTTTACTTTC	ATTTCCCATT
901	TCTTTGGCTA	TTTTCCCATT	CTTCTTTGCT	TCCGAACGCT	CCGCTGCAGT	GCGTTATCTG	CGTTTGTTTA	ATCGGGTTTG	ATTACAGTGA	AATGAAGCTG
1001	ACCGTCCAGC	CATTGGGAAG	GGCCTAATAG	CCAACACTAC	TGCGAGGGTG	AAATTATGAG	TGTAAATAAT	CCACACATCT	CCCACAAACA	CACACTCACG
1101	CGTGTTTCTG	ATCGGAAACC	ATCATCATCG	TCAACATCAT	CATGAGAAAA	AAGCATCAGT	TTCGTACGGG	AAAGTACGTA	AAATCGTGCC	CCATCACGAC
1201	AGCTACGGGA	CTCACAAATG	CCGGCCACTT	TAAAAACCGG	GATTGTAGTA	TTGCCACACA	AGTGTGTTGT	TTTTACACGC	TGACAAAATA	TATTTATTTT
1301	CATTAATTTG	TCTCTAGCCG	GGCTCGCTGG	GCGCGGCAAA	GCCCGTAATG	CGCCTCGTAA	AAACCATCCG	GGTGTCCAGG	GCGGTATCGT	TAAAGTGCAT
1401	TTTAAAAGCA	ATTAAAGCCC	AGTAGCCGCA	GGCGCACTTT	TGCGTAGAGC	AGCTGGGTCA	GAACAATAAC	AAGCGGACCT	TAGCGACACA	GACACACACA
1501	CACCCGGTCG	GACGGGCACC	AGGGCCCGGT	GGACGTGCGC	GTCACACGTT	GGTTTTTTTC	GGGCGAGACT	TACACGGCGA	ATGAAATGAG	CAATCGCTTT
1601	AGCGAATAAT	TATGTCTTCA	GTAAGTGTGA	GCATTTGCAT	GAGGCTGATT	ATGCTAATAA	TGACCGTAAA	CATTCCTAAA	AGTGAATGAT	GTTGTTTGTG
1701	ACGGAATTGG	TTTAATACAA	AAATGCTTAT	TTATTAAGTG	TGTCCGGTAA	TTGCTGTTAG	CATCAAATTT	CATTTCGATC	CATAATAAAG	ATGTCTTTAA
1801	AAGTGTTTAA	AATGATTTTA	TTTCATTTAT	TCATTTTTAA	ACCGATGTAC	AATCATGTCT	AATCGGCAAT	TTTATACAGG	TATTCCCCGA	TATTAGAATA
1901	TTAGCGTAAG	TTGAATTTCG	AGATTTTCAG	TCAAATAATA	GCTATAATTT	CTAAATTATA	ATTAAATTAT	AGGTAAGTTT	TATAAAATTT	TGCTTGAATA
2001	ACTAAACAAG	TTATTTGATC	TGATTTCAAC	TGAAGATTAA	ААААСТСААА	GAAATGATAA	TATATTTTGC	ATTTTACAAT	TGATATGTCA	AATCAGTACA
2101	ATTTGCTCAA	AGAACTGTCA	AATTAGGTCA	AATAGCGTAT	AGCGAAATCG	CGTATATCGA	GTAAGGCGTA	TATCGGATGA	TATTAAATAT	GACGCCTTAG
2201	TTATTAAATT	ААААСААСАА	CATGAAATAT	TAAATTTGAT	TAATTATCAT	CTTTTTATAA	ACAAATCTCG	TAGATAAAAA	TCCTAGTCAT	GGCACCAATA
2301	ATGCTGTTCA	ATTGGCTTTT	AATAGACCCA	ACATTCTCAT	ACGGAAAGTA	AAGTGCCATA	TAGACATATA	TAGCCGAAGA	TAACAAACAA	AATATAATAA
2401	ATACATATCT	TTGCATTAAT	TTTTCTCATT	TTGATTTTTG	AAACTCATTA	AATATGACGA	CTGGAAATTC	ATGTCACTTT	GGTGGATTCC	GACTATCGTG
2501	TGATAGAGCA	TGTAATGAAA	TTATAAATGA	AATACAGACT	CTGATAATTT	AGATGTTCAA	TCGATTGACA	ATTTTCATCA	AGAATTCGTG	TTCACTTATG
2601	TCTTGTATTT	TCAGTGTATT	AAGGATTTTT	CAAATATTTG	ATCGTTAATT	ATTCATACGG	TGAACATTTT	TTTTATTTTA	TTATCACGAT	TTAGTTATTT
2701	TTGCTTTTAA	ACGTTATAGA	AAGCATTGAT	AGTATGGTTT	GCGTGTTCGT	TAAAACAATC	TGACAATATT	GAACGCGTTA	CAGGAAAGAA	AGCGATGAAA

2801	TTCTTCATTC	GAATAAAAAT	TTCATGTTAA	GCTTTCATAA	ATACTTAATT	GTGTTAAGAA	ACACGAAACA	GACAAACAGA	CATTATTTCC	TTGCAAAATA
2901	CTTTTAGATG	AAAAGAAAAA	CTTGATTCAT	ATTTCATATT	ACTTATTTCA	TATGACAGGA	AAAAAGAGAC	GTATTAATAG	CTGCTAAATA	ACAGATAAAT
3001	GAATAAATCA	ATATCAATGT	ATCAAAATAA	CAGTGCAATT	AAGCGAATAT	TTAGGGCAGC	AATGCTTCTA	GTTGAAAACA	TGCTGCTTTA	TTGTTTTGCA
3101	атаааасаат	Сатсаататт	Састтаааат	TCATCCAACG	CATTGCCCCC	TTGCTAGACG	СФТААТССТТ	CATAAAGTAC	GTAACTCGTT	CCTCGAGCTG
3201	CTTACCAAAT	CCAACCCCCT	AAACCACCCA	AACCCACAAA	CULCCOVCC	TCCCAATACA	CACAACACCA	CCCCCCCTCCA	CCCNAACCCT	CAATCACCCC
2201	NNNNGGAAAI	CGAACCGGCI	AAACCACCGA	AACGGAGAAA	GIIGCCACCI	IGCCAAIAGA	GAGAAGAAGCA	GCGGCGICCA		CAAICACCCC
3301	AAAAACCACG	GTAAGTGATT	AGGAGGTGAC	GTACGCAGAA	AGGTGAAAAA	CGGGAACAGG	GTTAGTGCAC	ACAAACACGG	ACGAAACAAA	AAAAAAGCA'I'
3401	AAGCCCTGCA	TCAGTACCAC	GGTAGCACCA	AGTGCACCCC	TTTGCCCCAT	ATGTTGTCTT	GCTAGACGCA	CTTTCGATTT	CAAATACACA	CACACACACA
3501	TATACTTGTC	CACGGAAGAC	AAAAATCAAT	TTTCCTTCTT	GTGACCTTCC	CTGGGCAAGA	CATGCTTCTT	GGGTCGTTCC	CCGTTCACGC	CGAATCGTTC
3601	CCCCGTGCTC	AGTCTGTCTC	GCTTCACTTC	CGAAACTTTA	CCTCTGGAAA	AGCCCTACCC	AGGCCCTGCC	CGACCCAAGG	ACGGTTACTA	TGGACACGGC
3701	GAGAGCCACG	GAAGCGAGAG	ATAGAAATAG	ATCAACCTAC	GCCATGAGAC	AGAGAGAATG	AGACCTACCC	CCACACCTCA	GCAACTTACG	ACCACCCCTC
3701	GAGAGCCACG	GAAGCGAGAG	AIAGAAAIAG	AICAACGIAG	GGCAIGAGAC	AGAGAGAAIG	AGAGCIAGCG	GCAGAGCICA	GCAACIIACG	AGCAGGCGIG
3801	AAAGAGAGCG	CGTTGGAATA	GGGAAGAGGA	AAAAAAAGG'I'	TCTGTTTTCT	TCTGTTTCAC	CCTCACGCAG	TGCGTTGTAG	CTACCCCTAG	GGATGCATGG
3901	GATGGTTGGT	AGTACTGGCA	GGTCCCAAAA	AATCTCAGAT	CAGTTTTTAA	ATCGCCAGTG	TGCAGTGTGG	TAGTCCTTCT	GGAAAGCCTG	CTGGTCCCGC
4001	ACGGAGTTTT	GTCAAGTCGT	AAGCGTATGT	GTTTTGTACG	AGTTCGAGTG	TACGTACGCG	GAACGTTTTC	GGAGCAGAGC	AGCAATACAT	TGCAGGCCAA
4101	AAGTCCCGGT	TGTTACGAAA	GTGTTCAGTT	GCAGGAGTCG	TTTAAAGACG	GTGAAACGTT	ACACTGTTGT	GCTTGTGATT	GAAATATGCT	ACAAATCGAA
4201	TCGAAGAAAG	ΔͲͲϹͲϹͲͲͲΔ	ACAAACTCAA	<u>እር እጥርጥጥርርጥ</u>	GGAATCTTTG	AGTGATTTCC	AAACTCTTCT	AAAAAAAAA	AAACTGAAAG	TGAAGAGACA
4201		ATIOIOTTIA	TOTAL	ACATCITCCT	TOCARIOITIO	ACTOATTICC			AAACIOAAAO	IGANGAGACA
4301	CAAAAACGAG	AGACACAATT	TUGTUAUGGA	CCCGTCCCCG	TUUUUGIGAT	GCIGIAGICG	GCCTGTGGTC	AGCAGCAGGA	CATCGTCGAA	CUTTITGGIC
4401	TACATTTTCT	TCGCCACATT	TGCATCGTGC	AACCATTCAG	GTAAGAATGT	CACACAGCAC	ACAGAAACCC	AATCGGTTGG	AGGTCGAATC	TCACTTTTAC
4501	AACACAAACT	CATGCCGGCA	CAAACGTACG	AGCCCGAAAG	ATGATTGACC	TACTTCCACG	CCCCCCCGTC	CCTGCATCCT	TTGGCATTCG	AAGGACAACT
4601	CTTTTGTCCT	TAAGATTTTA	CAGCGCAACA	TTGGGCCCCC	GCAGTACGGT	GGGATGGTGC	AATTAGGAAT	TGGCAAACTT	TATGATGGCT	TCCCAAACTT
4701	TGACTTTTCA	AACCTTTCCT	CTCCATTGGC	астстсстст	TGTTGGGTAG	GTAGGTGACT	CAGGGACATC	GTCGAGTCTA	CCCGTTCGTA	TATCCTCTCA
1001	CARACUACCO	mcccmacmmm	CCCTCACCCC		CCCCCCCCC	CUMUCCCCAN	CCCAACCCAC	mcccaaacmc	AAAMMCCCCCC	CIICACCIIIIIC
4001	GAAAGIACCC	ICGGIACIII	GGGIGAGGCG	IATACAAGIC	GGGGGGCGGII	CITIGGCCAA	CCGAACGGAG	IGGGAAAGIC	AAAIICGGGG	GICAGCIIIC
4901	GTTTCGGGGT	CTCGTTTGCG	TTCGGCCCAA	GGGTACGGCT	GTAAAATCGA	TTCAAATTCA	GAGAGGGAAA	AAGGGGAAGC	TTGTAAAAAA	ATAAAACTTA
5001	AAAATAACAC	ATCTCCAGCA	TCCCATGGTG	ATTTTGGGAA	CGGGGGACTTG	TGCTTGGGAC	GGTTGCAATG	CTTGCGAGGG	AAAGTGCATT	GTTGTCACAA
5101	TTGTTGCCAT	CGACCTTACC	CTTTACACCG	AAACGGGACG	CACCATTTGG	GGCGATGGCA	AAAGGGTCGC	AAAGGTTCGC	TCTTAGTTCT	TGCTACGCTC
5201	GTAAAGTCGA	ልጥርጥጥጥልጥጥጥ	ACGAGCCGTG	AGGAAAGTGT	ͲͲልႺልͲͲͲͲϹ	GGGATATTG	TGACGTGGGT	GACCCATAGA	ATGGGAAACA	ΔͲϹϾϹͲͲϹϹͲ
5201	CTIME ICTOR				TIMOMITITC	CCCCANAMCA	NMCCCCCC	mcmmcqammc		CINNAGCAC
2201	CIACGCGIAC	TTTCAAAGAA	CGAAIGGGIA	CAIGIGGAAA	IGGCIGGAAC	GGGGAAAIGA	AAIGGCCGII	IGIIGCALIG	TAAAAACCGA	CAAAAACGAG
5401	GTAGAACCCA	GAGTTGTATA	GGGTGTAATT	GAAATTTTAA	TGTTAAAAAC	AGACGATAAT	GTTTGGAAAT	TATTAAAACT	GCAAAGCCTC	ACAATGTTTC
5501	AAGCTAACAA	AACGTTCATT	TTTGTTGAAA	ATATGTAATT	CTTCATCATA	CTATTAAAAG	AATTGAATTC	ATTTCAATGG	AATATCATGA	GAATATTCCT
5601	AGAATACAGT	AAATTAATTC	CAATTTTCCT	TGAAGTATAC	ACTCATTATG	TGATGCTTCT	CGGCGACGGT	ACATAATGAA	ACGGCTCCCC	CACAAAATAT
5701	AACGACAAAC	AATAGCGTTT	TGTGGAGGAG	GATGAGGCAA	AAAAAGCCAA	AACCACAAAG	AAATATTCTG	GATTACGAGA	ACCCTCCCAA	ATATACAGAA
5001	AACACAACAA	CCAACAAAAA	AAACCAACAC	ACANACCUMC	mmcmcmccca	ACACACAAAA	CCCAACTTOTO	CIICCONTOCCO	CCCCCCCCC	mccammmcc
5001	ANGAGAAGAA	GGAAGAAAAA	AAACGAACAC	ACAAACCIIC	TICIGICGCA	ACACACAAAA	CGCAACITIC	CIGCCAIGCC	GGGIGGGIIA	ICGAILITCC
5901	AAACCCACCA	ACCAAGCA'I'I'	GAGCGTTGTG	AGAGGAGGCA	AAAAAACGGC	AGGCAAAAAA	GGAATAAAAG	ATTCCGCAAA	AATCCTCCAA	GACAAAGGAG
6001	GAACGTACGC	ATTTTCCACC	CCAACGAAGG	AGAAGGTAAA	AGGGGGGAGG	GTGGTGTAGG	AGAAGAAGAA	AGAAGAAAAA	CTCCTTCACG	CAAGAGGGGA
6101	CAAACAAAGC	CGGCAAAATG	GGAAAAATGC	TGCTCGGTTG	TTGTGTTTTG	GTTTCGGTTT	CCATTCCAAG	GGAAATCGTT	CGTTGCTAAT	GGAAGAAGAA
6201	GAAGAAGTAG	CCCCCAAAAA	AACTCTCTTC	TCTCGTACTA	AACGAGCACA	GATACATTCT	TGCAAGCAAC	ACACACACAT	AACACTAGGA	GAAGCCGGAA
6301	ΔΑΤΑGΤΑΑΑΑ	CACACAGCGC	GTTCAATTAT	CGTTTTTCGG	GATGGAATAG	GGGGGGAGGCT	TGTGGTAAAG	ATCGTTCCCA	AACGCAGGGA	AATAAAAAGC
6301 6401	AATAGTAAAA	CACACAGCGC	GTTCAATTAT	CGTTTTTCGG	GATGGAATAG	GGGGGGAGGCT	TGTGGTAAAG	ATCGTTCCCA	AACGCAGGGA	AATAAAAAGC
6301 6401	AATAGTAAAA TGCGATGAAA	CACACAGCGC ATAAAAATAA	GTTCAATTAT TCACCCACAC	CGTTTTTCGG ACACACACAT	GATGGAATAG ACACGCGCAT	GGGGGAGGCT GAGTGTGAGC	TGTGGTAAAG AGCAGCACAT	ATCGTTCCCA TTTGCCCCCT	AACGCAGGGA TTTTTGTCTT	AATAAAAAGC CTCGGGACAT
6301 6401 6501	AATAGTAAAA TGCGATGAAA TTAATTTGGC	CACACAGCGC ATAAAAATAA TGTGGGGAGG	GTTCAATTAT TCACCCACAC ACTTCTCACC	CGTTTTTCGG ACACACACAT CTCGCAAACA	GATGGAATAG ACACGCGCAT CATATACTTT	GGGGGAGGCT GAGTGTGAGC TCACCGACCC	TGTGGTAAAG AGCAGCACAT TCTATCACTG	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG	AACGCAGGGA TTTTTGTCTT CGTGTTTATA	AATAAAAAGC CTCGGGACAT TACCGAGGGT
6301 6401 6501 6601	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCGTGTGGC	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG	GGGGGAGGCT GAGTGTGAGC TCACCGACCC CACTCGCGAG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAAACACACA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA
6301 6401 6501 6601 6701	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC	CACACAGCGC ATAAAAATAA TGTGGGGGAGG CAAAAAAAAA CTACTCACAC	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCGTGTGGC AAATACGTAC	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA	GGGGGAGGCT GAGTGTGAGC TCACCGACCC CACTCGCGAG AAACGAAAAA	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAAACACACA CAGCATGCCA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT
6301 6401 6501 6601 6701 6801	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCGTGTGGC AAATACGTAC TTTTTGCAGA	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT	GGGGGAGGCT GAGTGTGAGC TCACCGACCC CACTCGCGAG AAACGAAAAA AATGGCGATG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA GCGGCAGCAA	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAAACACACA CAGCATGCCA CACAACAACA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCAGCAAA
6301 6401 6501 6601 6701 6801 6901	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGGCAGATCA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA	GGGGGAGGCT GAGTGTGAGC TCACCGACCC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA GCGGCAGCAA CCGAACCACT	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAAACACACA CAGCATGCCA CACAACAACA CGACTACCAT	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCAGCAAA ATCATCCGCT
6301 6401 6501 6601 6701 6801 6901 7001	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCACCAT	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCGTGTGGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCCGACTCC	GGGGGAGGCT GAGTGTGAGC TCACCGACCC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA GCGGCAGCAA CCGAACCACT GACGACCCACT	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAAACACACA CAGCATGCCA CACAACAACA CGACTACCAT CTCCCCCACCA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCACCACCAC	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6601 6701 6801 6901 7001	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGCAT	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCGTGTGGGC AAATACGTAC TTTTTGCAGA ACCGCAACGC ACCTGCACCCA	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCG	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GGCCGGAGTGC	GGGGGAGGCT GAGTGTGAGC TCACCGACCC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACCACT GACGACCACC	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAAACACACA CAGCATGCCA CACAACAACA CGACTACCAT CTCGCGAGGCG	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6601 6701 6801 6901 7001 7101	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGCAT TCCCCCCGCAGA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GGCGGAGTGC AGAGTGGATTC	GGGGGAGGCT GAGTGTGAGC TCACCGACCC CACTCGCGAG AAACGACAAAA AATGGCGAATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA GCGGCAGCAA CCGAACCACT GACGACGCAC AAAGGACGAC	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAAACACAC CAGCATGCCA CACCACCAT CTCGCGAGCG ACAGTCTCAT	AACGCAGGGA TTTTTGCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC	AATAAAAGC CTCGGGACAT TACCGAGGCT ACGGTACTA ATGGTTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC
6301 6401 6501 6601 6701 6801 6901 7001 7101 7201	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGCAT TCCCCGCAGA TCCGGCTTTAA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCGCCTGGCT	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGGGCT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GGCGGAGTGC AGAGTGGATT GCTGATCCCA	GGGGGAGGCT GAGTGTGAGC TCACCGACGC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA GCGGCAGCAA CCGAACCACT GACGACGACC AAAGGAGCAA CACAGCTCGC	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAAACACACA CAGCATGCCA CACAACAACA CGACTACCAT CTCGCGAGCG ACAGTCTCAT CGCTCAGCAG	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CAGCAGCAGC	AATAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTACTG CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6601 6701 6801 6901 7001 7101 7201 7301	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGCAGCAGCAT TCCCCCCAGA TCGGCTTTAA CCAGCAACGC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCAACAC	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCGCCTGGCT CCAACCGAAT	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCAACGGCTG CTCAGTGGGCA ATTCGGGGCAT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GGCGGAGTGC AGAGTGGATT GCTGATCCCA GGTACGGCA	GGGGGAGGCT GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCT	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA GCGGCAGCAA CCGAACCACT AAAGGAGCAA CACAGCTCGC GCCAACCTCC	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAACACACA CAGCATGCCA CACCATGCCA CACCACACACA CTCGCGAGCG ACAGTCTCAT CGCTCAGCAG CTAACGCCGA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CACACATCATC ATTCACACTC GTCGGGCGGC CACCAGCAGC CCCACACCCC	AATAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6601 6701 6801 7001 7101 7201 7301 7401	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGGCAGATCA GCAGCAGCAT TCCCCGCAGA TCGCCGCAGA CCAGCAACGC GTGACGCCAC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCAACAC CAAAGTCCCC	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCACCGAAT CAACGAGTCG	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCGCT GTACGGCAAT CTCGAAACAC	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGAGTGC AGAGTGGATT GCTGATCCCA GGTACGGCA CCACCAAAGA	GGGGGAGGCT GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGGC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CGGAACCACT GACGACGACC AAAGGAGCAA CACAGCTCGC GCCAACCTCC GGCTCGGACT	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAAACACACA CAGCATGCCA CAGCATGCCA CGACTACCAT CGCCGAGCG ACAGTCTCAT CGCTCAGCAG CTAACGCCGA GTCAGGATGG	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CACCACACCC CCCACACCCC CTCGTACGAT	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCAGCAAA ATCATCCGAT CAACCCAACC
6301 6401 6501 6601 6701 6801 7001 7101 7201 7301 7401 7501	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGCAT TCCCCGCAGA TCCGCCTTTAA CCAGCAACGC GTGACGCCAC ACGAGGACGG	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAA CTACTCACAC TTTTCCCCAA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCAACAC CAAAGTCCCC CATCCGCAAG	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCGCCTGGCT CAACGAGTCG CCGAAGGTGA	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG CTCAGTCGGCCG CTCAGTGGTA ATTCGGGGCAT CTCGAAACAC ACTCGCACGG	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTCA GGCGGAGTGC AGACTGGATT GCTGATCCCA GGTACGGCA CCACCAAGA TCAGGTGAAG	GGGGGAGGCT GAGTGTGAGC CCACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCT GGAGGACGCC AAGTTCCGCCT	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACCACT GACGACGACT CACAGCTCGC GCCAACCTCC GCCACCACT GCCAGCAGTG GCAAGCAGTG	ATCGTTCCCA TTTGCACGTT GTATGTGTG GAAACACA CAGCATGCCA CAGCATGCCA CGACTACCAT CTCCCGAGCG ACAGTCTCAT CGCTCAGCAGG GTGAGGATGG GTGAGGATGG GGAGTTTGTC	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CAGCAGCAGC CCCCACACCCC CTCGTACGAT GCGGTAACGA	AATAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTATTGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6601 6701 6801 7001 7101 7201 7301 7401 7501 7501	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGCAT TCCCCCCACAA TCGGCTTTAA CCCAGCAACGC GTGACGCCAC CTGCGACGCAT	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCACACAC CAAAGTCCCC CATCCGCAAG	GTTCAATTAT TCACCCACAC ACTCTCACC ACCGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCGCCTGGCT CCAACGGATCG CCAACGGTGA ACATCAAGCC	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CCAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGGCA CTCGAAACAC ACTCGCACAG CCCAAAACAC	GATGGAATAG ACACGCGAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GGCGGAGTGC AGAGTGGAT GCTGATCCCA GGTACGGCA CCACCAAAGA TCAGGTGAAG	GGGGAGGGC GAGTGTGAGC TCACCGCGAC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCT GGAGGACGCC AAGTTCCCGT	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACACACT GACGACGCAC CCGAACCACT GACAGCTCGC GCCAACCTCC GCCTCGGACT GCTAGGACG	ATCETTCCCA TTTGCCCCT TGTATCTGTG GAACACACA CACACACACA CACACACACA CACAACACA CACACACACA CACACACACA CACACACACA CTCGCGAGCG GACTCTCAT CGCTCAGCAG GGACTTTGTC CGACTACACC	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCGGC CCACACCCC CCCCACACCCC CTCGTACGAT GCGGTAACGA	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC ACCAACAGCA ACCAACGAG GGCAGCGAGG AGCTCAGCTT
6301 6401 6501 6601 6701 6801 7001 7101 7201 7301 7401 7501 7601	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGGCAGATCA GCAGCAGCAT TCCCCGCAGA TCGGCATTAA CCAGCAACGC GTGACGCCAC ACGAGGACGG CTGGGAGCACG CTGGGAGCACG	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTTCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG CTAGCAACAC CAACATCACC CAACGCCCCC ACGCGCAAG	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCACCGCAGTT CCAACGAATC CCGAAGGTGA ACATCAAGCC	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCGCA GTACGGCAAT CTCGAAACACG GGAAAAGATG GGAAAAGATG	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGAGTGC AGAGTGGATT GCTGATCCCA GGTACGGCA TCAGGTGAAG CTCACCTGCC CTCACCTGCC	GGGGGAGGCT GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACACC TTTTGAGGCA CCACCACCAC AGGTGGAGGC GGAGGACGCC AAGTTCCGCT CAAAGTGTCCC CACCACACAC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CGGAACCACT GACGACGACC GACGACCACC GCCAACCTCC GGCTCGGACT GCAAGCAGTG GTTCGTGACG	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAACACACA CAGCATGCCA CACCACACACA CACACACACA CCACTCCACAC CTCCCGACGACG CTAACGCCGA CTAACGCCGA CTGACGACGACG CGACGTTGTC CGACTTCCCC	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CCACACCCC CCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCACCAAA ATCATCCGAT CAACCCACC TACAACCCTC ACCAACAGCA ACCGATGGAC GGCAGCGAGG AGCTCAGCTT GTACCATCTG
6301 6401 6501 6601 6701 6801 7001 7101 7201 7301 7401 7501 7601 7701	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGGATCA TCCCCGCAGA TCGGCTTTAA CCAGCAACGC GTGACGCCAC ACGAGGACGG CTGGGAGCAT CGCAACCACC	CACACAGCGC ATAAAAATAA TGTGGGAGG CAAAAAAAA CTACTCACAC TTTTCCCCAA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCAACAC CAAAGTCCCC CATCCGCAAG ACGCGCAGCC ACGCGCTCCAA	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCGCCTGGCT CAACGAATCC ACGGCTGGATCA ACATCAAGCC GCCGTTCCAG	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCGGTG CTCAGTGGTA ATTCGGGGCT GTACGGCAGT CTCGAAACAC ACTCGCACGG GGAAAAGATG TGTCCGAAGT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTCA GGCGGAGTGC AGACTGGATT GCTGATCCCA GGTACGGCA CCACCAAGA TCAGGTGAAG CTCACCTGCC	GGGGGAGGCT GAGTGTGAGC CCACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGGC GGAGGACGC AAGTTCCGCT CAAAGTGTCC CTCCGCTCAAC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACCACT GACGACCACT GACGACGACC GCCAACCTCCC GGCTCGGACT GCAAGCAGTG GCTAGCAGTG GTTCGTGACG	ATCGTTCCCA TTTGCCCCCT GTATGTGTG GAAACACACA CAGCATGCCA CAGCATGCCA CGACTACCAT CTCCCGAGCG ACAGTCTCAT CGCTCAGCAG GTGAGGATGG GGACTTTTCTC GAGTACAAGC GGACTTTGTC	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CAGCAGCAGCAGC CCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA	AATAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTATTGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6601 6701 6801 7001 7101 7201 7201 7301 7401 7501 7501 7701 7801	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGGCAGCATCA GCAGCAGCAT TCCCCCGCAGA TCGGCTTTAA CCCAGCAACGC GTGACGCCAC ACGAGGACGG CTGGGAGCAT GCCAACCACC TGTTCCAGTA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCACACAC CAAAGTCCCC CATCCGCAAG ACGCGCAGCC ACGCGTCCCA CCGGTGTGCG	GTTCAATTAT TCACCCACAC ACTCTCTCACC ACCGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCGACCGGAT CCAACGGATCG CCAACCGAAGTGA ACATCAAGCC GCCTTCCAG GATTGTAACT	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGCTG TCATCGCCCG CTCAGTGGTA ATTCGGGGCA CTCGAAACAC ACTCGCACGG GGAAAAGATG TCTCCGAAGT ACGCCACCAA	GATGGAATAG ACACGCGAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GGCGGAGTGC AGAGTGGAT CCACCAAGG CCACCAAGA CCACCAAGA CTCACGGCCACA GTACTGCCAC	GGGGGAGGC GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCT GGAGGACGCC AAGTTCCGCT CAAAGTGTCC CTGCGTCAAC TCGCTCAAGC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACACACT GACGACGCAC GACGACGCAC AAAGGAGCAA GCCAACCTCC GGCTCGGACT GCAAGCAGTG GTTCGTGAGG TGCATCACG TGCATCTGCG	ATCGTTCCCA TTTGCCCCT TGTATCGTGTG GAAACACACA CACACACACA CACACACACA CACACACACA CACACACACA CACACACACA CACACACACA CTGCGCAGCG GAGTATCTCC GAGTACAGC CAAGTATCGCG	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCAGCAGC CCACACCCC CCCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA CATCAAGCCGG	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC ACCAACAGCA ACCATCGAC GGCAGCGAGG AGCTCAGCTT GTACCATCTG ACCTGGACG
6301 6401 6501 6601 6701 6801 7001 7101 7201 7301 7401 7501 7601 7701 7801 7901	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGCAT TCCCCGCAGA TCGGCATTAA CCAGCAACGC GTGACGCCAC ACGAGGACGG CTGGGAGCAT GGCAACCACC TGTTCCAGTA GAACCTGGAC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCAACAC CAACGTCCCC AGCGCTCCAA GCGGTCCCAA GCGGTGCGCG GGTACACCGA	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCACCGCTGGCT CCAACGAGTCG CCGAAGGTGA ACATCAAGCC GCCGTTCCAG GATTGTAACT ACCCGCTACC	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCCG GTACGGCAAT CTCGAAACAC ACTCGCACGG GGAAAAGATG GGCACCAA GATCATCGAT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GGCGGAGTGC AGAGTGGATT GCTGATCCCA GGTACGGCAA CCACCAAAGA TCAGGTGAAG CTCACCTGCC GTAGCTACAG GTACTGCCAC GTGTACGGGA	GGGGGAGGCT GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAA AATGGCGATG ACGATCAAAT AGCAGGACACC TTTTGAGGCA CCACCACCAC AGGTGGAGGC GGAGGACGCC AAGTTCCGCT CAAAGTGTCCC CTGCGTCAACC TGGCTCAACC CACGTCGTGG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CGGAACCACT GACGACGACA CAAGGAGCAA GCCAACCTCC GGCTCGGACT GCCAACGACTG GTTCGTGACG AAATCAATGC TGCATCGGAAAGCG	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAACACACA CAGCATGCCA CACGACTACCAT CTCCCGACCA CACACTCCAT CGCTCACCAG CTAACGCCA CTAACGCCA CGAGTTGTC GAGTACAGCC TGAACTCGCA AAACCCCAGA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CCACACCCC CCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA ACATGAAGTCG AGAACGCCGA	AATAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC ACCAACAGCA ACCGATGGAC GGCAGCGAGG AGCTCCAACGT CACCCATCTG CACTCGAACG ACGTGGTGCT TAAGCATCTG
6301 6401 6501 6701 6801 6901 7001 7101 7201 7301 7401 7501 7701 7801 7901 8001	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGCCAGATCA GCAGCAGCAT CCGCCAGAA CCGGCATTAA CCAGCAACGC CTGGAGCACG CTGGAGCACG CTGGAGCAC GCAACCACGA GAACCTGGAC	CACACAGCGC ATAAAAATAA TGTGGGAGG CAAAAAAAA CTACTCACAC TTTTCCCCAA ACAGCAGCAA GCCGTGGACAG TCCACTCACA CACACTCACA CAAAGTCCCC CATCGCAAG ACGCGCAGCC ACCGCTCCAA CCGGTGTGCG GGTACACCG AACAGCCCA	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCGCCTGGCT CAACGAATCCA CCAACGATGA ACATCAAGCC GCCGTTCCAG GATGTAACT ACCCGCTACC GCAGGAGCAG	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCGG CTCAGTGGTA ATTCGGGGCAT CTCGAACAC ACTCGCACGG GGAAAAGATG TGTCCGAAGT ACGCCACCAA GATCATCGAA	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTCA GGCGGAGGATT GCTGATCCCA GGTACGGCA CCACCAAGA TCAGGTGAAG CTCACCTGCC GTACGCACG GTACGGCA GTACTGCCAC GTGTACGGGA CACTCAACA	GGGGGAGGCT GAGTGTGAGC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGGC AAGTTCCGCT CAAAGTGTCC CTGCGTCAAC CACGTCGTGG GCAAGAATTG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACCACT GACGACCACT GACGACCACA CACAGCTCGC GCCACCTCCG GCTCCGGACT GCAAGCAGTG GCTCGTGACG AAATCAATGC TGCATCTGCG ACCGAACCCCG	ATCGTTCCCA TTTGCACCCCT GGAACACACA CAGCATGCCA CAGCATGCCA CAGCATGCCA CGACTACCAT CGCTCAGCAGCG CTACCGACG CTACCGACG GTGAGGATGG GGACTTCTCC GAGTACAAGC CGACTTTGTC GAGTACAAGC CAACTCCCA AAACCGCAGA GAGATCAATC	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CAGCAGCAGC CCCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA CATGAAGTCG AGAACGCCGA GCCCGCACCA	AATAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTATGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6601 6701 6801 6901 7001 7101 7301 7401 7501 7501 7501 7501 7801 8001 8101	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CTCTTTTCTC CGGCAGCACAT TCCCCGCAGA TCGGCTTTAA CCAGCAACGC GTGACGCAC GTGACGCAC CGCAACCACC TGTTCCAGTA GAACCTGGAC CTCAAGCAAG TAGTCAGCA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCACCAC CAAAGTCCCC CATCCGCAAG ACGCGCAGCC AGCGCTCCAA ACCGGTGTGCG GGTACACCGA AACAGGCTCA	GTTCAATTAT TCACCCACAC ACCTCTCACC AACTTCTCACC TTTTTCCAGA ACGGCACGCA ACCTGCACCA TATGGATTCA CCGCCTGCACCA TATGGATTCA CCAACGGATCCA GCCGTCCCAG GATTGTAACT ACCCGCTACC GCAGGACCAC	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGACT CAAACGGCGTG CTCAGTGGCCG CTCAGTGGCA ATTCGGGCAT CTCGGAACAC ACTCGCACGG GGAAAAGATG TCTCCGAAGT ACGCCACCAA GATCATCGA	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTCGATTTA GCCGGAGTGC AGAGTGGATTGC GGTACGGCA CCACCAAAGA TCAGGTGAAG GTACTGCCAC GTACTGCCAC GTACTACAGA CACCCACAGA CACTCAAACA	GGGGGAGGC GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCT GGAGGACGCC CAGGTCCACC TCGCTCAAC CCACGTCGTGG GCAAGAATTG GGAAGTGCC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACACCACT GACGACGCAC AAAGGACGACA GCCAACCACC GGCTCGGACT GCAAGCAGCG GTTCGTGAAG GTTCGTGAAG TGCATCTGCG ACCAACTCCG AACCAACTCCG AAACGAATCG	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAACACACA CACACACACA CACACACACA CACACACA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCGCGC CCACACCCC CCCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA CACAAGCCGG AGAACGCCGA GCCCGCACCC	AATAAAAAGC CTCGGGACAT TACCGAGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCAC ACCAACAGCA GGCAGCGAGG AGCTCAGCT GTACCATCTG CACTCGAACG CCATCGAACG CTAAGCATCTG CTAAGCATCTG CAACGCAGC
6301 6401 6501 6601 6701 7001 7101 7201 7401 7501 7501 7601 7701 7801 7901 8001 8101	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGCAT TCCCCGCAGA TCGCCGCAGA CCAGCAACGC GTGACGCCAC ACGAGGACGG CTGGGAGCAT GGCAACCACC TGTTCCAGTA GAACCTGGAC CTCAAGCAAG TAAGTCAGCA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCAACAC CAACGCCCCA AGCGCTCCAA GGTACACCGA ACAGGCTCA AACAGGCTCA GGCATGGCA	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCACCGCAGAT CCAACGAAT CCAACGAAT CCCAACGAAT CCCGACGAGCG GACTGTAACT ACCCGCTACCA GCAGGAGCAG ACGGCACCGA	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGGCT GTACGCCACG GGAAACACT ACTCGCACGG GGAAAGATG TGTCCGAAGT ACGCCACCAA GATCATCGAT CAATCTCAAC CGACTCCGTC	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGAGTGC AGAGTGGATT GCTGATCCCA GGTACGGCA CCACCAAAGA GTACGGCA GTACTGCCAC GTAGCTACAG GTACTGCCAC GTGTACGGGA CACTCAAACA CACCCCTCCC	GGGGGAGGCT GAGTGTGAGC TCACCGACGA CACTCGCGAG AAACGAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CACCACCACAG AGGTGGAGCC CAGGTGAGCC CAAGTTCCGCT CAAAGTGCCC TGCGTCAAGC CACGTCCAGC GCAAGAATTG GCAAGAATTG GCAAGTGCCC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CCGAACCACT GACGACGCAC AAAGGAGCAA CACAGCTCGC GCCACGCACTCC GCCACGCACTCC GCCAGCACTCC TGCATCGGCACG ACCGAACCCCG ACCAACTCCG AAACGAATGC AAACGAATGC	ATCGTTCCCA TTTGCCCCT TGTATGTTG GAACACCACA CAGCATGCCA CACCACACACA CACACACACA CACATCCAT CGCTCACCAC CTCACGACG ACAGTCTCAC GTGACGACG GAGTTCACACC CAAGTACTGCC AAACCGCCA AAACCGCCA GAGATCATCGC GAGATCATCG CTCACCCCCA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CCACACCCC CTCGTACGAT GCGGTAACGA CACAAGCCGA CACAAGCCGG AGAACGCCGA ACCCCCTCTCCC ACCCCCCCCC	AATAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTATGT CAGCACCAA ATCATCCGCT CAACCCACCC TACAACCCTC ACCAACAGCA ACCGATGGAC GGCAGCGAGG AGCTCAGCTT GTACCATCTG CACTCGAACG CACAGCACCTC CTAACGCACC
6301 6401 6501 6601 6701 6801 6901 7101 7201 7201 7301 7301 7501 7701 7801 7901 8001 8101 8201	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CTCTTTTCTC CGGCAGACACA GCAGCAGCAT TCCCCGCAGA CCGGCACACC GTGACGCCAC CTGGAGCACC GCGAGCACC TGTTCCAGTA GAACCTGGAC CTCAGCAAC CTCACGCAC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCAGCAA CACGCAGCAA CACGTGGACAG TCCACTCACA CTAGCACACAC CAAAGTCCCC CATCCGCAAG AGCGCTCCAA CCGGTGTGCG GGTACACCG AACAGGCTCA CGGCATGGCA CCGCTCCGCAC	GTTCAATTAT TCACCCACAC ACTCTCACC ACCGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCGCCTGCACC CCAACGGAT CCAAGGAGTGA ACATCAAGCC GCCGTTCCAG GATTGTAACT ACCCGCTACC GCAGGAGCAG ACAGGCACCGA	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCGG CTCAGTGGTA ATTCGGGGCAT CTCGAACAC ACTCGCACGG GGAAAAGATG TGTCCGAAGT ACGCCACCAA GATCATCGAT CAATCTCCAAC CGACTCCGTT GACTGCGCGCG	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTCA GGCGGAGTGC AGACTGGATT GCTGATCCCA GGTACGGGA CCACCAAGA TCAGCTGCAC GTACGGCA GTACTGCCAC GTACTGCCAC GTGTACGGGA CACTCAACAC CACCCACCGCC GCCAACGGCA	GGGGAGGGC GAGTGTGAGC TCACCGCGAC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCT GGAGGACGCC AGGTCCCGTC TCGCTCAACC CCACGTCGTCG GCAAGAATTG GGAAGGACGC GTCTGCCCACT	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACACCACT GACGACGCAC CCGAACCACT GACGACGCAC GCCAACCTCC GGCTCGGACT GCTCGTGACG AAATCAATGC TGCATCTGCG ACCGAAACCG ACCGAAACCG CACCGAAACCG CACCGAAACCG CACCGAATCG GTTCCCGTACC GCTCCCGTACC	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAAACACACA CACACACACA CACACACACA CACAACACA CACAACA	AACGCAGGGA TTTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCAGCAGC CCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA CACAAGCCGG AGAACGCCGA ACCCTCTCC ACTTCCACAT	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC ACCAACAGCA ACCAACGAG GGCAGCGAGG AGCTCAGCT GTACCATCTG CACTCGAACG CACTCGAACG CACTCGAACG CTAACGCACCT CTAACGCACCT GTTAGCCACC
6301 6401 6501 6601 6701 6801 6901 7001 7201 7301 7301 7501 7501 7501 7501 7801 8001 8101 8201 8301	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGCCAGATCA GCAGCAGCAT TCCCCCCGCAGA TCGGCTTTAA CCAGCAACGC ACGAGGACGG CTGGAGCACG CTGGAGCACG CGCAACCACC TGTTCCAGTA GAACCTGGAC CTCAAGCAAG TAAGTCAGCA CTTCCCGAAC CAGCAGCCG	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAC CACCATCATC GCGTGGACAG CTAGCACTCAC CAACGCACCAC ACGCGCACGCC AGCGCTCCAA GGGCATGGCG GGTACACCGA GGCATGGCG GCTTGCAGCA	GTTCAATTAT TCACCCACAC ACTCTCTCACC AACTGTGTGGC AAATACGTAC TTTTGCAGAC ACGGCAACGC ACCTGCACCA TATGGATTCA CCAACGCATC CCAACGAGTTC CCAACGAGTGA ACATCAAGCC GCCGTTCCAG GACTGTAACT ACCCGCTACC GCAGCACCAG ACAGCACCAA GCAGCACCAG	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA TCGCAGACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCGCA GTAACGCACA CTCGCACGA GGAAAAGATG TGTCCGAACA GATCCCACA GATCATCGAC CGACTCCGTT GAATCCCCC	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGACTGC AGAGTGGATT GCTGATCCCA GGTACGGCAA CCACCAAAGA CTCACCTGCC GTAGTACGGA CACCCTCCA GCCACCGCA CCCCCCT	GGGGGAGGCT GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGACAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAA AGGTGGAGCC GGAGGACGCC CAGGTCAAC CCACGTCGTGG GCAAGATTG GGAAGACGCC GTCTGCCACT CCACGCAGAAC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CGGAACGACT GACGACGCAC GACGACGCAC GCCAGCTCC GGCTCGGACTC GGCTCGGACT GCAAGCAGTG AAATCAATGC TGCATCAGCG AAACGAATGC GTTCCCGTAC CTGATGAACC	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAAACACACA CACCATGCCA CACAACACA CACATCCAT CTCGCGAGCG ACAGTCTCAT CTCGCGAGCG CTAACGCCGA GTGACGTTGTC GACTACAGCC CAACTATCGC AAACCGCAGA GAGATCATCC CTGAACGCCGA AACCGCAGA GCGATCATCC CTGAACCTCA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA AGAACGCCGA AGAACGCCGA AACCTCTTCC ACTTCCAGAT CGAAAAGCCG	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCACCAAC TACAACCCACC TACAACCCACC CAACCCAACAGCA ACCGATGGAC GGCAGCGAGG GGCAGCGAGG CACTCGAACG ACGTGGTGCT TAAGCATCTG CTAACGCACC GTTTGCCGAC AACGGCAGTG
6301 6401 6501 6601 6701 6801 7001 7101 7201 7301 7301 7401 7501 7701 8001 8101 8201 8301 8401	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGCAT TCCCCGCAGA TCGCCGCAGA CCAGCAACCACC ACGAGGACGG CTGGGAGCAT GGAACCAGCA GTCAAGCAAG CTCCAGAA CTTCCCGAAC CAGCAGCCG CAACGAGCAG CAACGAGCAG	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCAACAC CAACGCCCCA AGCGCTCCAG GGTACACCGA ACCGGCTCCA ACCGCCTCGCA GGCCTCGCAC CGGCTCGCAC	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCAACGCAGAT CCAACGAAT CCAACGAAT CCCAACGATCA GCCGTACCA GATTGTAACT ACCCGCTACCC GCAGGACCAG ACATGTTCAA GCACGACACAG CAGGACAATG	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGGCAT GTACGCACGA GGAAAAGATG TGTCCGAAGT ACTCCCACGA GATCATCGAT CAATCTCAAC CGACTCCCCT ATCTTCGGT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGAGTGC AGAGTGGATT GCTGATCCCA GGTACGGCA CCACCAAAGA GTACTGCCAC GTGACCACG GTGTACGGGA CACCCAACGGA CACCCACCCT GCCAACGGCA	GGGGGAGGCT GAGTGTGAGC TCACCGACGAG CACTCGCGAG AAACGAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CACCACCACAG AGGTGGAGCT GAAGGAGCC CAGGTCAAGC CACGTCAAGC CACGTCAAGC GCAAGAATTG GGAAGAATTG GGAAGTAGCC GTCTGCCACT CCAGCGGCT	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CCGAACCACT GACGACGCAC AAAGGAGCAA CACAGCTCGC GCCAGCACTCC GCCAGCACTCC GCCAGCACTCG AAATCAATGC ACCGAACCCG AACCAACTCG AACCAACTCG GTTCCCGTAC GTCCCGTAC CTGATGAACC CCCACCTGGG	ATCGTTCCCA TTTGCCCCT TGTATGTTG GAAACACA CAGCATGCCA CAGCATGCCA CAGCATGCCA CGACTACCAT CTCGCGAACG ACAGTCTCAT CGTCACGCGA CTAACGCCGA CGAGTTCGT GAGTACAGCCG CAAGTCATGCG AAACCCGCA GAGATCAATC GCTGACCGCG CTGAACCCCC CTGAACCCCG AGCTGCCCGC CGAGGACCTCA	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA CACAAGCCGG AGAACGCCGA AGCCCGCACCA ACCTCTTCC ACTTCCAGAT CCGAAAGCTG CCGAAGCACA	AATAAAAGC CTCGGGACAT TACCGGGGT ACGGTTACTA ATGGTTATGT CAGCACCAAC ATCATCCGAC CAACCCAACC
6301 6401 6501 6601 6701 6801 7001 7101 7201 7201 7201 7401 7501 7501 7501 8001 8101 8201 8301 8401 8501	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CTCTTTTCTC CGGCAGCACAT TCGGCTTAA CCAGCAACGC GTGACGCAC GTGACGCAC CTGGGAGCAC GCGAGCAC CTCAAGCAAC CTCAAGCAAG CTCCCGAAC CTCCAGAA CTTCCCGAAC CACCAGCA CAACCAGCAC CAACGAGCAG CAACGAGCAG GGCGGCTGAC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCAGCAA CACGCAGCAA CACGTGGACAG TCCACTCACA CTAGCACACAC CAAAGTCCCC CATCCGCAAG GGTACACCGA AACAGGCTCCAA ACGGCTCCGAC GGGCATGGCA CGGCTCCGCC CGTTGCAGCA CGGCTCCGACA	GTTCAATTAT TCACCCACAC ACTCTCACC ACCTGTGGG AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCGCCTGCACCA CCACCGAAT CCAACGAGTGA ACATCAAGC GCATTCAAG GATGTAACT ACCCGCTACC GCAGGACCAG ACAGCACCAG ACAGCACCAATG CAGGCACCAG CAGGACAATG CAGGCACCAG	CGTTTTTCGG ACACACACAT TTGCCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT TCATCGCCCG TCAGTGGTA ATTCGGGCTG GTACGGCAAT CTCGAAACAC ACTCGCACGG GGAAAAGATG TGTCCGAAGT ACGCCACCAA GATCATCGAT GAGTGCGGCG CAATCCTCCC ATCTTCCGGT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTCGATTTA GGCGGAGTGC AGAGTGGAT CCACCAAGG CTCACCAGGCA CTCACCACG GTACTGCCAC GTACTGCCAC GTACTGCACA GTACTGCACA GTCCAACGGCA CACCCCTCCC GCCAACGGCA CCCCCCCCC GCCAACGGCA CCCCACGCACCTT GCCGACCAGC	GGGGGAGGGC GAGTGTGAGC TCACCGCGAC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCC GAAGGACGCC CACGTCCAAC CCACGTCCAGC GCAAGAATTG GGAAGTGGCC GTCTGCCACT CCAGCGGCT ACAGTGAATC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACACACT GACGACGCAC GACGACGCAC GACGACGCAC GCCAACCTCC GCCAACCTCC GCCAACCACG GTTCGGGACG AACTCAATGC TGCATCACG AACCAACTCG AACCAACTCG GACGAATGG GTTCCCGTAC CTGATGAACC CCACCCTGGG GGAACGACC	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAAACACACA CACACACACA CACACACACA CACACACACA CACACACACA CACATCTCAT CTCGCGGAGG GGAGTACAGCG GAGTACAGCG GAGTACAACC CAAGTATCGCG GAGATCAATC GCTGAACCCCA AAGCTCCCCA AAGCTCACCC CTGAACCTCCA AGCTCGCGGG CGAGGAGCTG CAGGCTACAC	AACGCAGGGA TTTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCAGCGC CAGCAGCAGC CCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA ACCATCTCGA GCCGCACCA AACCTCTTCC ACTACCAGAT CGAAAGCTG CTGAAGCACA CGCTGCAGGC	AATAAAAAGC CTCGGGACAT TACCGAGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC ACCAACAGCA GGCAGCGAGG GGCAGCGAGG GGCAGCGAGG AGGTGGTGCT TAAGCATCTG CTAACGCCAC GACTCGACC GTTTGCCGAC TTAGGCGCG ATCATTCGAT
6301 6401 6601 6701 6801 7001 7101 7201 7301 7401 7401 7401 7501 7501 7801 8001 8101 8301 8301 8401	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGCCAGATCA GCAGCAGCAT CCGCCAGCAGAT TCCCCCGCAGA TCGGCATTAA CCAGCAACGC ACGAGGACGG CTGGAGCACG CTGGAGCACG CAGCAGCAG CAGCAGCAG CAGCAGCAG CAGCAGCCG CAACGACCA CAACGACCAC CAACGAGCAG GGCGGCTGAC GAGTCCCCGA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAAC ACAGCAGCAAC CTAGCACCACA CTAGCACCACA CTAGCACCACA ACGGCACCCA AGCGCTCCAA GGGCATCGCG GGTACACCGA GGCTCCGCAC CCGCTGCAGCA CGGCTCCGCAC AGTTCGGCAGCA CATCGGCATGC	GTTCAATTAT TCACCCACAC ACTCTCTCACC ACCGTGTGGC AAATACGTAC TTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCAACGCATC CCAACGCAGTCG CCGAAGGTGG ACCGCTACCA GCCGTCCAG GACTGTAACT ACCCGCTACC GCAGCACCAG ACGGCACCAG ACAGCACCAG CCGCAGCAGCAG CGGCGGCACCA CGCGTCGGACG GTCCAATCCG	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCCG CTCAGTGGCACG GTCAGTGCACGG GGAAAAGATG TGTCCGAAGT CACTCCGACG GAATCCTCAC CGACTCCCGTT CAATGCGCCC ACTTTTCGGT CAATGCGCCC	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGACTGC AGAGTGGATT GCTGATCCCA GGTACGGCAA CCACCAAAGA CTCACCTGCC GTAGTACGGA GTACTGCCAC GTGTACGGA CACCCTCCC GCCAACGGCA CCCCCCCT GCGACCACCG CGACCACCGG	GGGGGAGGCT GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGACAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA TTTTGAGGCA CGACGACCACCA GGAGGACGCC CAGGTGAGCC CACGTCAAC CCACGTCGTGG GCAAGATTG GGAAGATGG GCAAGTGCC GTCTGCCACT CCAGCAGAAC TCCAGCGGCT ACAGTGAAC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CGGAACCACT GACGACGCAC AAAGGAGCAA GCCAACTCCG GCTCGGACTCC GCAACGCATCG GAAGCAAGCC AACCAACTCCG AACCGAAAGCG AAACGAATGG ATCCCGTAC CTGATGAACC CCACCTGGG GGAGCGACC	ATCGTTCCCA TTTGCCCCT TGTATCGTGTG GAAACACACA CACCATCGCA CACAACACA CACATCCAT CTCGCGAGCG ACAGTCTCAT CTCGCGAGCG CTAACGCCGA GTGACGTTGTC GACTACACGC AAACTATCGC AAACCGCAGA GCGATCATCC CTGAACTCACCA GCGAGTCATCC CTGAACCTCA AGCTGGCGGG CGGAGCACG CAGGCTACAC	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CCCACACCCC CCCGTACGAT GCGGTAACGA ACCATCTCGA AGAACGCCGA AGAACGCCGA AACCTCTTCC ACTTCCAGAT CGAAAAGCTG CTGAAGCACA CGCTGCAGGC TGCCTTTAAG	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCACCAAC TACAACCCACC TACAACCCACC TACAACCCTC ACCAACAGCA GCCAGCGAGG GGCAGCGAGG GGCAGCGAGC CACTCGAACG CACTCGAACG CACACCAGCT GTACGCCAC ACGGCAGCG TACGCAGC ACGGCGGC CACACCAGC TACGCCAGC CACACCAGC TACGCCGC CACACCAGCG CACACCAGCG CACACCAGC CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGCG CACACCAGC
6301 6401 6501 6701 6701 7001 7101 7301 7301 7501 7601 7701 8001 8101 8301 8401 8301 8401 8501 86701	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGCTTGC TTCTTTTCTC CGGCAGATCA GCAGCAGCAT TCCCCGCAGA TCGGCAGCACA CCAGCAACCACG CTGGGAGCACC CGCAACCACC CTCAAGCAAG CTCCCGAAC CACCAGCCG CAACGACCAG CAACCAGCAG CAACCAGCAG CAACCAGCCG CAACGACCAC GAGCTCCCGA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCAACAC CAACGCCCCA AGCGCCCCA GGTACACCGA ACAGGCTCA GGCATGCAC CGGTTCGCAGCA CGGCTCGCAC AGTCGCACGAC	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCAACCGAAT CCAACCGAAT CCAACGAATG CCGAAGGTGCA GATTGTAACG GACTGTACCA GCAGGACCAG ACAGCACCAA ACATGTTCAA GCACGCACCAG CCGAGGACCAG GTCCAATCCG TCGGGTGCGCT	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATCGCTGTGT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGGCCG CTCAGTGGTA ATTCGGGGCA CTCGAACACA ACTCGCACGA GGAAACACCA CGACACCAT CAATCTCAC CGACTCCCC ATCTTCCGT CAATGCGCTC CGGACACCTT CAAACGAAT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGAGTGC AGAGTGGATT GCTGATCCCA GGTACGGCA CCACCAAAGA GTACGGCA GTACTGCCAC GTGTACGGGA CACCCAACGGA CACCCACCC GCCAACGGCA CCCCCAGGAGG CGCACCACCAC	GGGGGAGGC GAGTGTGAGC TCACCGACG CACTCGCGAG AAACGAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CACACACACA GGAGGAGCC CAGCGCACAC CACGTCAAGC GCAAGAGTGCACC GTCTGCCACT GCACGACAACC TCCAGCGAGAAC TCCAGCGAGAC TACGGGCACAC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CCGAACCACT GACGACGACC AAAGGAGCAA CACAGCTCGC GCCACGCACCT GCCACGCACCT GCAAGCAGTG GTTCGTGACG AACCAACTCCG AACCAACTCCG AACCAACTCCG AACCAACTCCG AACCAACTCCG AACCAACTCCG CAACCCCGGG GTTCCCGTAC CTGATGAACC CCGACCTGGG GGAGACGACC CGGAGCGACC	ATCGTTCCCA TTTGCCCCT TGTATGTTG GAAACACACA CAGCATGCCA CAGCATGCCA CAGCATGCCA CAGCATCCAT CGCTCAGCAG CTAACGCCGA CTAACGCCGA CGAGTTCATCGC CAAGTCATCGCG AAACCGCCA GAGATCAATC GCTGACGCGA GAGATCAATC GCTGAACCGCG AGCGCCGCG CGAGGACCTG CAGCTCACC CTGAACCTCAC	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CACACACCC CCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA GCACACCCG ACAACGCCGA ACACCCCTTCCA ACTCCAGAT CGAAAGCCG CACAAGCCG ACTTCCAGAT CGAAAGCCG CTGAAGCACA CGCGGCACCA	AATAAAAGC CTCGGGACAT TACCGGGGCACAT ACGGTTACTA ATGGTTATGT CAGCACCAAC TACAACCCAC TACAACCCAC TACAACCCAC ACCGATGAC GCCAGCGAGG AGCTCAGCTT GTACCATCTG CACTCGAACG CACACCAGCT GTTACCACCAC GTTACCACCAC GTTACCGCGC TAACGCAGCT GTTACCGCGC ACGGCGGC ACCATTCGAT CTGGAGCGGC
6301 6401 6501 6701 6801 7001 7101 7201 7301 7401 7501 7801 7801 7801 8001 8301 8301 8401 8401 8501 8501	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CTCTTTTCTC CGGCAGCACAT TCCCCGCAGA TCGGCTTTAA CCAGCAACGC GTGACGCAC CTGAGGACGCA CGCAACCACC CTCAAGCAAG CTCCCGAAC CTCCCGAAC CACCAGCAG CAACGAGCAG GACGCCGAC CAACGAGCAG GACGTCCCGG TTACAGAACC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCACACAC CAAAGTCCCC CATCCGCAAG ACGCGCGCAGC ACGCGCGCACCA ACAGGCTCCAA CGGCTCGCGC GTTACACGA CGGCTCCGCAC AGTCCGCACA CGGCTCCGCAC AGTCCGCACAC	GTTCAATTAT TCACCCACAC ACCTCTACCA ACCTGTGTGGC AAATACGTAC TTTTTGCAGA ACCGCACGCA CCGCCTGCACCA TATGGATTCA CCACCGAAT CCAACGAGTGA ACATCAAGCC GCCGTTCCAG GATGTAACT ACCCGCTACC GCAGGACAGCA ACATGTTCAA GCAGCACCAA CAGGCACCAG CAGGACAATG CCGCTGGCACCG TCGGCTGGCC	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA TCACGGCGTG TCATCGCCCG CTCAGTGGTA ATTCGGGCAT CTCGGAGCAT CTCGGAACAC ACTCGCACGG GGAAACACT ACGCCACCAA GATCATCGAAGT CAATCCCACG CAACTCCCCC ATCTTTCGGT CAATGCGCTC CAGCACCTT CAGACACTT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGAGTGC AGAGTGGATTTA GCTGATCCCA GGTACGGGCA CCACCAAAGA CTCACCTGCC GTACTGCACGGA CACTCAAACA CACCCCTCCC GCCAACGGCA CCCCCCCCC GCCAACGCA CCCCCGCACCTCC GCGACCACGG GGACCACCAC GGACCACCAC	GGGGGAGGC GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCC GGAGGACGCC CACGTCCAGC GCAAGATTG GCAAGACTCC CCGCCTCAAC TCCGCCCACT CCACGCGCC TCCAGCGGCT ACAGTGAACC TCCAGCGGCACG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACACAC CGAACACAC CGAACCACT GACGACGCAC AAAGACTCCC GCCAACCTCC GGCTCGGACG GTTCGTGACG AACCAACTCCG AACCAACTCCG AACCAACTCCG GACCACATCG GACCACCTCG GAACCACTCC CCACCCTGGG GGAGCCACC CGGAGTCGCC CTGCGGAGACC	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAACACACA CACCATGCCA CACCATGCCA CACCATCCAT CTCGCGAGCG ACAGTCTCAT CGCTCAGCAG GTGAGGATG GAGTACAGC GAGTACACC CAAGTATCGCA CAAGTATCGCC CTGAACCCCAGA GAGACCACC CTGAACCCCAGA GAGACCACC CTGAACCCCAGA GCTGATCGCCG CGCAGCCACC GTGAGGCCG GGCGCCACT	AACGCAGGGA TTTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCAGCAGC CCACACCCC CCCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA GCCGCACCG AGAACGCCGA AACCTCTTCC ACTTCCAGAT CGAAAGCTG CTGAAGCACG CTGAAGCACA CGCTGCACGCC TGCCTTTAAG ACCCCGTGA	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6701 6801 7001 7101 7201 7301 7301 7301 7301 7301 7801 8001 8301 8301 8301 8301 8301 8301 8	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGGCAGCACAG CGGCAGCAGCAT TCCTCTTCTCC CGGCAGCAGCAT TCCCCGCAGA CCAGCAGCACG CTGGCAGCACG CTGGCAGCACG CGCAACCACC TGTTCCAGTAA GAACCTGGAC CTCCAGCAAC CAGCAGCAG CAGCAGCCG CAACGAGCAG GGCGCCTGAC CAACGAGCAG GGCGCCGAGACAG TTACAGAACC CACCCGAGAAC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAAC ACAGCAGCAAC ACAGCAGCAAC GCGTGGACAG CTAGCACTCACA ACGGCGCAGCC AGCGCTCCAA GGGCACGCG GGTACACCGA ACAGGCTCG AGGCCTCGCAC AGGCTCGCAC AGGTCCGCAC AGGTCCGCAC AGGTCCCCAC CATCGGCATC GGCCACTTCG GGCCACTTCG GGCCACTTCG CCCCCGCAC	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCACCGCTGGCT CCAACGAATC CCGACGAGTGG ACGTCTCAG GACTGTAACC GCCGTTCCAG GACGTCCAG ACGGCACCGA ACATGTTCAA GCAGCAGCAG CGGGTGGGACG GTCCGATCG TCGGGTGCGCT CCCCCAAGCA	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCCG CTCAGTGGTA GTACGCACGA CTCGCACGA GGAAAAGATG GGACCACCAA GATCCTCCAC CAATCTCCCC ATCTTCGGT CAAACGAACT CGGACACCTT CAAACGAACT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGACTGC AGAGTGGATT GCTGATCCA GGTACGGCA CCACCAAAGA CCACCAAAGA CTCACGTGCAC GTATGGCAC GTGTACGGGA CACCCCTCCC GCCAACGGCA CCCCCGCACCTT GCGGACCACG GGACACACAC GGACCACACG GGACACACAC	GGGGGAGGGT GAGTGTGAGC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA TTTTGAGGCA CCACCACCAC AGGTGGAGCAC CCACGACACC CAGGTGAGC CAGGTCAAC CCACGTCGTGG GCAAGAATTG GGAAGAGTGGC GTCTGCCACT CCAGCAGAAC TCCAGCGGCT ACAGTGAAC TACGAGCACG TACGAGCACG ACTGCGACAC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CCGAACCACT GACGACGACC CACAGCTCG GCTACGGACTCC GCTACGGACTCC GCAACCACTCC GCAACCACTCCG AACCAAACCCG AACCAACTCCG AACCAACTCCG CACCAACTCCG CACCAACTCCG CACCAACTCCG CACCAACTCCG CACCAACTCCG CACCCTCCGG GGAGCGACC CCGCACCTCCG CGGAGCCACC CCGAGCTCCCG	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAAACACACA CAGCATGCCA CACAACACA CACACACACA CACATCCCAT CTCCCGACGACG ACAGTCTCACA CTCACGAGATCG CGAGTACACCCCA CAGATCACCCCA AACCCCCAGA GAGATCAACC CTGAACACCCCA AGCTGCGCGGG CGAGACACCTCA AGCTGCCGGG CGAGCACCTCA CAGATCACCCCA CAGATCACCCCA CAGCACCCCA CAGCACCCCA CAGCACCCCA CAGCACCCCA CAGCACCGCAGG CGAGCACCACC CAGCACCACCAC	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CCCACACCCC CCCGTACGAC CCCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA ACATCAGCCGA AGAACGCCGA AACCTCTTCC ACTTCCAGAT CGAAAAGCTG TGCATTTAG ACACCGGTGA ACACCGGTGA	AATAAAAAGC CTCGGGACACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCCACCAC CAACCCAACC
6301 6401 6501 6601 7001 7001 7201 7201 7301 7401 7501 7401 7501 8001 8001 8301 8301 8501 8601 8601 8601 8601 8601 8601 8601 86	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGCCAGATCA GCAGCAGCAT CCGCCAGACA TCGGCTTTAA CCAGCAACGC GTGACGCAC GTGACGCAC CGCAACCACC TGTACCAGAA GACCTGGAC CTCAAGCAAG CTCCAGCAA CTCCCGAAC CAGCAGCCG CAACGAGCCG CAACGAGCCG CAACGAGCCG CAACGAGCCG CAACGAGCCG CAACGAGCCG CAACGAGCCG CAACGAGCCG CAACGAGCCG CAACGAGCAG CACTCGAGAA TAACACGGGT	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCACAC CACCATCATC GCGTGGACAG TCCACTCACA CTAGCACACC CAAAGTCCCC AGCGCTCCAA CGGGCTCCAA GGGCATGGCA GGCATGGCA ACAGGCTCCGAC AGCGCTCCGAC AGTCGGCAC GGTCCGCAC GGCCACTTCG GCCCTGCGC GCCCCGCC ACGACGCACT	GTTCAATTAT TCACCCACAC ACTTCTCACC AACTGTGTGGC AAATACGTAC TTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCAACGGATCA CCGACGGATCA GCAGCAGCAG GATTGTAACT ACCCGCTACCG GCAGGACCAG ACAGCACCAG ACAGCACCAG ACAGCACCAG CAGGACACGA CACGACACGA CACGCACCAAGA CACCCAACGA CACCCAACGA	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATCGCTGTGT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGGGCA CTCAGTGGTA CTCGAACACA ACTCGCACGG GGAAAAGATG CACTCCCAT CAATCCCAC CAATCCCCC ATCTTTCGGT CAATGCGCC CGACACCTC CGACACCTC CGACACCTC CGACACCTC CAATGCGCGC CGACACCTC CAAAGAAAT ACGTTCACTC AACATGTGCG	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGAGTGC AGAGTGGATT GCTGATCCCA CGACCAAAGA CCACCAAAGA GTACGGCA GTACTGCCAC GTGTACGGGA CACTCAACAG CACCCACCCC GCCAACGGCA CCCCACGGC CCCACGGACCACC TCCCAGGAGG GGACCACCAC GGACACACAC	GGGGGAGGGT GAGTGTGAGG CACTCGCGAG AAACGAACA AATGGCGAC ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCT GGAGGACGCC AGGTGGAGCC CTGCGTCAAC TCGCTCAAGC GCAAGAATTG GGAAGTGGCC GTCTGCCACT CCAGCGACATT CCAGCGACATC TACGGCGACA TACGGGCACA TACGGACAG ACTGCGACAT	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACGACA CCGAACCACT GACGACGACC AAAGGAGCAA CACAGCTCGC GCCAGCACTCC GCCAGCACTCC GCCAGCACGAC TGCATCTGGA AACGAAAGCG ACCGAACTCCG AACGAACCC GAACACCCG GTTCCCGTAC CTGATGAACC CCGACCTGGG GGAGACGACC CGGAGACGACC CGGAGACGACC CTGAGAAACC	ATCGTTCCCA TTTGCCCCT GAACACCAC CACACACCA CACACACAC CACACACAC	AACGCAGGGA TTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC ATTCACACTC GTCGGCAGCAGC CCCACACCCC CTCGTACGAT GCGGTACGAT GCGGTACGAT ACCATCTCGA ACCATCTCCGA ACCACTCTTCC ACTACAGACG CTGAAGCCG CTGAAGCACA CGCTGCAGGC TGCACCGAGAG ACACCGGTGA ACCCCTTAAG ACCCCGTGAAGCAC TGCACCGCAGAG	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC ACCAACAGCA GGCAGCGAGG AGCTCAGCT GACTCGAACAG CACTCGAACAG CACTCGAACG CACTCGAACG CACACCAGCT GTTACGCCAC AACGGCAGTG TTACGGCACTG ATCATTCGAT CGGCACTGGT CCACATGGGC GCGCACTGAT
6301 6401 6501 6701 6701 7001 7101 7201 7301 7301 7301 7301 7801 7801 8001 8101 8201 8301 8401 8401 8401 8401 8401 8701 8401 8401 8401 8401 8401 8401 8401 84	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CTCTTTTCTC CGCCAGAACAA CGAGCAGCAT TCCCCGCAGA TCGGCTTAA CCAGCAACGC GTGACGCCAC ACGAGGACGC CTGGAGCAC CTCAAGCAAC CTCAAGCAAC CTCCAAGCAAC CTCCCGAAC CACCAGCCG CAACCAGCCG CAACCAGCCG CAACCAGCCG CAACCAGCCG CAACCAGCCG CAACCAGCCG CAACCAGCCG CAACCAGCCG CAACCAGCCG CAACCAGCCG TTACAGAACC CCCCGACAAC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAAC CACCATCATC GCGTGGACAG TCCACTCACA CTAGCACACAC CAAAGTCCCC CATCCGCAAG ACGCGCAGCC GGTACACCGA ACGGCTCGAGC ACGCTCGGCG GGTCCGCACG CATCGGTATC GGCCACTCG GTCCCCTGCGC GTCCCCTGCGG ACGCCCTGCG GTCCCCTGCGG ACGCCCTGCG GCCACTCG GGGACGCCCG	GTTCAATTAT TCACCAACA ACCTCACACA ACCTCTCACCA AACTGCAGCA ACTGCACCA ACTGCACCA TATTGGATCA CCGCCTGCACCA TATGGATTCA CCAACGAATC CCAACGATCA GCCATCCAG GATTGTAACT ACCCGCTACCG GCCGTTCCAG GACTGTACA ACATGTTCAA GCAGCAACAG CAGGACAATG CGGTGGGACG GTCCAATCCG TCGCCCAAGGAC GTCCAACTGC AGATGGTGCC AGATGGTGCC	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATGCAGACTG TCATCGGCCG CTCAGTGGTA ATTCGGGCAT CTCGAGTGGTA ATTCGGACAT CTCGAAACAC ACTCGCACGG GGAAACAGAT CACCCACCAA GATCATCGAC CAATCCCAC CGACTCCCC CACTCTCCCC TCTTTCCGGT CAATGCGCTC CGGACACCTT CAAACGAAT ACGTCACTC CAATGCGCTC CGGACACCTT CAAACGAAT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCTGGACCACA GGTACGGACTG CCACCAAAGA TCAGGTGAAG GTACTGCCAC GTACTGCCAC GTACTGCCAC GTCACAGGGA CACCCCCCC GCCAACGGCA CCCACGGACCAC TCCCAGGACGAG GGACCACACG GGACCACACAC ACCGCGCACTG GCGACAAGAG ATCGGGGATG	GGGGGAGGC GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGAC GGAGGACGCC AAGTGCGACC GGAGGACGCC CACGTCCAGC GCAAGATTG GCAAGTGCC GTCTGCCACT CCACGTCGTGG GCAAGAATTG GCAAGAGACTG CACGTCGCACAT TCACGACAGC ACTGCGACATG ACAGTGGACATG TACGGCCATG TACGCGCATG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CGGACGCAC CGAACGACG CGAACGACG GCCAACCTCC GCCAACCTCC GCCAACCTCC GCCAGCACTG CAAGCAGTG AAATCAATGC TGCATCTGCG AACCGAAATCCG AACCGAACTCCG AACCGAACTCCG CCAACCTCGG GTTCCCGTAC CCGAGACGACC CGGAGTCCCC TGCAGGACACAC	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAAACACACA CACCATGCCA CACAACACA CACAACACA CACATCCCAT CTCGCGAGCG CGACTACCAT CTCGCCAGCAG GAGATTTGTC GAGTACACAC CAAGTATGCG AAACCGCAGA GAGATCATCC CTGAACTCCA AGCTGGCGGG CGAGGACTG GCGAGGACTG GCGGCTACAC GACGATGACC TCCTGCACAT TACCTGATAA	AACGCAGGGA TTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCAGCAGC CCACACCCC CCCCACACCCC CCCCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA GCCGCACCG AGAAAGCCGG AGAACGCCGA GCCCGCACCG CTGAAAGCAG CTGAAAGCAG CTGAAAGCAG CTGAAAGCAG CTGAAAGCAG CTGCATTAAG ACACCGGCAGA GTATTTATTT	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6701 6701 7001 7101 7201 7301 7301 7301 7301 7301 7301 7301 73	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGGCAGATCA GCAGCAGCATCA CCAGCAGCATCA CCAGCAGCAT TCCCCGCAGA CCAGCAACCAC CGGAACCACC TGTCCCAGTA GAACCTGGAC CTCCAGCAAC CTCCAGCAAC CAGCAGCCG CAACGAGCAG GGCCGCTGAC GACTCCGGAA TATCACGGAT CCCCCGAAGAC	CACACAGCGC ATAAAAATAA TGTGGGGGGAG CAAAAAAAAA CTACTCACAC TTTTCCCCAC ACAGCAGCAA CACCATCATC GCGTGGACAG TCACCACTCACA CTAGCAACAC CAACGTCCCC AGCGCTCCAA CCGCTGCGCG GGTACACCGA AGCGCTCCGAC GGCTTGCAGCA CGGCTCCGAC AGGCTCCGAC CATCGGCATC GGCCACTTCG CACCGCTCCGC CATCGCACGC CACCGCCGC CCCCGCCCCG	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCTGTGGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCACCGCTGCAT CCAACGAAT CCAACGAATC CCGAAGGTGA CCGAAGGTGG ACCTGTCAAGCC GCCGTTCCAG GACTGTCAAG ACATGTTCAA CCGCCAGCAG CGCGCGCACGA GTCCAATCCA TCGGGTGGCCT CGCCCAAGGA GTTCAAGTGC AGATGGTGGC	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATCGCTGTGT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCCG CTCAGTGGTA ATTCGGCACG GGAACACAT CTCGAACACA ACTCGCACGA GGAAAAGATG GAATCCACCA CAATCTCCAC CGACTCCCC ATCTTCCGGT CAAATGCGCTC CGGACACCTT CAAATGCGCTC CGAACGAAAT ACGTCACTC AACTGCACCC ACGTCACTC ACGTCACTC CAAACGAAAT	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGACTGC AGAGTGGATT GCTGATCCA CCACCAAAGA TCAGGTGAAG CTACACTGCC GTAGTGCAC GTACTGCAC GTACTGCAC GTACTGCAC GCCACCGGCA CCCCCAGGAG CCCCCAGGAG CCCACCACCT GCGACCACCG GGACACACAC GACTGCAAGT ATCGGGATG GATGGGATG	GGGGGAGGCT GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA TTTTGAGGCA GGAGGACGCC AAGTTCCGCT CAAGTGCAAC CCACGTCGTGG GCAAGAATTGC GCAAGAATGGC GTCTGCCACT CCAGCGGACAC TCCAGCGGACAC TACCGCGACAT TAAGGGCACG AATGCGCACAT TAATGCATA	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CGGAACACAC GCGACACCACT GACGACGACC AAAGGAGCAAC GCCAACCTCC GGCTCGGACTCC GCCACGCACTCC GCCACCTGGA AAATCAATGC TGCATCAGCG AACCAACTCCG AACCAACTCCG AACCAACTCCG CTGATGAACC CTGATGAACC CCGCCTGGG GGAGACGACC CGGAGACGACC TGCGTTCCGG ACCGGCTCCTG GCGGCACACC CGGACACACC CGGACACACC	ATCETTCCCA TTTGCCCCT TGTATGTTG GAACACCACA CACCATGCCA CACCATGCCA CACAACACA CACATCACAT	AACGCAGGGA TTTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CCCACACCCC CCCGTACGAC CCCCACACCCC CCCGTACGAT GCGGTAACGA ACCATCTCGA ACATCAGACGCGA AGAACGCCGA ACACCCTTCAGAT CGCAAAAGCTG TGCCTTTAAG ACACCGGTGA GTATTATTT CAGAAGCTAG	AATAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCACCAAC TACAACCCTC CAACCCAACC
6301 6401 6501 6701 6701 7001 7101 7201 7301 7401 7501 7801 7801 8301 8301 8301 8301 8301 8301 8301 8	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGCCAGATCA GCAGCAGCAT CCCCCGCAGA TCGGCTTTAA CCAGCAACGC GTGACGCAC GTGACGCAC CTGGAGCAC CTGGAGCAC CTCAGCAAC CTCAGCAAC CTCCAGAA CTTCCCGAAC CACCAGACCAG	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCAGCAA CACGCAGCAA CACGTGGACAG TCCACTCACA CTAGCACACAC CAAAGTCCCC CATCCGCAAG AGCGCTCCAA AGCGCTCCAA AGCGCTCCACA CGGCTGCGCC GGTACACCGA CGGCTCCGCC GGTCCGCCC GGCCACTTCG GGCCACTTCG GGCCACTTCG GGCCACTTCG GGCCACTTCG GGCCACTTCG CACGCCGCC ACGACGACGT GGGACGCCG AGACAACGT	GTTCAATTAT TCACCCACAC ACTCCACCAC ACTCTCACCC ACCTGTGGGC AAATACGTACC TTTTTGCAGA ACGGCACGCA CCACCGACAC CCACCGAATCCA CCACCGAAGGTGA ACATCAAGCC GCAGTGTAACT ACCCGCTACCG GCAGTGTCAAG CGCAGCACCAG ACAGCACCACA CAGGCACCAG CCGGTGGACG CCGGTGGACG GTCCAATCCG CGCGGTGGCC TCGGCTGACCG TCGCCCAAGGA GTTCAAGTGC CGGTGGGCG TTCGATGGCC	CGTTTTTCGG ACACACACAT TTGCCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CCAACGGCTG TCATCGCCCG CTCAGTGGTA ATTCGCGGCT CTCGGAACAC ACTCGCACGG GGAAACACT CTCGCACGAG GGAAACAG GGCACCCAT CAATCCCCACG CAATCCCCGTT GAGTGCGGCG CAACTCTCCC CGGACACCTT CAAACGAAT ACGTCACGC ACTTTCCGG GCGTACCGA CTTATTTCCC	GATGGAATAG ACACGCGAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTCGATTTA GGCGGAGTGC AGAGTGGAT CCACGGCA CCACCAAGG GTACTGCAC GTACTGCAC GTACTGCAC GTACTGCAC GTACTGCAC GTACTGCAC GTACTGCAC GTACTGCAC GCCAACGGCA CACCCACCC GCCAACGGCA CCCCCCCC	GGGGGAGGGT GAGTGTGAGC TCACCGCGAC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCT GGAGGACGCC CACGTCCAGC GCAAGATGGCC GCAAGATGGCC GCAGGCACAT CCAGCGGCT ACAGTGAATC TAAGGGTGAG TACAGGACACG TAGCGACATG CACGACCACG TAGCGACATG CACGACCACG TATGCCCATA	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACACCACT GACGACGCAC GACGACGCAC GACGACGCAC GCCAACCTCC GGCTCGGACT GCAAGCAGTG GTTCGTGAG AAATCAATGC TGCATCTGCG AACCAACTCC GAACCACCC CGGAGACGAC CTGATGAACC CGGAGTCCCC CGGAGCACC CGGAGCCCC TGCGTTCCGG ACCGGCTTCT CGGGCTCCC CATAAATAAT	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAACACACA CACACACACA CACACACACA CACACACA	AACGCAGGGA TTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCAGCAG CCACATCATC CCGCGCAGCC CCCCCACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA CATGAAGCCGG AGAACGCCGA CACCACCCC CTCACAAGCCG CTCACAAGCCG CTCACAGCCCA CGCTGCAGGC TGCCTTTACA GCCCGCAAG GTATTTACTT CAGAAGCTAG GATTTACACA	AATAAAAAGC CTCGGGACAT TACCGAGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCACC ACCAACAGCA GGCAGCGAGG GGCAGCGAGG AGCTCAGCT GTACCATCTG CACCGCACTG CTAACGCCACC CACACCAGCT GTTGCCGAC ACGTGGTGCT CTAGGCGGC CACACCAGCT CTGGAGCGGC CACATCGGC CCACATGGGC CCACATGGGC CCGCACTGAT ATTGGTAAGAG TCGGCAGAGA
6301 6401 6501 6701 6701 7001 7101 7201 7301 7301 7301 7301 7301 7301 8001 8301 8301 8301 8301 8301 8301 8	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CTCTTTTCTC CGCCAGCAGCAT TCCCCGCAGA TCGGCTTTAA CCAGCAACACC GTGACGCCAC ACGAGGACGAC GTGACGCCAC TGTTCCAGTA GAACCTGGAC CTCAAGCAAC CTCAAGCAAC CACCAGGCCG CAACCACCC CACCAGCCAC CACCAGCCAC CACCAGCCGAC GAGTTCCCGG TTACAGAACC CACCAGCCG TACCCGGATAG GGACAGGCGA ACGTTTCACGGA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCACTCACA CTAGCACCAC ACGGCTGCACA GGCACTCCAA GGCACTCCGA GGTACACCGA GGCACTCGGCA CCGCTCGGCA CGCTCGGCA CGCTCGGCA CATCGGCAGC CATCGGCAGC CATCGGCAGC CACCGCTCGCG ACGCCCCG AGGCACTCG GGGACGCCCG AGGCACACGT AGGCCACTG GGACACCCG AACGCTCAGG CCCCCG CCCCCGCG CACCCCCG AACGCCCG AACGCCCCG AACGCCCCG AACGCCCCG AACGCCCCG AACGCCCCG CCACTCCGCCCG AACGCCCCG AACGCCCCG AACGCCCCG AACGCTCCAAACGTG AACGCTAACG	GTTCAATTAT TCACCAACA ACTCTCACCA ACCTGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGCA ACCTGCACCA TATGGATTCA CCACCGAAT CCAACGAGTGA ACGCCTGGCT CCAACGAGTGA ACATCAAGCC GCCTTCCAG GATTGTAACT ACCCGCTACC GCAGTCCAG ACGGCACCAG ACGGCACCAG CGGGTGGACG GTCCAATCCG TCGGGTGGCT CGCCCAAGGA GTTCAAGTGC GAAGACAGC GGAAGAAGACG GAAGAAGACAC	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA TCGCGTGTG CAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCCCG CTCAGTGGTA CTCGAACAC ACTCGCACGG GGAAAAGATG TGTCCGACGA GATCATCGAC CGACTCCCGTT CAATCTCCC CGGACACCTT CAAATGCGCCC ACGTTCACCC ACGTTCACCC ACGTTCACCC ACGTCACCGA CGGCGTACCGA CGCATACCGA CTTATTTCCC GCGAGTACCGA	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGACTGC AGAGTGGATT GCTGATCCCA GGTACGGCA CCACCAAAGA TCAGGTGAAG GTACTGCCAC GTGTACGGGA CACCCACCGC GCCACCACCGC GCCACCACGCA CGCCACACGC GGACACACCC GGACACACCC GGACACACCC GGACACACCC GGACACACAC	GGGGGAGGC GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CACCACCAC AGTGGAGCC GGAGGACGCC AAGTTCCGCT CAAGAGTGAGC CTCGCTCAAC CCACGTCGTGG GCAAGAATTG GGAAGTGCAC TCCAGCGGCT ACAGTGCACAC TCAGCGGCACT TAAGGGCACG AATGCGCATG TATATGCATA	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CGGACGCAC CGAACCACT GACGACGCAC AAAGGAGCAC GCTCGGACTGC GCTCGGACTGC GCTCGGACTGC GCTCGGAAGCG AAACCAACTCCG AAACCAACTCCG AAACCAACTCCG CAGCATCAGC CCAACCACG CCAACCTCGG GGAGCACCC CGGAGTCGCC CTGAGGAAGCC CTGAGGAACACC CGGAGCACACC CATAAATAAT AAAAGCAACG	ATCETTCCCA TTTGCCCCT TGTATGTTG GAACACACA CACCATGCCA CACACACACA CACACACACA CACACACACA CACATCTCAT CTCGCGAGCG CGACTCTCAT CGACTCTCACAG GAGATTTGTC GACTACACAC CAAGTATACAGC CTGAACGCAGA GAGATCATC CTGAACTCCA CAGCTGACGCG GAGGCTCACAC GACGGCACAG GACGATCTGC TCCTGCACAT TACCTGATAA TCGCAATGG CAATGAC	AACGCAGGGA TTTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCGCGC CACACCCC CCCCACCCC CCCCACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA ACACTCTCCA ACTCCCGCACCA CCGCACCCG CGAAAGCCGA CCGCACCGCACA CCGCGCACGG CGAAAGCTG CGCACGCGCAG GCTATTACACAAT TGCCCGCAAG GTATTTACACAA CACTACAGCTAG CACAAGCTAG CACAAGCTAG CACAAGCTAG CACAAGCTAG CACAAGCTAG CACAAGCTAG CACAAGCTAG	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6401 6501 6701 6701 7701 7701 7701 7701 7701 77	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGGCAGATCA GCAGCAGCAT TCCCCGCAGA CCAGCAGCAC TCCCCGCAGA CCAGCAACAC CGCACCAC CGCAACCAC CGCAACCAC CTCCAGCAAC CAGCAGGCCG CAACGAGCAG GGCCGCTGAC CACCCGAAA CACTCCAGAA CACTCCAGAAA CACCTCGAAA CACCTCGAAA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCAC ACAGCAGCAA CACCATCATC GCGTGGACAG TCACTCACA CTAGCAACAC CAACGCCCC AGCGCTCCAA CCGGTGCGCG GGTACACCGA CGGCTCCGAC AGCGCTCCGAC AGCGCTCCGAC AGCGCTCCGAC AGCGCTCCGAC AGTCGCAGCA CGGCTCCGAC AGTCGCACGC CATCGCACGC CATCGCACGC CACCCCGCC GGCACGCCG AGACAACGTG AGACAACGTG AACGTCAGAG	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCACCGCTGCT CCAACGGAGTCG CCGAAGGTGG ACCTGCACCA GCCGTTCCAG GCCGTTCCAG GCAGGACCAG ACATGTTCAA CCAGCAGCAG GTCCGATCCA GTCGACCAGCAG GTCCAATCCG TCCGGTGGCC GTCCAATCG TCCGCCAAGGA AGATGTTCAAGTGC AGATGATGACCAG GACAGAGAGA ACATGTTCAACTGC CAGCAACAGC CAGAGAAGAG ACTGATCACCG TTCGATGACCAGCAGA	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATCGCTGTGT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCCG CTCAGTGGTA CTCGAACACA ACTCGCACGG GGAAAAGAG GGACACCTC CAATCTCCAC CGGACACCTC CGGACACCTC CAATCTCCC ATCTTCGGT CAATGCGCG CGGACACCTC CAACGCACC CGGACACCTC CAACGCACCC CGGACACCTC CAACGCACCC ACTTTCCGC GGCGTACCG GGCGTACCG CCAACTCCCC ACTTTTCCC GCAATTTCCC CGCACTCCCACCC ACTTTCCCC	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGACTGC AGAGTGGATT GCTGATCCA AGATGGACAC CCACCAAAGA CCACCAAAGA CACCCACACG GTGACGGCA CCCCCAGAGG CGACCACCAC GCGACCACCT GCGACCACCG GGACACACAC ATCGGGATG ATCGGGATG ATCGGGATG GACAGCATTG TTTTAGTGGT	GGGGGAGGGT GAGTGTGAGC CACTCGCGAG AAACGAAAA AATGGCGAC CCACTCGCGAG ACGATCAAAT AGCAGGACAC TTTTGAGGCA TTTTGAGGCA GGAGGACGCC AAGTTCCGCT CAAGTGCAAC CCACGTCGTGG GGAGGACGCC GTCTGCCACG GCAGGAGAATTG GGAAGTGCGCC GTCTGCCACT TCCAGCGGCT ACAGCGACAT TACAGGACACG ACTGCGCACAT TGACGACACG AATGCGCACAT TATATGCATAA	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CGGAACACAC GCGACACCACT GACGACGACC AAAGGAGCAAC GCCAACCTCC GCCACGCACCTCC GCCACGCACTCC GCCACGCACTCC GCCACCTGGA AACGAATGG ACCGAACCCCG AACCAACTCCG AACCGAACCCC GGAGACGACC CTGATGAACC CCGCACCTGGG TCCGCTCCGG ACCGACCCCGG GGAGACGACC CTGAGGACCCC CGGACACCCCGG ACCGACTCCG ACCGACTCCG CTCAGGCTTCCG ACCGACACCC CATAAATAAT AAAAGCAACG CCAACAACG	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAACACCAC CACCACACACA CACCACACACA CACCACACACA CACACACACA CACACTCCAT CTCCCGACG CACGTCACCAG CTAACGCCGA CTAACGCCGA CACATTGCT GACTACACCC CACATACACCC CACATACACCC CACACACCCCA ACCCCCACA CACACACCCCA ACCCCCACA CACACACCCCA ACCCCCACA CACACACCCCA CACACACCCCA CACACACCCCA CACCCCCACA CACCCCCACA CACCCCCACA CACCCCCACT CACCCCCACT TCCCCACATG CACCCCACATG CACACACCCCA	AACGCAGGGA TTTTGTCTT CGTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGGC CCACATCATC GTCGGGCGGC CCCACACCCC CCCGTACGAT GCGGTACGAT GCGGTACGAT GCGCTACGAT GCACAGCCGA AACCTCTTCC CGAAAAGCTG CGAAAAGCTG TGCCTTAAG ACCTCGCAGAT TGCCCGCAGA GTATTTATT CCACAGCTAG CATTACACA ACAGTAATGC CACATCACA	AATAAAAAGC CTCGGGACACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCAGCAAA ATCATCCGCT CAACCCAACC
6301 6401 6501 6701 6701 7001 7101 7201 7301 7401 7301 7401 7501 7701 7801 7801 8001 8301 8401 8401 8401 8401 8401 8401 8401 9001 9101 9201 9301	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CTCTTTTCTC CGCCAGACACA CCAGCAGCAT TCCCCCCACAA CCAGCAACACC GTGACGCCAC ACGAGCACACC CTGAGCACACC CTCAAGCAAC CTCCAGAA CTTCCCGAAC CTCACGACAC CACCAGCA CACCAGCAC CACCAGCAC CACCAGCAC CACCAGCAC CACCAGCAC CACTCCAGAA CTTCCCGAAC CACTCCAGAC CACCAGCCG CAACGACCAG CACTCCAGAA CTCCCGAAAA CTCCCGAAAA CTCCCGAAAA CACCTGACAG CACTCCAGAC	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCAGCAA CACGCATCATC GCGTGGACAG TCCACTCACA CTAGCACACAC CAAGGCTCCACA CTAGCGCAGCC AGCGCTCCCA ACGGCTCCGCA GGTACACCGA CGGCTCGGCA CGGCTCGGCA CGGCTCGGCA CATCGGCAGC GTCCGCACAC GGCACTCCG CATCGGCACG CATCGGCACG CATCGGCACG CACCCTCCG AGGCACCCCG AGGCACCCCG AGGCACACCG AGGCTCAGACG CGGCTCCGCC ACGACGCCG CGGACGCCCG AGGCACACCG AGGCTCAGACG CGATCCAAAG CAATCCAAAG CAATCCAAAG	GTTCAATTAT TCACCCACAC ACCTCTACCA ACCTGTGTGGC AAATACGTAC TTTTTGCAGA ACCGCACGCA CCACCGACACGC ACCTGCACCA TATGGATTCA CCACCGAAGTGA ACATCAAGCG GCCTTCCAG GATGTAACT ACCCGCTACC GCAGGACAGCA CAGGACAATG CGCGGCACGA GACGCACAGCA GTCCAATCCG GTCCAATCCG GCAGCACCAG GTCCAATCCG GCGCCAAGTG CGCCCAAGTGC CGCCAAGTGC CGCAGCACGA ACATGTTCAA CGCCCAAGTGC CGCAAGTGCC CGCAAGAGACA ACATGGTGACC CGCAAGCACAA	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CAAACGGCTG TCATCGCCCG CTCAGTGGTA ATTCGGGCAT CTCGAAGCAC ACTCGCACGG GGAAACAT CTCGCACGAG GGAAACAT CACCCCCAT GACTCCCCA ACTCTCCCC CACTCCCC CACTCCCCT CAACGCGCC CACTCCCCT CAACGCCCC ATCTTCCGG CAACTCCCC CCGGACACCTT CAACGCCCCC ACCTTCCGC GGCGTACCGA GCCGTACCGA CTTATTTCCC GCAGTTCGG GCAAAAGAAA	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGAGTGC AGAGTGGATG AGAGTGCAAG CCACCAAAGA CCACCAAAGA CTCACTGCCAC GTACTGCCAC GTACTGCCAC GTACTGCCAC GTACTGCACGGA CACCTCCC GCCAACGGCA CACCCCCCC GCCAACGGCA CCCCCGCACCTCC GCGACCACCG GGACCACCG GGACCACCG GGACCACCG GGACCACCG GAGTGCAAGT CCCAGGAAGAG ATCGGGGATG ATCGGGGATAG GAGTGCAAACG GAATCAAACA	GGGGGAGGC GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCC GGAGGACGCC AGGTGGAGCC CTGCGTCAAC TCGCTCAAC CCACGTCGTGG GCAAGATGGC GCAAGATGGC TCCAGCGGCA TCCAGCGGCA TCCAGCGGCA TCCAGCGGCA TCCAGCGCATG TCACGGCCATG TCACGGCCATG TCACGACCGC ATGCCGCATG CATGTCGTAA CATGTGCAAA TTGTCGTAAA	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA GCGGCAGCAC CCGAACACCACT GACGACGCAC GACGACGCAC GCCAACTCC GGCTCGGACT GCAACCTCC GGCTCGGACG GTTCGTGAG AACCAACTCCG AACCAACTCCG AACCAACTCCG GACACACTCC CGCAACACCC CGCAACACCC CGCAACACCC CGCAACACCC CGCAACACCC CGCAGCACCC CGCAGCACCC CGCAGCACCC CGCGACCCCG ATCGGCTTCCC GGCGACCACAC CCAAGAACAT ATAAACAACG CCAAGAACGAT	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAAACACACA CAGCATGCCA CAGCATGCCA CAGCATCCAT CTCGCGAGCG ACAGTCTCAT CGCTCAGCAG GTGAGGATGG GAGTACAACC CTAACGCCGA AAGTATCACC CTGAACTCACC CTGAACTCACC CTGAACTCACC CTGAACCCCAG AGGCTCACCC GAGGACTCACC GAGGAGCTG CAGGCTCACC TTACGGCGAG GGCGATGAT TCCTGCACAG TCCTGCACAG TCCACCATGG CAGATATGACC AGACGAACAG TTGGCCAATG CGAGGCAAAG TTGGCCGAAGAC	AACGCAGGGA TTTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCAGCAGC CCACATCATC CCGCACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA CACAAGCCGG AGAACGCCGA ACCTCTTCC ACTTCCAGAT CGCAAAGCTG CTGAAGCACG CTGCATTAAG ACACCGGTGA GCTATTATTT CAGAAGCTAG CATTACACA ACATTACACA	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC ACCAACAGCA GGCAGCGAGG AGCTCAGCT GTACCATCG CACTCGAACAGC CACTCGAACAGC TTAGCGCGC GTTAGCGCACC GTTAGCCGC ACGTAGCGC TTAGGCGGC CACCAGCT TTAGGCGCG CACCAGCAGT CACACGCAGTG TCGGAGCGGC GGCCACTGAT ATTGTAAGAG CGGCACTGAT ACAACTAAG CAAACTAAG
6301 6401 6501 6701 6701 7001 7101 7201 7301 7301 7501 7701 7801 8001 8301 8301 8501 8501 8501 8901 9101 9101 9101 9301 9401	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGCCAGATCA GCAGCAGCAT TCCCCCGCAGA CCAGCAACACG GTGACGCCAC ACGAGGACGG CTGGACGCAC ACGAGGACGG CTGCAGCACC CTCCAGCAA CACCAGGACGC CAGCAGCAGC CAGCAGCAGC GAGTCCCGAAC CACCAGGACG GAGTCCCGG TTACAGGACC CACCGGATAC GGACAGCCG ACCCGAAA AAGTTACTAT CCCGTTTTCA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG CTAGCACTCAC CAACGCACCAC CAACGCACCAC ACGGCACCCA GGTACACCGA GGTACACCGA GGCACTCGCGC GGTTCGGCAC CATCGGTATC GGCCACTCGCAC AGTCCGCAC CACCGGTATC GGCCACTCG ACGCCCCG AGACACCGC AGACACCGC AGACACCGC AGACAACGT AACTAGTAGT AGTCCAAG GCCTCGCAC	GTTCAATTAT TCACCAACA ACTCTCTCACA ACCGTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCACGCTGCAT CCAACGCAAT CCAACGAGTGG CCGAAGGTGGA CCGACGACCAG ACATCTAAGCC GCCGTTCCAG GACTGTAACT ACCCGCTACC GCAGCACCAA ACATCTTCAA GCAGCACCAA GTCCAATCCG TCGGGTGGCAC GCCCCAAGGA GTCCAATCCG GCCCAAGCA GTCCAAGCG CAGGGGCCCA GTCGACAGCA GTCCAAGCG CAGAGCACAG ACTGGTTGACCG GACAGCACAG ACGGCCACAGT TGGTTGACGAGA	CGTTTTTCGG ACACACACAT CTCGCAAACA ATCGCTGTGT ACCGTTTTAA TCGCGGATG TCATCGCCCG CTCAGTGGTA ATTCGGCCCG CTCAGTGGTA ATTCGGCCCG GTCAGTGGAACAC ACTCGCACGG GGAAAAGATG CACTCCGACGA GATCATCGAC CGACTCCCGTT CAATGCGCCG CGACACCTT CAAACGAAAT ACGTCACCG ACTTTTCGGT CAATGCGCCC ACGTTCACTC ACATGTCACTC ACATGTCCCC ACATGTCCCC ACATGTCCCC ACATGTCCCC ACATGTCCCC ACATGTCCCC CGACACCTT CAAACGAAAT ACGTCACCCACA CTTATTTCCC GCGAAAAAGG ACAGCAAAA ACTGTTACGA	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGACTGC AGAGTGGATT GCTGATCCA AGGTACGGAA CCACCAAAGA CCACCAAAGA CACCCTCCC GTACTGCCAC GTGTACGGA CACCCCTCC GCCAACGGA CACCCCTCC GCCACCACGG GGACACACAC GGACACACAC GGACACACAC	GGGGGAGGGT GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGAAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CTTTGAGGCA CTTGGAGCC GGAGGACGCC AAGTTCCGCT CAAGTGCAGC CACGTCGTGG GCAAGAATTG GGAAGACGCC GCTCGCCACG GCAAGAATG GGAAGAAGC GCAGCAGAAC TCCAGCGGCT CCAGCGGCA TACGAGCACG ACTGCGACATG TACGAGCACG ATGCGCACATG TACAGGACACG TATATGCATAA CATGTAGAAC	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CGGAACGACA CGGAACCACT GACGACGCAC AAAGGAGCAA CACAGCTCGC GCCTCGGACTCC GGCTCGGACTCC GCACGAACGCG AAACCAACTCCG AACCGAAAGCG AACCGAACTCCG CACCACTCCG CACCACTCCG CCACCCTGGG GGTCCCCTGAC CCGAGCACCC CGGAGCGACC CCGAGCTCCCC CGGAGCGACC CCGCTTCCGG ATCGGCTTCCG CGGACCACC CGGACCACC CCGACCACCC CGGACCACC CCGACCACCC CGGAGCACACC CCACACCCC CGGACCACC CCGACCACCC CGGACCACC CCGACCACC CCGACCACC CCGACCACCC CGGACCACC CCACACCCC CGGACCACC CCGACCACCC CGGACCACC CCACACCCC CGGACCACC CCGACCACC CCGACCACCC CGGACCACC CCACCACCCC CGGACCACC CCGACCACCC CCGACCACCC CCGACCACCC CCGACCACCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCGACCCCC CCCCCCCC	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAACACACA CACCATGCCA CACAACACA CACATCCAT CTCGCGAGCG ACAGTCTCAT CTCGCGAGCG CTAACGCCGA GTGAGGATGG GAGATTGTC GACTACAGCC CTGAACTCCA AACCGCAGA GCGATCATCC CTGAACTCCA CGCGCAGCG CGGGGTCACAC GGCGGTCACAC GGCGGTCACAC GCCGCTCACT CTCCTGCACAT TACCTGATAA TTGGCCGAGG TTGGCCGAGG TTGGCCGAGG TTGGCCGAGG TTGGCCGAGG TTGGCCGAGG TTGGCCGAGG TTGGCCGAGG TTGGCCGAGG TTGGCCGAGG TTGGCCGAGG	AACGCAGGGA TTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCGCGC CCCCACCCC CCCCACCCC CCCCACCCC CCCCACCCC CCCCACCCC CCCCACCCC CCCCACCCC CCCCACCCC CCCCACCCC CCCCCC	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCCGCCAC CAACCCAACC
6301 6401 6501 6601 7001 7001 7201 7201 7401 7501 7401 7501 7501 8001 8001 8301 8301 8301 8301 8301 8501 8601 8501 8601 9101 9101 9201 9301 9401	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGCCGCAGA CGCGCAGATCA GCAGCAGCAC TCCGCCTTAA CCAGCAACGC GTGACGCAC CGCAACCACC TGTACGAGACAG CTCAGGAACAC CTCAGCAAG CTCCCGACA CAGCAGCCAG CACCAGCCAG GGCGGCTGAC GAGTTCCCGA GAGTCCCGACA TAACACGGGT TACCGGAAAC CACCGGAAAA TATCACGGGT CCCCTTTTCA AAGTTACTAG	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCACCAC CACCATCATC GCGTGGACAG TCCACTCACA CTAGCACACAC CAAAGTCCCC AGCGCTCCAA AGCGCGCCCA AGCGCTCCAA GGCATGGCA AGCGCTCCGAC GGCTCCGAC GGCTCCGAC GGCCACTTCG GGCACTCCG AGGCCCCGC CACGCCCG CACGCCCG CACGCCCGC AGGACGCCG AGACAACGT AGACACGTAG CAATCCAAAG CCCTCGCACAC	GTTCAATTAT TCACCCACAC ACTTCTCACC ACCTGTGGC AAATACGTAC TTTTTGCAGA ACGGCAACGC ACCTGCACCA TATGGATTCA CCACCGCACG CCGAAGGTGG CCGAAGGTGG CCGAAGGTGG CCGAAGGTGCA GCCGTCCCAG GCCGTCCCAG GCAGGACACG CCGCCACCGA ACATGTTCAA CCGCCAACG GTCCAATCC GCCCAATCG TCCGGTGGCC GTCCAATCG GTCCAATCG GTCCAATCG CGCCAAGGA GTCCAATCG GTCCAATCG CAGGACAAGA GTCCAACGG CAGGACACG CAGGACACGG CAGGACACGA CAGGCCCAGG CGCCAAGGA CGCCAAGGA CGCCAAGGA CGCCAAGGA CGCCAAGGA CGCCAAGGA CGCCAAGGA CGCCAAGGA CGCCAAGGA CGCACAGA CGCCAAGGA CGCACAGA CGCCAAGGA CGCACAGACACA CGCACAGACACA CGCACACACA	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATCGCTGTGT CAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCCGA CTCAGTGGTA ATTCGGCACG GGAAACACC ACTCGCACGG GGAAAAGAG GATCATCGAC CGACTCCCCT CAATCTCCCC ATCTTCGGT CAATCCCCC CGGACACCTC CAACTCTCCC CGGACACCTC CAACTCTCCC CGGACACCTC CAACTCTCCC CGGACACCTC CAACTCTCCC CGGACACCTC CAACTCCCCG CGCGTACCGA CTTATTCCC CGCAGTACCGA ACGACACAA ACGGCAAAAGA ACGGCAACA ACGGCAACA ACGACACAA	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGACTGC AGAGTGGATT GCTGATCCCA CCACCAAAGA TCAGGTGAAG CTCACCTGCC GTATCGGGA GTACTGCCAC GTGTACGGGA CACCCAACGG GCAACGCCAC GCCAACGGCA CCCCCAGAGG CGACCACCAC GCGACCACCT GCGACACACC GGACACACC GGACACACAC ATCGGGATG ATCGGGATG ATCAGGGATG GAGCAAACG GAATCAAAGA ACTTCTAGC	GGGGGAGGGC GAGTGTGAGC TCACCGACC CACTCGCGAG AAACGACAAAA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCT GGAGGACGCC CACGTCGTCGAC TCGCTCAAGC GCAAGAATTG GGAAGTGGCC GTCTGCCCACT CCAGCGACATT CCAGCGACATT TCCACGGCCAT TCCACGGCCAT TGACGGCCACT TGACGACCAC ATTGCCGCACA TTGTCGCACAT TGACGACAG ATTCGTAGAA CATCTAGAAA ATACTTTTGG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CGGAACACAC GCGACACCACT GACGACGACC AAAGGAGCAAC GCCAACCTCC GGCTCGGACTCC GCCACGCACTCC GCCACGCACG AAAGCAATGG ACCGAACCCCG AACCAACTCCG AACCGAACCCC GGACACGACC CTGATGAACC CCGACCTGGG GGACACGACC CGGAGTCCGG ACCGACCTCGG GCACCTCGG CTCCGGTCCCG CGGACACGACC CGGACACGACC CGCACCACG CCACACACCG CACAACAACG ACCGACACCC CGGACACACC CGCACCACG CCACACACCG CACAACAACG ACCAACAACG ACCGACACCC CGCACACCCCG CACAACAACG ACCAACAACG ACCAACAACG ACCAACAACG CACAACAACG CACAACAACG CACAACAACG CACAACAACG CCAACAACG CCAACAACG CCAACAACG CCAACAACC CCAACAACAACG CCAACAACCACC CCAACAACAACG CCAACAACG CCAACAACCA CCAACAACCA CCAACAACCA CCAACAA	ATCGTTCCCA TTTGCCCCT GAACACCAC CACACACACA CACACACACA CACACACA	AACGCAGGGA TTTTGTCTT CCTGTTTACTG GTCACAGCAG CCACATCATT TTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCAGCAGC CCCACACCCC CTCGTACGAT GCGGTACGAT GCGGTACGAT ACCATCTCGA ACCATCTCGA ACCATCTCCG ACAAGCCGG AGAAAGCCGA ACCCTCTCCAGAT CGAAAAGCTG CTCAAGCACA CGCTGCAGGC TGCCCGCAAG GTATTTACTT CAGAAGCTAG CATTTACACA ACAGTAATGC CTCCTCGGAT	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC ACCAACAGCA GGCAGCGAGG AGCTCAGCT GTACCACTCG CACTCGAACAG CACTCGAACAG CACTCGAACG CACTCGACGC CACACCAGCT GTTACGCCAC AACGGCAGTG TTACGCCAC CACATCGAT CGGCACTGAT CTGGCAGGG CCACATGAGC CACATCAAC AGGAAGTAAA AGGAATAAC AGCAATTCT
6301 6401 6501 6701 6701 7001 7101 7201 7301 7401 7501 7401 7501 7801 7801 8001 8101 8201 8301 8401 8401 8401 8401 8401 9001 9101 9201 9401 9501 9401	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CTCTTTTCTC CGCCAGACACA CCAGCAGCACA CCAGCACACG GTGACGCAC CTGACGCAC CTGACGCAC CTCAAGCAAC CTCAAGCAAC CTCAAGCAAC CTCAAGCAAC CTCCCGAAC CACCAGCCG CAACGACCAC CACACGACCAC CACACGACCAC CACTCCAGAA CTTCCCGAAC CACCTCGACAC CACTCCAGAA CACCTCGACAC CACTCCAGAA CTTCCCGAAAA ACGTTTCCAG TAACCCGGAAAA AAGTTACTAT CCCGTTTTCT	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCAGCAA CACCATCATC GCGTGGACAG TCCACTCACA CTAGCAACAC CAAAGTCCCC CATCCGCAAG ACGCGCACGC GGTACACCAA ACAGGCTCCAA CGGCTCGCGC GCTCGCGCA GGCACTCGGCAG ACGCTCGGCAG ACGCCCTGCGC GCCCCTGCGC ACGCCACTTCG ACGCACTTCG ACGCCACTCG CACCCCGCCG ACGCCCCGCCG ACGACGACGT AGGCACACCG AGGTACACGA ACGTAGTAAG CCACTCCAAAG CCACTCCAAAG CCACTCCCAAG CATCCCAAAG CCACTCCCAAG CATCCCAAAG CCACTCCCAAG CATCCCAAAG	GTTCAATTAT TCACCCACAC ACCTCTCACC AACTCTCACC AACTGCTGCGC AAATACGTAC TTTTTGCAGA ACGGCAACGC CCACCGACAC CCACCGAAT CCACCGAT CCAACGAGTGA ACATCAAGC GCCGTTCCAG GATGTAACT ACCGCCACGA ACATGTTCAA CGCGGCACCGA GCGCGGCACGA GCGCTGGCACG GCGCTGGCACG GCGCCAAGCG GTCCAATCCG GCGCAGCAGCA GTCCAATCCG GCGCAGCAGCA GTCCAAGTGC GGAAGAGCA ACATGTTGACC GGAAGAAGAG ACGGCACAGA TGGTTAAGAG TTCGGGCACAGT	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATCGCAGACT CAAACGGCTG TCATCGCCCG CTCAGTGGTA ATTCGCAGTG CTCGAGTGGTA CTCGGAACAC ACTCGCACGG GGAAACACT ACCCCCACGA GATCATCGAA GATCCTCGC CAATCCCCAT CAATCCCACG CAATCTCCCC CGGACACCTT CAACGCGCC CGGACACCTT CAACGCGCC CGGCTACCGA GGCGTACCGA CTTATTTCCC GCAGTTTGGG GGAAAAGGA ACGAGCAAAA ATGTTTACGA	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCG GGCGAGTGC AGAGTGACTTTA GCTGGATCCA GTACGGCA CCACCAAAGA TCAGGTGAAG GTACTGCCAC GTACTGCCAC GTACTGCCAC GTACTGCACG GTACTGCACG GCAACGGCA CCCCCCC GCCAACGGCA CCCCCCCC	GGGGGAGGC GAGTGTGAGG CACTCGCGAG AAACGAACA AATGGCGATG ACGATCAAAT AGCAGGACAC TTTTGAGGCA CCACCACCAG AGGTGGAGCC GGAGGACGCC AGGTGGAGCC CACGTCCAGC GCAAGATTG GCAAGTCCC GCAGTCGTCG GCAAGATTG GCAAGACGC TCCAGCGGCT TAAGGGTGAG ACTGCCACT TAAGGGTGAG ACTGCGACACT TGACGACCGC AATGCGCATG TATATGCATG CACGTCGAACA TTGTCTGTAG ATATCGTAGG GCGAAGAACA ATACTTTGG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CCGAACACAC CGAACACAC CGAACACCAC GACGACGCAC GACGACGCAC GGCTCGGACT GCAACCTCC GGCTCGGAC TGCATCTGCG AACCAACTCCG AACCAACTCCG AACCAACTCCG CCAACCCGG GTTCCCGTAC CCGACACCG CGCAGCACCC CGGAGTCGCC CGGAGTCGCC TGCGTTCCGG ATCGGCTTCT CGGGACACAC CGGGACACAC CCAACACGC TGCGTTCCGG ATCGGCTTCT CGGGACACAC CAACAACAC CAACAACAC CAACAACAC CAACAA	ATCGTTCCCA TTTGCCCCT TTGCCCCC GAAACACACA CAGCATGCCA CAGCATGCCA CAGCATCCAT CTCGCGAGGG CGACTACCAT CGCTCACCAG GGAGTACCAT GAGTCTCACA GAGGACTG CGAGTACACAC CAGCTACACAC GTGAACCCCAC CTGAACCCCAC CTGAACCCCAC GCGAGGACTG GCCGATCACCC GTAAGGACGG GGCGATCATC CAGCCTACAC GTAAGGACG GGCGCCAGT TCCTGCACACA TTCCCGCACAT TCCCGCACAT TCCGCACACA TTGGCCGAG TTAGGCGAG TTAGGCGAG TTAGGCGAG TTAGGCGAG CACTCTCTCC	AACGCAGGGA TTTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGCAGCAGC CCACATCATC CCACACCCC CTCGTACGAT GCGGTAACGA ACCATCTCGA CATCAAGCCGG AAACGCCGA ACCTCTTCC ACTTCCAGAT CGAAAGCCG CTGAAAGCAG TGCCTTTAAG ACCCGCGACA GTATTTATTT CAGAAGCTAG GTATTTATTT CAGAGCAGAT CCTCCGGAT CACACGAT	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC ACCAACAGCA GGCAGCGAGG AGCTCAGCT GTACCATCTG CACTCGAACAGC TAAGCACTCG CACTCGACAGC TTACGCCGC CACTCGACAGC TTACGCCGC CACACAGCAGT CTGGAGCGGC GTTACGCCGC CACACTGAT CCACATGGG CGCACTGAT ATTGTAAGA CAAACTAAG AGGAAGTAAA TTTCGCCCAC CACACCACTG
6301 6401 6401 6501 6701 7001 7101 7201 7301 7301 7501 7301 7501 7301 7501 7301 8001 8301 8301 8301 8301 8401 8501 9001 9101 9201 9301 9401 9501 9601	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CGCCAGATCA GCAGCAGCAT TCCTTTTCTC CGCCGCAGA TCGGCTTTAA CCAGCAACACG CTGGAGCACG CTGGAGCACG CTGCACCACC TGTTCCAGTA GAACCTGGAC CTCCAGCAAC CAGCAGGCCG CAACGAGCAG GGCGGCTGAC GAGTCCCGAA TATCACGGAC TACCGAGACC CACCCGAAA AAGTTACTAG GACACCGAAA AAGTTACTAT CCCGTTTTT AACACTAAAT CAGTTACTGG	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCCA ACAGCAGCAAC ACAGCAGCAAC CACCATCATC GCGTGGACAG CTAGCACTCAC CAACGCACCAC AGCGCACCCA GGTACACCGA GGTACACCGA GGCACTCGCG GGTACACCGA CGCTTCGCAGC CATCGGTATC GGCCACTTCG AGGCCACTCG CACCCGGTATC GGCACTCGCG AGGCACACCG AGGCACACCG AGGCACACCG AGGCACACCG AGGCACACCG AGGCACACCG AGGCACACCG AGGCACACCG AGGCACACCG AGGCACACCG AGGCACACGC AGGCACACGC AGGCACACG CACTCGAAG AGTACACAG GCCTCGCAG GCACTCGAAG CAATCCAAAG CAATCCAAAG	GTTCAATTAT TCACCCACAC ACTCTCTCACC ACTCTCTCACC ACCGTGTGGC AAATACGTAC TTTTGCAGA ACGGCACGC TATGGATTCA CCACCGCTGCT CCAACGCAAT CCACGCTGCC GCCGTCCAG GCCGTCCAG GCCGTCCAG GACTGTAACT ACCCGCTACC GCAGCACCAG ACGGCACCAG CGGGGCACCAG GTCCAATCCG GTCCAATCCG GTCCAATCCG GTCCAATCCG GTCCAATCCG GTCCAATCCG GTCCAAGCGC ACATGTGACAGA ACTGGTGGACAG ACTGGTTGACCG GACGACAGAG ACGGCACAGT TGGTTAACAG TCGGCTCACG CACAGCACAG	CGTTTTTCGG ACACACACAT CTCGCAACA ATCGCTGTGT ACCGTTTTAA ATCGCTGTGT CAAACGGTTG TCATCGCCCG CTCAGTGGTA ATTCGGCCCG CTCAGTGGTA CACCGCACGA CTCGCACGG GGAAACACC CGACCCCCA CACTCCCACCA CACTCCCCACCA CACTCTCCCC ACTCTTCCGT CAAACGAACT CACACCCCT CGGACACCTT CAAACGAACT CACATCTCCC CGGACACCTT CAAACGAACT CACATCTCCC CGCACTCCCC CGCACTCCCC CGCACTCCCC CGCACTCCCC CCACTTCCCC CCACTTCCCC CCCACTTCCCC CCCACTTCCCC CCCACTTCCCC CCCCCCCC	GATGGAATAG ACACGCGCAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GCCGGACTGC AGAGTGGATT GCTGATCCCA CCACCAAAGA TCAGGTGAAG GTACGGCA GTACTGCCAC GTACTGCCAC GTACTGCCAC GCCACCACGG GCACCACCGC GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACCG GGACACACAC	GGGGGAGGGT GAGTGTGAGC CACTCGCGAG AAACGACAC CACTCGCGAG AAACGACAAT ACGACGACAC TTTTGAGGCA CCACCACAAT AGCAGGACACC TTTTGAGGCA CGACGACGCC AAGTGCAGCC CAGGTCGAGC GCAGGTCCAAC CCACGTCGTGG GCAAGAATTG GCAAGATGGGCA GCACGTCGTGG GCACGTCGACAC TCCCACCGCGCT CCAGCGGCAC TACGGCACAC TGACGACACG ATGCGCACAG ATGCGCACAG ATGCCGCACA TGACGACACA TATTCGTAGAA CATGTAGAAC TATTCGTAGG CTAGCACACA TTGTCGTAGG CTAGCACACA TTGTCGTAGACA TATTCGTAGG CTGCTCTTAG AGACTTTTAG	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA CGGAACACAC GGCAGCACAC GACGACCACT GACGACGCAC GCAAGCTCGC GCCACGCACTCC GGCTCGGACTCC GCAAGCAGTG AAATCAATGC TGCATCAGCG AACGAAAGCG AACGAAAGCG AACGAACTCCG CACCCTGGG GGAGCGACC CGGAGCGACC CGGAGTCGCC CGGAGCGACC CGGAGCCACC CGGAGCCACC CGGAGCCACC CGGAGCCACC CGGAGCCACC CGGAGCCACC CGGAGCCACC CGGAGCCACC CGGAGCCACC CGCACCACCCG CGCACCACCCG CGGAGCCACC CGGAGCCACC CGGAGCCACC CGAGACCACC CGAGACGACTCCG TTAGGATTAT TTAACTTTC AATTTACTTC AATTTACTTC	ATCGTTCCCA TTTGCCCCT TGTATGTGTG GAAACACAA CAGCATGCCA CACAACAACA CAGCATGCCA CACAACACA CACATCCACT CTCGCGAGGG CAAGTCTCACT CTCACCAGA GTAACGCCGA GAGTTTGTC GAGTACAGCC GAGTACTGCA CAAGTATTGCG AAACCGCAGA GAGTCATCACC CTGAACCTCA AGCTGGCGGG GGGGTCACAC GCGGCTACAC CTCACCACTG CACCATGCC CACCATGCC CACTATGC CACTATGCC AGACTACGCAAG TTGCCCGACAG TTGGCCGACG TTGGCCGACG TTGGCCGACG TTACTGGT CACCACCCCTAC	AACGCAGGGA TTTTTGTCTT CCTGTTTATA ACCATCCATT TGTTTTACTG GTCACAGCAG CCACATCATC ATTCACACTC GTCGGGCGCC CCCACATCATC CCCACACCCC CCCCTCGTACGAT GCGGTACGAT GCGGTACGAT CCCCCCCC CACAAGCCGG AGAACGCCGA ACATCTCCAGAT CGCAAAGCCGG CGCATTTCCAGAT TGCCCGCAAG GTACTTTACGAA ACACTGTGAGCTA CGCAGAGCTAG CTCCCCGAAG CTTTTACACA ACACGAGCACA ACACGCGTGCA CACAGCAGTAT CACAGCAGCACA CCTCCCGGAT TGACGGTGCA CTCCACGATC CTCCCGGAT	AATAAAAAGC CTCGGGACAT TACCGAGGGT ACGGTTACTA ATGGTTTTGT CAGCCGCT CAACCCAACC
6301 6401 6501 6601 7001 7101 7201 7301 7401 7501 7401 7501 7801 7801 8301 8301 8301 8301 8301 8301 8301 8	AATAGTAAAA TGCGATGAAA TTAATTTGGC GTAGCAACAG AAGGGCTTGC CTCTTTTCTC CGCCAGATCA GCAGCAGCAT TCCCCGCAGA TCGGCTTTAA CCAGCAACGC GTGACGCCAC CGCAACCACC TGTACCAGAA CTCCCGAGA CTCCCGACA CTCCCGACA CACCAGAGCAG GGCGGCTGAC GACCAGCAG GGCGCTGAC GACTACCAGA TACCCGAAAA TATCACGGGT TACCGGAAAC CACTCAGAAAA TACCCGAAAA AAGTTTTCAG AAAAACTGA	CACACAGCGC ATAAAAATAA TGTGGGGAGG CAAAAAAAAA CTACTCACAC TTTTCCCCA ACAGCACCAC CACCATCATC GCGTGGACAG TCCACTCACA CTAGCACACAC CAAAGTCCCC CATCCGCAAG ACGCGCACTCCA ACGCGCTCCAA CGGCTCCCAC GGCATGGCA CGGCTCCGAC AGTCGGCAC GGCCACTTCG GGCCACTTCG GGCCACTTCG GGCACGCCG CATCCGACA CGGCCCCGCC ACGACGCCG CACCACTCG CACCACTCG CACCACTCG CACCACTCG CACCACTCG CACCACTCG CACCACTCG CACCACTCG CACCCCACTCG CACCCCACTCG CACCCCCAC CACCCCCAC CACCCCCAC CACCCCCAC CACCCCCAC CACCCCCAC CCCCCC	GTTCAATTAT TCACCCACAC ACTCCACCAC ACTCTCACC ACCTGTGGGC TTTTTGCAGA ACGGCACGCA CCACCGACACGC CCAACCGATTCA CCACCGATCCAG GCCTTCCAG GCCTTCCAG GCATGTAACT ACCCGCTACCG GCAGGCACCAA ACATCAAGCC GCAGGCACCAA ACATGTTCAAC GCGCTGGCACGA GTCCAATCCA GTCCAATCG GCCATCCACG TCGGCGGCACGA CAGGCACCAG TCGGCTGAGCAG ACTGACCAGCA CAGGCACCAG TTGGTTAACG CCCATCGCCT CGCAAGCAGCA ACTGACCAGCA CAGGCACCAG TTGGTTAACG CCCATCGCCC CAAGGAGCACAG TTCGGCCCT CAAGGAGCACAG CCATTTTCCTT	CGTTTTTCGG ACACACACAT TTGCCAAACA ATCGCTGTGT ACCGTTTTAA ATGCAGAACT CCAACGGTTG TCATCGCCCG TCCAGTGGTA ATTCGGGCA CTCAGTGGCA CTCGAAGGCA CTCGCACGG GGAAAACAC ACTCGCACGA GGCACCCAT CAATCCCCAT GAGTCCGCGC CAATCCCCGTT CAACGCGCC CGACACCT CCGACACCT CAACGCGCC CGACACCT CAACGCGCC CGACACCT CAACGCGCC CGACACCT CAACGCGCC CGCACACCT CAACGCGCC CGCACACCT CAACGCCCC CGCACACCT CAACGCCCC CGCACACCT CAACGCCCC CGCACACCT CAACGCGCC CGCACACCT CCCCCCCC CGCACCCCC CGCACACCT CACATCTCCC CCCCCCCCCC	GATGGAATAG ACACGCGAT CATATACTTT GAGTGAATCG GGGGAGGATA TTGACTCCGT GTTCGATTTA GGCGGAGTGC AGAGTGGAT CCACGGCA CCACCAAAGA CCACCAAAGA CTCACGGCAC GTACTGCCAC GTACTGCCAC GTACTGCCAC GTACTGCACAC GTACTGCACAC GCACCACCAC GCCAACGGCA CCCCCCCCC GCGACCACCAC GGACGCACAGG GAGTGCAAGG GAGTGCAAGG GAGTGCAAGG GAGTGCAAGG GAGTGCAAGG GAGTGCAAGG GAGTGCAAGG GAGTGCAAGG GAGTGCAAGG GAATCAAAGA ACTCTTAGC GAATCAAAGA ACTCTTAGC GAATCAAAGA	GGGGGAGGGC GAGTGTGAGG CACTCGCGAG AAACGAACA AATGGCGACC CACTCGCGAG AAACGACAAAA AGCAGGACAC TTTTGAGGCA CCACCACGA GGAGGACGCC AGGTGGAGCC CACGTCCAGC GCAAGACATG GGAAGTGCC CTGCGTCAAC TCGCTCAAGC GCAAGACAGC TCCACGGCA TCCACGGCCA TCCACGGCCA TCCACGGCACA TGACGACCAC TACAGTGAAC TTGCCTGTAA ACTGCGCACA TTGCCTGTAA ACTGCGACAA TTGTCGTAA ACTGCTAAA ATACTTTGG GCGTAGAACA ATACTTTGG TCGTCGTAA ACTGCTAGAACA	TGTGGTAAAG AGCAGCACAT TCTATCACTG GGGAGAAAGT AACATGAACA GCGGCAGCAC CCGAACACCAC GACGACGCAC GACGACGCAC GCAACCTCC GGCTCGGACT GCAACCACTC GGCTCGGACAC TGCATCTGCG AACCAACTCCG AACCAACTCCG GACGACTCCG GAACAACTCC CCGAACACCC CTGAGAAAGCG CTGATGACC CTGAGGACCC CTGAGGACCC CGGAGCCCCC CTGAGGACCC CTGAGGACCC CTGAGGACCC CTGAGGACCC CTGAGGACCC CTGAGGACCC CTGAGGACCC CATAAATAAT AAAAGCAACG GCAAGAACGAT TTAGGATTCT TATGACTTCC	ATCGTTCCCA TTTGCCCCCT TGTATGTGTG GAACACACA CACACACACA CACACACACA CACACACA	AACGCAGGGA TTTTTGTCTT CCTGTTTACTG GTCACAGCAG GTCACAGCAG CCACATCATC ATTCACACTC ATTCACACTC GTCGGCAGCAG CCACATCATC CAGCAGCAGC CCCCACCCC CTCGTACGAT GCGGTACGAT GCGGTACGAT GCACACCCG ACAACCCGG ACAACGCCG ACAACGCCG ACACCGCTCCAGAT GCCGCCACAG CTGCACGCAG CTGCACGGTGA CGCTGCAGGC TGCCCGCAAG GTATTTACTT CAGAAGCTAG CACTTACCACA ACACTACTACCA ACACCGCTGA CACACCGCTGA CGCTGCAGGC TGCCCGCAAG CTCACACACA CCCCCCCAAG CATTTACACA ACACTACTAC CACAGTAATGC CTCCTCGGAT TGACGGTGCA ACACGGTAC CTCACGAACTCC CTCAAGAAATC ACAGTAATCC	AATAAAAAGC CTCGGGACAT TACCGAGGT ACGGTTACTA ATGGTTTTGT CAGCAGCAAA ATCATCCGCT CAACCCACC TACAACCCTC ACCAACAGCA GGCAGCGAGG GGCAGCGAGG ACCTCAGCT GTACCATCTG CACTCGAACAGC CACTCGACAGC CTAACGCCACT GTTTGCCGAC CACACCAGCT GTTGGCAGCG CACATCAGC CACATCGAC CACATGGCGC CACATCAGC CCACATGAGC CCACATGAGC CCACATCAGC CCACATCAAC AGGAAGTAAAG AGGAAGTAAAG AGGTAACTTC CACATCCATTA CCACATCATC CACATCATC CACATCCATTA AGGTAACTCA

Source: GenBank entry AAAB01008979 (563001..573000). Alignment with cDNA sequences: Positions 4367..4440 correspond to the first exon of the *Anopheles hunchback* P1 transcript (SEQ21), and positions 6830..6835 correspond to parts of the second exon of the *Anopheles hunchback* transcript. One putative polyadenylation signal was found in the selected part of the genomic sequence (9594..9599).

# **Bibliography**

- Altschul, S., Boguski, M., Gish, W. and Wootton, J. (1994). Issues in searching molecular sequence databases. *Nat Genet* 6, 119-29.
- Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-402.
- Anderson, D. (1966). The comparative embryology of the Diptera. Annu Rev Entomol 11, 23-46.
- Anderson, D. (1972). The development of holometabolous insects. In *Developmental systems: insects*, (ed. S. Counce and C. Waddington), pp. 165-242. London, New York: Academic Press.
- Andres, A. and Thummel, C. (1994). Methods for quantitative analysis of transcription in larvae and prepupae. *Methods Cell Biol* 44, 565-73.
- Arora, K., Levine, M. and O'Connor, M. (1994). The *screw* gene encodes a ubiquitously expressed member of the TGF-beta family required for specification of dorsal cell fates in the *Drosophila* embryo. *Genes Dev* 8, 2588-601.
- Ashe, H. (2005). BMP signalling: synergy and feedback create a step gradient. Curr Biol 15, R375-7.
- Ashraf, S., Hu, X., Roote, J. and Ip, Y. (1999). The mesoderm determinant *snail* collaborates with related zincfinger proteins to control *Drosophila* neurogenesis. *EMBO J* 18, 6426-38.
- Benson, D., Karsch-Mizrachi, I., Lipman, D., Ostell, J. and Wheeler, D. (2006). GenBank. Nucleic Acids Res 34, D16-20.
- Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M. and Nüsslein-Volhard, C. (1988). The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J* **7**, 1749-56.
- Birnstiel, M., Busslinger, M. and Strub, K. (1985). Transcription termination and 3' processing: the end is in site! *Cell* **41**, 349-59.
- **Boer, P. and Hickey, D.** (1986). The *alpha-amylase* gene in *Drosophila melanogaster*: nucleotide sequence, gene structure and expression motifs. *Nucleic Acids Res* **14**, 8399-411.
- Bonneton, F., Shaw, P., Fazakerley, C., Shi, M. and Dover, G. (1997). Comparison of *bicoid*-dependent regulation of *hunchback* between *Musca domestica* and *Drosophila melanogaster*. *Mech Dev* 66, 143-56.
- Boulay, J., Dennefeld, C. and Alberga, A. (1987). The *Drosophila* developmental gene *snail* encodes a protein with nucleic acid binding fingers. *Nature* 330, 395-8.
- Brady, G. and Iscove, N. (1993). Construction of cDNA libraries from single cells. *Methods Enzymol* 225, 611-23.
- Bucher, G., Scholten, J. and Klingler, M. (2002). Parental RNAi in *Tribolium* (Coleoptera). *Curr Biol* **12**, R85-6.
- Bullock, S., Stauber, M., Prell, A., Hughes, J., Ish-Horowicz, D. and Schmidt-Ott, U. (2004). Differential cytoplasmic mRNA localisation adjusts pair-rule transcription factor activity to cytoarchitecture in dipteran evolution. *Development* 131, 4251-61.
- Calvo, E., Walter, M., Adelman, Z., Jimenez, A., Onal, S., Marinotti, O. and James, A. (2005). Nanos (nos) genes of the vector mosquitoes, Anopheles gambiae, Anopheles stephensi and Aedes aegypti. Insect Biochem Mol Biol 35, 789-98.
- Campos-Ortega, J. and Hartenstein, V. (1997). The embryonic development of *Drosophila melanogaster*. Berlin, Heidelberg, New York: Springer Verlag.

- Carney, G. and Bowen, N. (2004). p24 proteins, intracellular trafficking, and behavior: *Drosophila mela-nogaster* provides insights and opportunities. *Biol Cell* 96, 271-8.
- Casanova, J. and Struhl, G. (1989). Localized surface activity of *torso*, a receptor tyrosine kinase, specifies terminal body pattern in *Drosophila*. *Genes Dev* **3**, 2025-38.
- Cha, B., Koppetsch, B. and Theurkauf, W. (2001). In vivo analysis of *Drosophila bicoid* mRNA localization reveals a novel microtubule-dependent axis specification pathway. *Cell* **106**, 35-46.
- Collins, K. and Wiegmann, B. (2002). Phylogenetic relationships of the lower Cyclorrhapha (Diptera: Brachycera) based on 28S rDNA sequences. *Insect Syst Evol* 33, 445-56.
- Copf, T., Schröder, R. and Averof, M. (2004). Ancestral role of *caudal* genes in axis elongation and segmentation. *Proc Natl Acad Sci U S A* 101, 17711-5.
- Crauk, O. and Dostatni, N. (2005). Bicoid determines sharp and precise target gene expression in the *Droso-phila* embryo. *Curr Biol* **15**, 1888-98.
- Crevel, G. and Cotterill, S. (1995). DF 31, a sperm decondensation factor from *Drosophila melanogaster*: purification and characterization. *EMBO J* 14, 1711-7.
- Crevel, G., Huikeshoven, H. and Cotterill, S. (2001). Df31 is a novel nuclear protein involved in chromatin structure in *Drosophila melanogaster*. J Cell Sci 114, 37-47.
- Cummings, L., Riley, L., Black, L., Souvorov, A., Resenchuk, S., Dondoshansky, I. and Tatusova, T. (2002). Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes. *FEMS Microbiol Lett* 216, 133-8.
- Curtis, D., Apfeld, J. and Lehmann, R. (1995). *nanos* is an evolutionarily conserved organizer of anteriorposterior polarity. *Development* 121, 1899-910.
- Dave, V., Zhao, C., Yang, F., Tung, C. and Ma, J. (2000). Reprogrammable recognition codes in *bicoid* homeodomain-DNA interaction. *Mol Cell Biol* 20, 7673-84.
- Dearden, P. and Akam, M. (1999). Developmental evolution: Axial patterning in insects. Curr Biol 9, R591-4.
- Doyle, H., Harding, K., Hoey, T. and Levine, M. (1986). Transcripts encoded by a homoeo box gene are restricted to dorsal tissues of *Drosophila* embryos. *Nature* 323, 76-9.
- Driever, W. and Nüsslein-Volhard, C. (1988). A gradient of *bicoid* protein in *Drosophila* embryos. *Cell* 54, 83-93.
- **Driever, W. and Nüsslein-Volhard, C.** (1989). The *bicoid* protein is a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. *Nature* **337**, 138-43.
- **Driever, W., Thoma, G. and Nüsslein-Volhard, C.** (1989). Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the *bicoid* morphogen. *Nature* **340**, 363-7.
- Drysdale, R., Crosby, M. and Consortium, F. (2005). FlyBase: genes and gene models. *Nucleic Acids Res* 33, D390-5.
- **Dulac, C. and Axel, R.** (1995). A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195-206.
- Dunham, I., Dewar, K., Kim, U.-J. and Ross, M. (1997). Bacterial Cloning Systems. In *Genome Analysis: A Laboratory Manual*, (ed. B. Birren E. Green and S. Klapholz): Cold Spring Harbor Laboratory Press.
- Elbetieha, A. and Kalthoff, K. (1988). Anterior determinants in embryos of *Chironomus samoensis*: characterization by rescue bioassay. *Development* **104**, 61-75.
- Ephrussi, A., Dickinson, L. and Lehmann, R. (1991). Oskar organizes the germ plasm and directs localization of the posterior determinant *nanos*. *Cell* 66, 37-50.
- Ephrussi, A. and Lehmann, R. (1992). Induction of germ cell formation by oskar. Nature 358, 387-92.

- Falciani, F., Hausdorf, B., Schröder, R., Akam, M., Tautz, D., Denell, R. and Brown, S. (1996). Class 3 Hox genes in insects and the origin of *zen*. *Proc Natl Acad Sci U S A* **93**, 8479-84.
- Ferguson, E. and Anderson, K. (1992). Decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the *Drosophila* embryo. *Cell* **71**, 451-61.
- Finkelstein, R., Smouse, D., Capaci, T., Spradling, A. and 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, 1516-27.
- Fleig, R. and Sander, K. (1988). Honeybee morphogenesis: embryonic cell movements that shape the larval body. *Development* 103, 525-34.
- Foe, V. and Alberts, B. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *J Cell Sci* 61, 31-70.
- Frohnhöfer, H. and Nüsslein-Volhard, C. (1986). Organization of anterior pattern in the *Drosophila* embryo by the maternal gene *bicoid*. *Nature* **324**, 120-5.
- Gabay, L., Seger, R. and Shilo, B. (1997). MAP kinase *in situ* activation atlas during *Drosophila* embryogenesis. *Development* 124, 3535-41.
- Galindo, K. and Smith, D. (2001). A large family of divergent *Drosophila* odorant-binding proteins expressed in gustatory and olfactory sensilla. *Genetics* **159**, 1059-72.
- Gans, M., Audit, C. and Masson, M. (1975). Isolation and characterization of sex-linked female-sterile mutants in *Drosophila melanogaster*. *Genetics* **81**, 683-704.
- Gavis, E. and Lehmann, R. (1992). Localization of nanos RNA controls embryonic polarity. Cell 71, 301-13.
- Gehring, W., Affolter, M. and Bürglin, T. (1994). Homeodomain proteins. Annu Rev Biochem 63, 487-526.
- Goltsev, Y., Hsiong, W., Lanzaro, G. and Levine, M. (2004a). Different combinations of gap repressors for common stripes in *Anopheles* and *Drosophila* embryos. *Dev Biol* 275, 435-46.
- Goltsev, Y., Peterson, B., Hsiong, W., Lanzaro, G. and Levine, M. (2004b). Evolution of gene expression in Diptera. Distinct features of patterning gene expression in *Anopheles gambiae* early embryo. *A Dros Res Conf* 45.
- Gregor, T., Bialek, W., de Ruyter van Steveninck, R., Tank, D. and Wieschaus, E. (2005). Diffusion and scaling during early embryonic pattern formation. *Proc Natl Acad Sci U S A* **102**, 18403-7.
- Grimaldi, D. and Engel, M. (2005). Evolution of the insects. New York: Cambridge University Press.
- Grosskortenhaus, R., Pearson, B., Marusich, A. and Doe, C. (2005). Regulation of temporal identity transitions in *Drosophila* neuroblasts. *Dev Cell* 8, 193-202.
- Hancock, J., Shaw, P., Bonneton, F. and Dover, G. (1999). High sequence turnover in the regulatory regions of the developmental gene *hunchback* in insects. *Mol Biol Evol* **16**, 253-65.
- Handel, K., Grünfelder, C., Roth, S. and Sander, K. (2000). *Tribolium* embryogenesis: a SEM study of cell shapes and movements from blastoderm to serosal closure. *Dev Genes Evol* 210, 167-79.
- Hanes, S. and Brent, R. (1989). DNA specificity of the *bicoid* activator protein is determined by homeodomain recognition helix residue 9. *Cell* 57, 1275-83.
- Hanes, S., Riddihough, G., Ish-Horowicz, D. and Brent, R. (1994). Specific DNA recognition and intersite spacing are critical for action of the *bicoid* morphogen. *Mol Cell Biol* **14**, 3364-75.
- Holt, R. Subramanian, G. Halpern, A. Sutton, G. Charlab, R. Nusskern, D. Wincker, P. Clark, A. Ribeiro, J. Wides, R. et al. (2002). The genome sequence of the malaria mosquito Anopheles gambiae. Science 298, 129-49.
- Houchmandzadeh, B., Wieschaus, E. and Leibler, S. (2002). Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature* **415**, 798-802.

- Hülskamp, M., Lukowitz, W., Beermann, A., Glaser, G. and Tautz, D. (1994). Differential regulation of target genes by different alleles of the segmentation gene *hunchback* in *Drosophila*. *Genetics* **138**, 125-34.
- Hülskamp, M., Pfeifle, C. and Tautz, D. (1990). A morphogenetic gradient of *hunchback* protein organizes the expression of the gap genes *Krüppel* and *knirps* in the early *Drosophila* embryo. *Nature* **346**, 577-80.
- Hülskamp, M., Schröder, C., Pfeifle, C., Jäckle, H. and Tautz, D. (1989). Posterior segmentation of the *Drosophila* embryo in the absence of a maternal posterior organizer gene. *Nature* **338**, 629-32.
- Irish, V., Lehmann, R. and Akam, M. (1989). The Drosophila posterior-group gene nanos functions by repressing hunchback activity. Nature 338, 646-8.
- Isshiki, T., Pearson, B., Holbrook, S. and Doe, C. (2001). *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* **106**, 511-21.
- Jiménez, G., Guichet, A., Ephrussi, A. and Casanova, J. (2000). Relief of gene repression by *torso* RTK signaling: role of capicua in *Drosophila* terminal and dorsoventral patterning. *Genes Dev* 14, 224-31.
- Johannsen, O. and Butt, F. (1941). Embryology of insects and myriapods. New York: McGraw-Hill Book Company, Inc.
- Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C. and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*: II. Zygotic loci on the third chromosome. *Roux's Arch Dev Biol* 193, 283-95.
- Kalthoff, K. (1979). Analysis of a morphogenetic determinant in an insect embryo (*Smittia* spec., Chironomidae, Diptera). In *Determinants of spatial organization*, (ed. S. Subtelny and I. Konigsberg), pp. 79-126. New York: Academic Press.
- Kalthoff, K. (1983). Cytoplasmic determinants in dipteran eggs. In *Time, space, and pattern in embryonic development*, (ed. W. Jeffery and R. Raff), pp. 313-348. New York: Alan R Liss, Inc.
- Kalthoff, K. and Sander, K. (1968). Der Entwicklungsgang der Mißbildung "Doppelabdomen" im partiell UVbestrahlten Ei von *Smittia parthenogenetica* (Dipt., Chironomidae). *Wilhelm Roux' Arch* **161**, 129-46.
- Kambadur, R., Koizumi, K., Stivers, C., Nagle, J., Poole, S. and Odenwald, W. (1998). Regulation of POU genes by castor and *hunchback* establishes layered compartments in the *Drosophila* CNS. *Genes Dev* 12, 246-60.
- Kandler-Singer, I. and Kalthoff, K. (1976). RNase sensitivity of an anterior morphogenetic determinant in an insect egg (*Smittia* sp., Chironomidae, Diptera). *Proc Natl Acad Sci U S A* **73**, 3739-43.
- Karpen, G. and Spradling, A. (1992). Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics* 132, 737-53.
- Kim-Ha, J., Smith, J. and Macdonald, P. (1991). *oskar* mRNA is localized to the posterior pole of the *Drosophila* oocyte. *Cell* 66, 23-35.
- Klingler, M., Soong, J., Butler, B. and Gergen, J. (1996). Disperse versus compact elements for the regulation of runt stripes in *Drosophila*. *Dev Biol* 177, 73-84.
- Kosman, D., Mizutani, C., Lemons, D., Cox, W., McGinnis, W. and Bier, E. (2004). Multiplex detection of RNA expression in *Drosophila* embryos. *Science* **305**, 846.
- Kramer, P. (2000). cDNA library construction from single cells. Current Protocols In Neuroscience.
- Lehmann, R. and Nüsslein-Volhard, C. (1986). Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of *oskar*, a maternal gene in *Drosophila*. *Cell* **47**, 141-52.
- Lehmann, R. and Nüsslein-Volhard, C. (1987a). *hunchback*, a gene required for segmentation of an anterior and posterior region of the *Drosophila* embryo. *Dev Biol* **119**, 402-17.
- Lehmann, R. and Nüsslein-Volhard, C. (1987b). Involvement of the *pumilio* gene in the transport of an abdominal signal in the *Drosophila* embryo. *Nature* **329**, 167-70.

- Lehmann, R. and Nüsslein-Volhard, C. (1991). The maternal gene *nanos* has a central role in posterior pattern formation of the *Drosophila* embryo. *Development* **112**, 679-91.
- Lehmann, R. and Tautz, D. (1994). In situ hybridization to RNA. Methods Cell Biol 44, 575-98.
- Lewis, E. (1978). A gene complex controlling segmentation in Drosophila. Nature 276, 565-70.
- Li, Y., Brown, S., Hausdorf, B., Tautz, D., Denell, R. and Finkelstein, R. (1996). Two orthodenticle-related genes in the short-germ beetle *Tribolium castaneum*. Dev Genes Evol 206, 35-45.
- Liang, Z. and Biggin, M. (1998). Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in *Drosophila*. *Development* 125, 4471-82.
- Liu, P. and Kaufman, T. (2005). Short and long germ segmentation: unanswered questions in the evolution of a developmental mode. *Evol Dev* 7, 629-46.
- Liu, Y. and Montell, D. (2001). Jing: a downstream target of slbo required for developmental control of border cell migration. *Development* 128, 321-30.
- Lukowitz, W., Schröder, C., Glaser, G., Hülskamp, M. and Tautz, D. (1994). Regulatory and coding regions of the segmentation gene *hunchback* are functionally conserved between *Drosophila virilis* and *Drosophila melanogaster*. *Mech Dev* **45**, 105-15.
- Lynch, J., Brent, A., Leaf, D., Pultz, M. and Desplan, C. (2006). Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp Nasonia. Nature 439, 728-32.
- MacDonald, P. (1990). bicoid mRNA localization signal: phylogenetic conservation of function and RNA secondary structure. Development 110, 161-71.
- Macdonald, P. (1992). The *Drosophila pumilio* gene: an unusually long transcription unit and an unusual protein. *Development* 114, 221-32.
- Macdonald, P. and Struhl, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature* **324**, 537-45.
- Marcey, D., Watkins, W. and Hazelrigg, T. (1991). The temporal and spatial distribution pattern of maternal exuperantia protein: evidence for a role in establishment but not maintenance of *bicoid* mRNA localization. *EMBO J* 10, 4259-66.
- Margolis, J., Borowsky, M., Shim, C. and Posakony, J. (1994). A small region surrounding the distal promoter of the *hunchback* gene directs maternal expression. *Dev Biol* 163, 381-8.
- Margolis, J., Borowsky, M., Steingrímsson, E., Shim, C., Lengyel, J. and Posakony, J. (1995). Posterior stripe expression of *hunchback* is driven from two promoters by a common enhancer element. *Development* 121, 3067-77.
- Markstein, M., Markstein, P., Markstein, V. and Levine, M. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* **99**, 763-8.
- McAlpine, J. (1989). Phylogeny and classification of the Muscomorpha. In *Manual of Nearctic Diptera*, vol. 3 (ed. J. McAlpine and D. Wood), pp. 1397-1518. Ottawa, Canada: Agriculture Canada Monograph.
- McGregor, A. (2005). How to get ahead: the origin, evolution and function of *bicoid*. *Bioessays* 27, 904-13.
- McGregor, A., Shaw, P. and Dover, G. (2001a). Sequence and expression of the *hunchback* gene in *Lucilia sericata*: a comparison with other Dipterans. *Dev Genes Evol* 211, 315-8.
- McGregor, A., Shaw, P., Hancock, J., Bopp, D., Hediger, M., Wratten, N. and Dover, G. (2001b). Rapid restructuring of *bicoid*-dependent *hunchback* promoters within and between Dipteran species: implications for molecular coevolution. *Evol Dev* **3**, 397-407.

- **Mlodzik, M., Fjose, A. and Gehring, W.** (1985). Isolation of *caudal*, a *Drosophila* homeo box-containing gene with maternal expression, whose transcripts form a concentration gradient at the pre-blastoderm stage. *EMBO J* **4**, 2961-2969.
- **Mlodzik, M. and Gehring, W.** (1987a). Expression of the *caudal* gene in the germ line of *Drosophila*: formation of an RNA and protein gradient during early embryogenesis. *Cell* **48**, 465-78.
- **Mlodzik, M. and Gehring, W.** (1987b). Hierarchy of the genetic interactions that specify the anteroposterior segmentation pattern of the *Drosophila* embryo as monitored by *caudal* protein expression. *Development* **101**, 421-35.
- **Mohler, J.** (1977). Developmental genetics of the *Drosophila* egg. I. Identification of 59 sex-linked cistrons with maternal effects on embryonic development. *Genetics* **85**, 259-72.
- Moses, K., Ellis, M. and Rubin, G. (1989). The *glass* gene encodes a zinc-finger protein required by *Drosophila* photoreceptor cells. *Nature* **340**, 531-6.
- Moulton, J. and Wiegmann, B. (2004). Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta). *Mol Phylogenet Evol* **31**, 363-78.
- Moussian, B. and Roth, S. (2005). Dorsoventral axis formation in the *Drosophila* embryo--shaping and transducing a morphogen gradient. *Curr Biol* 15, R887-99.
- Murata, Y. and Wharton, R. (1995). Binding of pumilio to maternal *hunchback* mRNA is required for posterior patterning in *Drosophila* embryos. *Cell* 80, 747-56.
- Niessing, D., Driever, W., Sprenger, F., Taubert, H., Jäckle, H. and Rivera-Pomar, R. (2000). Homeodomain position 54 specifies transcriptional versus translational control by Bicoid. *Mol Cell* 5, 395-401.
- Novotny, T., Eiselt, R. and Urban, J. (2002). Hunchback is required for the specification of the early sublineage of neuroblast 7-3 in the *Drosophila* central nervous system. *Development* **129**, 1027-36.
- Nüsslein-Volhard, C. (1979). Maternal effect mutations that alter the spatial coordinates of the embryo of *Drosophila melanogaster*. In *Determinants of Spatial Organisation.*, (ed. S. Subtelney and I. Konigsberg), pp. 185-211. New York: Academic Press.
- Nüsslein-Volhard, C., Frohnhöfer, H. and Lehmann, R. (1987). Determination of anteroposterior polarity in Drosophila. Science 238, 1675-81.
- Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Droso-phila*. *Nature* 287, 795 801.
- Nüsslein-Volhard, C., Wieschaus, E. and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*: I. Zygotic loci on the second chromosome. *Roux's Arch Dev Biol* 193, 267-282.
- Orr, H. (2005). The genetic theory of adaptation: a brief history. Nat Rev Genet 6, 119-27.
- Padgett, R., St Johnston, R. and Gelbart, W. (1987). A transcript from a *Drosophila* pattern gene predicts a protein homologous to the transforming growth factor-beta family. *Nature* 325, 81-4.
- Pankratz, M. and Jäckle, H. (1993). Blastoderm segmentation. In *The development of Drosophila mela-nogaster*, vol. 1 (ed. M. Bate and A. Martinez-Arias), pp. 467-516. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Paroush, Z., Wainwright, S. and Ish-Horowicz, D. (1997). Torso signalling regulates terminal patterning in Drosophila by antagonising Groucho-mediated repression. Development 124, 3827-34.
- Patel, N., Hayward, D., Lall, S., Pirkl, N., DiPietro, D. and Ball, E. (2001). Grasshopper *hunchback* expression reveals conserved and novel aspects of axis formation and segmentation. *Development* **128**, 3459-72.
- Perondini, A., Gutzeit, H. and Sander, K. (1987). Double abdomen induction by UV in *Bradysia tritici* (syn. *Sciara ocellaris*, Sciaridae): sensitive stages and conditions for photoreversal. *Roux's Arch Dev Biol* 196, 268-72.

- Perrimon, N., Mohler, D., Engstrom, L. and Mahowald, A. (1986). X-linked female-sterile loci in Drosophila melanogaster. Genetics 113, 695-712.
- Pignoni, F., Baldarelli, R., Steingrímsson, E., Diaz, R., Patapoutian, A., Merriam, J. and Lengyel, J. (1990). The *Drosophila* gene *tailless* is expressed at the embryonic termini and is a member of the steroid receptor superfamily. *Cell* 62, 151-63.
- Pultz, M., Westendorf, L., Gale, S., Hawkins, K., Lynch, J., Pitt, J., Reeves, N., Yao, J., Small, S., Desplan, C. et al. (2005). A major role for zygotic hunchback in patterning the Nasonia embryo. Development 132, 3705-15.
- **Raff, R.** (1996). The shape of life: Genes, development, and the evolution of animal form. Chicago: The University of Chicago Press.
- Rafiqi, A. M., Lemke, S., Ferguson, S., Stauber, M. and Schmidt-Ott, U. (in preparation).
- Raftery, L. and Sutherland, D. (2003). Gradients and thresholds: BMP response gradients unveiled in *Droso-phila* embryos. *Trends Genet* 19, 701-8.
- Ray, R., Arora, K., Nüsslein-Volhard, C. and Gelbart, W. (1991). The control of cell fate along the dorsalventral axis of the *Drosophila* embryo. *Development* 113, 35-54.
- Reichert, H. and Simeone, A. (1999). Conserved usage of gap and homeotic genes in patterning the CNS. *Curr Opin Neurobiol* 9, 589-95.
- Rivera-Pomar, R. and Jäckle, H. (1996). From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet* 12, 478-83.
- Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. and Jäckle, H. (1996). RNA binding and translational suppression by *bicoid*. *Nature* 379, 746-9.
- Rohr, K., Tautz, D. and Sander, K. (1999). Segmentation gene expression in the mothmidge *Clogmia albipunctata* (Diptera, psychodidae) and other primitive dipterans. *Dev Genes Evol* 209, 145-54.
- Rosenberg, U., Schröder, C., Preiss, A., Kienlin, A., Côté, S., Riede, I. and Jäckle, H. (1986). Structural homology of the product of the *Drosophila* Krüppel gene with *Xenopus* transcription factor IIIA. *Nature* 319, 336-9.
- Roth, S., Stein, D. and Nüsslein-Volhard, C. (1989). A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell* 59, 1189-202.
- Rubin, G. and Spradling, A. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218, 348-53.
- Rushlow, C., Colosimo, P., Lin, M., Xu, M. and Kirov, N. (2001). Transcriptional regulation of the *Drosophila* gene *zen* by competing Smad and Brinker inputs. *Genes Dev* **15**, 340-51.
- Rushlow, C., Doyle, H., Hoey, T. and Levine, M. (1987). Molecular characterization of the *zerknüllt* region of the Antennapedia gene complex in *Drosophila*. *Genes Dev* 1, 1268-79.
- Rushlow, C., Han, K., Manley, J. and Levine, M. (1989). The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in *Drosophila*. *Cell* **59**, 1165-77.
- Rushlow, C. and Levine, M. (1990). Role of the *zerknüllt* gene in dorsal-ventral pattern formation in *Drosophila*. Adv Genet 27, 277-307.
- Sambrook, J. and Russel, D. (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor: Laboratory Press.
- Sander, K. (1976). Specification of the basic body pattern in insect embryogenesis. *Adv Insect Physiol* 12, 125-238.
- Schmidt, O., Zissler, D., Sander, K. and Kalthoff, K. (1975). Switch in pattern formation after puncturing the anterior pole of *Smittia* eggs (Chironomidae, Diptera). *Dev Biol* 46, 216-21.

- Schmidt-Ott, U. (2000). The amnioserosa is an apomorphic character of cyclorrhaphan flies. *Dev Genes Evol* 210, 373-6.
- Schmidt-Ott, U. (2005). Insect serosa: a head line in comparative developmental genetics. Curr Biol 15, R245-7.
- Schmidt-Ott, U., Sander, K. and Technau, K. (1994). Expression of *engrailed* in embryos of a beetle and five dipteran species with special reference to the terminal regions. *Roux's Arch Dev Biol* 203, 298-303.
- Schröder, C., Tautz, D., Seifert, E. and Jäckle, H. (1988). Differential regulation of the two transcripts from the *Drosophila* gap segmentation gene *hunchback*. *EMBO J* **7**, 2881-7.
- Schröder, R. (2003). The genes *orthodenticle* and *hunchback* substitute for *bicoid* in the beetle *Tribolium*. *Nature* 422, 621-5.
- Schröder, R. and Sander, K. (1993). A comparison of transplantable *bicoid* activity and partial *bicoid* homeobox sequences in several *Drosophila* and blowfly species (Calliphoridae). *Roux's Arch Dev Biol* 203, 34-43.
- Schwalm, F. (1987). Insect Morphogenesis. Basel: S. Karger AG.
- Schüpbach, T. and Wieschaus, E. (1986). Maternal-effect mutations altering the anterior-posterior pattern of the *Drosophila* embryo. *Roux's Arch Dev Biol* **195**, 302-17.
- Schüpbach, T. and Wieschaus, E. (1989). Female sterile mutations on the second chromosome of *Drosophila melanogaster*. I. Maternal effect mutations. *Genetics* 121, 101-17.
- Seeger, M. and Kaufman, T. (1990). Molecular analysis of the *bicoid* gene from *Drosophila* pseudoobscura: identification of conserved domains within coding and noncoding regions of the *bicoid* mRNA. *EMBO J* 9, 2977-87.
- Shaw, P., Salameh, A., McGregor, A., Bala, S. and Dover, G. (2001). Divergent structure and function of the *bicoid* gene in Muscoidea fly species. *Evol Dev* 3, 251-62.
- Shaw, P., Wratten, N., McGregor, A. and Dover, G. (2002). Coevolution in *bicoid*-dependent promoters and the inception of regulatory incompatibilities among species of higher Diptera. *Evol Dev* **4**, 265-77.
- Shinmyo, Y., Mito, T., Matsushita, T., Sarashina, I., Miyawaki, K., Ohuchi, H. and Noji, S. (2005). *caudal* is required for gnathal and thoracic patterning and for posterior elongation in the intermediate-germband cricket *Gryllus bimaculatus*. *Mech Dev* **122**, 231-9.
- Simpson-Brose, M., Treisman, J. and Desplan, C. (1994). Synergy between the *hunchback* and *bicoid* morphogens is required for anterior patterning in *Drosophila*. *Cell* **78**, 855-65.
- Smith, K. (1969). Diptera: Lonchopteridae. London: Royal Entomological Society.
- Sommer, R., Retzlaff, M., Goerlich, K., Sander, K. and Tautz, D. (1992). Evolutionary conservation pattern of zinc-finger domains of *Drosophila* segmentation genes. *Proc Natl Acad Sci U S A* **89**, 10782-6.
- Sommer, R. and Tautz, D. (1991). Segmentation gene expression in the housefly *Musca domestica*. *Development* **113**, 419-30.
- Sonoda, J. and Wharton, R. (1999). Recruitment of Nanos to *hunchback* mRNA by Pumilio. *Genes Dev* 13, 2704-12.
- Spencer, F., Hoffmann, F. and Gelbart, W. (1982). Decapentaplegic: a gene complex affecting morphogenesis in *Drosophila melanogaster*. Cell 28, 451-61.
- Sprenger, F., Stevens, L. and Nüsslein-Volhard, C. (1989). The Drosophila gene torso encodes a putative receptor tyrosine kinase. Nature 338, 478-83.
- Sprenger, F., Trosclair, M. and Morrison, D. (1993). Biochemical analysis of *torso* and D-raf during *Droso-phila* embryogenesis: implications for terminal signal transduction. *Mol Cell Biol* **13**, 1163-72.
- St Johnston, D., Beuchle, D. and Nüsslein-Volhard, C. (1991). *Staufen*, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell* 66, 51-63.

- St Johnston, D., Driever, W., Berleth, T., Richstein, S. and Nüsslein-Volhard, C. (1989). Multiple steps in the localization of *bicoid* RNA to the anterior pole of the *Drosophila* oocyte. *Development* **107** Suppl, 13-9.
- St Johnston, D. and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68, 201-19.
- Stathopoulos, A. and Levine, M. (2002). Dorsal gradient networks in the *Drosophila* embryo. *Dev Biol* 246, 57-67.
- Stauber, M., Jäckle, H. and 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**, 3786-9.
- Stauber, M., Prell, A. and Schmidt-Ott, U. (2002). A single Hox3 gene with composite *bicoid* and *zerknüllt* expression characteristics in non-Cyclorrhaphan flies. *Proc Natl Acad Sci U S A* **99**, 274-9.
- Stauber, M., Taubert, H. and Schmidt-Ott, U. (2000). Function of *bicoid* and *hunchback* homologs in the basal cyclorrhaphan fly *Megaselia* (Phoridae). *Proc Natl Acad Sci U S A* **97**, 10844-9.
- Steward, R. (1987). Dorsal, an embryonic polarity gene in *Drosophila*, is homologous to the vertebrate protooncogene, c-rel. *Science* 238, 692-4.
- Steward, R. (1989). Relocalization of the *dorsal* protein from the cytoplasm to the nucleus correlates with its function. *Cell* 59, 1179-88.
- Steward, R., McNally, F. and Schedl, P. (1984). Isolation of the dorsal locus of Drosophila. Nature 311, 262-5.
- Strimmer, K. and von Haeseler, A. (1996). Quartet Puzzling: A Quartet Maximum-Likelihood Method for Reconstructing Tree Topologies. *Mol Biol Evol* 13, 964-9.
- Struhl, G. (1989). Differing strategies for organizing anterior and posterior body pattern in *Drosophila* embryos. *Nature* **338**, 741-4.
- Struhl, G., Struhl, K. and Macdonald, P. (1989). The gradient morphogen *bicoid* is a concentration-dependent transcriptional activator. *Cell* 57, 1259-73.
- Tautz, D. (1988). Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal morphogenetic centres. *Nature* **332**, 281-4.
- Tautz, D., Lehmann, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jäckle, H. (1987). Finger protein of novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. *Nature* 327, 383-9.
- Tautz, D. and Pfeifle, C. (1989). A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. *Chromosoma* 98, 81-5.
- Thisse, C., Perrin-Schmitt, F., Stoetzel, C. and Thisse, B. (1991). Sequence-specific transactivation of the *Drosophila twist* gene by the dorsal gene product. *Cell* 65, 1191-201.
- Thummel, C. and Pirrotta, V. (1992). New pCaSpeR P element vectors. Dros Inf Serv 71, 150.
- Treier, M., Pfeifle, C. and Tautz, D. (1989). Comparison of the gap segmentation gene hunchback between Drosophila melanogaster and Drosophila virilis reveals novel modes of evolutionary change. EMBO J 8, 1517-25.
- Treisman, J. and Desplan, C. (1989). The products of the *Drosophila* gap genes *hunchback* and *Krüppel* bind to the *hunchback* promoters. *Nature* **341**, 335-7.
- Treisman, J., Gönczy, P., Vashishtha, M., Harris, E. and Desplan, C. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell* 59, 553-62.
- Vogt, R., Prestwich, G. and Lerner, M. (1991). Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. *J Neurobiol* 22, 74-84.

- Wakimoto, B., Turner, F. and Kaufman, T. (1984). Defects in embryogenesis in mutants associated with the *antennapedia* gene complex of *Drosophila melanogaster*. *Dev Biol* **102**, 147-72.
- Wang, C. and Lehmann, R. (1991). Nanos is the localized posterior determinant in *Drosophila*. *Cell* 66, 637-47.
- Wharton, R. and Struhl, G. (1991). RNA regulatory elements mediate control of *Drosophila* body pattern by the posterior morphogen *nanos*. *Cell* 67, 955-67.
- Wieschaus, E., Nüsslein-Volhard, C. and Jürgens, G. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*: III. Zygotic loci on the X-chromosome and fourth chromosome. *Roux's Arch Dev Biol* 193, 296-307.
- Wimmer, E., Carleton, A., Harjes, P., Turner, T. and Desplan, C. (2000). Bicoid-independent formation of thoracic segments in *Drosophila*. Science 287, 2476-9.
- Wolff, C., Schröder, R., Schulz, C., Tautz, D. and Klingler, M. (1998). Regulation of the *Tribolium* homologues of *caudal* and *hunchback* in *Drosophila*: evidence for maternal gradient systems in a short germ embryo. *Development* 125, 3645-54.
- Wolff, C., Sommer, R., Schröder, R., Glaser, G. and Tautz, D. (1995). Conserved and divergent expression aspects of the *Drosophila* segmentation gene *hunchback* in the short germ band embryo of the flour beetle *Tribolium. Development* 121, 4227-36.
- Yajima, H. (1964). Studies on embryonic determination of the harlequin-fly, *Chironomus dorsalis*. II. Effects of partial irradiation of the egg by ultra-violett light. J Embryol Exp Morphol 12, 89-100.
- Yeates, D. and Wiegmann, B. (1999). Congruence and controversy: toward a higher-level phylogeny of Diptera. *Annu Rev Entomol* 44, 397-428.
- Zdobnov, E., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R., Christophides, G., Thomasova, D., Holt, R., Subramanian, G. et al. (2002). Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. Science 298, 149-59.
- Zhao, C., Dave, V., Yang, F., Scarborough, T. and Ma, J. (2000). Target selectivity of *bicoid* is dependent on nonconsensus site recognition and protein-protein interaction. *Mol Cell Biol* 20, 8112-23.

# CV

Steffen Joachim Lemke

1983-1987	Grundschule Am Engelnberg, Wuppertal-Elberfeld
1988-1996	Gymnasium Am Kothen, Wuppertal-Barmen
1996-1997	Zivildienst
1997-2000	Undergraduate studies at the Georg August Universität Göttingen
1999	Vordiplom in Molecular Genetics, Plant Physiology, Organic Chemistry and Physical Chemis-
	try
2000-2006	Graduate studies in the MSc/PhD Molecular Biology Program
	at the Georg August Universität Göttingen.
2002	MSc with Urs Schmidt-Ott: Department of Developmental Biology (Herbert Jäckle) at the
	Max-Planck-Institut für biophysikalische Chemie, Göttingen, Germany. Topic: Initial Analysis
	of the Evolution of Bicoid-dependent hunchback Regulation.
2001-2006	PhD with Urs Schmidt-Ott: Department of Developmental Biology (Herbert Jäckle) at the Max-
	Planck-Institut für biophysikalische Chemie, Göttingen, Germany. Since 2003: Department of
	Organismal Biology and Anatomy, The University of Chicago, Chicago, IL, USA. Topic: Evo-
	lution of the Bicoid-dependent <i>hunchback</i> regulation in Diptera.

Meeting Presentations

Botanikertagung, 2000, Jena, Germany. Lemke, S., Reumann, S. "Identification of Novel Proteins from Plant Peroxisomes by Bioinformatic Analyses"

44th Annual Drosophila Research Conference, 2003, Chicago, IL. Lemke, S. J., Prell, A. H., Stauber, M., Schmidt-Ott, U. "Evolution of transcriptional control of the *Drosophila* gap gene *hunchback*"

45th Annual Drosophila Research Conference, 2004, Washington, DC. Lemke, S. J., Rafiqi, A. M., Prell, A. H., Stauber, M., Schmidt-Ott, U. "Evolution of transcriptional control of the *Drosophila* gap gene *hunchback*"

Publications

- Reumann, S., Ma, C., <u>Lemke, S</u>., Babujee, L. (2004). AraPerox. A database of putative *Arabidopsis* proteins from plant peroxisomes. Plant Physiol. 2004 136: 2587-608.
- Walther, K., Krauss, M., Diril, M. K., <u>Lemke, S.</u>, Ricotta, D., Honing, S., Kaiser, S., Haucke, V. (2001). Human stoned B interacts with AP-2 and synaptotagmin and facilitates clathrin-coated vesicle uncoating. EMBO Rep. 2: 634-640.

CV