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:


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 hunchback 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 hunchback 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
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Results</td>
<td>24</td>
</tr>
<tr>
<td>3.1</td>
<td>Identification of <em>bicoid</em> orthologues from <em>Platypeza</em> and <em>Lonchoptera</em></td>
<td>24</td>
</tr>
<tr>
<td>3.2</td>
<td>Subtractive screening for <em>bicoid</em>-like genes: a new method</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Subtractive screening for <em>bicoid</em>-like genes in <em>Clogmia</em></td>
<td>27</td>
</tr>
<tr>
<td>3.4</td>
<td>Cloning of dipteran <em>hunchback</em> genes</td>
<td>29</td>
</tr>
<tr>
<td>3.5</td>
<td>Functional comparison of early dipteran <em>hunchback</em> regulation</td>
<td>31</td>
</tr>
<tr>
<td>3.6</td>
<td>Expression studies in <em>Episyrphus</em></td>
<td>33</td>
</tr>
<tr>
<td>3.6.1</td>
<td><em>Episyrphus hunchback</em> shares expression characteristics of cyclorrhaphan and non-cyclorrhaphan dipterans</td>
<td>33</td>
</tr>
<tr>
<td>3.6.2</td>
<td>The extraembryonic anlage of <em>Episyrphus</em> extends to the anterior pole</td>
<td>34</td>
</tr>
<tr>
<td>3.6.3</td>
<td><em>Episyrphus orthodenticle</em> is not expressed in pre-blastoderm embryos</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>Discussion</td>
<td>37</td>
</tr>
<tr>
<td>4.1</td>
<td>Do all cyclorrhaphan dipterans have <em>bicoid</em>?</td>
<td>37</td>
</tr>
<tr>
<td>4.2</td>
<td>Does <em>Episyrphus</em> reflect the primitive patterning mechanism of cyclorrhaphan flies?</td>
<td>38</td>
</tr>
<tr>
<td>4.3</td>
<td>How did early anterior <em>hunchback</em> regulation change in dipteran evolution?</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Summary and Conclusions</td>
<td>40</td>
</tr>
<tr>
<td>A</td>
<td>Appendix</td>
<td>41</td>
</tr>
<tr>
<td>A.1</td>
<td>Figures and Tables</td>
<td>41</td>
</tr>
<tr>
<td>A.1.1</td>
<td>Figures</td>
<td>41</td>
</tr>
<tr>
<td>A.1.2</td>
<td>Tables</td>
<td>59</td>
</tr>
<tr>
<td>A.2</td>
<td>Supplemental Figures</td>
<td>64</td>
</tr>
<tr>
<td>A.3</td>
<td>Sequences</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>79</td>
</tr>
</tbody>
</table>
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 *Drosophila* 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 Diptera, 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 plasma-membrane folds inwards between the nuclei, the syncytial blastoderm turns into a cellular blastoderm (Anderson, 1966), although in *Drosophila melanogaster* 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 germ band (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 germ band 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 germ band re-

---

1 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 \(^2\) (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) and represses its translation (Rivera-Pomar et al., 1996). Thus, Bicoid induces a Caudal gradient complementary to the Bicoid gradient (Macdonald and Struhl, 1986; Mlodzik and Gehring, 1987a; 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-

\(^2\) 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 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-β (TGF-β) protein (Padgett et al., 1987), which, together with the TGF-β 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 (Schmidt-Ott, 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_{2}H_{2} 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; Staub er 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* homo-
logue 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)⁺ 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
Introduction

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-
Introduction

I

tempt to map the emergence of bicoid to this transition, previous studies have covered a vari-
ety 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 non-
cyclorrhaphan dipterans, I developed a new screening method for anterior localized transcripts
and explored the non-cyclorrhaphan Clogmia albipunctata (Psychodidae, Figure 1). This di-
rect approach failed to identify an anterior localized transcript in Clogmia, and a bicoid homo-
logue in the cyclorrhaphan Episyrphus balteatus (Syrphidae, Figure 1). Thus, to complement
this quest for the potential hunchback activator in non-cyclorrhaphan dipterans and Episyr-
phus, I have also generated reporter constructs to directly compare the early regulation of dip-
teran hunchback homologues in transgenic Drosophila. This approach also extends previous
work, in which the regulation of hunchback homologues from Drosophila virilis, Musca, Cal-
liphora, 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 hunch-
back homologues from four basal cyclorrhaphans (Episyrphus, Megaselia, Platypeza and Lon-
choptera), and four non-cyclorrhaphan dipterans (Empis livida, Haematopota pluvialis, Clog-
mia, and Anopheles), which represent mostly paraphyletic dipteran branches (Figure 1). The
results provide additional support for Bicoid-dependent hunchback regulation in flies with bi-
coid, 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 ex-
pression 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 *Epi-syrphus 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 mosquitoes) 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+ 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 µg of total RNA was extracted from 50 to 100 µl fly tissue with RNAwiz (Ambion) according to the manual. Poly A+ RNA was enriched using the Oligotex mRNA midi kit (Qiagen), with an average yield of about 2 µg poly A+ RNA. Enriched poly A+ RNA was used to prepare the cDNA according to the user manuals. The Anopheles 5’ SMART
RACE cDNA template was prepared from 1.2 µg 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

*Platypera bicoid* was amplified by PCR on genomic DNA with the degenerate primer pair 5’-YTGGGYMMAGCYCAGGTSAARATWTGGTT/5’-TYTTBGGGYGTAYAHGGYT-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 *Platypera bicoid* cDNA, 5’ and 3’ rapid amplification of cDNA ends (RACEs) were performed (Table 1), and the products were cloned into pCRIII-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 *Platypera 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 *Platypoza*, *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* were designed based on the published sequence of *Megaselia hunchback* (Stauber *et al.*, 2000); primers to isolate *Anopheles hunchback* cDNA were designed based on the published genome sequence of *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’-CTGCATTGAATCCCGTTCTGC). 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 *Episyphus 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 *NotI* fragment in front of the *Drosophila* hsp43 basal promoter 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 *SphI*-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 *NotI* and cloned into the *NotI* 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 *NotI* 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’-CTCGTAGCTAGCTGGCATAGCGCGTGAAGTGC). The product was cloned into pGEM-T Easy, then excised with *NotI* and cloned into the *NotI* 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, in-
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’-CGGCACAACGATACTGATACACAGAA) 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 NotI 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’- GACGCGTTCGATTAACCGGATATAAA) and a primer specific for the second exon of the transcript immediately upstream of the ORF (5’-TTCAATTTAACTGCGATGGAGGC). 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 NotI 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-GTCATGTCTGATCTTATA) and a primer specific for the second exon of the transcript at the start of the ORF (5’-ACTATTAATTGCTGTGTGTGATTC). 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 NotI and cloned into the NotI site in front of the minimal *even-skipped* promoter 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 NotI 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-GATGGGTATGTTTGTAG) 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 NotI and subcloned in the NotI 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 range PCR from the phage DNA using a gene-specific primer (5’-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’-TGGCTTAGATATAGTCATTACCT/5’-ATCTCGAGTGACTGAAAGAATAGAAA) 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-CGGAGTCAAAAGTTC) 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 NotI and subcloned into the NotI site of pCaSpeR-hsp43-lacZ, yielding 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 µl 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 Na₂HPO₄, 2 mM KH₂PO₄, pH 7.4), and 500 µl 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/µl. 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 *Episyphus 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 µl of cDNA lysis buffer (1X MMLV buffer [Invitrogen] with 0.5% NP40 [USB], containing 24 µM pd(T)₄₄ [IDT], 0.2 U/µl SuperRNaseIn [Ambion], 0.3 U/µl RNA-guard [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 µl 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)₄₄ oligonucleotide at room temperature (22 °C) for 2 minutes. First-strand 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 NotI site within the linker region (AL1mod: 5’-AGCGGCGCGGAATCC(T)). 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+ 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 μCi/μl α[P32] 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 µl 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 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 hour-old *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
RESULTS

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 colonies 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.*).
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_{2}H_{2} 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 cytoplasm. CG6459 like hybridized exclusively to the anterior fraction of one preparation but not at all to the cDNA pools
RESULTS

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 sev-
eral 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 promoter (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 promoter (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, Haematopota, Clognia, 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 Clognia, 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, Haematopota and Clognia 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.
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
**RESULTS**

*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 Episyrsphus hunchback throughout early development and compared it to both lower and higher dipterans. In the freshly laid egg, the maternal transcripts of Episyrsphus hunchback are evenly distributed (Figure 14 A) but disappear from the posterior half during blastoderm formation (Figure 14 B, C). A distinct increase of Episyrsphus 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, Episyrsphus 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 Episyrsphus 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, Episyrsphus 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 Episyrsphus extends to the anterior pole

To further explore the hypothesis that Episyrsphus 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* 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 (*orthodentine*) 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 non-cyclorrhaphan 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.,
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 ashi-zan 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 bicaud and culicomorphan mosquitoes, where a bicaud-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 bicaud 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 *zerknüllt* at the anterior pole of the blastoderm and the expression of *hunchback* in the *zerknüllt* 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 *zerknüllt* 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.
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 comprises about 125,000 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 Pco BCD are shaded red; amino acids identical with the homeodomain of Llu BCD are shaded green. Amino acids identical with the homeodomains of both, Pco BCD and Llu BCD, are shaded yellow.
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 Platypedia (this work); Llu HB, Hunchback of Lonchopetra (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).
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 are based on cDNA sequences, for Empis, only the N-terminal portion of 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 polyadenylation 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.
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, Platyleza, 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.
Figure 12. Functional analysis of *hunchback* regulatory DNA by reporter gene expression in *Drosophila* embryos. For each dipteran species, *hunchback* regulatory DNA upstream of the ORF was cloned in front of a reporter gene and integrated into the genome of *Drosophila*. Unless mentioned otherwise, expression of the reporter gene was assayed by whole-mount *in situ* hybridization at two time points: onset of cellularization (respectively left panel), and mid-cellularization (respectively right panel). Embryos are shown in lateral view, unless indicated otherwise; anterior is to the left. (A, B) *Drosophila*, endogenous *hunchback* expression. (C, D) *Megaselia*, endogenous *hunchback* expression. (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 *Platypoza 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 germ band 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 zerknült*. Whole-mount *in situ* hybridizations showing embryos at (A) pre-blastoderm, (B-D) syncytial blastoderm, (E) the onset of gastrulation, (F) the beginning of germ band extension, and (J-L) during germ band 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.
**Figure 17. Protein alignment of Episyrphus Orthodontic 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.
Table 1. cDNA isolation of dipteran bicaud and orthodenticle homologues: templates, primers and products. cDNA has been isolated by PCR on cDNA prepared with SMART RACE cDNA Amplification Kit.

<table>
<thead>
<tr>
<th>Homologue</th>
<th>Template*</th>
<th>Primer pair†</th>
<th>Length (bp)</th>
<th>Sequences§</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Platypeza bicaud</strong></td>
<td>5'</td>
<td>Pco-SMART 5' cDNA (l)</td>
<td>GTGGCGCGCTTAGATTCGAAAGTGAG/10xUPM</td>
<td>1568</td>
</tr>
<tr>
<td></td>
<td>5'</td>
<td>Pco-SMART 5' cDNA (l)</td>
<td>GTGGCGCGCTTAGATTCGAAAGTGAG/10xUPM</td>
<td>ne: CTGGAAACTTTCGAGTGCA/GCAGG/10xUPM</td>
</tr>
<tr>
<td></td>
<td>3'</td>
<td>Pco-SMART 3' cDNA (l)</td>
<td>CGAGTTCGGACCAAGGTCAAC/10xUPM</td>
<td>1568</td>
</tr>
<tr>
<td></td>
<td>5'</td>
<td>Pco-SMART 5' cDNA (l)</td>
<td>CAAAAATAGGCGCAATGTT/TTGATGCAAGTG/10xUPM</td>
<td>1568</td>
</tr>
<tr>
<td><strong>Lonchoptera bacoid</strong></td>
<td>5'</td>
<td>Llu-SMART 5' cDNA (a)</td>
<td>GTTACCTGATACTCAGCTAGAG/10xUPM</td>
<td>2338</td>
</tr>
<tr>
<td></td>
<td>3'</td>
<td>Llu-SMART 3' cDNA (a)</td>
<td>AAACCACTTAAATATCTAGCTCAGCTAGAG/10xUPM</td>
<td>ne: AAACCACTTAAATATCTAGCTCAGCTAGAG/10xUPM</td>
</tr>
<tr>
<td></td>
<td>3'</td>
<td>Llu-SMART 3’ cDNA (a)</td>
<td>TCAGGACACATCGCGATAGCTAGC/10xUPM</td>
<td>2338</td>
</tr>
<tr>
<td><strong>Episyrphus orthodente</strong></td>
<td>5'</td>
<td>Eba-SMART 5' cDNA (e: 0-5 hrs) M&amp;S</td>
<td>CATCTAAATTGGCGCGCTGTTGTAGATG/10xUPM</td>
<td>1603</td>
</tr>
<tr>
<td></td>
<td>3'</td>
<td>Eba-SMART 3’ cDNA (e: 0-5 hrs) M&amp;S</td>
<td>CATCTAACGACGGCGATAGCAGAGAG/10xUPM</td>
<td>1603</td>
</tr>
</tbody>
</table>

* Origin of the tissue for mRNA isolation: adult females (a), larvae (l) and embryos (e). The age of the embryos in hours at 25 °C is indicated following the colon. † Primer sequences in 5’-3’ direction. ne, nested RACE; 10xUPM/NUP, adaptor primers of SMART RACE Kit. § All sequences have been listed in the Appendix A.2. M&S Cloning by Ab. Matteen Rafiqi. M Cloning/preparation by Michael Stauber.
Table 2. cDNA isolation of dipteran hunchback homologues: templates, primers and products. cDNA has been isolated by PCR on cDNA prepared with SMART RACE cDNA Amplification Kit, or on cDNA prepared with Marathon cDNA Amplification Kit, respectively. The cDNA of Clogmia hunchback has been isolated from a maternal Lambda-ZAP cDNA library.

<table>
<thead>
<tr>
<th>Homologue</th>
<th>Template*</th>
<th>Primer pair†</th>
<th>Length (bp)</th>
<th>Sequences§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episyphus hunchback</td>
<td>5' <em>Eba</em>-SMART 5' cDNA (e: 0-5 hrs) MS</td>
<td>GATACACCGACGAGTGTGACTTCC/10xUPM ne: AGCCCTGGTGGAGTAAGTGGATTA/NUP</td>
<td>2876</td>
<td>SEQ04 (P1)</td>
</tr>
<tr>
<td></td>
<td>5' <em>Eba</em>-SMART 5' cDNA (e: 0-5 hrs) MS</td>
<td>GGGAAATATTTTAAACCGCACCAGAAGTGGATTA/NUP</td>
<td>372</td>
<td>SEQ05 (P2)</td>
</tr>
<tr>
<td></td>
<td>5' <em>Eba</em>-SMART 3' cDNA (e: 0-5 hrs) MS</td>
<td>GGCGGAAATATTTTAAGCCCTGGTGGAGTAAGTGGATTA/NUP</td>
<td>1052</td>
<td>SEQ06 (P3)</td>
</tr>
<tr>
<td>Megaselia hunchback</td>
<td>5' <em>Mab</em>-SMART 5' cDNA (e: 0.5-4 hrs) MS</td>
<td>ATCCACATGCGAAACGCTTTG/10xUPM ne: AGCCCTGGTGGAGTAAGTGGATTA/NUP</td>
<td>873</td>
<td>SEQ08 (P2)</td>
</tr>
<tr>
<td>Platypeza hunchback</td>
<td>5' <em>Pco</em>-SMART 5' cDNA (l)</td>
<td>ATGACGCGAGTGTGACTTCC/10xUPM ne: AGCCCTGGTGGAGTAAGTGGATTA/NUP</td>
<td>2106</td>
<td>SEQ10 (P1)</td>
</tr>
<tr>
<td></td>
<td>5' <em>Pco</em>-SMART 3' cDNA (l)</td>
<td>CGGACGCGAGTGTGACTTCCGGTGCCCGATTAAC/UPM ne: GATACACCGACGAGTGTGACTTCC/NUP</td>
<td>485</td>
<td>SEQ11 (P2)</td>
</tr>
<tr>
<td>Lonchoptera hunchback</td>
<td>5' <em>Llu</em>-SMART 5' cDNA (a)</td>
<td>TAACTTGTATGTTGCTGGGAAATGATGCA/10xUPM ne: GATACACCGACGAGTGTGACTTCC/NUP</td>
<td>3119</td>
<td>SEQ13 (P1)</td>
</tr>
<tr>
<td>Empis hunchback</td>
<td>3' <em>Llu</em>-SMART 3' cDNA (a)</td>
<td>ATGACGCGAGTGTGACTTCCGGTGCCCGATTAAC/UPM ne: GATACACCGACGAGTGTGACTTCC/NUP</td>
<td>1477</td>
<td>SEQ15 (P1)</td>
</tr>
<tr>
<td>Haematopota hunchback</td>
<td>5' <em>Hpl</em>-Marathon cDNA (o) (Stauber et al., 2002)</td>
<td>GGTTTCGAGCCATCATGTTGCTGGGAAATGATGCA/10xUPM ne: GATACACCGACGAGTGTGACTTCC/NUP</td>
<td>2451</td>
<td>SEQ17 (P1)</td>
</tr>
<tr>
<td>Clostridia hunchback</td>
<td>5' <em>Hpl</em>-Marathon cDNA (o) (Stauber et al., 2002)</td>
<td>ATGACGCGAGTGTGACTTCCGGTGCCCGATTAAC/UPM ne: GATACACCGACGAGTGTGACTTCC/NUP</td>
<td>2802</td>
<td>SEQ19 (P1)</td>
</tr>
<tr>
<td>Anopheles hunchback</td>
<td>5' <em>Age</em>-SMART 5' cDNA (a)</td>
<td>TACCATGCGCTGATCGGCTGGTGCCCGATTAAC/UPM ne: GATACACCGACGAGTGTGACTTCC/NUP</td>
<td>78</td>
<td>SEQ21 (P1)</td>
</tr>
</tbody>
</table>

* Origin of the tissue for mRNA isolation: adult females (a), ovaries (o), larvae (l) and embryos (e). The age of the embryos in hours at 25 °C is indicated following the colon. † Primer sequences in 5'-3' direction, ne, nested RACE; AP1/AP2, adaptor primers of Marathon Kit; 10xUPM/NUP, adaptor primers of SMART RACE Kit. § All sequences have been listed in the Appendix A.3. Cloning/preparation by Michael Stauber. Cloning by Alexander Prell. The position of the first exon relative to the ORF is indicated in parentheses for cDNAs of transcripts with alternative 5' UTRs: The transcript with the first exon most proximal to the start of the ORF has been assigned to “P1”, the next proximal “P2”, etc.
Table 3. RNA probes for whole mount *in situ* hybridization.

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

* 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 comprehensible 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.

<table>
<thead>
<tr>
<th>cDNA</th>
<th>Putative molecular function</th>
<th>Expect value <em>(Anopheles)</em></th>
<th>Expect value <em>(Drosophila)</em></th>
<th>Degree of conservation</th>
<th>Abundance in screen</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Obp99a like</em></td>
<td>pheromone binding</td>
<td>( 2 \times 10^{-13} )</td>
<td>( 5 \times 10^{-10} )</td>
<td>+++</td>
<td>12 (5/7)</td>
</tr>
<tr>
<td><em>exu like</em></td>
<td>RNA localization</td>
<td>( 6 \times 10^{-6} )</td>
<td>( 2 \times 10^{-3} )</td>
<td>++</td>
<td>4 (-/-/4)</td>
</tr>
<tr>
<td><em>Df31 like</em></td>
<td>chromatin remodeling</td>
<td>( 5 \times 10^{-2} )</td>
<td>( 7 \times 10^{-3} )</td>
<td>++</td>
<td>1 (1/-/1)</td>
</tr>
<tr>
<td><em>CG14764 like</em></td>
<td>unknown</td>
<td>( 5 \times 10^{-12} )</td>
<td>( 7 \times 10^{-10} )</td>
<td>+++</td>
<td>1 (1/-/1)</td>
</tr>
<tr>
<td><em>CG1967 like</em></td>
<td>post-Golgi transport</td>
<td>( 2 \times 10^{-30} )</td>
<td>( 3 \times 10^{-32} )</td>
<td>+++</td>
<td>1 (1/-/1)</td>
</tr>
<tr>
<td><em>CG6459 like</em></td>
<td>mitochondrial</td>
<td>( 7 \times 10^{-5} )</td>
<td>2.8</td>
<td>+</td>
<td>7 (7/-/1)</td>
</tr>
<tr>
<td><em>jing like</em></td>
<td>transcription factor</td>
<td>1.7</td>
<td>4.5</td>
<td>-</td>
<td>1 (1/-/1)</td>
</tr>
<tr>
<td><em>His3:CG31613 like</em></td>
<td>histone</td>
<td>( 2 \times 10^{-32} )</td>
<td>( 1 \times 10^{-31} )</td>
<td>+++</td>
<td>1 (-/-/1)</td>
</tr>
<tr>
<td><em>Graf like</em></td>
<td>Rho GTPase</td>
<td>no hit</td>
<td>4.9</td>
<td>-</td>
<td>4 (3/1)</td>
</tr>
<tr>
<td><em>Amy-d like</em></td>
<td>sugar metabolism</td>
<td>no hit</td>
<td>4.9</td>
<td>-</td>
<td>6 (6/-/1)</td>
</tr>
</tbody>
</table>

Chosen for analyses of transcript localization, based on putative molecular function and high degree of conservation.
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.

<table>
<thead>
<tr>
<th>Species</th>
<th>P1 leader</th>
<th>P1 intron</th>
<th><em>hunchback</em> fragment</th>
<th>Lines</th>
<th>Blastoderm expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episyrphus</td>
<td>-5839 to -5277</td>
<td>-5276 to -39</td>
<td>-5382 to +24</td>
<td>4 (4)</td>
<td>yes</td>
</tr>
<tr>
<td>Megaselia</td>
<td>-1117 to -907</td>
<td>-906 to -20</td>
<td>-8000 to +2868</td>
<td>2 (2)MS</td>
<td>yes</td>
</tr>
<tr>
<td>Platypeza</td>
<td>-2035 to -1765</td>
<td>-1764 to -12 (P2)</td>
<td>6173 to +39</td>
<td>3 (4)</td>
<td>yes</td>
</tr>
<tr>
<td>Lonchoptera</td>
<td>-5119 to -4565</td>
<td>-4564 to -88</td>
<td>-640 to +958</td>
<td>2 (3)</td>
<td>yes</td>
</tr>
<tr>
<td>Empis</td>
<td>-5882 to -5450</td>
<td>-5449 to -10</td>
<td>-5892 to +105</td>
<td>3</td>
<td>no</td>
</tr>
<tr>
<td>Haematopota</td>
<td>-1937 to -1646</td>
<td>-1645 to -15</td>
<td>-9000 to +105</td>
<td>3 (2)</td>
<td>yes</td>
</tr>
<tr>
<td>Clogmia</td>
<td>-3050 to -2730</td>
<td>-2729 to -9</td>
<td>-6872 to -3</td>
<td>3 (3)AP</td>
<td>yes</td>
</tr>
<tr>
<td>Anopheles</td>
<td>-2464 to -2393</td>
<td>-2392 to -2</td>
<td>-5205 to +31</td>
<td>4</td>
<td>no</td>
</tr>
</tbody>
</table>

* estimate, fragment was not completely sequenced. °S Cloning and fly lines established by Michael Stauber. °AP 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+ 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 anterior-most zerknüllt expression. Arrows hint at the anterior-most 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'-ATCTAATCC) 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 SEQ09. (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 bicond, cDNA.**

```
| 1 | AGTTCAGGTC | CGCGCACAGA | ACCGACGCTT | GACATGATT | TCACACAGCA | TCACACAGCA |
| 101 | ACAATACCC | TCAAAACTG | CAATTTCGCA | GCGTTCCCG | GACCCGCTGC | GATTTTCCG |
| 201 | TATCATGCTT | ATCTATGGCA | ATCGATGCA | ATCGACACAA | ACTGATGCA | ATCGATGCA |
| 301 | AGATGACTT | CGCAAATGTT | ATCGGATCC | ATCGGATCC | ATCGGATCC | ATCGGATCC |
| 401 | ACAATATAT | AGATGACTT | CGCAAATGTT | ATCGGATCC | ATCGGATCC | ATCGGATCC |
| 501 | TAAATCTGCA | GAGGCTCCAG | TGAAGACGTT | GTCCAGAGTG | GTATTAGACG | ATACAGGCT |
| 601 | CGAGGACGCT | ATCCTACGAC | TATCGGATTT | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |
| 701 | TCTGCTCTAC | TCACACAGCA | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |
| 801 | CTATGACAA | ACACTCAAG | TATCGGATTT | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |
| 901 | ATCTGACGG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |
| 1001 | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |
| 1101 | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |
| 1201 | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |
| 1301 | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |
| 1401 | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |
| 1501 | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG | ATGTTTGGTG |

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 bicond, cDNA.**

```
| 1 | TGGCAAGTG | AGTAAATTGT | TTATTTGTTA | GTTACAAGTC | ATGTTTGGTG | ATGTTTGGTG |
| 101 | AGATGACTT | CGCAAATGTT | ATCGGATCC | ATCGGATCC | ATCGGATCC | ATCGGATCC |
| 201 | TGGCAAGTG | AGTAAATTGT | TTATTTGTTA | GTTACAAGTC | ATGTTTGGTG | ATGTTTGGTG |
| 301 | AGATGACTT | CGCAAATGTT | ATCGGATCC | ATCGGATCC | ATCGGATCC | ATCGGATCC |
| 401 | TGGCAAGTG | AGTAAATTGT | TTATTTGTTA | GTTACAAGTC | ATGTTTGGTG | ATGTTTGGTG |
| 501 | TGGCAAGTG | AGTAAATTGT | TTATTTGTTA | GTTACAAGTC | ATGTTTGGTG | ATGTTTGGTG |
| 601 | TGGCAAGTG | AGTAAATTGT | TTATTTGTTA | GTTACAAGTC | ATGTTTGGTG | ATGTTTGGTG |
| 701 | TGGCAAGTG | AGTAAATTGT | TTATTTGTTA | GTTACAAGTC | ATGTTTGGTG | ATGTTTGGTG |
| 801 | TGGCAAGTG | AGTAAATTGT | TTATTTGTTA | GTTACAAGTC | ATGTTTGGTG | ATGTTTGGTG |
| 901 | TGGCAAGTG | AGTAAATTGT | TTATTTGTTA | GTTACAAGTC | ATGTTTGGTG | ATGTTTGGTG |
| 1001 | TGGCAAGTG | AGTAAATTGT | TTATTTGTTA | GTTACAAGTC | ATGTTTGGTG | ATGTTTGGTG |

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 | GATTTCATTA | CTTTGCGGAG | TCTTTGATCA | GCTTCAAGCA | GCTAATAAAAT | GCTAATAAAAT |
| 101 | ATTTTGATAC | TTATTTATGT | GCCGCGCGCA | GGTGATATTA | ATGAAATAAT | ATGAAATAAT |
| 201 | ATTTTGATAC | TTATTTATGT | GCCGCGCGCA | GGTGATATTA | ATGAAATAAT | ATGAAATAAT |
| 301 | ATTTTGATAC | TTATTTATGT | GCCGCGCGCA | GGTGATATTA | ATGAAATAAT | ATGAAATAAT |
| 401 | ATTTTGATAC | TTATTTATGT | GCCGCGCGCA | GGTGATATTA | ATGAAATAAT | ATGAAATAAT |
| 501 | ATTTTGATAC | TTATTTATGT | GCCGCGCGCA | GGTGATATTA | ATGAAATAAT | ATGAAATAAT |
| 601 | ATTTTGATAC | TTATTTATGT | GCCGCGCGCA | GGTGATATTA | ATGAAATAAT | ATGAAATAAT |
| 701 | ATTTTGATAC | TTATTTATGT | GCCGCGCGCA | GGTGATATTA | ATGAAATAAT | ATGAAATAAT |
| 801 | ATTTTGATAC | TTATTTATGT | GCCGCGCGCA | GGTGATATTA | ATGAAATAAT | ATGAAATAAT |
| 901 | ATTTTGATAC | TTATTTATGT | GCCGCGCGCA | GGTGATATTA | ATGAAATAAT | ATGAAATAAT |

Source: Three independent genomic DNA templates (1..374; 49..608), 3’ RACE product (866..1568) and an additional PCR product (38..1539), all amplified from a larval cDNA template of adult females.
Source: 5’ RACE product (1.763) and 3’ RACE product (740..1603), amplified from an embryonic cDNA template.

SEQ04  *Episyrphus hunchback*, cDNA, P1 transcript.

<table>
<thead>
<tr>
<th>1</th>
<th>AGCTTGTTTG</th>
<th>GAGACGTACA</th>
<th>AACCAAGGAG</th>
<th>ATGCTGCTCA</th>
<th>TCGGCTGTTA</th>
<th>GTAATTATTT</th>
<th>TCAATATCTA</th>
<th>AAATATTAAA</th>
<th>TTATTAATAT</th>
<th>ACATATATTTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>TACAAAAGGC</td>
<td>AATCTCTGAT</td>
<td>ATCTACATG</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
</tr>
<tr>
<td>301</td>
<td>CCAACCAAGG</td>
<td>CAGCTTGAG</td>
<td>CATGATACA</td>
<td>TTTAAAATAA</td>
<td>AAACACTTTT</td>
<td>GTTATATAG</td>
<td>TTATTAATAT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
</tr>
<tr>
<td>401</td>
<td>ATATTTATCT</td>
<td>AAGAATAGG</td>
<td>TGTATATATG</td>
<td>ATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>501</td>
<td>TGCCTAAGG</td>
<td>GCAGCCGACG</td>
<td>ACCATCTCA</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>601</td>
<td>GTACCTGAG</td>
<td>TGGTCCGTAC</td>
<td>GCACAGGACC</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>701</td>
<td>ACCACCTCA</td>
<td>CACCAAGCAG</td>
<td>ATGCTGCTCA</td>
<td>TCGGCTGTTA</td>
<td>GTAATTATTT</td>
<td>TCAATATCTA</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
<td></td>
</tr>
<tr>
<td>801</td>
<td>AGGCTACTG</td>
<td>AAAAAATAGG</td>
<td>GCTGGCAGG</td>
<td>ATGCTGCTCA</td>
<td>TCGGCTGTTA</td>
<td>GTAATTATTT</td>
<td>TCAATATCTA</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
</tr>
<tr>
<td>901</td>
<td>GCATCTGAG</td>
<td>TGGTCCGTAC</td>
<td>GCACAGGACC</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>1001</td>
<td>GTTTTCTTAC</td>
<td>CACCAAGCAG</td>
<td>ATGCTGCTCA</td>
<td>TCGGCTGTTA</td>
<td>GTAATTATTT</td>
<td>TCAATATCTA</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
<td></td>
</tr>
<tr>
<td>1101</td>
<td>ATATTTATCT</td>
<td>AAGAATAGG</td>
<td>TGTATATATG</td>
<td>ATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>1201</td>
<td>TGCCTAAGG</td>
<td>GCAGCCGACG</td>
<td>ACCATCTCA</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>1301</td>
<td>GTACCTGAG</td>
<td>TGGTCCGTAC</td>
<td>GCACAGGACC</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>1401</td>
<td>GCATCTGAG</td>
<td>TGGTCCGTAC</td>
<td>GCACAGGACC</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>1501</td>
<td>GTTTTCTTAC</td>
<td>CACCAAGCAG</td>
<td>ATGCTGCTCA</td>
<td>TCGGCTGTTA</td>
<td>GTAATTATTT</td>
<td>TCAATATCTA</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
<td></td>
</tr>
<tr>
<td>1601</td>
<td>ATATTTATCT</td>
<td>AAGAATAGG</td>
<td>TGTATATATG</td>
<td>ATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>1</th>
<th>GAAGCTTCTA</th>
<th>AATTGTGTTC</th>
<th>CTTCGCTGAC</th>
<th>ATAAAATTAA</th>
<th>ATATACATC</th>
<th>TCCGGACTAA</th>
<th>CAGAAAGCGG</th>
<th>TCTCAACGCTC</th>
<th>AAACATGGCA</th>
<th>TTCAATTACG</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>CCAGAGCCCA</td>
<td>ATTTACTTCC</td>
<td>TATACATGCA</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>201</td>
<td>AGGCTACTG</td>
<td>AAAAAATAGG</td>
<td>GCTGGCAGG</td>
<td>ATGCTGCTCA</td>
<td>TCGGCTGTTA</td>
<td>GTAATTATTT</td>
<td>TCAATATCTA</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
</tr>
<tr>
<td>301</td>
<td>ATATTTATCT</td>
<td>AAGAATAGG</td>
<td>TGTATATATG</td>
<td>ATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>401</td>
<td>TGCCTAAGG</td>
<td>GCAGCCGACG</td>
<td>ACCATCTCA</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>501</td>
<td>GTACCTGAG</td>
<td>TGGTCCGTAC</td>
<td>GCACAGGACC</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>601</td>
<td>GCATCTGAG</td>
<td>TGGTCCGTAC</td>
<td>GCACAGGACC</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>701</td>
<td>GCAGCTGCTC</td>
<td>CATGATACA</td>
<td>TTTAAAATAA</td>
<td>AAACACTTTT</td>
<td>GTTATATAG</td>
<td>TTATTAATAT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
<td></td>
</tr>
<tr>
<td>801</td>
<td>AGGCTACTG</td>
<td>AAAAAATAGG</td>
<td>GCTGGCAGG</td>
<td>ATGCTGCTCA</td>
<td>TCGGCTGTTA</td>
<td>GTAATTATTT</td>
<td>TCAATATCTA</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
</tr>
<tr>
<td>901</td>
<td>GCAGCTGCTC</td>
<td>CATGATACA</td>
<td>TTTAAAATAA</td>
<td>AAACACTTTT</td>
<td>GTTATATAG</td>
<td>TTATTAATAT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
<td></td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th>1</th>
<th>AGCTGGCTAG</th>
<th>GTGCTCTTTC</th>
<th>TTTTAAAAC</th>
<th>CCAGACATTC</th>
<th>GCTGCCTTGT</th>
<th>GCGAGAAAA</th>
<th>ATTTAAATTA</th>
<th>CGAAGAGAAT</th>
<th>TGCAAGAGAT</th>
<th>TTATTTGTTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>CAGGAGCCCA</td>
<td>ATTTACTTCC</td>
<td>TATACATGCA</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>201</td>
<td>AGGCTACTG</td>
<td>AAAAAATAGG</td>
<td>GCTGGCAGG</td>
<td>ATGCTGCTCA</td>
<td>TCGGCTGTTA</td>
<td>GTAATTATTT</td>
<td>TCAATATCTA</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
</tr>
<tr>
<td>301</td>
<td>ATATTTATCT</td>
<td>AAGAATAGG</td>
<td>TGTATATATG</td>
<td>ATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>401</td>
<td>TGCCTAAGG</td>
<td>GCAGCCGACG</td>
<td>ACCATCTCA</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>501</td>
<td>GTACCTGAG</td>
<td>TGGTCCGTAC</td>
<td>GCACAGGACC</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>601</td>
<td>GCATCTGAG</td>
<td>TGGTCCGTAC</td>
<td>GCACAGGACC</td>
<td>TATCTATCTA</td>
<td>TACATCTGCA</td>
<td>CCTGCATCTC</td>
<td>TCTGGAGGA</td>
<td>AAATATATT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
</tr>
<tr>
<td>701</td>
<td>GCAGCTGCTC</td>
<td>CATGATACA</td>
<td>TTTAAAATAA</td>
<td>AAACACTTTT</td>
<td>GTTATATAG</td>
<td>TTATTAATAT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
<td></td>
</tr>
<tr>
<td>801</td>
<td>AGGCTACTG</td>
<td>AAAAAATAGG</td>
<td>GCTGGCAGG</td>
<td>ATGCTGCTCA</td>
<td>TCGGCTGTTA</td>
<td>GTAATTATTT</td>
<td>TCAATATCTA</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
</tr>
<tr>
<td>901</td>
<td>GCAGCTGCTC</td>
<td>CATGATACA</td>
<td>TTTAAAATAA</td>
<td>AAACACTTTT</td>
<td>GTTATATAG</td>
<td>TTATTAATAT</td>
<td>AAATATTAAA</td>
<td>TTATTAATAT</td>
<td>ACATATATTTT</td>
<td></td>
</tr>
</tbody>
</table>

Source: 5’ RACE product, amplified from an embryonic cDNA template.
SEQ 08 Megaselia hunchback, cDNA, partial 2nd transcript.

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 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.

Source: Two different PCR products (1.532 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).

APPENDIX
Source: Genomic DNA of phage Mab-hb ph2a, which was partially sequenced in a subcloned SphI (1..2309) and an XhoI (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..4801 correspond to the first exon of the P1 transcript (SEQ10). Positions 1927..4801 correspond to the first exon of the P1 transcript (SEQ11). Positions 1927..4801 correspond to the first exon of the P1 transcript (SEQ12).

SEQ10 Platypeza hunchback, cDNA, P1 transcript.

<table>
<thead>
<tr>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic DNA of phage Mab-hb ph2a, which was partially sequenced in a subcloned SphI (1..2309) and an XhoI (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..4801 correspond to the first exon of the P1 transcript (SEQ10). Positions 1927..4801 correspond to the first exon of the P1 transcript (SEQ11). Positions 1927..4801 correspond to the first exon of the P1 transcript (SEQ12).</td>
</tr>
</tbody>
</table>

SEQ11 Platypeza hunchback, cDNA, partial P2 transcript.

<table>
<thead>
<tr>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic DNA of phage Mab-hb ph2a, which was partially sequenced in a subcloned SphI (1..2309) and an XhoI (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..4801 correspond to the first exon of the P1 transcript (SEQ10). Positions 1927..4801 correspond to the first exon of the P1 transcript (SEQ11). Positions 1927..4801 correspond to the first exon of the P1 transcript (SEQ12).</td>
</tr>
</tbody>
</table>

SEQ12 Platypeza hunchback, genomic.

<table>
<thead>
<tr>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic DNA of phage Mab-hb ph2a, which was partially sequenced in a subcloned SphI (1..2309) and an XhoI (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..4801 correspond to the first exon of the P1 transcript (SEQ10). Positions 1927..4801 correspond to the first exon of the P1 transcript (SEQ11). Positions 1927..4801 correspond to the first exon of the P1 transcript (SEQ12).</td>
</tr>
</tbody>
</table>
Source: Genomic DNA, phage Pco-1. 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 position of the second exon (SEQ11). Positions 6403..8399 are presumably common to the second

---

**APPENDIX**

71
Source: Two PCR products (1.4557, 2726..5470) amplified from different 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.

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.

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 PI transcript (SEQ15), and positions 5854..5942 correspond to parts of the second exon of the Empis hunchback transcript.

SEQ17 Haematopota hunchback, cDNA transcript.

Source: OpenAlign2 sequence alignments.

APPENDIX
SEQ18 *Haematopota hunchback*, genomic.

| Source: | Genomic DNA, phage Hpl-hb. 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 one putative NRE sequence. Positions 2801..1091 correspond to the first exon of the *hunchback* genomic DNA, phage Hpl.
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Source:</td>
<td>One putative NRE sequence (5022..5043) was identified.</td>
</tr>
</tbody>
</table>
SEQ19 Clogmia hunchback, cDNA, P1 transcript.

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

SEQ20 Clogmia hunchback, genomic.

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

Source: 5' RACE product, amplified from a cDNA sample of adult adult female Anopheles mosquito.

SEQ22 Anopheles hunchback, genomic.

Source: 1 Clogmia hunchback, cDNA, partial P1 transcript.
Bibliography


BIBLIOGRAPHY


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 Chemistry
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.


Meeting Presentations

Publications