An evolutionary perspective on germ cell specification genes in insects

An evolutionary perspective on germ cell specification genes in insects

A dissertation presented

by

Benjamin Ewen-Campen

to

The Department of Organismic and Evolutionary Biology

In partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biology

Harvard University
Cambridge, Massachusetts

April 2014

Dissertation Advisor: Professor Cassandra Extavour Benjamin Ewen-Campen

An evolutionary perspective on germ cell specification genes in insects

ABSTRACT

This dissertation investigates the embryonic specification of a specific group of cells: the germ cells. Germ cells, which give rise to sperm and egg, are the only cells in sexually-reproducing animals that directly contribute hereditary information to the next generation. Germ cells are therefore a universal cell type across animals, and represent a profound novelty that likely arose near the base of the animal phylogeny. Yet despite their conserved, essential function in all animals, there is surprising diversity in the mechanisms that specify these cells during embryonic development. In this dissertation, I address the diversity of germ cell specification mechanisms in insects. I focus on two species, the milkweed bug *Oncopeltus fasciatus* (Hemiptera) and the cricket *Gryllus bimaculatus* (Orthoptera), which both branch basally to the Holometabola (those insects which undergo metamorphosis, including the well-studied fruit fly *Drosophila melanogaster*), and thus provide important phylogenetic breadth to our understanding of germ cell specification across insects. Using functional genetic approaches, I show that germ cell specification in both *Oncopeltus* and *Gryllus* differs fundamentally from germ cell specification in *Drosophila*. Specifically, I provide evidence that germ cells arise via inductive cell signaling during mid-embryogenesis, rather than via maternally-supplied cytoplasmic determinants localized in the oocyte, as is the case for *Drosophila*. These data suggest that *Drosophila* employs an evolutionarily derived mode of germ cell specification. In further support of this hypothesis, I show that several of the genes required for *Drosophila* germ cell specification perform other functions in both *Oncopeltus* and *Gryllus*. I demonstrate that one of these genes, *oskar*, which is the only gene both necessary and sufficient for germ cell specification in *Drosophila*, instead functions in nervous system of the cricket, both during embryonic development and in the adult brain. I suggest that the evolution of the derived mode of

germ cell specification seen in *Drosophila* may have involved co-opting *oskar* into the germ cell specification pathway from an ancestral role in the nervous system.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Several people directly contributed data to this dissertation. Ryo Wakuda, Kanta Terao, Yukihisa Matsumoto, and Makoto Mizunami at the University of Hokkaido, Japan, conducted the *Gryllus* behavioral experiments presented in Chapter 5, and I thank them for allowing these data to be included in this dissertation. Others who directly contributed data appear as co-authors on individual chapters. In addition, I am grateful to all members of the Extavour Lab for providing invaluable training, support, feedback, and friendship. I thank Franz Kainz, Friedemann Linsler, Evelyn Schwager, and Frederike Alwes for welcoming me to the lab and teaching me to work with embryos and with molecular biology. I thank Didem Sarikaya, Andre Greene, Tripti Gupta for help with *Drosophila*, as well as Diane Duplissa for maintaining the Fly Room. I thank Tamsin Jones, Seth Donoughe, and Taro Nakamura for their constant help with crickets and ideas. I thank Victor Zeng for his bioinformatic help, and Yue Meng for keeping the crickets alive, happy, and reproductive. I thank the talented undergraduates Omar Delannoy-Bruno, Nat Clarke, Ian Dunn, and Chris Probert for their help. I am enormously indebted to my fellow graduate student John Srouji, who has guided me through all of my cloning and protein work, and who has been extraordinarily helpful, patient, motivating, and hilarious throughout.

I thank Mimi Velazquez and Barbara Hanrahan for keeping our lab running, and for being wonderful co-workers. I am exceedingly grateful to Chris Preheim in the OEB offices for being a friend to all graduate students, and for always being there to answer any question. I thank Amir Karger, formerly of the Research Computing Staff, for devoting many of his valuable hours to teaching me the basics of bioinformatics.

My research has benefitted greatly from the input of my committee, including Arkhat Abzhanov, Naomi Pierce, Elena Kramer, Alex Schier, and Casey Dunn. I am especially indebted to Arkhat, with whom our lab has shared very interesting biweekly lab-meetings for the past six years, and to Casey, who has welcomed me to his lab at Brown several times and who was

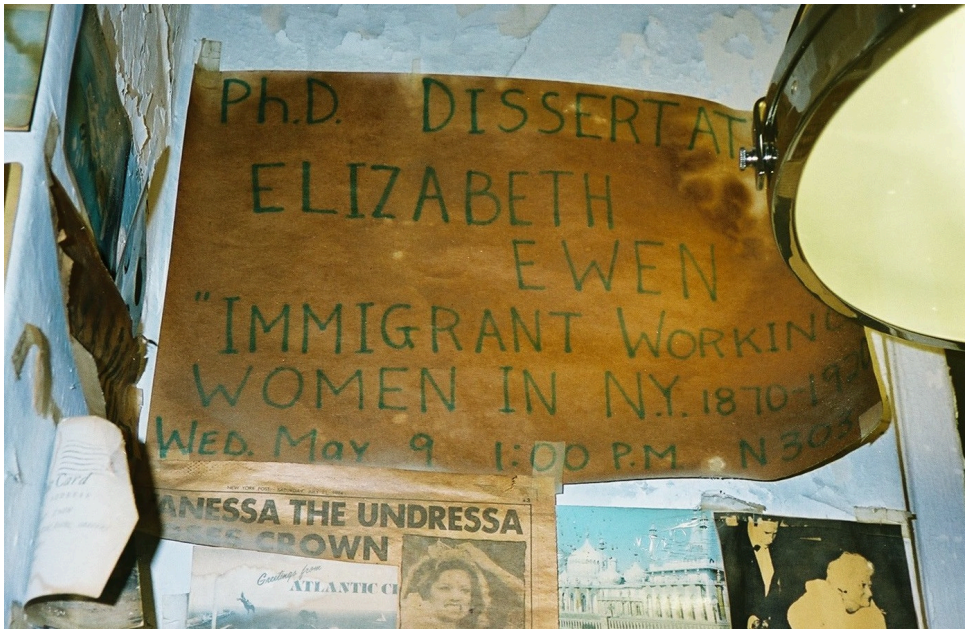
particularly helpful during my work on transcriptomes. I also thank Gonzalo Giribet, with whom I taught for two years, and serves as an inspiration for how enjoyable it can be to do rigorous and new science.

Scott Gilbert and Doug Emlen, two former advisors, have continued to play an outsized role in motivating me and inspiring my research, and I thank them for guiding me during the beginning of my education in biology, and for continuing to be outstanding mentors.

I owe a special debt of gratitude to my advisor, Cassandra Extavour, who has never ceased to encourage me and to motivate me in my research and in many other aspects of my life. In addition to her hands-on training, feedback, and constant availability as an advisor, I have greatly benefitted from many inspiring and motivating conservations about the many ways that biology exists within society, not apart from it.

I offer a wordless gesture of unspeakable love and thankfulness towards my friends, scientific heroes, and life partners Arpiar Saunders and John Tuthill, who are pure beams of light energy. This research would not have been possible without the love and friendship of those with whom I share my life, including Toby David, Tevye Kelman, Aaron Strong, Jonathan Edwards, John R. Williams, Ike Sriskandarajah, Pennie Taylor, Morten Kjaergaard, Mehmet Fisek, James Crall, Sarah Bard, Boat Brainaird, Christopher Edley III, Maude Baldwin, Fenna Krienen, Ben Bradlow, Max Geller, Blake Roberts, Jeremy Butman, Garret Edwards, Zoe Jet Ellis, Michael Casper, Bob Weisz, Ben Jordan, Josh Penn, and others. I give love and thanks to my parents, Jim Campen and Phyllis Ewen, and my sister and her husband, Georgia Ewen-Campen and Devon Huang, who have always been my collective role models. And lastly, I thank my girlfriend, Alexei Fraser, who has always given me love, inspiration, and perspective, and who sets my heart a-fluttering.

Dedicated to the memory of my grandmother, Jane Campen, and the memory of my aunt, Liz Ewen. Both were teachers, trouble-makers, and family-builders.



Announcement of Liz Ewen's dissertation defense (1979), "Immigrant Working Women in New York 1870-1920," on display in the Ewen Library (aka my uncle's bathroom).

INTRODUCTION

… Life is a continuous stream …
The individual body dies, it is true, but the germ cells live on.

- E.B. Wilson, 1900, p 8-10

Germ cells: bridging the generation gap

Every sexually reproducing animal begins as a single cell: a fertilized egg. This egg cell, prior to being fertilized, originates within the mother's body as the result of a dividing germ cell in her ovary. Of course, the germ cells within the mother's ovary, like all the other cells in her body, trace back to a single fertilized egg cell yet again, which itself arose from a dividing germ cell within the ovary of the previous generation. This direct cellular link continues back through the generations, every organism tracing to a single cell division event in its mother. Germ cells thus directly connect the generations, serving as the physical and genetic link between parents and offspring. As cell biologist E.B. Wilson wrote in 1900, "the body is, as it were, an offshoot from the germ cell," (Wilson, 1900).

Given the seeming immortality of the germ cells (or "PGCs", for Primordial Germ Cells as they are called when they first form in the embryo), the mechanisms that make these cells different from the mortal somatic cells have intrigued biologists for centuries. The restriction of reproductive potential to a small, dedicated group of germ cells appears to be a universal process across sexually reproducing animals, one that probably arose at the dawn of animal multicellularity (Buss, 1987). The indispensable function of germ cells might lead one to assume that germ cell specification during development is a highly conserved process. However, despite the homology of germ cells in all metazoans, the mechanisms that specify germ cells during development have turned out to be remarkably diverse in different taxa (Extavour and Akam, 2003), raising questions about how this process has evolved across the animal phylogeny. This

dissertation is concerned with how PGC specification mechanisms have evolved in a particular group of animals: the insects.

Germ cell specification in *Drosophila*

Arguably our most detailed understanding of PGC specification in any animal comes from the fruit fly, *Drosophila melanogaster*. Over a century of embryological research has demonstrated that in this species, as in many other holometabolous insects (a monophyletic group of insects that undergo complete metamorphosis) PGCs are specified via cytoplasmic components that localize to the oocyte posterior while the egg cell is developing within the mother's ovary (reviewed in Mahowald, 2001). This specialized cytoplasm, termed "germ plasm," was first observed using histological techniques in a variety of holometabolous insects around the turn of the 20th century, including *Drosophila* in 1923 (**Figure 1.1A**) (Huettner, 1923). The germ plasm was shown to be asymmetrically inherited during the earliest cell divisions into a small group of so-called "pole cells" at the embryonic posterior (Hegner, 1914; Mahowald, 2001).

Beginning in the 1930s, several lines of experimental evidence in *Drosophila* demonstrated that if germ plasm was destroyed, for example by UV irradiation, it would lead to adults that lack germ cells, indicating that germ plasm is necessary for germ cell formation (**Figure 1.1B**) (reviewed in Mahowald, 2001). The sufficiency of germ plasm to autonomously induce PGC fate was decisively established in the 1970s, in a series of technically virtuosic germ plasm transplantation experiments conducted in Anthony Mahowald's laboratory. In these experiments, germ plasm from a donor embryo was transplanted to ectopic locations in a host embryo, where it was found to induce ectopic pole cells (**Figure 1.1C**) (Illmensee and Mahowald, 1976a; 1974). To demonstrate conclusively that these ectopic pole cells were functional, Illmensee and Mahowald took the additional step of transplanting the resulting ectopic pole cells

into the posterior of a second host embryo, where they formed functional germ cells (Illmensee and Mahowald, 1976b; 1974).

Together, these findings established that germ plasm was necessary and sufficient for PGC formation. However, the molecular components of germ plasm remained a mystery until the first genetic screens for maternal effect genes in *Drosophila*. These screens uncovered a group of "grandchildless" mutants, representing genes that function in the mother's ovary to correctly specify the PGCs in her offspring (Ephrussi et al., 1991; Lehmann and Nüsslein-Volhard, 1991; Schüpbach and Wieschaus, 1986). In subsequent years, detailed analysis of these and additional genes has provided an exquisitely detailed understanding of the molecular mechanisms that orchestrate PGC specification in *Drosophila*.

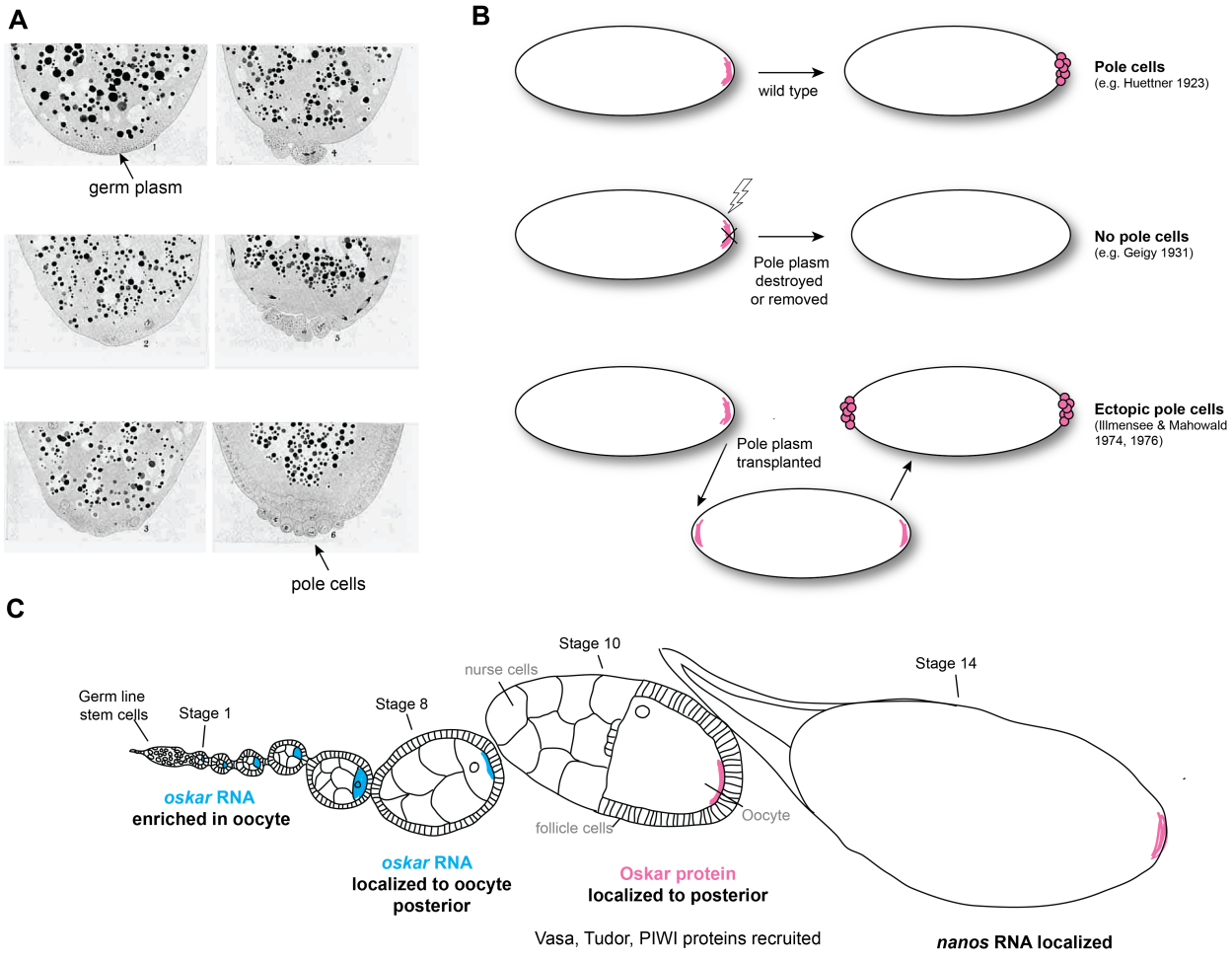


Figure 1.1 (*previous page*). Germ cell formation in *Drosophila melanogaster*. (A) Huettner's (1923) original description of germ plasm and pole cell formation in *Drosophila*. Germ plasm is visible as dark, punctate staining at the posterior cortex, which is subsequently taken up by the pole cells during cellularization. Posterior is down. **(B)** Summary of the experiments demonstrating that germ plasm is both necessary and sufficient for PGC formation. Posterior is to the right. (Reviewed in Mahowald, 2001). **(C)** Simplified summary of the molecular basis for germ plasm assembly during *Drosophila* oogenesis. Posterior is to the right. Ovariole traced from (Horne-Badovinac and Bilder, 2005). Figure adapted from (Lehmann and Ephrussi, 1994), with additional data from (Bardsley et al., 1993; Harris and Macdonald, 2001; Lasko and Ashburner, 1990).

During oogenesis, each *Drosophila* oocyte develops in conjunction with 15 clonally-related "nurse cells," which remain connected to the oocyte via cytoplasmic bridges (**Figure 1.1C**). These nurse cells synthesize mRNAs and proteins that are transferred into the oocyte, many of which are subcellularly localized within the oocyte and thereby establish the axes of the embryo. Among these nurse cell-derived factors is *oskar* RNA, which is transcribed by the nurse cell nuclei and transported via microtubules as an untranslated RNA to the oocyte posterior (**Figure 1.1C**) (Kim-Ha et al., 1995; 1991; Zimyanin et al., 2008). Upon posterior localization, *oskar* is translated into two protein isoforms, dubbed "Long Osk" and "Short Osk," which are thought to anchor Oskar to the posterior and to recruit downstream germ plasm components, respectively (**Figure 1.1C**) (Markussen et al., 1995). Via unknown biochemical mechanisms – Oskar is a novel protein lacking any functionally characterized domains – Short Osk recruits additional germ plasm components including Vasa, Tudor, PIWI-family proteins, and *nanos*, the latter of which has additional roles in abdominal patterning (Harris and Macdonald, 2001; reviewed in Mahowald, 2001). Following cellularization, the so-called "pole cells" at the posterior of the embryo inherit these cytoplasmic factors, which induce germ cell fate and repress somatic differentiation programs (reviewed in Seydoux and Braun, 2006).

Interestingly, while many of the germ plasm factors are necessary for germ cell formation, only Oskar has been shown to be sufficient for inducing germ cell formation in ectopic locations (Ephrussi and Lehmann, 1992; Smith et al., 1992). Given the sufficiency of Oskar to assemble

functional germ plasm, this gene is considered to represent the most upstream member of the germ cell specification pathway in *Drosophila*.

Patterns of germ cell specification mechanisms across Metazoa

The mode of germ cell specification in *Drosophila*, generically termed "cytoplasmic inheritance," represents one of the two modes of germ cell specification that have been described in animals. This mode, whereby maternally supplied cytoplasmic factors specify PGC fate, occurs in nearly all widely studied model organisms for developmental biology. The nematode *Caenorhabditis elegans*, the zebrafish *Danio rerio*, the frog *Xenopus laevis*, and likely also the chicken *Gallus gallus,* all form germ cells via cytoplasmic inheritance of maternally provided germ plasm.

In contrast, the mouse *Mus musculus* specifies PGCs via "zygotic induction," whereby cell-cell signaling induces PGC specification in the absence of a maternally supplied germ plasm. Accordingly, cell transplantation experiments in mouse embryos demonstrate that, as late as the initiation of gastrulation, distal embryonic cells that are experimentally grafted proximally into the location of normal PGC formation are capable of responding to local signals and differentiating into PGCs (Tam and Zhou, 1996). Thus, lineage restriction of the PGCs involves inductive signals and does not occur until well after the activation of the zygotic genome (reviewed in Saitou and Yamaji, 2012). Interestingly, once germ cells are specified in the mouse, they express orthologs of many of the genes found in *Drosophila* germ cells, including Vasa (Fujiwara et al., 1994; Toyooka et al., 2000), Nanos (Tsuda, 2003), Tudor-domain proteins (Chuma et al., 2006; Hosokawa et al., 2007), and PIWI proteins (Carmell et al., 2007; Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004), although several of these genes have sex-specific functions in the mouse.

The signaling pathways involved in mouse germ cell specification have been characterized in great detail via genetic analyses and cell culture experiments. During the second half of embryonic day 5, several BMP ligands are secreted from the extraembryonic ectoderm and visceral endoderm and synergistically signal via SMAD1/5 to upregulate the expression of a trio of transcription factors, *Blimp1*, *Prdm14*, and *Tcfap2c,* in a small subset of embryonic cells (reviewed in Saitou and Yamaji, 2012). These three transcription factors are necessary for PGC development, and are sufficient to induce functional PGCs *in vitro* from cultured epiblast-like cells (Magnúsdóttir et al., 2013; Nakaki et al., 2013). Together, these transcription factors upregulate hundreds of PGC-specific genes and down-regulate hundreds of soma-specific genes (reviewed in Saitou and Yamaji, 2012). Of the BMP ligands required for PGC specification, BMP4 is uniquely capable of causing epiblast cells to adopt functional PGC fate *in vitro*, a process that also requires Wnt3 signaling from nearby cells (Ohinata et al., 2009) and acts at least partially through the transcription factor *brachyury* to regulate the transcription of *Blimp1*, *Prdm14*, and *Tcfap2c*, which together upregulate hundreds of additional genes involved in PGC development (Aramaki et al., 2013).

Although the majority of widely studied model organisms specify PGCs via cytoplasmic inheritance, two lines of evidence suggest that zygotic induction may be the ancestral mode of PGC specification in Metazoa, and that the cytoplasmic inheritance mode has evolved independently multiple times in various lineages. First, an analysis of embryological literature covering nearly all animal phyla suggests that zygotic induction is in fact the more common mode of specification (Extavour and Akam, 2003). Further, this analysis revealed that cytoplasmic inheritance tends to be nested within clades whose basally branching members are thought to employ zygotic induction (Extavour and Akam, 2003; Extavour, 2007). Thus, given the phylogenetic distribution of PGC specification modes, parsimony suggests that cytoplasmic inheritance is a derived character that has arisen multiple times.

The second line of evidence that cytoplasmic inheritance has evolved independently several times is that although there is conservation of several of the molecular components of germ plasm across animals (reviewed in Ewen-Campen et al., 2010), the molecular mechanisms that assemble these components into functional germ plasm in different taxa are highly disparate, driven by genes that show no signs of homology. For example, although Oskar protein is necessary and sufficient to recruit all of the cytoplasmic factors necessary to generate functional germ cells in *Drosophila* (Ephrussi and Lehmann, 1992), this gene is a novel, insect-specific gene, and its role in germ cell specification is unique to holometabolous insects (Lynch et al., 2011).

In the zebrafish *Danio rerio*, which lacks *oskar*, formation of germ plasm in the oocyte and its asymmetric distribution during early cell cleavages requires *bucky-ball*, a novel vertebrate-specific protein with unknown molecular function (Bontems et al., 2009; Marlow et al., 2008). In *C. elegans*, the *pgl-1* and *pgl-3* genes are not *oskar* orthologs, but have a function analogous to *oskar*, as they can promote the assembly of germ plasm granules (P granules), including Vasa orthologs, in ectopic cellular contexts (Hanazawa et al., 2011; Updike et al., 2011; Wang and Seydoux, 2013). However, it is noteworthy that when P granules are formed ectopically, they are insufficient to confer germ cell fate on the cells containing them (Gallo et al., 2010), unlike analogous experiments with *Drosophila* or *Xenopus* (Ephrussi and Lehmann, 1992; Illmensee and Mahowald, 1976b; 1974; Tada et al., 2012). In the context of wild type *C. elegans* embryogenesis, the localization of germ plasm involves dynamic P granule dissasembly in the anterior and reassembly in the posterior, mediated via MEX5/6 and PAR-1, respectively, in a process with no clear molecular homology to germ plasm assembly in any other model system (Gallo et al., 2010).

In sum, the patchy phylogenetic distribution of germ plasm, together with the fact that distinct developmental and molecular mechanisms are used to form germ plasm in those species where it is observed, suggests that the cytoplasmic inheritance mode of PGC specification has evolved independently multiple times.

There are profound mechanistic differences between the cytoplasmic inheritance and zygotic induction modes of PGC specification. For those species that specify PGCs via cytoplasmic inheritance, the operative molecular mechanisms are those which maintain early embryonic cells in an essentially undifferentiated, totipotent state (reviewed in Seydoux and Braun, 2006). In *Drosophila*, for example, transcription is globally repressed in PGCs, a process mediated by the *pgc* protein, a small peptide which inhibits the transcriptional activity of RNA polymerase II by interfering with a critical phosphorylation event (Hanyu-Nakamura et al., 2008). A remarkably similar mechanism has independently evolved in *C. elegans*, where the germ cell-specific protein PIE-1 interferes with RNA polymerase II phosphorylation, effectively repressing transcription in the germ cells (Wang and Seydoux, 2013). There is also evidence that transcription is globally repressed at the level of chromatin modifications in both species (reviewed in Seydoux and Braun, 2006).

In contrast, for those species that specify PGCs inductively, mechanisms exist for re-establishing totipotency *de novo* in the PGCs. In the mouse, mechanisms exist to de-differentiate the PGCs towards the totipotent state that they held several days prior. Namely, PGCs must "undo" several aspects of somatic differentiation, including a DNA demethylation across the entire genome to remove imprinting and re-activate a recently de-activated X-chromosome, the upregulation of "pluripotency" factors including Nanog, Oct4 and Sox2, the active repression of somatic differentiation genes, and a variety of global changes in histone modifications (reviewed in Saitou and Yamaji, 2012).

Given the fundamental differences in the molecular mechanisms at play in these two modes of PGC specification, it is interesting to ask how this process has evolved across the animal tree. Namely, how could the cytoplasmic inheritance mode have arisen (likely multiple

times) from an ancestral mode of zygotic induction[1]? In order to address this issue, comparative studies of related species that differ in their PGC specification mechanisms are needed.

Insects: a case study in the evolution of PGC specification mechanisms

Insects provide an attractive opportunity to study the evolution of PGC specification mechanisms because basally branching taxa are thought to specify PGCs using zygotic induction, whereas derived lineages (including *Drosophila melanogaster*) form PGCs via cytoplasmic inheritance (Extavour and Akam, 2003). Thus, comparative developmental studies of basally branching taxa are poised to shed light on how cytoplasmic inheritance can evolve, a process which appears to have occurred repeatedly in a variety of phyla (Extavour and Akam, 2003; Extavour, 2007). Furthermore, an understanding of the specific signaling pathways that underlie zygotic induction in basally branching insects may provide new insight into how PGC identity can be conferred on embryonic cells, and whether this process has any conservation with other species that utilize the zygotic induction mode of PGC specification.

In the insects, the hypothesis that germ plasm is a derived character confined principally to Holometabola and their close relatives is based on surveys of embryological literature from the past 150 years (Extavour and Akam, 2003; Matsuda, 1976; Nieuwkoop and Sutasurya, 1981), coupled with experimental data from a small number of species (Lynch et al., 2011; Mahowald, 2001). Beginning in 1908, when Hegner experimentally removed the posterior cytoplasm from just-laid beetle embryos and observed the loss of germ cells (Hegner, 1908), cytoplasmic

[1] Some have also asked *why* these processes would evolve, i.e. what are the selective advantages of one specification mechanism versus the other. For example, Johnson and colleagues (2011) have suggested that in vertebrates, the uncoupling of PGC specification from somatic patterning in the cytoplasmic inheritance mode could allow for greater body plan evolution and thus increased speciation. In contrast, Buss (1988) found evidence for the opposite association across Metazoa (with many stated exceptions). In addition, Buss (1983) invokes the need to "police" potential cell-cell conflicts in any multicellular context as the operative selective pressure on the timing of PGC specification. From the standpoint of an experimental embryologist, these hypotheses seem difficult to test, and these ideas are not pursued further in this dissertation.

inheritance has been experimentally demonstrated in several additional species of Coleoptera (beetles), Diptera (flies and mosquitoes), and Hymenoptera (wasps, bees, and ants) (Ewen-Campen et al., 2013a; Hegner, 1914; Lynch et al., 2011; Mahowald, 2001). Moreover, pole cells have been cytologically identified in many additional holometabolous insects, including Lepidoptera (butterflies and moths), and several less well-studied orders such as Neuroptera (lacewings), Strepsiptera (twisted-winged flies), and Siphonaptera (fleas) (reviewed in Ewen-Campen et al., 2013a; Matsuda, 1976). Taken together, these data suggest that germ plasm and pole cells were present in the common ancestor of Holometabola (**Figure 1.2**).

Further, there is evidence that the germ plasm and pole cells in holometabolous insects are formed via homologous molecular mechanisms. Although functional molecular studies of *oskar*, *vasa*, and *nanos* have only been undertaken in *Drosophila melanogaster* and the wasp *Nasonia vitripennis*, these studies have revealed a largely conserved molecular program for germ plasm assembly in these species (Lynch et al., 2011; Lynch and Desplan, 2010). Given that *D. melanogaster* (Diptera) and *N. vitripennis* (Hymenoptera) phylogenetically bracket the holometabolous insects (Wiegmann et al., 2009), these studies support the hypothesis that germ plasm was present at the base of Holometabola, and suggest that its assembly involved *oskar* (**Figure 1.2**).

Interestingly, however, there are species within nearly all of the Holometabolous orders that lack cytologically discernable germ plasm and/or pole cells, including the moth-midge *Clogmia albipunctata* (Diptera) [but see (Jiménez-Guri et al., 2014)], the flour beetle *Tribolium castaneum* (Coleoptera), the honey-bee *Apis mellifera* (Hymenoptera), and the silkworm *Bombyx mori* (Lepidoptera) (reviewed in Ewen-Campen et al., 2013b; Lynch et al., 2011). Furthermore, in those species for which genomic data is available, the *oskar* gene is absent from the genome of those holometabolous species that lack germ plasm (Jiménez-Guri et al., 2013; Lynch et al., 2011). These observations have led to the hypothesis that the presence of *oskar* in the genome correlates with the cytoplasmic mode of PGC specification in insects, such that both the *oskar*

gene and the cytoplasmic inheritance mode of PGC specification have been repeatedly lost in several holometabolous lineages (Lynch et al., 2011).

In stark contrast to the Holometabola, there is no experimental evidence of a germ plasm in any of the species that branch basally to Holometabola (the hemimetabolous insects, which do not undergo metamorphosis), and PGC specification in these taxa remains essentially a mystery. Histological observations of these taxa have typically failed to identify a discernable germ plasm or pole cells (Anderson, 1972a; Ewen-Campen et al., 2013a; reviewed in Matsuda, 1976; Nieuwkoop and Sutasurya, 1981). Instead, the embryological literature on hemimetabolous insects suggests that PGCs are not identifiable until later stages of development, in some cases not until the embryo is fully segmented, implying that zygotic induction must specify PGCs (See Table S1 in Ewen-Campen et al., 2013a)**.**

The reported timing and location of PGC origin in hemimetabolous insects varies widely between species. In many Hemiptera ("true bugs"), cytologically-defined PGCs are first visible as a group of cells facing the yolk at the posterior of the blastoderm just prior to gastrulation (**Figure 3A**) (Butt, 1949; Heming and Huebner, 1994). At this stage, cellularization has occurred (Butt, 1949) and anterior segments have been specified, as revealed by *engrailed* expression (PZ Liu and T. Kaufman, 2004). These species are examples of relatively early PGC specification amongst the hemimetabola, but it is important to note this stage is still markedly later than PGC formation in *Drosophila*, and that germ plasm is not reported in either of these species, making it unclear if there could be a maternal contribution to PGC specification.

Reports from Orthoptera (crickets, grasshoppers, and locusts) differ from those in Hemiptera, and also vary widely between species. In the cricket *Achaeta domestica*, putative PGCs are reported at the egg posterior during the earliest stages of mesoderm formation [Heymons 1895, cited in (Matsuda, 1976)] whereas in two separate studies of the locusts *Concocephalus* (previous called *Xiphidium*) and *Locusta*, PGCs are reported to arise from mesodermal structures after the completion of segmentation (**Figure 1.3B**)

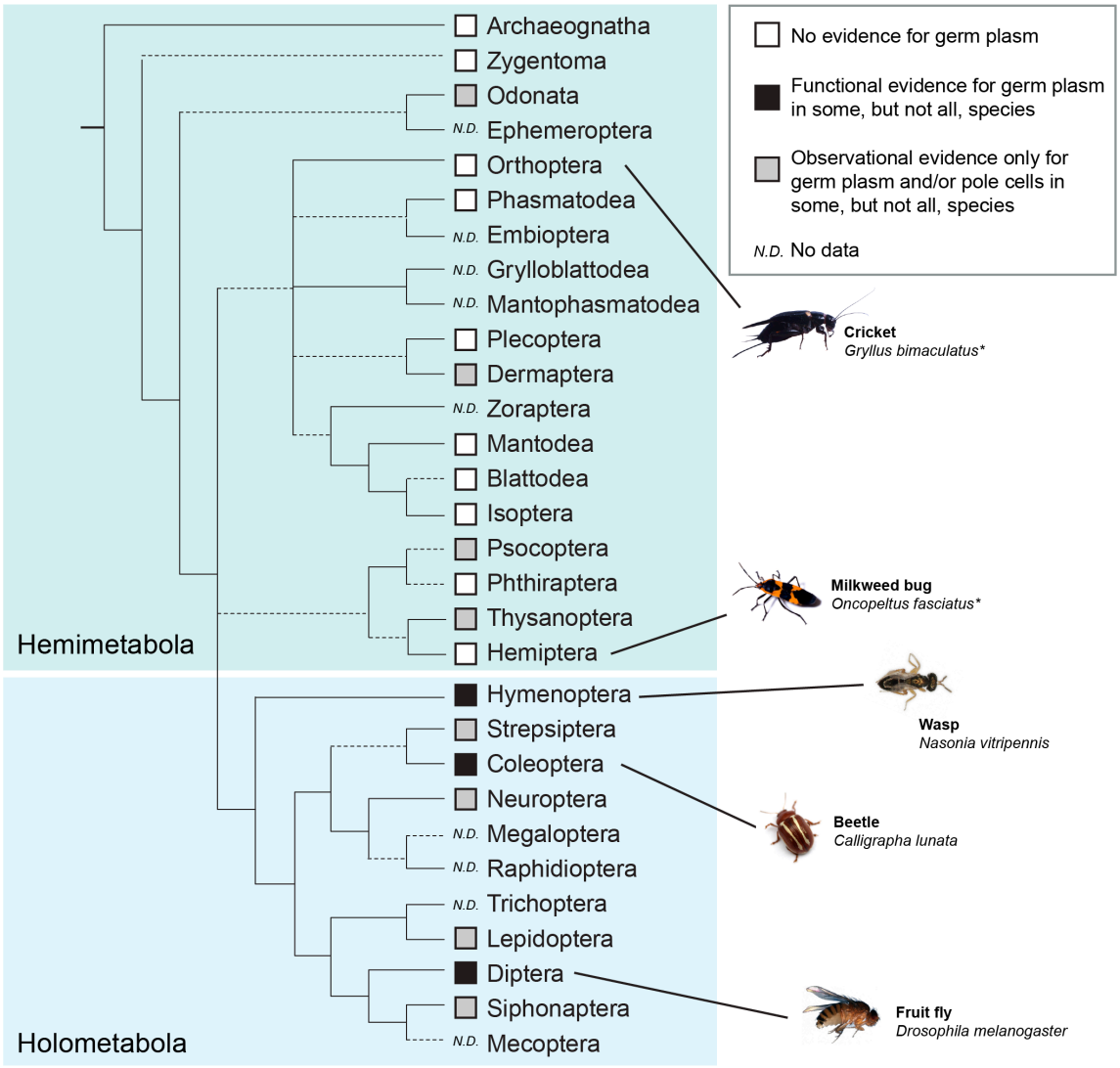
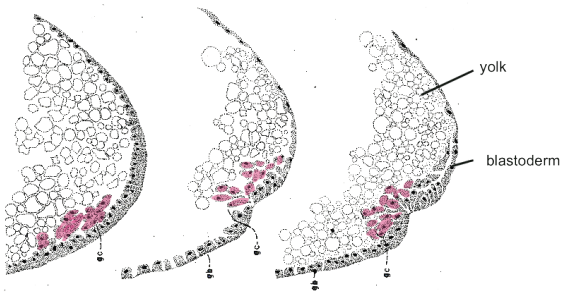


Figure 1.2. Distribution of germ plasm and pole cells across the insects. Germ plasm has only been demonstrated experimentally in holometabolous insects. Cytological evidence suggests that germ plasm is rare in hemimetabolous insects, and is absent from the closest hexapod outgroups, suggesting that zygotic induction is the ancestral mode of PGC specification in insects. Images are shown of species studied in this dissertation (*Gryllus* and *Oncopeltus*, indicated with astrices), or holometabolous species for which experimental evidence for germ plasm has been given, including *C. lunata,* one of the original beetle species studied by Hegner (1908). References are given in Supplemental Table 1 of (Ewen-Campen et al. 2013), with additional data for Odonata (Ando, 1962), Dermaptera (Singh, 1967), Neuroptera (Matsuda, 1976), and Strepsiptera (Matsuda, 1976). *N.D.* refers to orders for which, to my knowledge, germ cell origin has not been reported. Phylogeny from (Trautwein et al., 2012); dashed lines indicate uncertain relationships and/or possible polyphyly.
(Photo credits: *Gryllus* and *Oncopeltus* by David Behl, *Nasonia* from www.bioecologysrl.it, *Calligrapha* from bugguide.net, and *Drosophila* from http://www.lifesciencesfoundation.org/events-The_Drosophila_melanogaster_genome.html)

(Roonwal, 1937; Wheeler, 1893). A variety of PGC origin sites are reported for other hemimetabolous species, ranging from the "early" specification exemplified by Hempitera to the "late" specification exemplified by some Orthoptera. The distribution of these observations across species do not appear to follow a discernable phylogenetic pattern within the Hemimetabola (Extavour and Akam, 2003; Matsuda, 1976; Nieuwkoop and Sutasurya, 1981). In addition, PGC specification has not been reported, to my knowledge, for several of the smaller hemimetabolous orders, including Zoraptera, Mantophasmatodea, Grylloblattodea, Mantodea, and Ephemeroptera (Matsuda, 1976).

In spite of the notable variation observed between hemimetabolous species, there is general agreement amongst those who have reviewed this literature that PGC specification in holometabolous insects largely occurs via germ plasm and pole cells, whereas in hemimetabolous insects there is no evidence of germ plasm, implying that PGC specification must occur via other, unknown mechanisms (Anderson, 1972a; 1972b; Extavour and Akam, 2003; Matsuda, 1976; Nieuwkoop and Sutasurya, 1981). It should be noted, however, that a small number of hemimetabolous species are reported, based solely on histological observations, to specify PGCs very early, in a manner reminiscent of pole cell formation in *Drosophila* (**Figure 1.3C-F**). In one species of Thysanoptera (thrips), a cytologically identified germ plasm is localized to the oocyte posterior during late stages of oogenesis, and the cleavage nuclei that migrate into this germ plasm appear to become PGCs (**Figure 1.3C**) (Heming, 1979). In a species of Psocoptera (book lice), although germ plasm is not reported in the oocytes or just-laid eggs, PGCs form at the embryonic posterior just as the blastoderm forms, similar to pole cells (**Figure 1.3D**) (Goss, 1952). These two orders fall within the sister assemblage to Holometabola (**Figure 1.2**), and thus suggest that germ plasm and pole cells may have originated before the holometabolous radiation. In addition, there are also reports of germ plasm and pole cells in one species of Dermaptera (earwigs, **Figure 1.3E**) (Singh, 1967) and of pole cells (but not germ plasm) in one species of Odonata (dragonflies and damselflies, **Figure 1.3F**) (Ando, 1962), both of which branch much

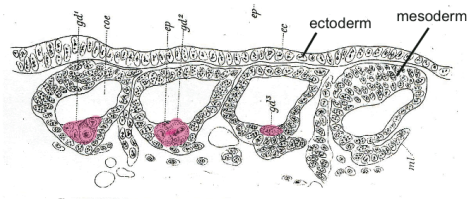
A



"Early" PGC specification at the blastoderm posterior
The milkweed bug *Oncopeltus fasciatus* (Hemiptera)

"The events that lead up to the formation of the embryo begin to occur early on the second day. At this time a clump of cells (gc [germ cells]) appears at the posterior pole of the egg situated within the yolk in close proximity to the inner surface of the blastoderm." (Butt 1947)

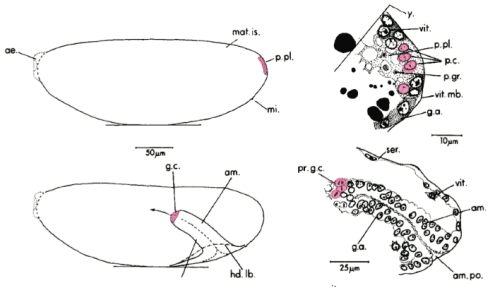
B



"Late" PGC specification in the segmental mesoderm of the abdomen
The locust *Xiphidium* (Orthoptera)

"In earlier stages, careful scrutiny failed to reveal any differentation of the mesoderm cells...it is not, therefore until the somites are established as distinct sacs that unmistakable primitive germ-cells make their appearance..."(Wheeler 1893)

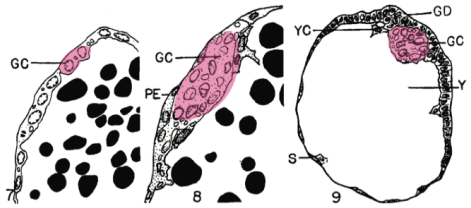
C



Putative germ plasm in the oocyte posterior
The thrip *Haplothrips verbasci* (Thysanoptera)

"In fully developed eggs of *H. verbasci*, the pole plasm (p. pl) is situated at the posterior end...Pole (germ) cells (p.c.) first become recognizeable late in stage B when a few cleavage energids enter the pole plasm...Although I do not have experimental evidence to prove the existence of pole plasm and germ cell determinants in eggs of *H. verbasci*, I believe that both occur." (Heming 1979)

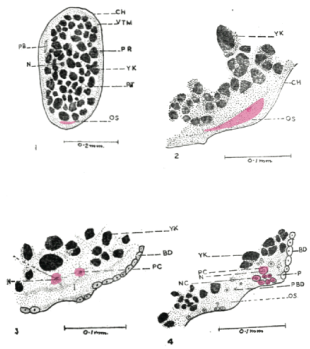
D



Putative pole cells during blastoderm formation
The book louse *Liposcelis divergens* (Psocoptera)

"The germ cells become differentiated when the primary epithelium is completed. As far as could be determined, no structures that could possibly be identified as oosomes or germ cell determinants were found in these eggs..." (Goss 1952)

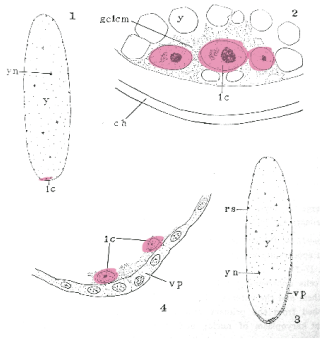
E



Putative germ plasm in the oocyte posterior
The earwig *Labidura riparia* (Dermaptera)

"In a section of freshly laid egg of *Labidura riparia* an area of distinct cytoplasmic differentiation has been observed near the posterior end...some of them [cleavage nuclei] along with their trailing cytoplasm move into the Oosome to give rise to the primordial germ cells..." (Singh 1967)

F



Putative germ plasm in the oocyte posterior
The dragonfly *Epiophlebia superstes* (Odonata)

"They are possible the primordial germ cells or "pole cells" which are found at the posterior of the eggs at the blastoderm formation...but final decision needs experimental prove, i.e. by cauterize method" (Ando 1962)

Figure 1.3 (*previous page*). Examples of histological descriptions of PGC formation in various hemimetabolous insects. Images and quotations are excerpted from primary descriptions of PGC formation, with germ plasm and/or PGCs false-colored in pink. **(A)** An example of "early" PGC formation seen in many hemimetabolous insects, at the posterior of the blastoderm, in the milkweed bug *Oncopeltus fasciatus*. **(B)** An example of "late" PGC formation in the segmented abdominal mesoderm, from the locust *Concocephalus* (formerly *Xiphidium*). **(C-F)** Four atypical examples of hemimetabolous species described as possessing germ plasm and/or pole cells, although note that none of these examples have been re-investigated using molecular techniques.

more deeply in the insect tree (**Figure 1.2**). It will be very interesting to re-examine these exceptional species using experimental techniques and molecular markers in order to test whether this mode is truly homologous to PGC specification in holometabolous insects such as *Drosophila*.

There is a major limitation of relying solely on histological and cytological data to determine the presence or absence of germ plasm: in some cases, the use of molecular markers can reveal a cryptic germ plasm that eluded prior histological studies. Indeed, in chicken, zebrafish, and *Amphioxus*, historical observations had suggested an inductive mode of PGC specification, but analysis of *vasa* gene expression revealed the presence of a maternally-supplied germ plasm (Gomez et al., 2011; Tsunekawa et al., 2000; Yoon et al., 1997). Thus, the use of molecular markers and functional genetic tests for PGCs and/or germ plasm is essential in order to test for the presence of germ plasm in any organism.

In the case of hemimetabolous insects, PGC gene expression studies have previously been limited to just three species, and have been far from conclusive. Studies of *vasa* expression in two orthopterans, the grasshopper *Schistocerca americana* and the cricket *Gryllus bimaculatus,* did not reveal a germ plasm in oocytes or early embryos, and left PGC specification an open question in these species (Chang et al., 2009; 2007; Mito et al., 2008). The parthenogenic embryos of the asexual-phase pea aphid *Acyrthosiphon pisum*, a hemipteran with a highly derived mode of viviparous development, appears to asymmetrically localize Nanos, but not Vasa, to the oocyte posterior and subsequently to the PGCs (Chang et al., 2009; 2007). It is not clear if this is the case for embryos developing in the sexual phase of *A. pisum*, as PGCs have not been

identified prior to segmentation stages (Miura et al., 2003). In addition, functional studies of "germ cell genes" have not previously been conducted in any hemimetabolous insects.

In sum, the mechanisms of PGC specification in basally branching insects remain unknown, and there is reason to believe it occurs in the absence of germ plasm. Thus, a molecular and functional understanding of PGC specification in hemimetabolous insects is necessary in order to understand how such a system may have evolved into the well-characterized mode observed in *Drosophila* and other holometabolous insects.

Outline of this dissertation

In this dissertation, I aim to add to our understanding of germ cell specification in basally branching insects, and thus to our knowledge of how germ cell specification mechanisms evolve. I focus on two hemimetabolous insects, the milkweed bug *Oncopeltus fasciatus* (Hemiptera) and the cricket *Gryllus bimaculatus* (Orthoptera), which occupy important phylogenetic positions in the insect phylogeny. *Oncopeltus*, a hemipteran, falls within the sister assemblage to Holometabola (along with Thysanoptera [thrips], booklice Psocoptera [booklice], and Phthiraptera [body lice]) (**Figure 1.2**) and therefore offer a relatively close comparison to such species as *D. melanogaster* and *N. vitripennis*. *Gryllus*, an orthopteran, branches far closer to the base of Insecta, offering a deeper phylogenetic comparison with Holometabola and with Hemiptera (**Figure 1.2**).

In addition to their informative phylogenetic positions, *Oncopeltus* and *Gryllus* are experimentally tractable for embryological studies, a non-trivial requirement for many evolutionarily interesting animals (Abzhanov et al., 2008). Both species breed in the laboratory at high density and low cost, laying hundreds to thousands of eggs each day, year-round. More importantly, protocols have been developed in both species for such essential embryological methods as gene expression analysis and functional knockdown via RNA interference (RNAi)

(Paul Liu and T. C. Kaufman, 2009a; 2009b; 2009c; Mito and Noji, 2008), which are essential for investigating PGC specification on a cellular and molecular level.

In Chapter 1 (Ewen-Campen et al., 2011), I describe a *de novo* embryonic and ovarian transcriptome database for the milkweed bug, *Oncopeltus*. For "non-model organisms" such as *Oncopeltus* and *Gryllus*, the rate-limiting step in developmental studies often continues to be the identification and cloning of individual genes-of-interest. Most commonly, individual genes are cloned using the time-consuming and failure-prone technique of degenerate PCR, limiting many studies to focusing on small numbers of candidate genes. With the advent of new sequencing technologies, however, it has become possible to assemble transcriptomic or genomic data *de novo* in the absence of a sequenced genome, and simultaneously identify thousands of genes simultaneously. In this paper, I present a simple conceptual framework for sequence assembly and annotation, and generate an easy-to-use database of over 10,000 genes expressed during oogenesis and embryogenesis of *Oncopeltus*. This work also established a cDNA synthesis protocol and analysis workflow which was later extended to the cricket *Gryllus bimaculatus* (Zeng et al., 2013) and the crustacean *Parhyale hawaiiensis* (Zeng et al., 2011), and laid the technical foundation for all subsequent work in this dissertation.

In Chapter 2 (Ewen-Campen et al., 2012), I focus on a particularly fascinating gene involved in the evolution of PGC specification: *oskar*. The only gene known to be both necessary and sufficient for germ cell specification in *Drosophila* (or any animal, for that matter), *oskar* sits atop the PGC specification pathway in *Drosophila* (Ephrussi and Lehmann, 1992). However, previous studies have suggested that this gene is absent from all hemimetabolous and non-insect genomes, and *oskar* was therefore believed to be a novel gene that evolved at the base of Holometabola (Lynch et al., 2011). In the course of sequencing the *Gryllus* transcriptome, I made the surprising discovery that *oskar* is present in this species, the first example of an *oskar* ortholog outside of Holometabola. I then showed that *Gryllus oskar* is involved in neural development rather than germ cell development. These results demonstrate that *oskar* evolved far

earlier in insect evolution than previously thought, and that its well-studied role in *Drosophila* germ cells is a relatively recent evolutionary innovation, possibly the result of co-option from an ancestral role in the nervous system.

In Chapter 3 (Ewen-Campen et al., 2013a), I describe in detail how PGCs arise in the cricket *Gryllus bimaculatus*. Using a variety of molecular markers, I provide evidence that PGCs first arise in close association with abdominal mesoderm, during abdominal elongation, long after cellularization has occurred, in the absence of a maternally-supplied germ plasm. I show that mesoderm is required for PGC specification using RNAi against *twist*, a transcription factor necessary for mesoderm formation. *twist* RNAi dramatically reduces PGCs, consistent with the hypothesis that PGCs arise from a pool of cells that also generate mesoderm. Lastly, I show that neither *vasa* nor *piwi* is required for PGC specification in *Gryllus*, suggesting that the roles of these genes in PGC specification in *Drosophila* may also be a derived character. In *Gryllus*, both *vasa* and *piwi* are instead involved in spermatogenesis in the adult male, reminiscent of their role in the mouse.

In Chapter 4 (Ewen-Campen et al., 2013b), I extend my studies of hemimetabolous PGC specification to the milkweed bug, *Oncopeltus fasciatus*. I perform an *in situ* hybridization screen of 19 genes with germ plasm expression in *Drosophila*, and show that none of these genes localizes posteriorly in oocytes or early embryos. I identify three *bona fide* molecular markers of PGCs (*vasa, tudor*, and *boule*), and show that transcripts all three of these markers are expressed ubiquitously in earliest stages of development, and only localize to germ cells after the cellularized blastoderm stage, just prior to gastrulation. These gene expression data argue against the presence of a germ plasm in *Oncopeltus*, and are consistent with previous histological observations of related species that suggested a post-cellularization origin of PGCs at the embryonic posterior (Butt, 1949; Heming and Huebner, 1994). Lastly, I show that, as in *Gryllus*, knockdown of these PGC markers does not disrupt PGC formation, and instead that *vasa* RNAi disrupts spermatogenesis.

In Chapter 5, I present unpublished results demonstrating the surprising finding that *Gb-oskar* functions in the adult brain during long-term olfactory learning and memory in the cricket *Gryllus*. In collaboration with the Mizunami lab (Hokkaido University), I show that RNAi against *Gryllus oskar* in adult crickets interferes with long-term (1 day) olfactory learning in an odor association assay, although short term (1 hour) associative learning is not disrupted. I show that *Gb-oskar* is expressed in a population of proliferative neuroblasts in the adult mushroom body, the anatomical substrate implicated in olfactory learning in insects. These data, together with recently published work on *oskar* in the *Drosophila* nervous system (Xu et al., 2013), demonstrate that *oskar* plays a conserved role in the nervous system in holometabolous and hemimetabolous insects, and support the hypothesis that the neural function may therefore be ancestral.

The research in this dissertation provides the first functional genetic investigation of PGC specification in basally branching hemimetabolous insects, and supports the hypothesis that the cytoplasmic inheritance mode of PGC specification known from *Drosophila* is a derived character, having evolved from a zygotic induction mode in ancestral insects. Furthermore, the evolution of germ plasm in insects may have involved the co-option of the *oskar* gene from an ancestral role in the nervous system to a novel function atop the PGC specification pathway.

REFERENCES

Abzhanov, A., Extavour, C., Groover, A., Hodges, S., Hoekstra, H., Kramer, E., Monteiro, A., 2008. Are we there yet? Tracking the development of new model systems. Trends Genet 24, 353–360.

Anderson, D.T., 1972a. The development of hemimetabolous insects, in: Counce, S.J., Waddington, C.H. (Eds.), Developmental Systems: Insects. Academic Press, New York, pp. 96–163.

Anderson, D.T., 1972b. The development of holometabolous insects, in: Counce, S.J., Waddington, C.H. (Eds.), Developmental Systems: Insects. Academic Press, New York.

Ando, H., 1962. The comparative embryology of Odonata with special reference to a relic dragonfly Epiophlebia superstes selys. The Japan Society For the Promotion of Science, Tokyo.

Aramaki, S., Hayashi, K., Kurimoto, K., Ohta, H., Yabuta, Y., Iwanari, H., Mochizuki, Y., Hamakubo, T., Kato, Y., Shirahige, K., Saitou, M., 2013. A mesodermal factor, T, specifies mouse germ cell fate by directly activating germ line determinants. Developmental Cell 27, 516–529.

Bardsley, A., McDonald, K., Boswell, R.E., 1993. Distribution of tudor protein in the Drosophila embryo suggests separation of functions based on site of localization. Development 119, 207–219.

Bontems, F., Stein, A., Marlow, F., Lyautey, J., Gupta, T., Mullins, M.C., Dosch, R., 2009. Bucky ball organizes germ plasm assembly in zebrafish. Curr Biol 19, 414–422.

Buss, L., 1987. The Evolution of Individuality. Princeton University Press, Princeton, NJ.

Buss, L., 1988. Diversification and Germ-Line Determination. Paleobiology 14, 313–321.

Butt, F.H., 1949. Embryology of the milkweed bug: Oncopeltus fasciatus (Hemiptera). Cornell University Agricultural Experiment Station 283, 2–43.

Carmell, M.A., Nakamura, T., Girard, A., Yoshizaki, M., van de Kant, H.J.G., Ogawa, S., Bourc'his, D., Okamoto, H., Bestor, T.H., Shinmyo, Y., de Rooij, D.G., Bando, T., Hannon, G.J., Ohuchi, H., Noji, S., Mito, T., 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germ line. Developmental Cell 12, 503–514.

Chang, C.-C., Huang, T.-Y., Cook, C.E., Lin, G.-W., Shih, C.-L., Chen, R.P.-Y., 2009. Developmental expression of Apnanos during oogenesis and embryogenesis in the parthenogenetic pea aphid Acyrthosiphon pisum. Int J Dev Biol 53, 169–176.

Chang, C.-C., Lin, G.-W., Cook, C.E., Horng, S.-B., Lee, H.-J., Huang, T.-Y., 2007. Apvasa marks germ-cell migration in the parthenogenetic pea aphid Acyrthosiphon pisum (Hemiptera: Aphidoidea). Dev Genes Evol 217, 275–287.

Chuma, S., Hosokawa, M., Kitamura, K., Kasai, S., Fujioka, M., Hiyoshi, M., Takamune, K., Noce, T., Nakatsuji, N., 2006. Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. Proc Natl Acad Sci USA 103, 15894–15899.

Deng, W., Lin, H., 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. Developmental Cell 2, 819–830.

Ephrussi, A., Dickinson, L.K., Lehmann, R., 1991. Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. Cell 66, 37–50.

Ephrussi, A., Lehmann, R., 1992. Induction of germ cell formation by oskar. Nature 358, 387–392.

Ewen-Campen, B., Donoughe, S., Clarke, D.N., Extavour, C.G., 2013a. Germ Cell Specification Requires Zygotic Mechanisms Rather Than Germ Plasm in a Basally Branching Insect. Current Biology 23, 835–842.

Ewen-Campen, B., Jones, T.E.M., Extavour, C.G., 2013b. Evidence against a germ plasm in the milkweed bug Oncopeltus fasciatus, a hemimetabolous insect. Biology Open 2, 556–568.

Ewen-Campen, B., Schwager, E.E., Extavour, C.G.M., 2010. The molecular machinery of germ line specification. Mol. Reprod. Dev. 77, 3–18.

Ewen-Campen, B., Shaner, N., Panfilio, K.A., Suzuki, Y., Roth, S., Extavour, C.G.M., 2011. The maternal and early embryonic transcriptome of the milkweed bug Oncopeltus fasciatus. BMC Genomics 12, 61.

Ewen-Campen, B., Srouji, J.R., Schwager, E.E., Extavour, C.G., 2012. oskar Predates the Evolution of Germ Plasm in Insects. Curr Biol 22, 2278–2283.

Extavour, C.G., Akam, M., 2003. Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. Development 130, 5869–5884.

Extavour, C.G.M., 2007. Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. Integr Comp Biol 47, 770–785.

Fujiwara, Y., Komiya, T., Kawabata, H., Sato, M., Fujimoto, H., Furusawa, M., Noce, T., 1994. Isolation of a DEAD-family protein gene that encodes a murine homolog of Drosophila vasa and its specific expression in germ cell lineage. Proc Natl Acad Sci USA 91, 12258–12262.

Gallo, C.M., Wang, J.T., Motegi, F., Seydoux, G., 2010. Cytoplasmic Partitioning of P Granule Components Is Not Required to Specify the Germ line in C. elegans. Science 330, 1685–1689.

Gomez, C., Las Heras, de, J.M., Wu, H.-R., Özbudak, E.M., Martinho, R.G., Chen, Y.-T., Wunderlich, J., Lehmann, R., Su, Y.-H., Baumann, D., Casanova, J., Luo, Y.-J., Lewis, J., Holland, L.Z., Pourquié, O., Yu, J.-K., 2011. Asymmetric localization of germ line markers Vasa and Nanos during early development in the amphioxus Branchiostoma floridae. Dev Biol 353, 147–159.

Goss, R.J., 1952. The early embryology of the book louse, Liposcelis divergens Badonnel (Psocoptera; Liposcelidae). Journal of Morphology 91, 135–167.

Hanazawa, M., Yonetani, M., Sugimoto, A., 2011. PGL proteins self associate and bind RNPs to mediate germ granule assembly in C. elegans. The Journal of Cell Biology 192, 929–937.

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P., Nakamura, A., 2008. Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. Nature 451, 730–733.

Harris, A.N., Macdonald, P.M., 2001. Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. Development 128, 2823–2832.

Hegner, R.W., 1908. Effects of removing the germ-cell determinants from the eggs of some chrysomelid beetles. Preliminary report. The Biological Bulletin 16, 19–26.

Hegner, R.W., 1914. The Germ-Cell Cycle in Animals. Macmillan Company, New York.

Heming, B., 1979. Origin and fate of germ cells in male and female embryos of Haplothrips verbasci (Osborn)(Insecta, Thysanoptera, Phlaeothripidae). Journal of Morphology.

Heming, B., Huebner, E., 1994. Development of the germ cells and reproductive primordia in male and female embryos of Rhodnius prolixus Stal (Hemiptera: Reduviidae). Canadian journal of zoology (Canada).

Horne-Badovinac, S., Bilder, D., 2005. Mass transit: Epithelial morphogenesis in theDrosophila egg chamber. Dev. Dyn. 232, 559–574.

Hosokawa, M., Shoji, M., Kitamura, K., Tanaka, T., Noce, T., Chuma, S., Nakatsuji, N., 2007. Tudor-related proteins TDRD1/MTR-1, TDRD6 and TDRD7/TRAP: domain composition, intracellular localization, and function in male germ cells in mice. Dev Biol 301, 38–52.

Huettner, A.F., 1923. The origin of the germ cells in Drosophila melanogaster. Journal of Morphology 39, 249–265.

Illmensee, K., Mahowald, A.P., 1974. Transplantation of posterior polar plasm in Drosophila. Induction of germ cells at the anterior pole of the egg. Proc Natl Acad Sci USA 71, 1016–1020.

Illmensee, K., Mahowald, A.P., 1976a. The autonomous function of germ plasm in a somatic region of the Drosophila egg. Exp Cell Res 97, 127–140.

Illmensee, K., Mahowald, A.P., 1976b. The autonomous function of germ plasm in a somatic region of the Drosophila egg. Exp Cell Res 97, 127–140.

Jiménez-Guri, E., Huerta-Cepas, J., Cozzuto, L., Wotton, K.R., Kang, H., Himmelbauer, H., Roma, G., Gabaldón, T., Jaeger, J., 2013. Comparative transcriptomics of early dipteran development. BMC Genomics 14, 123.

Jiménez-Guri, E., Wotton, K.R., Gavilán, B., Jaeger, J., 2014. A Staging Scheme for the Development of the Moth Midge Clogmia albipunctata. PLoS ONE 9, e84422.

Johnson, A.D., Richardson, E., Bachvarova, R.F., Crother, B.I., 2011. Evolution of the germ line-soma relationship in vertebrate embryos. Reproduction 141, 291–300.

Kim-Ha, J., Kerr, K., Macdonald, P.M., 1995. Translational regulation of oskar mRNA by bruno, an ovarian RNA-binding protein, is essential. Cell 81, 403–412.

Kim-Ha, J., Smith, J.L., Macdonald, P.M., 1991. oskar mRNA is localized to the posterior pole of the Drosophila oocyte. Cell 66, 23–35.

Kuramochi-Miyagawa, S., Anantharaman, V., Kimura, T., Zhang, D., Ijiri, T.W., Aravind, L., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., Lin, H., Matsuda, Y., Nakano, T., 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development 131, 839–849.

Lasko, P.F., Ashburner, M., 1990. Posterior localization of vasa protein correlates with, but is not sufficient for, pole cell development. Genes Dev 4, 905–921.

Lehmann, R., Ephrussi, A., 1994. Germ plasm formation and germ cell determination in Drosophila, in: Germ line Development: Ciba Foundation Symposium 182. John Wiley & Sons, pp. 282–296.

Lehmann, R., Nüsslein-Volhard, C., 1991. The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. Development 112, 679–691.

Liu, Paul, Kaufman, T.C., 2009a. In situ hybridization of large milkweed bug (Oncopeltus) tissues. Cold Spring Harbor Protocols 2009, pdb.prot5262.

Liu, Paul, Kaufman, T.C., 2009b. Morphology and husbandry of the large milkweed bug, Oncopeltus fasciatus. Cold Spring Harbor Protocols 2009, pdb.emo127.

Liu, Paul, Kaufman, T.C., 2009c. Dissection and fixation of large milkweed bug (Oncopeltus) embryos. Cold Spring Harbor Protocols 2009, pdb.prot5261.

Liu, PZ, Kaufman, T., 2004. hunchback is required for suppression of abdominal identity, and for proper germband growth and segmentation in the intermediate germband insect Oncopeltus fasciatus. Development 131, 1515–1527.

Lynch, J.A., Desplan, C., 2010. Novel modes of localization and function of nanos in the wasp Nasonia. Development 137, 3813–3821.

Lynch, J.A., Ozüak, O., Khila, A., Abouheif, E., Desplan, C., Roth, S., 2011. The phylogenetic origin of oskar coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola. PLoS Genet. 7, e1002029.

Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., Surani, M.A., 2013. A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat. Cell Biol. 15, 905–915.

Mahowald, A.P., 2001. Assembly of the Drosophila germ plasm. Int. Rev. Cytol. 203, 187–213.

Markussen, F.H., Michon, A.M., Breitwieser, W., Ephrussi, A., 1995. Translational control of

oskar generates short OSK, the isoform that induces pole plasma assembly. Development 121, 3723–3732.

Marlow, F.L., Marlow, F.L., Mullins, M.C., Mullins, M.C., 2008. Bucky ball functions in Balbiani body assembly and animal-vegetal polarity in the oocyte and follicle cell layer in zebrafish. Dev Biol 321, 40–50.

Matsuda, R., 1976. Morphology and evolution of the insect abdomen. Pegamon Press, Oxford.

Mito, T., Nakamura, T., Sarashina, I., Chang, C.-C., Ogawa, S., Ohuchi, H., Noji, S., 2008. Dynamic expression patterns of vasa during embryogenesis in the cricket Gryllus bimaculatus. Dev Genes Evol 218, 381–387.

Mito, T., Noji, S., 2008. The Two-Spotted Cricket Gryllus bimaculatus: An Emerging Model for Developmental and Regeneration Studies. CSH Protoc 2008, pdb.emo110.

Miura, T., Braendle, C., Shingleton, A., Sisk, G., Kambhampati, S., Stern, D.L., 2003. A comparison of parthenogenetic and sexual embryogenesis of the pea aphid Acyrthosiphon pisum (Hemiptera: Aphidoidea). J. Exp. Zool. B Mol. Dev. Evol. 295, 59–81.

Nakaki, F., Hayashi, K., Ohta, H., Kurimoto, K., Yabuta, Y., Saitou, M., 2013. Induction of mouse germ-cell fate by transcription factors in vitro. Nature 501, 222–226.

Nieuwkoop, P.D., Sutasurya, L.A., 1981. Primordial germ cells in the invertebrates: from epigenesis to preformation. Cambridge University Press, Cambridge.

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., Saitou, M., 2009. A Signaling Principle for the Specification of the Germ Cell Lineage in Mice. Cell 137, 571–584.

Roonwal, M.L., 1937. Studies on the embryology of the African migratory locust, Locusta migratoria migratorioides Reiche and Frm.(Orthoptera, Acrididae). II. Organogeny. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 227, 175–244.

Saitou, M., Yamaji, M., 2012. Primordial germ cells in mice. Cold Spring Harb Perspect Biol 4.

Schüpbach, T., Wieschaus, E., 1986. Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo. Dev Genes Evol 195, 302–317.

Seydoux, G., Braun, R.E., 2006. Pathway to totipotency: lessons from germ cells. Cell 127, 891–904.

Singh, J.P., 1967. Early embryonic development of gonads in Labidura riparia (Pallas) (Labiduridae: Dermaptera). Agra University Journal of Research 16, 67–76.

Smith, J.L., Wilson, J.E., Macdonald, P.M., 1992. Overexpression of oskar directs ectopic activation of nanos and presumptive pole cell formation in Drosophila embryos. Cell 70, 849–859.

Tada, H., Mochii, M., Orii, H., Watanabe, K., 2012. Ectopic formation of primordial germ cells by transplantation of the germ plasm: direct evidence for germ cell determinant in Xenopus. Dev Biol 371, 86–93.

Tam, P.P., Zhou, S.X., 1996. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. Dev Biol 178, 124–132.

Toyooka, Y., Tsunekawa, N., Takahashi, Y., Matsui, Y., Satoh, M., Noce, T., 2000. Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. Mech Dev 93, 139–149.

Trautwein, M.D., Wiegmann, B.M., Beutel, R., Kjer, K.M., Yeates, D.K., 2012. Advances in insect phylogeny at the dawn of the postgenomic era. Annu. Rev. Entomol. 57, 449–468.

Tsuda, M., 2003. Conserved Role of nanos Proteins in Germ Cell Development. Science 301, 1239–1241.

Tsunekawa, N., Naito, M., Sakai, Y., Nishida, T., Noce, T., 2000. Isolation of chicken vasa homolog gene and tracing the origin of primordial germ cells. Development 127, 2741–2750.

Updike, D.L., Hachey, S.J., Kreher, J., Strome, S., 2011. P granules extend the nuclear pore complex environment in the C. elegans germ line. The Journal of Cell Biology 192, 939–948.

Wang, J.T., Seydoux, G., 2013. Germ cell specification. Adv. Exp. Med. Biol. 757, 17–39.

Wheeler, W.M., 1893. A contribution to insect embryology. Journal of Morphology VIII, 1–160.

Wiegmann, B.M., Trautwein, M.D., Kim, J.-W., Cassel, B.K., Bertone, M.A., Winterton, S.L., Yeates, D.K., 2009. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. BMC Biol. 7, 34.

Wilson, E.B., 1900. The Cell in Development and Inheritance, 2nd ed. Macmillan Company, London.

Xu, X., Brechbiel, J.L., Gavis, E.R., 2013. Dynein-dependent transport of nanos RNA in Drosophila sensory neurons requires Rumpelstiltskin and the germ plasm organizer Oskar. J. Neurosci. 33, 14791–14800.

Yoon, C., Kawakami, K., Hopkins, N., 1997. Zebrafish vasa homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells. Development 124, 3157–3165.

Zeng, V., Ewen-Campen, B., Horch, H.W., Roth, S., Mito, T., Extavour, C.G., 2013. Developmental Gene Discovery in a Hemimetabolous Insect: De Novo Assembly and Annotation of a Transcriptome for the Cricket Gryllus bimaculatus. PLoS ONE 8, e61479.

Zeng, V., Villanueva, K.E., Ewen-Campen, B.S., Alwes, F., Browne, W.E., Extavour, C.G., 2011. De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiensis. BMC Genomics 12, 581.

Zimyanin, V.L., Belaya, K., Pecreaux, J., Gilchrist, M.J., Clark, A., Davis, I., St Johnston, D., 2008. In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization. Cell 134, 843–853.

Ewen-Campen et al. BMC Genomics 2011, 12:61
http://www.biomedcentral.com/1471-2164/12/61

CHAPTER 1

BMC Genomics

RESEARCH ARTICLE Open Access

The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*

Ben Ewen-Campen[1], Nathan Shaner[2], Kristen A Panfilio[3], Yuichiro Suzuki[4], Siegfried Roth[3], Cassandra G Extavour[1*]

Abstract

Background: Most evolutionary developmental biology ("evo-devo") studies of emerging model organisms focus on small numbers of candidate genes cloned individually using degenerate PCR. However, newly available sequencing technologies such as 454 pyrosequencing have recently begun to allow for massive gene discovery in animals without sequenced genomes. Within insects, although large volumes of sequence data are available for holometabolous insects, developmental studies of basally branching hemimetabolous insects typically suffer from low rates of gene discovery.

Results: We used 454 pyrosequencing to sequence over 500 million bases of cDNA from the ovaries and embryos of the milkweed bug *Oncopeltus fasciatus*, which lacks a sequenced genome. This indirectly developing insect occupies an important phylogenetic position, branching basal to Diptera (including fruit flies) and Hymenoptera (including honeybees), and is an experimentally tractable model for short-germ development. 2,087,410 reads from both normalized and non-normalized cDNA assembled into 21,097 sequences (isotigs) and 112,531 singletons. The assembled sequences fell into 16,617 unique gene models, and included predictions of splicing isoforms, which we examined experimentally. Discovery of new genes plateaued after assembly of ~1.5 million reads, suggesting that we have sequenced nearly all transcripts present in the cDNA sampled. Many transcripts have been assembled at close to full length, and there is a net gain of sequence data for over half of the pre-existing *O. fasciatus* accessions for developmental genes in GenBank. We identified 10,775 unique genes, including members of all major conserved metazoan signaling pathways and genes involved in several major categories of early developmental processes. We also specifically address the effects of cDNA normalization on gene discovery in *de novo* transcriptome analyses.

Conclusions: Our sequencing, assembly and annotation framework provide a simple and effective way to achieve high-throughput gene discovery for organisms lacking a sequenced genome. These data will have applications to the study of the evolution of arthropod genes and genetic pathways, and to the wider evolution, development and genomics communities working with emerging model organisms.

[The sequence data from this study have been submitted to GenBank under study accession number SRP002610 (http://www.ncbi.nlm.nih.gov/sra?term=SRP002610). Custom scripts generated are available at http://www.extavourlab.com/protocols/index.html. Seven Additional files are available.]

Background

New and emerging model organisms occupy an increasingly important part of the developmental biology and developmental genetics research landscape. While studying a huge diversity of animals has long been the norm in the classical fields of experimental embryology and functional morphology [see for example [1-3]], the molecular biology revolution and the advent of the "model system" concept [4] created demand for a small number of highly genetically manipulable organisms that could be intensively studied [5]. Research on these "big six" [sensu 6] genetic model organisms has led to enormous advances in our understanding of general principles of embryogenesis. However, placing these general principles in an evolutionary context requires broader taxonomic sampling. Many researchers have

* Correspondence: extavour@oeb.harvard.edu
[1]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
Full list of author information is available at the end of the article

highlighted the need for developing new model organisms for specific comparative, evolutionary and ecological questions [6-8]. It has also been suggested, however, that the single gene expression approach of the last several decades of evolutionary developmental biology ("evo-devo") has outlived its usefulness, and that what are needed are not more model organisms, but rather a smaller number of groups chosen for the ability to functionally manipulate genes [9,10]. Sophisticated gene expression techniques and even stable germline transgenesis have been developed in a large array of models outside of the "big six" [see for example [11,12]]. The ancient history of the small RNA processing machinery [13,14] means that gene knockdown is a feasible goal for most organisms, as long as the sequences of genes of interest are available.

While whole genome sequencing is an increasingly viable option for some organisms, many new models, particularly within the arthropods, lack the large community resources necessary to finance and maintain annotation of a genome. For these reasons, many researchers studying non-traditional model organisms have turned to Sanger-sequenced EST libraries [see for example [15,16]]. In principle this method of gene discovery can lead to high-throughput expression and functional genetic analyses of multiple genes [see for example [17]]. In practice, however, most non-traditional organism studies are still subject to a gene discovery bottleneck. This is largely because at the scale needed to uncover rare developmental transcripts, Sanger-based EST sequencing quickly becomes technically and financially prohibitive for many labs working on organisms with smaller research communities. In addition, those smaller-scale EST projects that have been carried out are often not publically available in easily searchable formats, and their potential contribution to the developmental and evolutionary biology fields is thus limited.

Next-generation sequencing (NGS) offers comparative and evolutionary developmental biologists a way to obtain orders of magnitude more developmental gene data than ever before, at a fraction of its former cost. Several studies have demonstrated the feasibility of NGS for identifying SNPs for population studies and gene sequences for use as phylogenetic markers [18-35]. Unfortunately, the lack of suitable protocols for cDNA preparation, and of established pipelines for analysis have left this tool under-utilized by many evo-devo researchers. Furthermore, according to some estimates [35], few of these studies have been carried out at a scale large enough to provide significant recovery of rare transcripts, and therefore of developmental genes. Here we present an optimized protocol for synthesizing cDNA for 454 Titanium pyrosequencing, as well as a simple workflow for *de novo* assembly of the data without a reference genome, annotation and analysis of the dataset, and a demonstration of its utility for comparative developmental genetics.

A large body of literature is dedicated to the development and genomics of holometabolous insects (insects undergoing complete metamorphosis between embryonic and adult stages). Tens of holometabolous insect genomes are now available, thanks largely to work on *Drosophila melanogaster*, other drosophilids, and dipteran disease vectors [36,37]. In contrast, relatively little is known about the development of hemimetabolous insects, which undergo incomplete metamorphosis. Although several of these insects are amenable to laboratory culture and a variety of experimental manipulations, molecular developmental studies are scarce, and gene discovery rates remain low. Notable exceptions among the Hemiptera are the aphid *Acyrthosiphon pisum* and the Chagas' disease vector *Rhodnius prolixus*, whose genomes are completed and in progress respectively [38,39]. However, the aphid genome has undergone extensive duplications and gene loss, possibly due to its unusual reproductive and ecological characteristics [38]. The mammalian blood feeding needs of *R. prolixus* make it a sub-optimal organism for developmental studies.

The milkweed bug *Oncopeltus fasciatus* (Figure 1A-D) has emerged as a promising hemipteran system for studying the molecular development of hemimetabolous insects [40-42]. It can be reared easily and cheaply in the laboratory, and has a long history as a laboratory animal for classical embryology and pattern formation studies [43-45]. More recently, robust protocols for *in situ* hybridization, live imaging of embryogenesis, and RNAi-mediated gene knockdown have been developed and successfully applied to the study of the evolution of development [see for example [46,47]].

Here we present the results of the sequencing and *de novo* assembly of the *Oncopeltus* ovarian and early embryonic transcriptome. We outline an assembly and analysis framework using a combination of existing tools and freely available custom-made command line computational tools, which we hope will make this approach to gene discovery accessible to comparative developmental biologists. We identify homologues of genes involved in all major signaling pathways and developmental processes, including biologically verified splicing isoforms for some genes. We also address the need for library normalization in these studies, and show that at large enough scales of NGS, large numbers of developmental genes can be discovered even with omission of a normalization step.

Results and Discussion

Assembling the ovarian and embryonic transcriptome of O. fasciatus

We prepared cDNA from ovaries and early to mid-staged embryos of *O. fasciatus*, covering oogenesis and all major

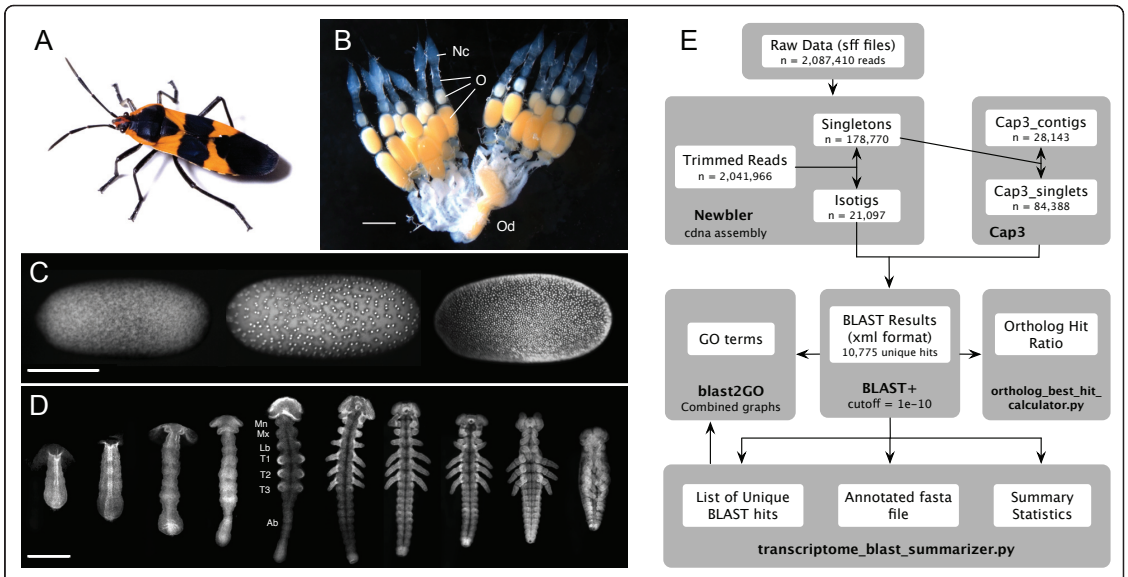


Figure 1 Introduction to *Oncopeltus fasciatus* and the workflow for producing a *de novo* transcriptome assembly. (*A*) An adult milkweed bug, *Oncopeltus fasciatus*. (*B*) Ovaries of adult female. Anterior is up. Oocytes (O) are visible in progressive stages of growth before reaching a common oviduct (Od). Oocytes are cytoplasmically connected to nurse cells (Nc) in the anterior of each ovariole. Scale bar = 1.0 mm. (*C-D*) The stages of *O. fasciatus* embryogenesis represented in this transcriptome. Embryos are stained with Sytox Green (Invitrogen) to visualize nuclei. Scale bars = 0.5 mm. (*C*) Development proceeds from left to right. Anterior is to the left. The cellularized blastoderm forms during the first ~20% of development (~0-24 hours at 28°C), as nuclei reach the surface of the yolk and repeatedly divide. (*D*) Germ band extension and segmentation occur from ~20-60% of development (~24-72 hours at 28°C). Development proceeds from left to right. Anterior is up. Mn = mandibular segment; Mx = maxillary segment; Lb = labial segment; T1-T3 = leg-bearing thoracic segments 1-3; Ab = abdomen. (*E*) The flow of information during this *de novo* transcriptome assembly project. Data files are represented as white boxes within grey boxes that indicate the computer programs used to generate these files. All of the computer programs used are freely available. Ortholog_best_hit_calculator.py and transcriptome_blast_summarizer.py are custom python scripts available at http://www.extavourlab.com/protocols/index.html (see text for details). Photograph in (*A*) courtesy of David Behl.

stages of embryonic patterning (Figure 1B-D). These cDNA samples were prepared using a protocol optimized for preparation of small or limiting samples for 454 pyrosequencing (see Materials and Methods). From these libraries, we generated a total of 2,087,410 sequence reads (Table 1). This includes reads generated using GS-FLX technology as well as both normalized (N) and non-normalized (NN) cDNA sequenced using the GS-FLX Titanium platform. As expected, the reads generated using GS-FLX Titanium technology were substantially longer than those generated using GS-FLX technology (Table 1, Figure 2A). However, the N sample gave an unexpectedly low number of reads, which were on average shorter than those generated by the NN sample (Table 1; Figure 2A). Given that a pilot run of one lane (1/8 plate) of this same normalized cDNA sample generated roughly equal number and size-distribution as a NN pilot study (Additional file 1), we suspect that a technical error reduced the sequencing efficiency of this plate. Despite the comparatively low yield of this normalized cDNA, it still generated more than 600,000 high quality reads that we therefore included in subsequent analyses.

We used the cDNA assembly algorithm of Newbler v2.3 (Roche) to screen the reads for adaptor sequence

Table 1 Sources of *O. fasciatus* sequence reads

Tissue	Normalized?	cDNA prep	454 Platform	No. Plates	No. Reads	Median Read Length	Accession #
Ovary	Y	SMART	GS-FLX	¼	65,394	225	SRR057570.2
Embryonic	Y	SMART	GS-FLX	¼	71,911	230	SRR057571.1
Ovarian and Embryonic	Y	Modified SMART	GS-FLX Titanium	1 + ¼	656,782	244	SRR057572.1
Ovarian and Embryonic	N	Modified SMART	GS-FLX Titanium	1 + 1/8	1,293,323	313	SRR057573.1
Total				**2 + 7/8**	**2,087,410**	**301**	SRP002610.1

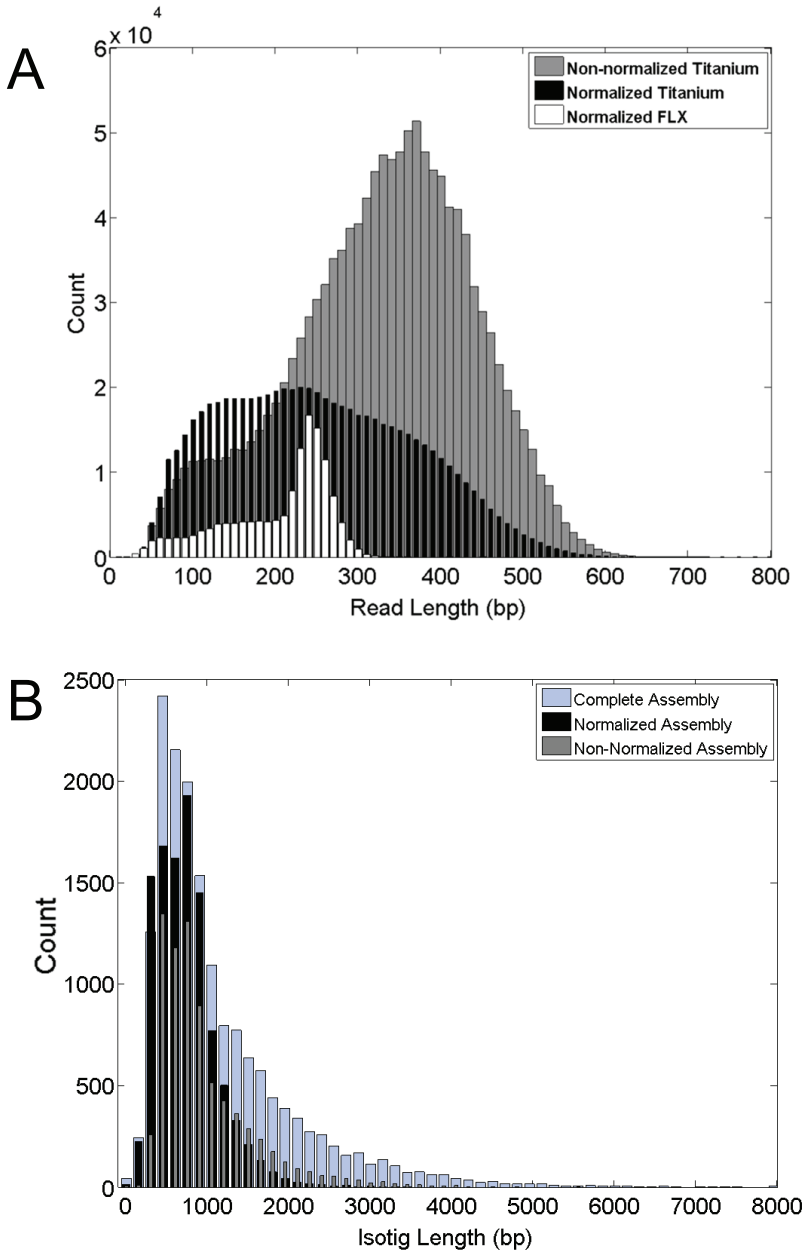


Figure 2 Effects of normalization and 454 sequencing chemistry on read length and isotig length. (*A*) Titanium sequencing chemistry (grey, black) generally results in longer read lengths when compared with FLX chemistry (white). However, the normalized sample run with Titanium chemistry (black) had shorter read lengths than the non-normalized sample (grey). This result is likely due to a technical error in that particular sequencing run, since a 1/8 plate run of the same sample showed a read length distribution comparable to that of the non-normalized sample (Additional file 1). (*B*) Isotig length distributions from assemblies of Titanium-sequenced data. The longest isotig per isogroup is shown. The number of bases in the non-normalized (grey) and normalized (black) samples has been equalized to eliminate possible bias due to the greater number and length of reads obtained from the run of the normalized sample (see (*A*)). The isotigs generated from the normalized cDNA tended to be shorter than those produced by the non-normalized cDNA (see also Table 2). Pooling all FLX and Titanium reads generates an assembly with more, longer isotigs (blue).

and then assemble the cleaned reads (see Note Added in Proof for a comparison with Newbler v2.5). After quality trimming and adapter screening, 2,041,966 reads (97.8%) were used in the assembly. Of these, 1,773,450 (86.9%) assembled either wholly or partially into contigs, and 178,770 (8.8%) remained as singletons. The remaining reads were excluded as either originating from repeat regions (9,875 reads; 0.05%), outliers (26,943 reads; 1.3%), or too short (<50 base pairs: 52,928 reads; 2.6%).

To our knowledge, Newbler v2.3 and higher are the only assembly programs that address alternative splicing and can output multiple isoforms per gene. Newbler v2.3 explicitly accounts for alternative splicing by creating a hierarchical assembly composed of three elements: contigs, isotigs, and isogroups. For consistency, we follow their terminology. Contigs are stretches of assembled reads that are free of branching conflicts. In other words, contigs can be thought of as exons or sets of exons that are always co-transcribed. Isotigs represent a particular continuous path through a set of contigs, i.e. a transcript. An isogroup is the set of isotigs arising from the same set of contigs, i.e. a gene. Different isotigs within an isogroup are thought to represent alternative isoforms of the same gene. Note that it is possible for an isogroup to contain only one isotig, and it is also possible for an isotig to be composed of only one contig.

After the initial Newbler assembly, we noticed substantial redundancy among the singletons. We therefore subjected the 178,770 unassembled singletons to a secondary assembly with CAP3 [48]. This secondary assembly reduced the number of singletons from 178,770 to 112,531 (28,143 cap3_contigs and 84,388 cap3_singlets). Thus, in total, our assembly generated a total of 133,628 sequences, including isotigs, cap3_contigs and cap3_singlets (Table 2).

Our data assembled into 22,235 contigs, organized among 21,097 isotigs (Figure 2B). The isotig N50 length was 1,735 bp (in other words, 50% of the bases are incorporated into isotigs ≥ 1,735 bp), and 14,460 (68.5%) of the isotigs contained only one contig. The 21,097 isotigs fell into 16,617 isogroups, of which 14,562 (87.6%) contain only one isotig (average number of isotigs per isogroup = 1.3).

The average coverage among contigs was 23.2 reads/bp (median coverage = 6.9 reads/bp) (Additional file 2). This coverage value is more than twice as high as the highest reported value from a *de novo* transcriptome assembly to date [summarized in [20]]. Such deep coverage should be helpful for overcoming the presence of insertion/deletion errors in the individual raw reads [49].

To test whether our assembly would have been aided by the inclusion of nucleotide sequence from *Rhodnius prolixus*, the most closely related hemipteran to *O. fasciatus* whose genome is currently being sequenced [39], we used the BLASTN algorithm to compare our isotigs (the longest isotig per isogroup) with the published ESTs of *R. prolixus* with an e-value cut-off of 1e-6. Consistent with previous observations of extremely low levels of conservation between insect genomes [50] we found that only 53 out of 16,617 isotigs had hits to *R. prolixus* ESTs on the nucleotide level. These results suggest that *de novo* sequencing and assembling efforts will be necessary for most insect species, even when sequence data are available for other members of the same order. We note, however, that a recent study [51] has shown that it may be possible to incorporate EST data from different species into a *de novo* assembly by using amino acid sequence rather than nucleotide sequence.

Validation of predicted alternate isoforms

To examine whether the alternative isoforms predicted by Newbler v2.3 are in fact present in developing embryos of *O. fasciatus*, we first focused on a gene of particular interest to developmental biologists, *nanos*. This conserved metazoan gene was first described as a loss of function mutation in *Drosophila melanogaster* [52], and is necessary for germ cell and posterior somatic development [reviewed in [53]]. Newbler v2.3 predicted this gene to encode two alternative isotigs within a single isogroup (Figure 3B). The two isotigs differ in that the

Table 2 *O. fasciatus* transcriptome assembly statistics

	Full Assembly	Normalized Assembly	Non-Normalized Assembly
Assembled reads (base pairs)	1,773,450 (508,738,047)	389,605 (84,353,140)	336,568 (108,372,883)
Isogroups ("genes")	16,617	10,581	7,591
Isotigs ("transcripts")	21,097	11,353	8,346
Isotig N50	1,735	846	1,162
Mean # isotigs per isogroup	1.3	1.1	1.1
Contigs ("exons")	22,235	11,839	8,731
Mean # contigs per isotig	1.9	1.2	1.3
Singletons (singletons after secondary CAP3 assembly)	178,770 (112,531)	110,265 (N/A)	52,585 (N/A)

To enable comparison, we equalized individual assemblies of Normalized and Non-Normalized samples to contain the same number of base pairs before assembly.

longer contains an additional 100-bp exon that is absent from the shorter (Figure 3B). We designed PCR primers against sequences present in both isotigs (Figure 3B arrows), which amplified two bands differing by ~100 bp from a pool of embryonic cDNA (Figure 3C). Sequencing of these two bands confirmed that they differ exactly as predicted by Newbler v2.3 (Figure 3D).

Importantly, a previous version of Newbler (v2.0), which does not account for alternative splicing, failed to join together the three fragments which were linked by Newbler v2.3 (Figure 3A). Because of this, Newbler v2.0 (and presumably other assemblers which do not address branching within contigs) predicted three separate contigs, only one of which could be identified as *nanos* with BLASTX, as the others fall in poorly conserved regions of the gene. Thus, the ability of Newbler2.3 to handle branching conflicts between reads allows this program to assemble longer continuous sequences, which are therefore in turn more easily annotated using BLAST.

To further characterize the accuracy of Newbler's predictions of alternative transcript isoforms, we randomly selected 10 isogroups that contained exactly two alternative isotigs differing by the presence/absence of a single contig (Additional file 3). As we did for *nanos*, we designed primers to flank the region differing between the two predicted isoforms (Additional file 3A), and performed RT-PCR on *O. fasciatus* embryonic cDNA. In eight of ten instances, we observed bands of the predicted sizes following agarose gel electrophoresis (Additional file 3B,C). However, in four of the eight positive cases, additional, unpredicted bands were present (Additional file 3). In one of the ten cases, we observed two RT-PCR products, but only one of them was of the predicted size (Additional file 3C, lane 6). Taken together, these results suggest that Newbler v2.3 has a low rate of false positives in the prediction of multiple splicing isoforms. Including our investigation of *nanos*, only one of 11 test cases (9.1%) produced a single RT-PCR product where Newbler v2.3 had predicted multiple products.

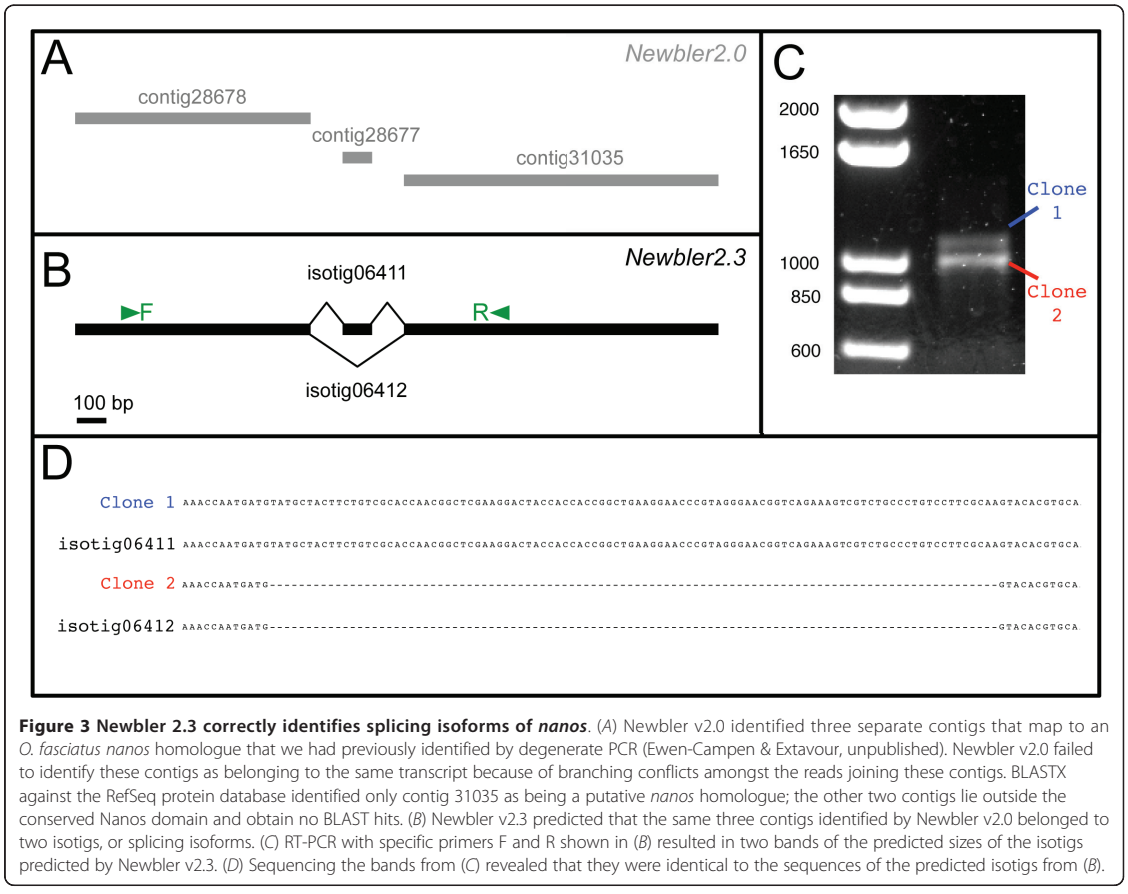


Figure 3 Newbler 2.3 correctly identifies splicing isoforms of *nanos*. (*A*) Newbler v2.0 identified three separate contigs that map to an *O. fasciatus nanos* homologue that we had previously identified by degenerate PCR (Ewen-Campen & Extavour, unpublished). Newbler v2.0 failed to identify these contigs as belonging to the same transcript because of branching conflicts amongst the reads joining these contigs. BLASTX against the RefSeq protein database identified only contig 31035 as being a putative *nanos* homologue; the other two contigs lie outside the conserved Nanos domain and obtain no BLAST hits. (*B*) Newbler v2.3 predicted that the same three contigs identified by Newbler v2.0 belonged to two isotigs, or splicing isoforms. (*C*) RT-PCR with specific primers F and R shown in (*B*) resulted in two bands of the predicted sizes of the isotigs predicted by Newbler v2.3. (*D*) Sequencing the bands from (*C*) revealed that they were identical to the sequences of the predicted isotigs from (*B*).

However, we observed that roughly half of the time, Newbler v2.3 failed to predict all of the isoforms identified via RT-PCR.

Transcriptome annotation

A BLASTN search of our dataset for the 93 existing GenBank accessions for *O. fasciatus* sequences yielded a hit result for 56% of the accessions, with an e-value cut-off of 1e-10. This result may be due in part to the short length of some of the GenBank sequences. Accordingly, we found that accessions with hits in the database were significantly longer (mean length 729 bp) than accessions without hits (mean length 397 bp) (unpaired Student's *t*-Test: $t = 2.89$, DF = 91, $p = 0.0048$). Of greater relevance to developmental applications of this dataset, however, was our finding that 85% of *O. fasciatus* developmental genes with existing GenBank accessions (n = 32) are represented in our transcriptome.

We then used BLASTX to map the 133,628 *O. fasciatus* sequences (isotigs, cap3_contigs and cap3_singletons) against the entire RefSeq Protein database with an e-value cut-off of 1e-10. To simplify these statistics, we report only the BLAST results for the longest isotig per isogroup, under the assumption that all isotigs within an isogroup share nearly identical BLAST results. Of 16,617 isotigs, 7,219 (43.4%) had at least one hit. Of the 28,143 cap3_contigs, 2,594 (9.2%) had hits, and of the 84,388 cap3_singlets, 2,367 (2.8%) had hits. These values are higher than comparable BLAST statistics of most other published studies of 454-generated *de novo* transcriptomes [24-26,30,32,33], likely because deeper sequencing increases the length of assembled sequences and thereby makes these sequences more likely to be identified via BLAST. The unidentifiable sequences likely originate from UTRs or non-conserved portions of protein-coding sequences. Of the top BLAST hits, 89.3% were genes from arthropod sequences (Additional file 4). Of the 12,180 *O. fasciatus* sequences with BLAST hits, 1,455 hit non-overlapping segments of the same top BLAST hit (i.e. potentially unassembled portions of the same transcript), and 825 hit overlapping segments of the same top BLAST hit (i.e. potential paralogs). Excluding those 1,455 potentially double-counted BLAST hits, our transcriptome identified a total of 10,775 genes. The assembled sequences generated in this study, as well as pre-computed BLAST results, are available as flat files from the authors upon request.

To explore and summarize the functional categories of the genes sequenced in this study, we obtained the Gene Ontology (GO) terms associated with the top 20 BLAST hits of each sequence using Blast2GO [54]. Among the 7,059 genes for which we obtained GO terms, we observed a wide diversity of functional categories represented on all levels of the Gene Ontology database (Figure 4). The *O. fasciatus* sequences fall into GO categories with a roughly similar distribution to that of the well-annotated *Drosophila melanogaster* genome, suggesting that our sequence data contain a large diversity of genes involved in a variety of biological processes, and do not contain any notable biases towards particular categories of genes.

Assessing coverage of the O. fasciatus *transcriptome*

We wished to know how thoroughly our sequencing efforts sampled the true diversity of transcripts present in our cDNA samples. This is a two-part question: first, of the genes truly expressed during *O. fasciatus* oogenesis and embryogenesis, how many did we identify? And second, of these identified genes, how thoroughly had we assembled their full-length transcripts?

To address the first question, we created eight separate assemblies of progressively larger sub-samples of our total reads and tallied the total number of genes identified via BLASTX. The number of newly discovered genes began to plateau after ~1.5M reads (1 7/8 plates in our case) (Figure 5 black line). However, the N50 isotig length continued to increase roughly linearly over this range of reads (Figure 5 grey line). These results suggest that additional sequencing of this sample is unlikely to identify substantially more genes, but may continue to lengthen the existing sequences. Although in the absence of a sequenced genome it is not possible to accurately estimate how many genes are in fact present in the *O. fasciatus* transcriptome, we note that while several developmental genes of interest were identified in this study, others were not. (Tables 3, 4 and see below). Because these data suggest that we have sequenced these specific cDNA samples quite deeply, some form of specific target enrichment may be necessary for future attempts to discover additional genes not identified in this dataset.

To address the second question, we employed a method proposed by O'Neil and colleagues [20] for addressing the question of how closely our sequences approached full-length transcripts. Their metric, the "ortholog hit ratio," compares the length of the newly discovered sequence that obtains a BLAST hit versus the full length of its top hit [20]. Thus, an ortholog hit ratio of one implies that a transcript has been assembled to its true full length, while values over one suggest insertions in the query sequence relative to its top BLAST hit. We note the caveat that many genes contain relatively poorly conserved regions that may fail to obtain a BLAST hit at all, causing the ortholog hit ratio to be an underestimate in these cases (Additional file 5). In our dataset, many of the *O. fasciatus* isotigs appear to be nearly fully assembled, while the singletons predictably tend to represent small portions of their top

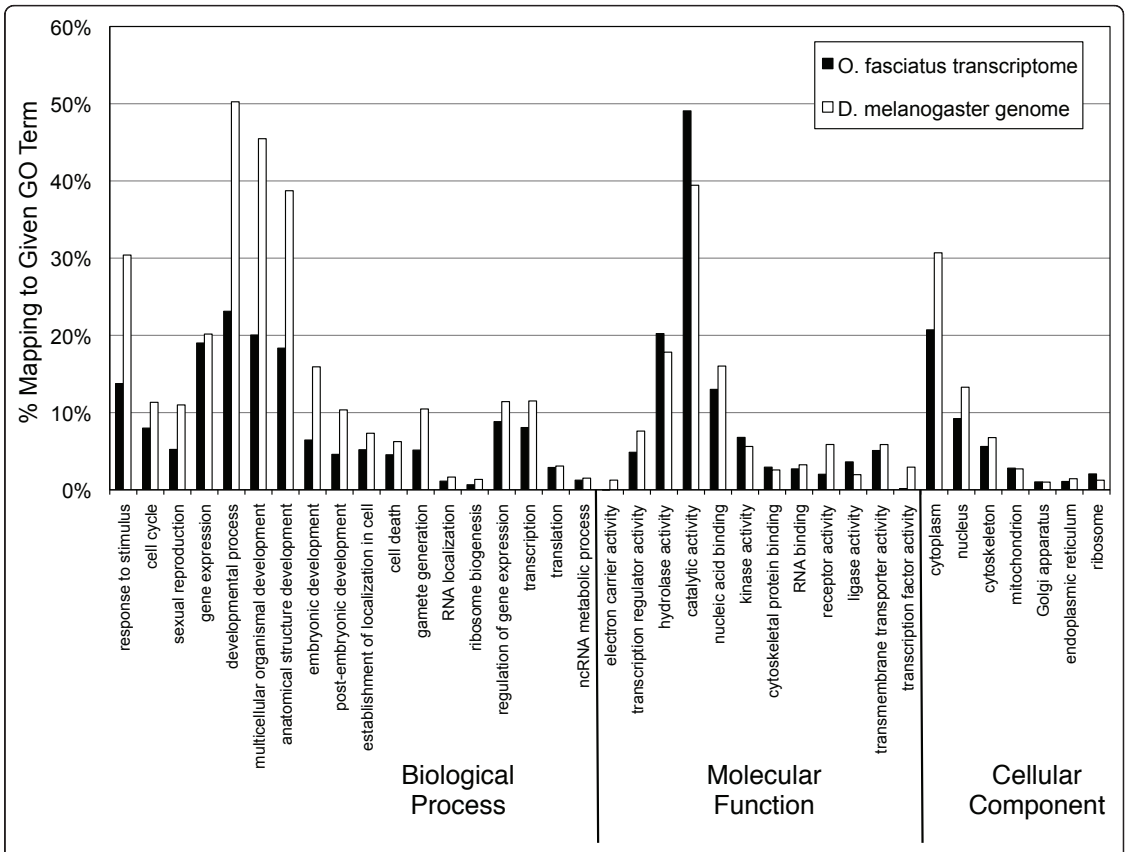


Figure 4 GO term distribution of BLAST hits from the *O. fasciatus* transcriptome compared with those from the *D. melanogaster* genome. Several GO categories are shown within the top-level divisions of Biological Process, Molecular Function, and Cellular Component. Column heights reflect the percentage of annotated sequences in each assembly that mapped to a given Biological Process GO term. The relative percentages of genes falling into GO categories are comparable between our *O. fasciatus* transcriptome (black) and the *D. melanogaster* transcriptome (white).

BLAST hit in RefSeq (Figure 6). In total, of the 7,219 isotigs with BLAST hits, 3,953 (54.8%) had ratios > 0.5 and 2,689 (37.2%) had ratios > 0.8.

We also asked, for those *O. fasciatus* sequences of developmental genes already present in GenBank that overlapped with transcriptome hits (n = 23), whether our transcriptome data provided any net gain in transcript sequence compared to the GenBank accession sequence. In 15/23 cases (68%), the transcriptome data extended the known sequence beyond that reported in GenBank by an average of 349 bp (range: 82-1,366 bp). In most cases, additional 3' sequence was obtained (Figure 7).

Assessing the value of cDNA normalization

Reducing the representation of highly abundant transcripts (i.e. normalizing the cDNA) is often considered essential to capture sequence from genes expressed at lower levels, including many important developmental genes [see for example [55-57]]. However, we hypothesized that current next-generation sequencing technologies could provide sufficiently deep sequence to render normalization largely unnecessary for construction of *de novo* transcriptomes for comparative developmental biologists. To address this question, we assessed the relative contribution of the N and NN cDNA to our final assembly using several strategies.

First, to test whether our normalization protocol successfully reduced the presence of highly abundant transcripts, we created separate assemblies from the N and NN cDNA samples (equalizing the total number of bases to reduce the contribution of additional sequence found in the NN sample). The N assembly contained a greater number of isotigs that were shorter on average

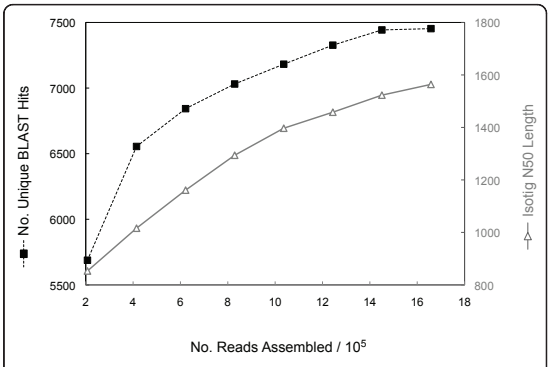


Figure 5 Assessing coverage of the *O. fasciatus* transcriptome. Randomly chosen subsets of increasing numbers of Titanium reads were used to generate progressively larger sub-assemblies. The number of reads in each sub-assembly (X axis) is plotted against the number of unique BLAST hits in each sub-assembly (left Y axis: black), and against the N50 isotig length (right Y axis: grey). For this analysis BLAST was performed against the SwissProt database. The number of unique BLAST hits plateaus when the assembly is composed of approximately 1.5 million reads. However, the N50 isotig length maintains an approximately constant rate of increase.

than those in the NN assembly (Figure 2B). Additionally, more singletons were generated in the N assembly relative to the NN assembly (Table 2). Further, similar to the results obtained by Bellin and colleagues [27], we observed the predicted decrease in the maximum number of reads per contig in the N assembly compared to the NN assembly (Figure 8A, B), demonstrating that the normalization procedure successfully reduced the sequencing of highly abundant transcripts. These statistics, which could be interpreted to suggest that the N reads generated an inferior assembly, may result from the shorter average length of reads in the N sample (Figure 2A). Indeed, Newbler rejected 7.9% (30,780) of the N reads as too short, compared to only 1% (3,935) of the NN reads. However, these assembly statistics could also indicate greater heterogeneity in the N sample, which would suggest that normalization might increase the number of new genes identified.

To discriminate between these possibilities, we explored the contribution of the N and NN reads to the genes discovered in our full assembly. We used BLASTN to map one plate's worth of raw reads from the N sample and from the NN sample (equalized to contain the same number of base pairs) against the complete assembled transcriptome, with an e-value cutoff of 1e-4. We then explored the GO annotation of those genes hit exclusively by only one of these two samples. We observed similar overall GO term distributions between the N and NN samples (Figure 8C). We found that a small number of GO terms (n = 20) were significantly differentially represented in the two samples, albeit generally with very few sequences in each GO term (Additional file 6). For example, we were surprised to see that three of the four terms statistically over-represented in the N sample were related to ribosome function (14/750 (1.9%) of the N hits were annotated with 'ribosomal subunit', compared to 1/1124 (0.09%) NN hits; FDR-corrected p-value = 0.006). In contrast, several terms related to active transmembrane transport were over-represented in the NN sample (Additional file 6) possibly indicating that normalization may have reduced the representation of genes involved in certain basic metabolic processes.

As an additional way to investigate the contribution of the N and NN samples to identifying specific genes of interest for our studies, we manually examined the results of mapping the N and NN samples to the fully assembled transcriptome. Of the 79 genes of interest that we investigated, four (5.1%) were uniquely present in the N sample, whereas nine (11.4%) were uniquely present in the NN sample, and the remaining 66 (83.6%) were present in reads of both the N and NN samples (Tables 3, 4). Although this may be an artifact of sequencing depth (i.e. low-abundance genes of interest may be present in only one of the two cDNA samples simply due to sampling effects rather than the normalization protocol *per se*), our data suggest that the normalized cDNA sample did not contribute disproportionately to gene discovery.

Gene discovery for developmental studies

The ultimate goal of this sequencing project was to identify a wide diversity of candidate genes involved in developmental processes. Traditionally, such gene discovery in "non-model" organisms has required degenerate PCR, which is labor-intensive, expensive, and prone to failure. The annotated transcriptome assembly we present here allows researchers to identify genes of interest via simple text searches, or via BLAST searches. To demonstrate the usefulness of these data for large-scale gene discovery, we report here the identification of several components from each of the seven widely studied metazoan signaling pathways (Table 3) as well as many genes involved in specific developmental processes (Table 4). We note that the majority of these gene fragments are of suitable length for immediate application of such widely used techniques as *in situ* hybridization and RNAi-based functional knockdown. In cases of functional experiments where full-length proteins are desirable, such as protein overexpression, RACE PCR will likely be required. Importantly, we note that many genes of interest were present among the singletons, many of which are long enough for immediate use as sequences for *in situ* hybridization probes or RNAi

Table 3 Selected signaling pathway genes identified in the *O. fasciatus* transcriptome

				Present in:	
Pathway	**# Hits**	**Hit ID (I/C/S)**	**Length (range)**	**Normalized**	**Non-Normalized**
HEDGEHOG					
cubitus interruptus	3	I,S	225-906	Y	Y
fused	2	I	516-1582	Y	Y
patched	2	C, S	225-418	N	Y
smoothened	2	I	1270-1604	Y	Y
JAK/STAT					
domeless	1	I	4028	Y	Y
hopscotch (janus kinase)	3	I, C	473-2644	Y	Y
Signal transducer and activator of transcription	4	I	444-3270	Y	Y
NFKB/TOLL					
cactus	7	I, C	629-1748	Y	Y
dorsal (Nuclear factor NF-kappa-B)	2	I	1308-3926	Y	Y
relish	1	I	2650	Y	Y
Toll	11	I, C, S	215-4323	Y	Y
NOTCH					
fringe	1	I	877	Y	Y
Hairless	1	I	1053	Y	Y
hairy (Enhancer of split/HES-1)	1	I	2530	Y	Y
mind bomb	7 (6†)	I,C,S	335-1185	Y	Y
Notch	1	S	235	Y*	N
Notchless	1	I	2035	Y	Y
Presenilin	1	I	1661	Y	Y
Serrate/Jagged	2	S	246-300	Y*	Y
strawberry notch	7	I,S	191-3519	Y	Y
Suppressor of Hairless	3	I,C	375-697	Y	Y
WNT					
armadillo	5	I,S	348-3001	Y	Y
dishevelled	2	I	954-1321	Y	Y
frizzled	3	C,S	194-500	N	Y
Wnt family (wingless, WNTs)	6	C,S	207-508	Y	Y
TGF-BETA					
decapentaplegic (BMP2/4)	1	C	547	Y	Y
glass bottom boat (BMP5/7)	2	I	510-737	Y	Y
SMADs (Mad, Smad2/3, Smad4/Medea)	7	I,C	276-2276	Y	Y
Type I Receptor (saxophone/thickveins/activin receptor type I)	5	I,C	236-2466	Y	Y
Type II Receptor (punt, wishful thinking)	3	I	259-5038	Y	Y
RECEPTOR TYROSINE KINASES					
Epidermal growth factor receptor	7 (5†)	I,C,S	229-715	N	Y
rhomboid	2	C	229-602	N	Y
HORMONE SIGNALING (ECDYSONE, NUCLEAR HORMONE)					
disembodied (ecdysteroidogenic P450)	1	I	1835	Y	Y
Ecdysone receptor	2	I,C	231-1393	Y	Y
E75	3	I,S	257-649	Y	Y

Table 3 Selected signaling pathway genes identified in the *O. fasciatus* transcriptome *(Continued)*

Ecdysone-induced protein 63E	1	I	1479	Y	Y
ecdysoneless	1	I	4158	Y	Y
Nuclear hormone receptor E78	1	I	3150	Y	Y
Nuclear hormone receptor HR3	2	I	529-737	Y	Y
phantom (cytochrome P450 306a1)	2	C	344-575	N	Y
shade (cytochrome 450 314A1)	1	I	2125	Y	Y
shadow (cytochrome 450 315A1)	1	I	1650	Y	Y
ultraspiracle nuclear receptor	1	C	245	Y*	N
without children	2	I	1155-1357	Y	Y

Hit ID indicates if gene hits were found among isotigs (I), Cap3-assembled contigs (C), or unassembled singletons (S). Sequence length (range) indicates the shortest and longest S, C or I hit sequences for each gene. These results were generated by BLASTing the raw reads from the N and NN samples against the full assembly. When multiple sequences were obtained via name search, they were tested to see whether they could be made to form a contig with Sequencher or CLC Combined Workbench (see Methods). Asterisk indicates hits only present in normalized GS-FLX reads. X($Y^{\dagger}$) indicates that the X sequences with hits could be assembled into Y contigs.

templates, emphasizing the importance of including these in NGS gene discovery studies.

Although we identified a diverse array of genes, some well-studied genes known to be expressed during embryogenesis were not easily identified in this study. For example, our BLAST results only contained three genes from the Hox cluster (*fushi tarazu, Antennapedia*, and *Abdominal-B*), although orthologs of all the canonical arthropod Hox genes are known to be present in *O. fasciatus* [58]. However, using the *O. fasciatus* Hox gene sequence fragments available from NCBI as a BLAST query against our transcriptome did reveal sequences for all Hox genes except *Sex combs reduced*. It is possible that these genes are expressed at very low levels during the developmental stages sampled here, suggesting that enrichment techniques may be necessary to more easily identify certain genes of interest. We do note, however, that *fushi tarazu*, the only Hox cluster gene not previously identified in *O. fasciatus*, was identified in both N and NN samples of this transcriptome dataset (Table 4).

Case study: gene discovery for endocrine regulation of development

In addition to surveying the transcriptome for genes involved in embryonic patterning and other developmental

Table 4 Selected developmental process genes identified in the *O. fasciatus* transcriptome

Process	# Hits	Hit ID (I/C/S)	Length (range)	Normalized	Non-Normalized
GERM PLASM					
Argonaute 3	2 ($1^{\dagger}$)	I	2042-2231	Y	Y
germ cell-less	2 ($1^{\dagger}$)	I	630-1817	Y	Y
maelstrom	1	I	994	Y	Y
nanos	1	I	1961	Y	Y
piwi/aubergine	1	I	2888	Y	Y
pumilio	2	I	424-2574	Y	Y
staufen	3	I	599-2100	Y	Y
Tudor	2	I	2719-3299	Y	Y
vasa	1	C	330	Y	Y
ANTERIOR-POSTERIOR DETERMINATION					
GAP					
hunchback	1	I	1429	Y	Y
Kruppel	1	S	250	N	Y
ocelliless (orthodenticle)	1	S	207	Y	N
TERMINAL GROUP					
huckebein	1	I	589	Y	Y
torso-like	2 ($1^{\dagger}$)	I,C	430-1868	Y	Y
PAIR RULE					

Table 4 Selected developmental process genes identified in the *O. fasciatus* transcriptome *(Continued)*

fushi tarazu	1	I	788	Y	Y
hairy (Enhancer of split/HES-1)	1	I	2530	Y	Y
odd skipped	1	C	346	N	Y
SEGMENT POLARITY					
armadillo	5	I,S	348-3001	Y	Y
cubitus interruptus	3	I,S	225-906	Y	Y
engrailed	1	S	227	Y*	N
fused	2	I	516-1582	Y	Y
pangolin	2	I,C	492-544	N	Y
patched	2	C, S	225-418	N	Y
Wnt family (wingless, Wnts)	6	C,S	207-508	Y	Y
DORSO-VENTRAL AXIS					
cactus	7	I, C	629-1748	Y	Y
decapentaplegic (BMP2/4)	1	C	547	Y	Y
gastrulation-defective	1	I	1773	Y	Y
nudel	4	I,S	322-1458	Y	Y
pipe	1	C	266	N	Y
short gastrulation	2	C	254-615	Y	Y
snake	1	I	1789	Y	Y
spätzle	2	I	993-3170	Y	Y
Toll	11	I, C, S	215-4323	Y	Y
MOLTING/METAMORPHOSIS					
cuticular proteins (including CP 49Ae and adult cuticle protein)	4	I,C	404-566	Y	Y
disembodied (ecdysteroidogenic P450)	1	I	1835	Y	Y
Ecdysone receptor	2	I,C	231-1393	Y	Y
E75	3	I,S	257-649	Y	Y
Ecdysone-induced protein 63E	1	I	1479	Y	Y
ecdysoneless	1	I	4158	Y	Y
ftz transcription factor 1	1	I	807	Y	Y
hormone receptor 4	2	I	1003-2114	Y	Y
juvenile hormone acid methyltransferase	5	I	548-2871	Y	Y
juvenile hormone binding protein	1	I	1099	Y	Y
juvenile hormone epoxide hydrolase	5	I,S	255-2859	Y	Y
juvenile hormone esterase	4	I	850-2382	Y	Y
juvenile hormone esterase binding protein	1	I	1057	Y	Y
Juvenile hormone-inducible protein	7	I	456-2757	Y	Y
Methoprene-tolerant	1	I	3415	Y	Y
Nuclear hormone receptor E78	1	I	3150	Y	Y
Nuclear hormone receptor HR3	2	I	529-737	Y	Y
phantom (cytochrome P450 306a1)	2	C	344-575	N	Y
shade (cytochrome 450 314A1)	1	I	2125	Y	Y
shadow (cytochrome 450 315A1)	1	I	1650	Y	Y
takeout	3	I	591-1011	Y	Y
ultraspiracle nuclear receptor	1	C	245	Y*	N
without children	2	I	1155-1357	Y	Y

Hit ID indicates if gene hits were found among isotigs (I), CAP3-assembled contigs (C), or unassembled singletons (S). Sequence length (range) indicates the shortest and longest S, C or I hit sequences for each gene. These results were generated by BLASTing the raw reads from the N and NN samples against the full assembly. When multiple sequences were obtained via name search, they were tested to see whether they could be made to form a contig with Sequencher or CLC Combined Workbench (see Methods). Asterisk indicates hits only present in normalized GS-FLX reads. $X(Y^{\dagger})$ indicates that the X sequences with hits could be assembled into Y contigs. Boldface indicates genes also present in Table 3.

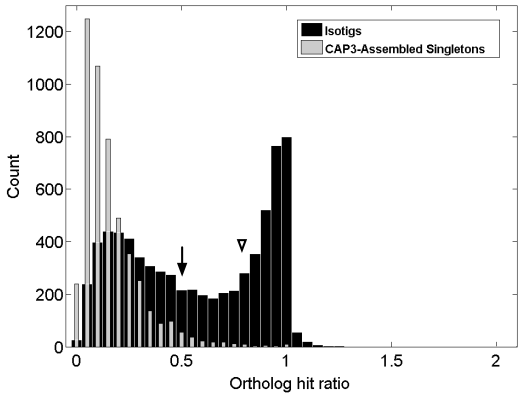


Figure 6 Ortholog hit ratio analysis of isotigs and CAP3-reassembled singletons. An ortholog hit ratio of one implies that a transcript has been assembled to its true full length. For isotigs (black), a majority (54.8%) appear to contain at least 50% of the full length transcript sequence (arrow), while over one-third (37.2%) appear to represent at least 80% of the full length transcript sequence (arrowhead). Most singletons (grey) represent much smaller percentages of full-length transcripts.

processes, we asked whether we could also identify genes known to be employed in biological processes during post-embryonic development of holometabolous insects. Recent studies have suggested that many of the genes used during holometabolous insect metamorphosis may also play important roles during embryogenesis in hemimetabolous insects [59,60]. To investigate this, we searched the *O. fasciatus* transcriptome for expression of key ecdysteroid- and juvenile hormone (JH)-related genes. We identified transcripts for many of the known ecdysteroid biosynthesis genes, including cytochrome P450 genes encoded by the *Drosophila* Halloween family, such as *shade* (CYP314A1), *shadow* (CYP315A1), *phantom* (CYP306A1) and *disembodied* (CYP302A1) (Table 4). We also detected expression of ecdysone response genes. In particular, we identified many of the ecdysone-regulated genes that play key roles during molting and metamorphosis, including *E75*, *HR3*, and *HR4* (Table 4). The presence of these genes in the ovaries and early embryos of *O. fasciatus* corroborates recent studies that implicate ecdysone-response genes in key developmental processes during embryogenesis [59-61]. As might be expected for a situation where ecdysone regulates embryonic development but not molting, transcripts encoding insect peptide hormones implicated in eclosion behavior, such as ecdysis-triggering hormone, eclosion hormone and crustacean cardioactive peptide, were not detected. JH biosynthesis and response genes were also isolated (Table 4). JH has been shown to play a role in promoting embryonic development and tissue maturation [62]. The expression of these genes, together with that of JH esterase and JH binding proteins, is consistent with previous studies implicating tight control of JH during embryogenesis [63].

Conclusions

We have used 454 pyrosequencing to create an early developmental transcriptome for the milkweed bug *O. fasciatus* in the absence of a reference genome. Although genomic sequence data will be necessary in the future for linkage or *cis*-regulatory analyses, at the early stages of establishing new model organisms, one of the most important goals is often gene discovery. In this regard, while no transcriptome generated in this way can realistically be "complete" in the sense of containing full length transcripts for all expressed genes, we propose that for many evolutionary developmental biology studies, the approach described here is a useful one for fast, high-throughput gene discovery. A high priority for comparative developmental biology research is gene expression and function analyses. By sequencing at great depth and testing a variety of cDNA preparation methods (normalized, non-normalized, embryo- and ovary-specific), we have generated tens of thousands of gene sequences of sufficient lengths for the commonly used developmental techniques of *in situ* hybridization and RNAi-mediated gene knockdown. These data can also be used for phylogenetic, population genetic, and functional genomic applications, provide a starting point for identification of genomic regulatory sequences, and assist with assembly of hemipteran genomes sequenced in the future.

Note added in Proof

While this article was in review, Kumar and Blaxter [64] published a comparison of *de novo* assemblers for 454 transcriptome data, and reported important shortcomings of Newbler v2.3 compared to other available assemblers. Specifically, the authors reported that Newbler v2.3 produced the smallest assembly (i.e. the smallest number of base pairs incorporated into contigs) of the assemblers tested. The authors argue that this poor performance is likely because Newbler v2.3 inexplicably discards portions of read overlap information. In contrast, a newer, currently unreleased version of Newbler, v2.5, produced the most complete assembly of all those tested. Kumar and Blaxter (2010) therefore strongly advise all *de novo* 454 transcriptome assembly projects which have used Newbler v2.3 to recompute their assemblies with Newbler v2.5.

To address this concern, we obtained a pre-release version of Newbler v2.5 from Roche and reassembled the *O. fasciatus* data, again using the -nosplit flag. In contrast to Kumar and Blaxter (2010), we observed much less dramatic differences between the assemblies

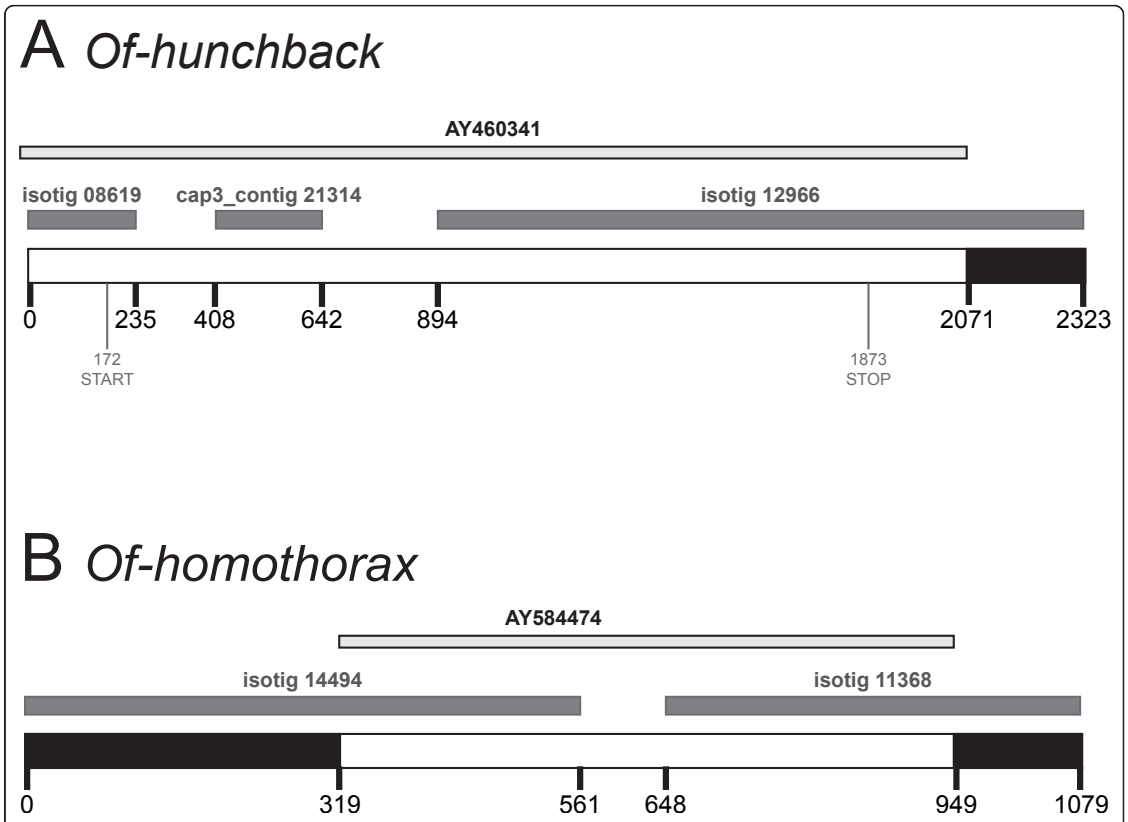


Figure 7 The *O. fasciatus* transcriptome adds sequence data to existing GenBank accessions, which in turn improves annotation of transcriptome sequences. *(A)* Extended contig for *Of-hunchback* (bottom), comprising the complete mRNA GenBank accession (top, light grey), two isotigs and one CAP3 contig from the transcriptome (middle, dark grey). The largest isotig provides an additional 252 bp of 3' UTR sequence to the GenBank sequence (black). Comparison with the GenBank sequence enabled isotig 08619 and cap3_contig 21314 to be assembled into the same contig. *(B)* Extended contig for *Of-homothorax* (bottom), with a partial mRNA GenBank accession (top, light grey) and two transcriptome isotigs (middle, dark grey). Both isotigs extend beyond the known GenBank sequence at the 3' and 5' ends, extending the known region by 449 bp in total (black). Both isotigs had been identified as *homothorax*, and because they did not overlap, they were classified as belonging to the same transcript rather than being paralogs. The GenBank sequence bridges an 87 bp gap between the isotigs, confirming that both sequences are fragments of a single gene.

produced by Newbler v2.3 and Newbler v2.5 (Additional file 7). For example, Kumar and Blaxter (2010) report that Newbler v2.5 increased their total assembly size by 39% compared to Newbler v2.3. For the *O. fasciatus* data analyzed here, Newbler v2.5 increased the total assembly size by less than 1% (Additional file 7). Further, we observed very similar numbers of isogroups, isotigs, and singletons between the two assemblies (Additional file 7). We did observe a 16% increase in the number of contigs reported by Newbler v2.5, but this difference was markedly less than the 80% increase observed in the data analyzed by Kumar and Blaxter (2010). After BLASTing all of the assembled isotigs and cap3-assembled singletons against the RefSeq database, we identified a total of 10,886 unique BLAST hits, compared to 10,775 genes identified using Newbler v2.3.

These results suggest that, although we did observe a modest increase in assembly size using Newbler v2.5, the analyses presented in the current study are largely robust against differences between currently available versions of Newbler. One possible explanation for the difference between these results and those observed by Kumar and Blaxter (2010), is the greater sequencing depth performed in the current study. If in fact the poor performance of Newbler v2.3 involves discarding information in regions of low coverage, the fact that our dataset includes ~2.4x more reads than that analyzed by Kumar and Blaxter (2010) may explain the reduced improvement that Newbler v2.5 provided our dataset.

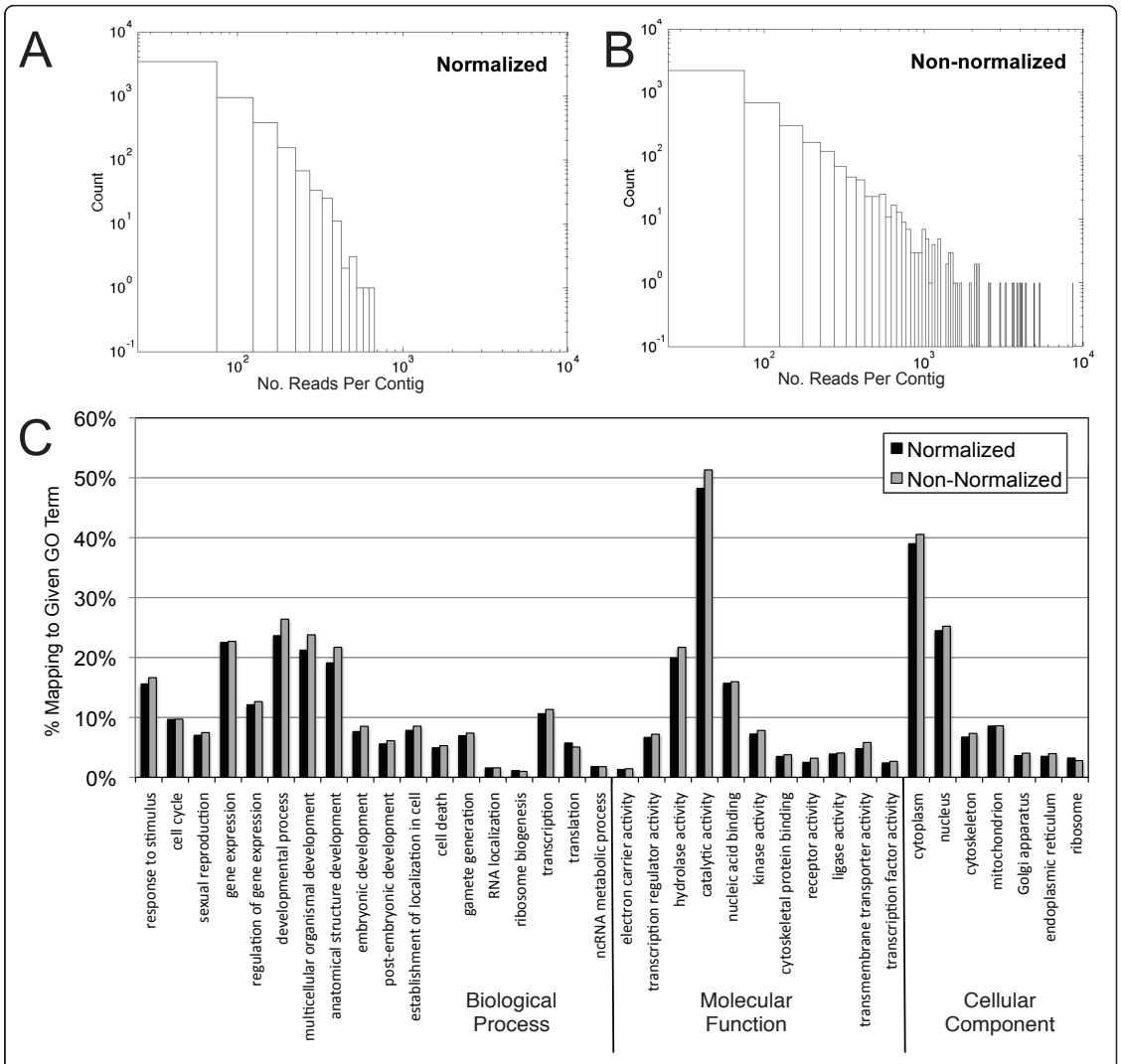


Figure 8 Normalization decreases coverage of highly abundant genes, but does not change the GO term distribution of contigs. In both samples, most contigs are composed of $<10^2$ reads. However, the non-normalized sample (*A*) contains contigs with many more reads per contigs than the normalized sample (*B*). In other words, normalization preferentially decreases the number of reads of those contigs with the most reads. *(C)* GO term distributions do not differ dramatically between pyrosequenced libraries of N versus NN cDNA. However, see Additional file 6 for exceptions. Column heights reflect the percentage of annotated sequences in each assembly that mapped to a given GO term. Note that the GO terms shown represent the results of mapping the N and NN reads against the complete assembly, rather than those obtained via independent assemblies of N and NN reads.

We also suggest that the reduced number of genes identified via BLAST observed by Kumar and Blaxter (their Table five) may result from the fact that the authors excluded singletons from their analyses. If Newbler v2.3 indeed fails to assemble regions of low coverage and instead retains those reads as singletons, many genes of interest may only be present as singletons. Indeed, we observed many genes of interest exclusively represented as singletons (Tables 3 and 4). Thus, for the purpose of gene discovery, we emphasize that future *de novo* transcriptome projects should analyze singletons as an important source of useful gene sequence.

Although our results do not appear to be greatly sensitive to which version of Newbler is used, we agree with Kumar and Blaxter (2010) that future transcriptome project should use utilize the most current

available version of Newbler, or whichever assembler algorithm they find most useful for their data.

Methods

Animal culture

The *O. fasciatus* specimens sequenced in this study were originally purchased from the Carolina Biological Supply Company (Burlington, NC) and were maintained in the laboratory on sunflower seeds under a 12h:12h light/dark cycle at 28°C.

cDNA Synthesis

For our pilot study using the GS-FLX platform, total RNA was isolated from mature ovaries (Figure 1B) and from mixed-stage embryos representing the first three days of development (roughly 60% of embryogenesis at 28°C; Figure 1C, D) using TRIzol (Invitrogen), following the manufacturer's protocols. For each RNA sample, approximately 5 μg of cDNA was prepared using the SMART cDNA library construction kit (Clontech, CA, USA). The cDNA was normalized using Evrogen's Trimmer-Direct cDNA Normalization kit (Evrogen, Moscow, Russia), and subsequently digested with SfiI to partially remove the SMART adapters. The size distributions of total RNA and cDNA were assessed on 1.0% agarose gels following each step of the protocol.

To prepare cDNA for sequencing on the GS-FLX Titanium platform, we followed a modified version of the SMART cDNA protocol [65] that has been optimized for cDNA quality and yield from small quantities of total RNA. A helpful guide that formed the initial basis for the optimization of this protocol was once available online from Evrogen, but has since been removed. At the time these libraries were prepared, Roche had not yet provided a specific protocol for cDNA library preparation for 454 pyrosequencing. Subsequently, the company has released a cDNA protocol that requires approximately 500 ng of purified mRNA (typically requiring isolation of 10 to 50 μg of total RNA). While useful for larger tissue samples, the Roche cDNA preparation protocol is difficult to apply to samples in which RNA quantity is limiting, as is the case with many non-model organisms. The protocol we present here does not require the loss-prone step of mRNA purification, and we have found that it produces sufficient quantities of high-quality cDNA when 5 μl of the RNA (18S and 28S bands) can be visualized on a 1% agarose gel stained with ethidium bromide. Compared with the original SMART protocol, we have optimized the primers, PCR conditions, and downstream purification steps to maximize the yield of double-stranded cDNA required for 454 pyrosequencing. We initially optimized this protocol for Roche's original 454 library preparation protocol (not specific to cDNA), which required input of double-stranded DNA amounts of 2.5-10 μg (in our experience, typically 10-20 μg prepared cDNA as measured by UV absorbance). However, newer protocols from Roche require only 500 ng double-stranded cDNA, limiting the need for a secondary amplification step, as described here, for samples with highly limiting quantities of total RNA.

After separately isolating total RNA from mature ovaries (Figure 1B) and from each of the first three days of embryogenesis (Figure 1C, D) as described above, each RNA sample was treated with DNAse to remove potential genomic contamination. Equal amounts of each sample were then pooled for use as a template for first strand cDNA synthesis. Due to concerns that the poly (T) primer used in the SMART kit could interfere with pyrosequencing, the 3'-primer used was modified in two ways: (1) the poly(T) was interrupted every fourth base by the inclusion of a cytosine [sensu 30]; and (2) the primer contained an *Mme*I site which allowed most of the poly(T) to be removed during digestion. This 3'-primer (PD243Mme-30TC, 5'-ATT CTA GAG CGC ACC TTG GCC TCC GAC TTT TCT TTT CTT TTT TTT TCT TTT TTT TTT VN-3') was used during first strand synthesis and for all subsequent amplification steps. Because *Mme*I also cleaves relatively commonly within eukaryotic genes, it may not always be desirable to use this enzyme for library preparation. As an alternative, we have additionally found that a similar 3' primer containing an *Sfi*I cleavage site (PD243-30TC, 5'-ATT CTA GAG GCC ACC TTG GCC GAC ATG TTT TCT TTT CTT TTT TTT TCT TTT TTT TTT VN-3') is also effective in producing cDNA that yields high-quality 454 data (data not shown).

For first-strand synthesis, 3 μg of total RNA (in 6 μl) and 2 μl 3' primer (12 μM) were mixed and denatured at 65°C for 5 minutes, then placed on ice. Reverse transcription reactions using SuperScript II (Invitrogen) in the manufacturer's recommended buffer were performed for 50 minutes at 42°C using twice the recommended concentration of enzyme, 1 μl of Protector RNAse inhibitor (Roche) to avoid RNA degradation, 2 μl 5' primer (12 μM), 2 μl 10 mM DTT, and 1 μl 10 mM dNTPs. Template-switching essential for the SMART technique was achieved using a 5' primer (PD242, 5'-AAG CAG TGG TAT CAA CGC AGA GTG GCC ACG AAG GCC rGrGrG-3') with three RNA nucleotides at its 3' end, which contains an *Sfi*I site. Reactions were then heat-inactivated for 15 minutes at 70°C and diluted 1:5 in milliQ water in preparation for PCR amplification. Contrary to some expectations, SuperScript III reverse transcriptase (Invitrogen) may be substituted in this protocol with equivalent results (data not shown).

To maximize yield during cDNA amplification, the first round of amplification was conducted using a 2:2:1

mix (v:v:v) of Hemo KlenTaq (New England Biolabs), Phusion (New England Biolabs), and PfuTurbo (Stratagene) polymerases. This mixture of enzymes was determined empirically to provide the highest yield of cDNA with a range of input first-strand concentrations. Cesium KlenTaq AC (DNA Polymerase Technologies) and the hot start versions of Phusion and PfuTurbo polymerases in the same ratio may be also substituted at this step without sacrificing yield; this may produce fewer PCR artifacts in the final cDNA preparation. Buffer conditions ($MgCl_2$ and DMSO) were also empirically optimized to maximize yield and minimize PCR artifacts. Reactions were performed in 100 μL total volume in 1X Phusion HF buffer, 1.5 μL polymerase mix, 5 μL first-strand cDNA (previously diluted 1:5 in H_2O), 1 μL 3' primer (PD243Mme-30TC, 12 μM), 1 μL 5' primer (PCRIIA, 5'-AAG CAG TGG TAT CAA CGC AGA GT-3', 12 μM), and a final concentration of 1% DMSO, 1.5 mM $MgCl_2$ (in addition to the $MgCl_2$ already present in the HF buffer), and 200 μM dNTPs. Reactions were cycled with the following program: 1 minute at 95°C, followed by 16-20 cycles of 30 seconds at 95°C (see below for determining optimal number of cycles), 30 seconds at 66°C, and 3 minutes at 72°C, and a final 10 minutes at 72°C. After cooling to room temperature, 10 μL 3M NaOAc pH 5.5 was added to each 100 μL secondary PCR reaction followed by purification with the QiaQuick PCR purification kit (Qiagen) using the manufacturer's recommended protocol. For all purification steps, samples were eluted with TM buffer (10 mM Tris-HCl pH 8.5, 1 mM $MgCl_2$) to prevent strand separation of double-stranded cDNA.

To produce sufficient cDNA for sequencing, Advantage 2 (Clontech) polymerase was used under the manufacturer's recommended conditions during the second round of amplification using the same primer concentrations and 1 μl of undiluted primary PCR product. We recommend testing a range of dilutions of the primary PCR product to obtain the desired quantity of amplified cDNA in 9-10 PCR cycles. In cases of highly limiting RNA concentration, we have also found that a secondary PCR reaction using a 1:1:1 mix of Phusion, Cesium KlenTaq AC, and Deep Vent (exo-) (New England Biolabs) polymerase in ThermoPol reaction buffer supplemented with 1.5 mM $MgSO_4$ and 1% DMSO produces the highest yield of secondary PCR product (note that this polymerase mix does not produce optimal results when used for first-round amplification). Secondary PCR reactions were cycled using the same parameters as the primary PCR but running for approximately 10 cycles.

To prevent overcycling during both rounds of PCR amplification, each reaction was prepared in duplicate, and one reaction was spiked with 1 μl of 1:750 SybrGreen I (Invitrogen). The spiked reactions were monitored in real time on an Mx3005P QPCR machine (Stratagene Inc.), and the samples were removed when amplification began to plateau. To increase the representation of double-stranded cDNA, two cycles of "chase PCR" were conducted following each round of cDNA amplification after the optimal number of cycles had been reached. Excess primers were added (1.5 μL of each, 12 μM primer per 100 μL reaction), and each reaction was subjected to two additional non-denaturing cycles of 1 minute at 77°C, 1 minute at 65°C, and 3 minutes at 72°C, followed by a 10 minute extension at 72°C.

Following the second round of amplification and PCR purification, the cDNA samples were double-digested with *Sfi*I and *Mme*I (40 and 26 units per 150 μl reaction, respectively). cDNA species <500 bp were then removed using Chroma Spin 400 columns (Clontech) which had been equilibrated with TM buffer following the manufacturer's protocol. It should be noted that the Chroma Spin column protocol suggested in the Clontech SMART cDNA kit is non-optimal, and that following the protocol provided with the separately purchased columns is less labor-intensive and produces a higher yield of size-selected cDNA. Equilibration of Chroma Spin columns is critical for maximizing the yield of double-stranded cDNA as required by the Roche library preparation protocols. Following size selection, cDNA was blunt-ended with the NEB Quick Blunting kit (New England Biolabs) and purified once more with the QiaQuick kit. After each step of cDNA synthesis, the size distribution was checked on 1.0% agarose gels, and the cDNA samples were quantified using a Qubit (Invitrogen), after observing that the NanoDrop 1000 (Thermo Scientific) did not reliably quantify ds-cDNA (C. Dunn, personal communication).

To prepare normalized cDNA for GS-FLX Titanium sequencing, 1 μl of the twice-amplified, purified cDNA sample described above was subjected to Evrogen's DSN-treatment protocol, followed by a single round of further amplification, *Sfi*I/*Mme*I digestion, and size selection. Approximately 5 μl of normalized and non-normalized cDNA were synthesized.

454 Titanium Pyrosequencing

For the pilot study using the GS-FLX platform, EnGenCore (University of South Carolina) conducted the final steps of library preparation, including nebulization, adaptor-ligation, and sequencing of each sample (¼ plate each). For sequencing using the Titanium platform, the samples were nebulized, adaptor-ligated, and pyrosequenced by the Institute for Genome Science and Policy DNA Sequencing Facility (Duke University).

Sequence Assembly

Raw reads were assembled using the cDNA assembly algorithm of Newbler v2.3 (Roche) with default assembly parameters. An adaptor-trimming step was included in the assembly (the "-v" flag), and the "-nosplit" flag was also used to reduce the generation of extremely short contigs that might otherwise have been created. All of the raw reads generated in this study have been submitted to the NCBI Short Read Archive (Study Accession Number: SRP002610.1).

Because redundancy was observed among the singletons generated by Newbler v2.3, the singletons were reassembled using CAP3 [48], with '-z' option set to 1. Prior to this secondary assembly, the singletons were screened for adaptor sequences using both cross_match [66-68] and a custom python script (Casey Dunn, personal communication), We note that Newbler can also be used to produce a .fasta and corresponding .qual files of trimmed reads using the '-tr' option. The final assembly thus consists of three types of sequences: Newbler-assembled sequences, cap3_contigs, and cap3_singlets, all of which were subjected to subsequent analyses.

Sequence Annotation

Sequences were first mapped against the RefSeq Protein database [[69], downloaded from ftp://ftp.ncbi.nih.gov/blast/db/ on April 27, 2010] using BLASTX. All BLAST searches were conducting using BLAST v2.2.23+ [70] with an e-value cut-off of 1e-10. We then used Blast2GO v1.2.7 [54] to retrieve the Gene Ontology (GO) [71] terms and their parents associated with the top 20 BLAST hits for each sequence. To avoid potentially double-counting sequences that might represent un-assembled portions of the same transcript, a custom python script ("transcriptome_blast_summarizer.py", available at http://www.extavourlab.com/protocols/index.html) was used to identify sequences with identical top BLAST hits prior to GO annotation. If multiple sequences hit non-overlapping portions of the same top BLAST hit, we used the conservative assumption that these sequences represented unassembled portions of the same transcript, and therefore only tallied the GO terms of one of these sequences. However, if multiple sequences hit overlapping portions of the same top BLAST hit, we considered these sequences potential paralogs and retained them all. Thus, the counts of sequences in each GO term only include one sequence per top BLAST hit, unless the multiple sequences mapped to overlapping portions of the same BLAST hit. These counts were used to compare the distribution of sequences among specific GO terms between the transcriptomes of *O. fasciatus* and the *Drosophila melanogaster* genome. For this comparison, we used a precomputed GO annotation of the *D. melanogaster* genome [72].

The FASTA formatted transcriptome data set file was examined in TextWrangler (v. 3.1, Bare Bones Software, Inc.). Candidate genes were sought via whole gene names and, where possible, via the gene name abbreviations, while avoiding irrelevant hits. The FASTA header annotation of transcriptome sequences includes the top 20 BLASTx hits to the RefSeq database as described above.

Sequencher (v4.8, Gene Codes Corporation; default settings: minimum 20 bp overlap between sequences, ≥85% sequence identity) and CLC Combined Workbench (v5.6.1, CLC Bio) were used to examine whether transcriptome sequences could be further assembled.

Estimating sequencing depth

To estimate how thoroughly our sequencing efforts sampled the *O. fasciatus* transcriptome, eight progressively larger subsets of the reads were independently assembled. The total number of genes was then identified via BLASTX. For these smaller assemblies, reads from one plate each of normalized and non-normalized reads were combined in random order and sampled without replacement. For each assembly, we BLASTed the longest isotig of each isogroup, and all of the singletons, against the SwissProt database [[73], downloaded from ftp://ftp.ncbi.nih.gov/blast/db/ on April 21, 2010]. We used the relatively small SwissProt database in order to reduce computation time. However, the absolute values of BLAST hits against this database are likely to be underestimates of those values that would have been obtained from a larger database such as RefSeq or nr. If multiple isotigs or contigs hit non-overlapping portions of the same top BLAST hit, only one of these sequences was counted. However, because frequent cases of identical, unassembled singletons were observed, we counted only one singleton per top BLAST hit, regardless of whether these hits overlapped or not.

We used a custom python script to calculate the ortholog hit ratio. This script, "ortholog_hit_ratio_calculator.py" is available at http://www.extavourlab.com/protocols/index.html).

Assessing the importance of cDNA normalization

To assess the relative contribution of cDNA normalization to the quality of our assembly, the screened, raw reads from both normalized (N) and non-normalized (NN) samples were mapped against the complete assembly of all reads using the BLASTN algorithm [70] with an e-value cut-off of 1e-4. Based on these results, the Fisher's Exact Test was used to identify over- and under-represented terms in each gene list. This test was performed using Blast2GO (two-tailed, removing double IDs so that only those genes hit uniquely by either N or NN reads were considered). The BLASTN results were also investigated

using text searches to find whether certain genes of interest were present in only one of the two cDNA samples.

Additional material

Additional file 1: Normalized sample did not perform equally in pilot and full sequencing runs. (*A*) For the normalized sample, the read lengths of the full plate sequencing runs (white) were shorter than those obtained by the 1/8 plate run (grey). (*B*) The read length distribution of the non-normalized sample was comparable for both 1/8 plate (grey) and full plate (white) sequencing runs.

Additional file 2: Distribution of average coverage (reads/bp) within contigs in the *O. fasciatus* transcriptome. The coverage within contigs is calculated by dividing the total number of base pairs contained in the reads used to construct a contig by the length of that contig. Note that Newbler v2.3 discards those contigs <100 bp.

Additional file 3: RT-PCR validation of bioinformatically predicted multiple isoforms. (*A*) Schematic of experimental design. Ten isogroups were randomly selected, each containing exactly two isotigs that differed by the presence/absence of a single contig. PCR primers were designed to flank the differing region. (*B*) Band sizes predicted by Newbler v2.3 for ten randomly selected isogroups containing exactly two isotigs. *(C)* Agarose gel following RT-PCR using primers against the sequences described in (*B*). Ladder sizes are given in base pairs on the left. Blue arrowheads: bands of the sizes predicted by Newbler v2.3; red arrowheads: bands not predicted by Newbler v2.3.

Additional file 4: Identity of taxa with top BLAST hits. "Isotigs" refers only to the longest isotig of each isogroup; "Singletons" refers to the Newbler-generated singletons after secondary CAP3 assembly. The category "other" is the summation of all those species obtaining very low numbers of BLAST hits.

Additional file 5: *O. fasciatus* assembly isotigs have ortholog hit ratios similar to predictions from fully genome-sequenced databases. When isotigs from the *O. fasciatus* transcriptome are BLASTed against the RefSeq protein database, ortholog hit ratios show a similar profile to those obtained when the complete *Acyrthosiphon pisum* gene prediction set (downloaded from http://www.aphidbase.com/aphidbase/downloads/) is BLASTed against the predicted gene set of *Drosophila melanogaster* (r5.28 downloaded from ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/) with an e-value cut-off of 1e-10.

Additional file 6: GO terms enriched in Normalized (N) and Non-Normalized (NN) cDNA samples. N (assembly generated from full plate of normalized cDNA) and NN (assembly generated from an equalized number of base pairs of non-normalized cDNA) reads were BLASTed against the full transcriptome assembly, and the results were used to generate "test" and "reference" sets for a Fisher's Exact Test. FDR: false discovery rate.

Additional file 7: Comparison of *de novo* transcriptome assemblies produced by Newbler v2.3 and Newbler v2.5. Number of BLASTx hits reflects a search against RefSeq Protein database with an e-value cut-off value of 1e-10.

Acknowledgements
Thanks to Casey Dunn and Freya Goetz for helpful discussions on cDNA preparation and bioinformatic analysis; Amir Karger, Jiangwen Zhang, and Suvendra Dutta of the Harvard FAS Life Science Computing team for help with bioinformatic analysis; Ana Conesa and Stefan Gotz for assistance with Blast2GO via the Blast2GO mailing list; Joe Jones and Lisa Bukovnik for their administration of the sequencing; Evelyn Schwager, Frederike Alwes, and other members of the Extavour lab for discussions of the results and manuscript. We thank David and Z Behl for the photograph of an *Oncopeltus* adult (Figure 1A). This work was partially supported by National Science Foundation (NSF) award IOS-0817678 to CE, an NSF Predoctoral Fellowship to BEC, DFG Collaborative Research Center grant 680 "The molecular basis of evolutionary innovations" to KP and SR, and the Wellesley College research fund to YS.

Author details
[1]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA. [2]Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039, USA. [3]Institute for Developmental Biology, University of Cologne, Cologne Biocenter, Zülpicher Straße 47b, 50674, Cologne, Germany. [4]Department of Biological Sciences, Wellesley College, 106 Central Street, Wellesley MA 02481, USA.

Authors' contributions
BEC helped design the research, performed the experiments, collected and analyzed the data, and wrote the manuscript. NS contributed new protocols and helped write the manuscript. KAP helped analyze the data and write the manuscript. YS helped analyze the data and write the manuscript, and obtained funding for the research. SR helped design the research and review the manuscript, and obtained funding for the research. CE proposed the idea for the research, helped design the research and analyze the data, wrote the manuscript and obtained funding for the research. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 7 October 2010 Accepted: 25 January 2011
Published: 25 January 2011

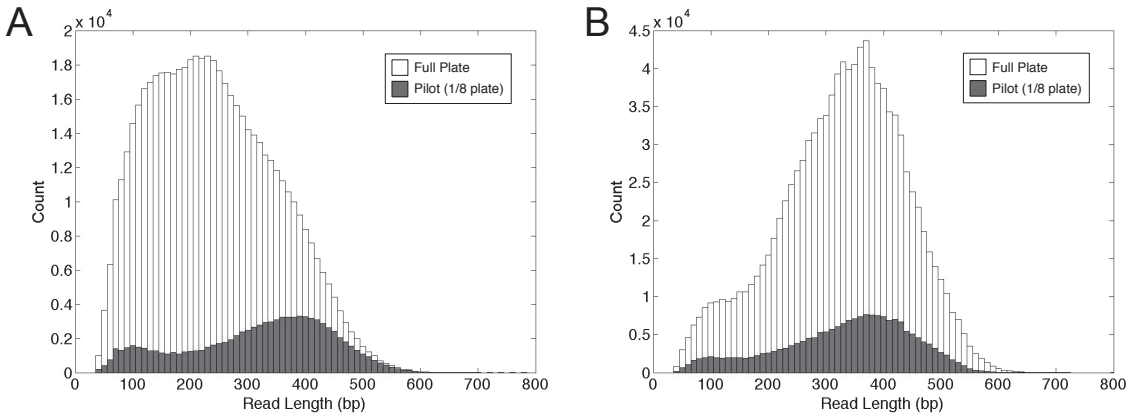
References

1. Kumé M, Dan K: **Invertebrate Embryology.** Belgrade: Prosveta; 1968.
2. Beklemishev WN: **Principles of Comparative Anatomy of Invertebrates: Promorphology.** Chicago: University of Chicago Press, 3 1969**1.**
3. Beklemishev WN: **Principles of Comparative Anatomy of Invertebrates: Organology.** Chicago: University of Chicago Press, 3 1969**2.**
4. Rosenblueth A, Wiener N: **The Role of Models in Science.** *Philosophy of Science* 1945, **12(4)**:316-321.
5. Hedges SB: **The origin and evolution of model organisms.** *Nat Rev Genet* 2002, **3(11)**:838-849.
6. Jenner RA, Wills MA: **The choice of model organisms in evo-devo.** *Nat Rev Genet* 2007, **8(4)**:311-319.
7. Bolker JA: **Model systems in developmental biology.** *BioEssays* 1995, **17(5)**:451-455.
8. Abzhanov A, Extavour CG, Groover A, Hodges SA, Hoekstra HE, Kramer EM, Monteiro A: **Are we there yet? Tracking the development of new model systems.** *Trends in Genetics* 2008, **24(7)**:353-360.
9. Sommer RJ: **The future of evo-devo: model systems and evolutionary theory.** *Nat Rev Genet* 2009, **10(6)**:416-422.
10. Slack JMW: **Emerging Market Organisms.** *Science* 2009, **323(5922)**:1674-1675.
11. Sarkar A, Atapattu A, Belikoff EJ, Heinrich JC, Li X, Horn C, Wimmer EA, Scott MJ: **Insulated piggyBac vectors for insect transgenesis.** *BMC Biotechnol* 2006, **6**:27.
12. Nunes da Fonseca R, von Levetzow C, Kalscheuer P, Basal A, van der Zee M, Roth S: **Self-regulatory circuits in dorsoventral axis formation of the short-germ beetle *Tribolium castaneum*.** *Dev Cell* 2008, **14(4)**:605-615.
13. Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP: **Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals.** *Nature* 2008, **455(7217)**:1193-1197.
14. Shabalina SA, Koonin EV: **Origins and evolution of eukaryotic RNA interference.** *Trends Ecol Evol (Amst)* 2008, **23(10)**:578-587.
15. Beldade P, Rudd S, Gruber JD, Long AD: **A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model.** *BMC Genomics* 2006, **7**:130.
16. Danley PD, Mullen SP, Liu F, Nene V, Quackenbush J, Shaw KL: **A cricket Gene Index: a genomic resource for studying neurobiology, speciation, and molecular evolution.** *BMC Genomics* 2007, **8**:109.
17. Lowe CJ, Wu M, Salic A, Evans L, Lander E, Stange-Thomann N, Gruber CE, Gerhart J, Kirschner M: **Anteroposterior patterning in hemichordates and the origins of the chordate nervous system.** *Cell* 2003, **113(7)**:853-865.
18. Schmid J, Müller-Hagen D, Bekel T, Funk L, Stahl U, Sieber V, Meyer V: **Transcriptome sequencing and comparative transcriptome analysis of the scleroglucan producer *Sclerotium rolfsii*.** *BMC Genomics* 2010, **11**:329.

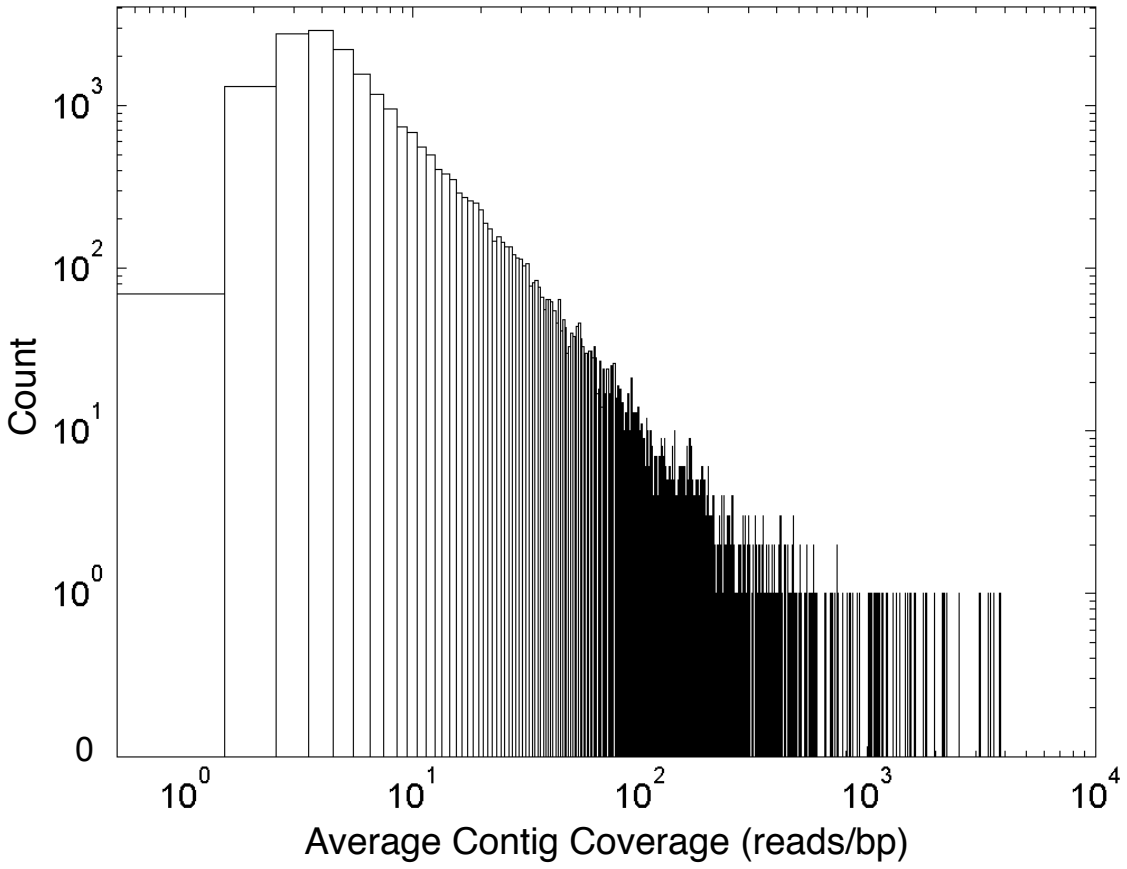
19. Timme RE, Delwiche CF: **Uncovering the evolutionary origin of plant molecular processes: comparison of Coleochaete (Coleochaetales) and Spirogyra (Zygnematales) transcriptomes.** *BMC Plant Biol* 2010, **10**:96.
20. O'Neil ST, Dzurisin JDK, Carmichael RD, Lobo NF, Emrich SJ, Hellmann JJ: **Population-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon*.** *BMC Genomics* 2010, **11**:310.
21. Zhang F, Guo H, Zheng H, Zhou T, Zhou Y, Wang S, Fang R, Qian W, Chen X: **Massively parallel pyrosequencing-based transcriptome analyses of small brown planthopper (*Laodelphax striatellus*), a vector insect transmitting rice stripe virus (RSV).** *BMC Genomics* 2010, **11**:303.
22. Toulza E, Shin MS, Blanc G, Audic S, Laabir M, Collos Y, Claverie JM, Grzebyk D: **Gene expression in proliferating cells of the dinoflagellate *Alexandrium catenella* (Dinophyceae).** *Appl Environ Microbiol* 2010, **76(13)**:4521-4529.
23. Bruder CE, Yao S, Larson F, Camp JV, Tapp R, McBrayer A, Powers N, Granda WV, Jonsson CB: **Transcriptome sequencing and development of an expression microarray platform for the domestic ferret.** *BMC Genomics* 2010, **11**:251.
24. Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA: **Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery.** *BMC genomics* 2010, **11**:180.
25. Roeding F, Borner J, Kube M, Klages S, Reinhardt R, Burmester T: **A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (*Pandinus imperator*).** *Mol Phylogenet Evol* 2009, **53(3)**:826-834.
26. Hahn DA, Ragland GJ, Shoemaker DD, Denlinger DL: **Gene discovery using massively parallel pyrosequencing to develop ESTs for the flesh fly *Sarcophaga crassipalpis*.** *BMC Genomics* 2009, **10**:234.
27. Bellin D, Ferrarini A, Chimento A, Kaiser O, Levenkova N, Bouffard P, Delledonne M: **Combining next-generation pyrosequencing with microarray for large scale expression analysis in non-model species.** *BMC Genomics* 2009, **10**:555.
28. Kristiansson E, Asker N, Förlin L, Larsson DGJ: **Characterization of the *Zoarces viviparus* liver transcriptome using massively parallel pyrosequencing.** *BMC Genomics* 2009, **10**:345.
29. Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, ffrench-Constant RH: **Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera.** *Insect Biochem Mol Biol* 2009, **39(5-6)**:403-413.
30. Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK, Willis BL, Matz MV: **Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx.** *BMC Genomics* 2009, **10**:219.
31. Garcia-Reyero N, Griffitt RJ, Liu L, Kroll KJ, Farmerie WG, Barber DS, Denslow ND: **Construction of a robust microarray from a non-model species (largemouth bass) using pyrosequencing technology.** *J Fish Biol* 2008, **72(9)**:2354-2376.
32. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH: **Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing.** *Mol Ecol* 2008, **17(7)**:1636-1647.
33. Novaes E, Drost DR, Farmerie WG, Pappas GJ Jr, Grattapaglia D, Sederoff RR, Kirst M: **High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome.** *BMC Genomics* 2008, **9**:312.
34. Cheung F, Win J, Lang JM, Hamilton J, Vuong H, Leach JE, Kamoun S, André Lévesque C, Tisserat N, Buell CR: **Analysis of the *Pythium ultimum* transcriptome using Sanger and Pyrosequencing approaches.** *BMC Genomics* 2008, **9**:542.
35. Papanicolaou A, Stierli R, Ffrench-Constant RH, Heckel DG: **Next generation transcriptomes for next generation genomes using est2assembly.** *BMC Bioinformatics* 2009, **10**:447.
36. Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, Millburn G, Osumi-Sutherland D, Schroeder A, Seal R, Zhang H: **FlyBase: enhancing *Drosophila* Gene Ontology annotations.** *Nucleic Acids Res* 2009, **37(Database issue)**:D555-559.
37. Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E, Hammond M, Hill CA, Konopinski N, Lobo NF, MacCallum RM, Madey G, Megy K, Meyer J, Redmond S, Severson DW, Stinson EO, Topalis P, Birney E, Gelbart WM, Kafatos FC, Louis C, Collins FH: **VectorBase: a data resource for invertebrate vector genomics.** *Nucleic Acids Res* 2009, **37(Database issue)**: D583-587.
38. Consortium IAG: **Genome sequence of the pea aphid *Acyrthosiphon pisum*.** *PLoS Biology* 2010, **8(2)**:e1000313.
39. Huebner E: **The *Rhodnius* Genome Project: The promises and challenges it affords in our understanding of reduviid biology and their role in Chagas' transmission.** *Comparative Biochemistry and Physiology, Part A* 2007, **148**:S130.
40. Liu P, Kaufman TC: **Dissection and fixation of large milkweed bug (*Oncopeltus*) embryos.** *CSH Protocols* 2009, **2009(8)**, pdb.prot5261.
41. Liu P, Kaufman TC: **In situ hybridization of large milkweed bug (*Oncopeltus*) tissues.** *CSH Protocols* 2009, **2009(8)**, pdb.prot5262.
42. Liu P, Kaufman TC: **Morphology and husbandry of the large milkweed bug, *Oncopeltus fasciatus*.** *CSH Protocols* 2009, **2009(8)**, pdb.emo127.
43. Lawrence PA: **The hormonal control of the development of hairs and bristles in the milkweed bug, *Oncopeltus fasciatus*, Dall.** *Journal of Experimental Biology* 1966, **44(3)**:507-522.
44. Butt FH: **Embryology of the Milkweed Bug, *Oncopeltus fasciatus* (Hemiptera).** *Cornell Experiment Station Memoir* 1949, **283**:2-43.
45. Lawrence PA: **Some new mutants of the large milkweed bug *Oncopeltus fasciatus* Dall.** *Genetical Research Cambridge* 1970, **15**:347-350.
46. Hughes CL, Kaufman TC: **RNAi analysis of *Deformed*, *proboscipedia* and *Sex combs reduced* in the milkweed bug *Oncopeltus fasciatus*: novel roles for Hox genes in the hemipteran head.** *Development* 2000, **127(17)**:3683-3694.
47. Panfilio KA: **Late extraembryonic morphogenesis and its *zen*(RNAi)-induced failure in the milkweed bug *Oncopeltus fasciatus*.** *Dev Biol* 2009, **333(2)**:297-311.
48. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome Res* 1999, **9(9)**:868-877.
49. Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, Lee WL, Russ C, Lander ES, Nusbaum C, Jaffe DB: **Quality scores and SNP detection in sequencing-by-synthesis systems.** *Genome Res* 2008, **18(5)**:763-770.
50. Zdobnov EM, Bork P: **Quantification of insect genome divergence.** *Trends Genet* 2007, **23(1)**:16-20.
51. Surget-Groba Y, Montoya-Burgos JI: **Optimization of *de novo* transcriptome assembly from next-generation sequencing data.** *Genome Res* 2010, **20(10)**:1432-1440.
52. Nüsslein-Volhard C, Frohnhöfer HG, Lehmann R: **Determination of anteroposterior polarity in *Drosophila*.** *Science* 1987, **238**:1675-1681.
53. Ewen-Campen B, Schwager EE, Extavour CG: **The molecular machinery of germ line specification.** *Mol Reprod Dev* 2010, **77(1)**:3-18.
54. Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21(18)**:3674-3676.
55. Mita K, Morimyo M, Okano K, Koike Y, Nohata J, Kawasaki H, Kadono-Okuda K, Yamamoto K, Suzuki MG, Shimada T, Goldsmith MR, Maeda S: **The construction of an EST database for *Bombyx mori* and its application.** *Proc Natl Acad Sci USA* 2003, **100(24)**:14121-14126.
56. Bogdanova EA, Shagin DA, Lukyanov SA: **Normalization of full-length enriched cDNA.** *Mol Biosyst* 2008, **4(3)**:205-212.
57. Robinson MD, Oshlack A: **A scaling normalization method for differential expression analysis of RNA-seq data.** *Genome Biol* 2010, **11(3)**:R25.
58. Angelini DR, Kaufman TC: **Functional analyses in the hemipteran *Oncopeltus fasciatus* reveal conserved and derived aspects of appendage patterning in insects.** *Dev Biol* 2004, **271(2)**:306-321.
59. Erezyilmaz D, Kelstrup H, Riddiford L: **The nuclear receptor E75A has a novel pair-rule-like function in patterning the milkweed bug, *Oncopeltus fasciatus*.** *Dev Biol* 2009, **334(1)**:300-310.
60. Erezyilmaz DF, Rynerson MR, Truman JW, Riddiford LM: **The role of the pupal determinant *broad* during embryonic development of a direct-developing insect.** *Dev Genes Evol* 2009, **219(11-12)**:535-544.
61. Piulachs MD, Pagone V, Belles X: **Key roles of the Broad-Complex gene in insect embryogenesis.** *Insect Biochem Mol Biol* 2010, **40(6)**:468-475.
62. Dorn A: **Precocene-induced effects and possible role of juvenile hormone during embryogenesis of the milkweed bug *Oncopeltus fasciatus*.** *Gen Comp Endocrinol* 1982, **46(1)**:42-52.
63. Orth AP, Tauchman SJ, Doll SC, Goodman WG: **Embryonic expression of juvenile hormone binding protein and its relationship to the toxic effects of juvenile hormone in *Manduca sexta*.** *Insect Biochem Mol Biol* 2003, **33(12)**:1275-1284.
64. Kumar S, Blaxter ML: **Comparing *de novo* assemblers for 454 transcriptome data.** *BMC Genomics* 2010, **11**:571.

65. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD: **Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction.** *BioTechniques* 2001, **30(4)**:892-897.
66. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8(3)**:195-202.
67. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8(3)**:175-185.
68. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res* 1998, **8(3)**:186-194.
69. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2007, **35(Database issue)**:D61-65.
70. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications.** *BMC Bioinformatics* 2009, **10**:421.
71. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25(1)**:25-29.
72. **B2G-FAR: A Species Centered GO Annotation Repository.** [http://bioinfo.cipf.es/b2gfar/showspecies?species=7227].
73. UniProt_Consortium: **The Universal Protein Resource (UniProt) in 2010.** *Nucleic Acids Res* 2010, **38(Database issue)**:D142-148.

doi:10.1186/1471-2164-12-61
Cite this article as: Ewen-Campen *et al.*: The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. *BMC Genomics* 2011 **12**:61.

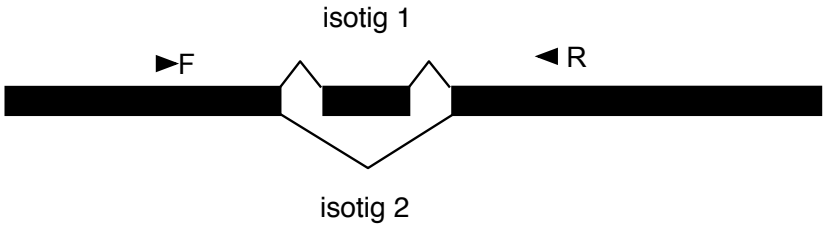


Additional File 1. Normalized sample did not perform equally in pilot and full sequencing runs. (*A*) For the normalized sample, the read lengths of the full plate sequencing runs (white) were shorter than those obtained by the 1/8 plate run (grey). (*B*) The read length distribution of the non-normalized sample was comparable for both 1/8 plate (grey) and full plate (white) sequencing runs.



Additional File 2. Distribution of average coverage (reads/bp) within contigs in the *O. fasciatus* transcriptome. The coverage within contigs is calculated by dividing the total number of base pairs contained in the reads used to construct a contig by the length of that contig. Note that Newbler v2.3 discards those contigs <100 bp

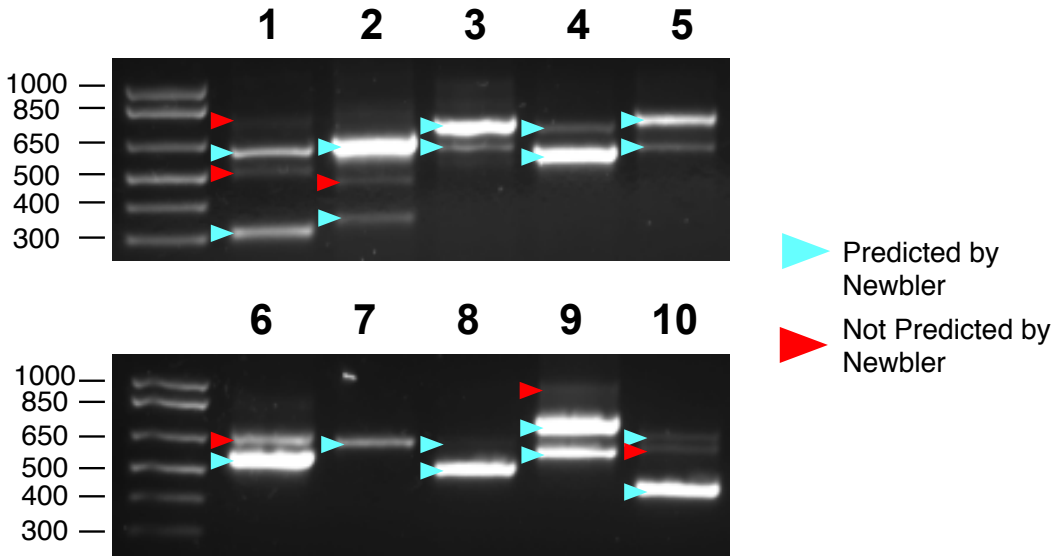
A



B

Lane	Isogroup	Predicted Band 1	Band 1 Present?	Predicted Band 2	Band 2 Present?	Additional Bands Present?
1	935	625	Y	319	Y	Y
2	948	657	Y	347	Y	Y
3	984	756	Y	628	Y	N
4	1045	759	Y	609	Y	N
5	1082	789	Y	623	Y	N
6	1133	701	N	579	Y	Y
7	1134	652	Y	519	N	N
8	1144	659	Y	515	Y	N
9	1162	759	Y	585	Y	Y
10	1179	700	Y	437	Y	Y

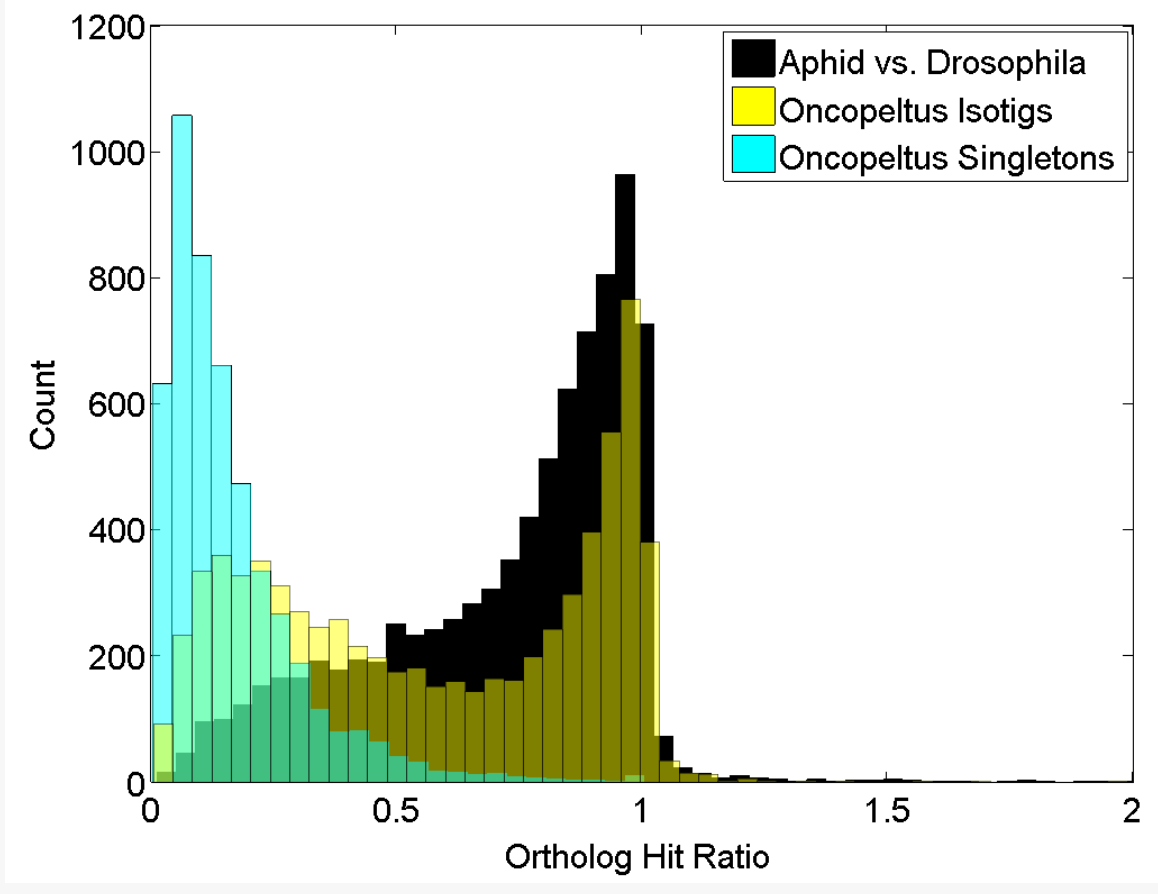
C



Additional File 3. RT-PCR validation of bioinformatically predicted multiple isoforms. (*A*) Schematic of experimental design. Ten isogroups were randomly selected, each containing exactly two isotigs that differed by the presence/absence of a single contig. PCR primers were designed to flank the differing region. (*B*) Band sizes predicted by Newbler v2.3 for ten randomly selected isogroups containing exactly two isotigs. *(C)* Agarose gel following RT-PCR using primers against the sequences described in (*B*). Ladder sizes are given in base pairs on the left. Blue arrowheads: bands of the sizes predicted by Newbler v2.3; red arrowheads: bands not predicted by Newbler v2.3.

Top BLAST hit taxa	Isotigs	Singletons (CAP3-assembled)	Total
Holometabola	4,311	2,650	6,961
Hemimetabola	2,156	1,640	3,796
Deuterostomes	358	234	592
Non-hexapod arthropods	62	58	120
Non-bilaterian metazoa	47	48	95
Non-metazoa	21	0	21
Others	264	331	595

Additional File 4. Identity of taxa with top BLAST hits. "Isotigs" refers only to the longest isotig of each isogroup; "Singletons" refers to the Newbler-generated singletons after secondary CAP3 assembly. The category "other" is the summation of all those species obtaining very low numbers of BLAST hits.



Additional File 5. *O. fasciatus* assembly isotigs have ortholog hit ratios similar to predictions from fully genome-sequenced databases. When isotigs from the *O. fasciatus* transcriptome are BLASTed against the RefSeq protein database, ortholog hit ratios show a similar profile to those obtained when the complete *Acyrthosiphon pisum* gene prediction set (downloaded from http://www.aphidbase.com/aphidbase/downloads/ *webcite*) is BLASTed against the predicted gene set of *Drosophila melanogaster* (r5.28 downloaded from ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/ *webcite*) with an e-value cut-off of 1e-10.

GO Term	FDR	# Normalized (total n = 750)	# Non-Normalized (total n = 1124)	Enriched
establishment of localization (GO:0051234)	0.0253	107	227	NN
transport (GO:0006810)	0.0283	106	224	NN
transporter activity (GO:0005215)	0.0475	50	121	NN
ATPase activity (GO:0016887)	0.0174	35	99	NN
establishment of localization in cell (GO:0051649)	0.0411	32	88	NN
vesicle-mediated transport (GO: 0016192)	0.0111	16	61	NN
active transmembrane transporter activity (GO:0022804)	0.0001	11	62	NN
ATPase activity, coupled to movement of substances (GO:0043492)	0.0001	7	51	NN
hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances (GO:0016820)	0.0001	6	52	NN
primary active transmembrane transporter activity (GO:0015399)	0.0001	6	49	NN
P-P-bond-hydrolysis-driven transmembrane transporter activity (GO:0015405)	0.0001	6	49	NN
ATPase activity, coupled to transmembrane movement of substances (GO:0042626)	0.0001	6	49	NN
GTPase activator activity (GO:0005096)	0.0480	6	31	NN
ATPase activity, coupled to transmembrane movement of ions (GO:0042625)	0.0475	2	20	NN
ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism (GO:0015662)	0.0325	1	18	NN
lipid transporter activity (GO:0005319)	0.0475	0	13	NN
cytosolic part (GO:0044445)	0.0031	21	4	N
structural constituent of ribosome (GO:0003735)	0.0282	17	4	N
ribosomal subunit (GO:0033279)	0.0059	14	1	N
large ribosomal subunit (GO:0015934)	0.0381	9	0	N

Additional File 6. GO terms enriched in Normalized (N) and Non-Normalized (NN) cDNA samples. N (assembly generated from full plate of normalized cDNA) and NN (assembly generated from an equalized number of base pairs of non-normalized cDNA) reads were BLASTed against the full transcriptome assembly, and the results were used to generate "test" and "reference" sets for a Fisher's Exact Test. FDR: false discovery rate.

	Newbler v.3	Newbler 2.5
Total bases assembled	19,921,298	20,096,403
Isogroups ("genes")	16,629	16,849
Isotigs ("transcripts")	21,097	20,985
Isotig N50	1,735	1,651
Mean # isotigs per isogroup	1.3	1.2
Contigs ("exons")	22,235	25,955
Mean # contigs per isotig	1.9	1.8
Singletons (singletons after secondary cap3 assembly)	178,770 (112,531)	168,807 (114,487)
Total # unique genes identified BLASTx	10,775	10,886

Additional File 7. Comparison of *de novo* transcriptome assemblies produced by Newbler v2.3 and Newbler v2.5. Number of BLASTx hits reflects a search against RefSeq Protein database with an e-value cut-off value of 1

Current Biology 22, 2278–2283, December 4, 2012 ©2012 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2012.10.019

CHAPTER 2

Report

oskar Predates the Evolution of Germ Plasm in Insects

Ben Ewen-Campen,[1] John R. Srouji,[2] Evelyn E. Schwager,[1] and Cassandra G. Extavour[1,*]
[1]Department of Organismic and Evolutionary Biology
[2]Department of Molecular and Cellular Biology
Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Summary

***oskar* is the only gene in the animal kingdom necessary and sufficient for specifying functional germ cells [1, 2]. However, *oskar* has only been identified in holometabolous ("higher") insects that specify their germline using specialized cytoplasm called germ plasm [3]. Here we show that *oskar* evolved before the divergence of higher insects and provide evidence that its germline role is a recent evolutionary innovation. We identify an *oskar* ortholog in a basally branching insect, the cricket *Gryllus bimaculatus*. In contrast to *Drosophila oskar*, *Gb-oskar* is not required for germ cell formation or axial patterning. Instead, *Gb-oskar* is expressed in neuroblasts of the brain and CNS and is required for neural development. Taken together with reports of a neural role for *Drosophila oskar* [4], our data demonstrate that *oskar* arose nearly 50 million years earlier in insect evolution than previously thought, where it may have played an ancestral neural role, and was co-opted to its well-known essential germline role in holometabolous insects.**

Results and Discussion

Animal germ cells can be specified either through the cytoplasmic inheritance of maternally deposited germ plasm or through inductive cell signaling [5]. In *Drosophila melanogaster*, germ cells form by incorporating germ plasm deposited and localized at the oocyte posterior. Germ plasm assembly requires *oskar* [6], which is necessary and sufficient for germ cell specification [1, 2]. Oskar is localized at the oocyte posterior, where it promotes the accumulation of conserved germline factors including Vasa, PIWI, and Tudor proteins [7] and induces posterior patterning by recruiting *nanos* mRNA [8]. Despite its essential role in *Drosophila* germ cell formation and axial patterning, *oskar* has only been identified in the genomes of a small number of holometabolous insects, all of which specify their germ cells via germ plasm [3]. In contrast, *oskar* is absent from the genomes of holometabolous insects that lack germ plasm, and neither *oskar* nor germ plasm have been identified in any insect taxa that branch basally to Holometabola [3]. The prevailing hypothesis is therefore that *oskar* arose as a novel gene at the base of Holometabola coincidentally with the evolution of insect germ plasm [3].

Here we report the first discovery of an *oskar* ortholog in a basally branching insect that lacks germ plasm, the cricket *Gryllus bimaculatus* (Orthoptera). We unexpectedly detected *Gb-oskar* in a combined *Gryllus* ovarian and embryonic de novo transcriptome. Three lines of evidence confirm that *Gb-oskar* is a bona fide *oskar* ortholog. First, Gb-Oskar is the reciprocal best protein BLAST hit against the protein products of all known *oskar* genes from flies, mosquitoes, ants, and the wasp *Nasonia*. Second, *Gb-oskar* encodes the LOTUS (aka OST-HTH) and SGNH hydrolase domains that characterize all *oskar* orthologs [3, 9, 10] (Figures 1A and see Figure S1 available online). Physicochemical conservation is pronounced within both the LOTUS and SGNH hydrolase domains of Gb-Oskar (37.1% and 36.3%, respectively), while amino acid identity is less strongly conserved (11.4% and 21.2%, respectively), consistent with previous observations of *oskar* orthologs [3, 11]. Third, phylogenetic reconstruction clearly places *Gb-oskar* with other known *oskar* genes and not within the *tdrd7* genes (Figure 1B), a conserved metazoan gene family that also contains the LOTUS domain [9, 10]. Thus, in contrast to previous hypotheses that *oskar* first arose in the lineage leading to the Holometabola [3], our analyses demonstrate that *oskar* was present at least ~50 million years earlier than previously thought [3] in the common ancestor of Orthoptera and Holometabola [12].

The presence of *oskar* in the genome of an insect branching basal to the Holometabola is surprising because neither germ plasm nor pole cells have been described in any of these taxa. Instead, germ cells are thought to arise from the mesoderm during midembryogenesis in orthopterans [13] and most other early diverging insects [14]. We therefore asked whether *Gb-oskar* plays a conserved role in germ cell formation in *Gryllus* or whether this function arose later during insect evolution. We examined the expression of *Gb-oskar* mRNA and protein (Figure S2) during oogenesis and embryogenesis. In stark contrast to *oskar* expression in *Drosophila* [15] and *Nasonia* [3], *Gb-oskar* mRNA and protein do not accumulate at the posterior of *Gryllus* oocytes and instead are distributed ubiquitously throughout all stages of oogenesis (Figures 2A–2D). In just-laid eggs, *Gb-oskar* mRNA does not localize posteriorly and is barely detectable by in situ hybridization (Figure 2E), although RT-PCR confirms that *Gb-oskar* is expressed throughout all stages of embryogenesis (Figure 2J). *Gryllus* primordial germ cells, marked by *Gb-piwi* and *Gb-vasa* transcript (Figures 2H and 2I) and protein (Figures 2H′ and 2I′; Figures S2G and S2H) expression, arise during abdominal segmentation stages, but expression of both *Gb-oskar* mRNA and Gb-Oskar protein remains at low levels in all embryonic cells throughout these stages and does not become enriched in primordial germ cells before or during their formation (Figures 2G and 2G′). The expression pattern of *Gb-oskar* therefore does not support a role for this gene in germ cell formation.

To directly test whether *Gb-oskar* is functionally required for germ cell formation in *Gryllus*, we knocked down *Gb-oskar* function during oogenesis and embryogenesis using maternal RNAi (mRNAi) and embryonic RNAi (eRNAi), respectively [16] (Figures 2, S3A, and S3B). In contrast to *Drosophila* and *Nasonia*, maternal knockdown of *Gb-oskar* did not reduce egg laying (Figure S3C), affect ovary morphology, or impede

*Correspondence: extavour@oeb.harvard.edu

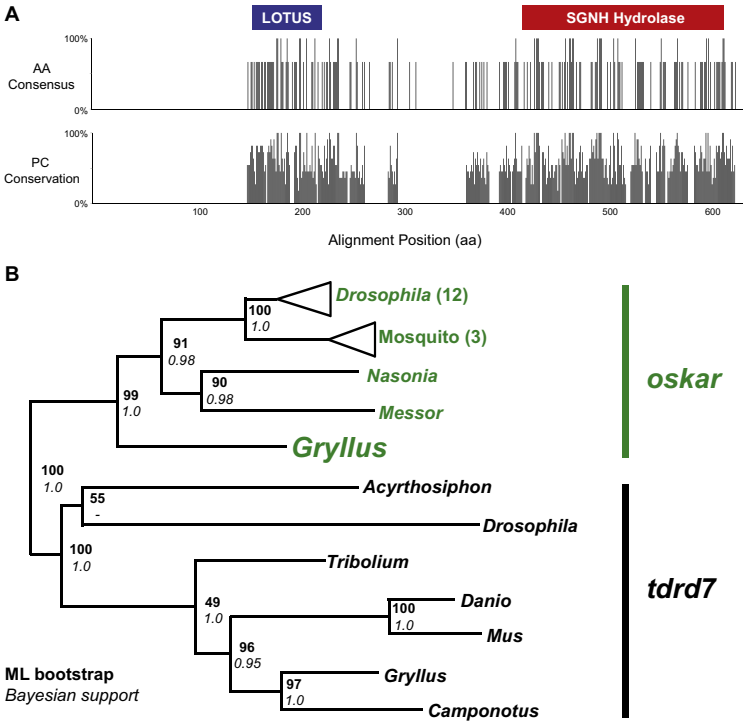


Figure 1. Domain Structure and Phylogenetic Analysis of *Gb-oskar*

(A) ClustalW protein alignment of Oskar orthologs from *Gryllus bimaculatus*, *Nasonia vitripennis*, and *Drosophila melanogaster* (see Figure S1 for alignment showing amino acids). Sequence identity at the amino acid level is not strongly conserved (top graph), but physicochemical conservation is more pronounced (bottom graph), specifically within the conserved LOTUS (blue) and SGNH hydrolase (red) domains in all three proteins, meaning that the chemical properties of these domains are conserved at a local level despite amino acid divergence and suggesting that these regions may represent functional domains of Oskar. (B) Phylogenetic reconstruction of *Gb-oskar* with known *oskar* (green) and *tdrd7* (black) orthologs. The best-scoring unrooted ML topology is shown: bold, ML bootstrap values; italics, Bayesian posterior probability.

the progress of oogenesis (Figure S3D). Embryonic *Gb-oskar* knockdown did not cause any of the posterior patterning defects seen in *Drosophila osk* mutants [6] (Figure S3E, Table S1), and these embryos were morphologically wild-type and hatched at normal rates (Figure S3F). These data show that in contrast to the known requirement for *oskar* in anterior-posterior (A-P) axial patterning in holometabolous insects [3, 6], *oskar* does not direct *Gryllus* axial patterning. Further, *Gb-oskar* eRNAi embryos produced germ cells that expressed Gb-Piwi and Gb-Vasa (Figures 2K–2N) and ultimately produced functional ovaries in adulthood (Figure 2P).

In contrast to the essential and conserved role that *oskar* plays in Holometabolous germ cell formation [3, 8], our analyses of *Gb-oskar* gene expression and function show that this gene is not required for germ cell formation in *Gryllus*. We therefore hypothesized that *Gb-oskar* has a distinct somatic function in *Gryllus* that may reflect an ancestral function for this gene. Consistent with this hypothesis, we observed that *Gb-oskar* has a specific somatic expression pattern during midembryogenesis: *Gb-oskar* mRNA and protein are enriched in neuroblasts along the A-P axis (Figures 3A–3C″). Neuroblasts are neural stem cells that arise from the ventral ectoderm and produce neurons of the CNS in *Drosophila* and other Pancrustacea (insects and crustaceans) [17]. *Gb-oskar* expression in the neuroblasts begins during the earliest stages of neurogenesis and persists throughout all stages examined. In addition to this embryonic expression pattern, *Gb-oskar* is also expressed in the adult brain (Figure 2J).

We examined embryonic nervous system development in *Gb-oskar* knockdown conditions and found that *Gb-oskar* eRNAi embryos exhibit morphological defects of the axonal tracts that are consistent with an impairment of neuroblast divisions: lateral connectives are often broken or reduced in width compared to controls (Figures 3D–3D″, $p < 0.001$), anterior commissure formation is significantly delayed or absent (Figures 3E–3E″, $p < 0.01$), and posterior commissure formation is similarly disrupted (Figures 3F–3F″, $p < 0.025$). These axonal defects suggest that *Gb-oskar* may be required for proper neuronal determination and are reminiscent of the axonal scaffold defects of *Drosophila miranda* and *prospero* mutants, which disrupt neuroblast divisions resulting in neuronal misspecification [18].

Aberrant neuroblast divisions can be assayed by examining the expression of *even-skipped*, which is expressed in a stereotyped subset of ganglion mother cells and neurons within each segment, including the aCC and pCC neurons that are homologous across insects [19]. We found that 34.5% of *Gb-oskar* eRNAi embryos displayed significant defects in aCC/pCC specification (Figures 3G–3G″, $p < 0.01$) that were never observed in controls, indicating that *Gb-oskar* is required for proper neuroblast division.

Our results demonstrate a role for *Gb-oskar* in the development of the cricket CNS, in contrast to its well-known role in germ plasm formation in holometabolous insects. These divergent functions of *oskar* suggest at least two possible evolutionary scenarios. First, *oskar*'s ancestral role in insects could be that of germ plasm assembly as seen in *Drosophila*, and the CNS function we report here could be a derived trait in the branch of the insect tree leading to *Gryllus*. However, several lines of evidence support a second scenario, whereby *oskar*'s neural function is ancestral to Orthoptera (crickets, locusts, and grasshoppers) and Holometabola, which diverged approximately 380 million years ago (Mya) [12], thus implying that its role in assembling germ plasm is a derived trait in higher insects (Figure 4). In support of this interpretation, we note that insect germ plasm itself is thought to be a derived trait within insects and unique to Holometabola [5], consistent with our observation that *Gryllus* possesses an *oskar* ortholog yet lacks germ plasm. Moreover, in *Drosophila* adults, *oskar* is expressed in the brain and is required for place learning, as are

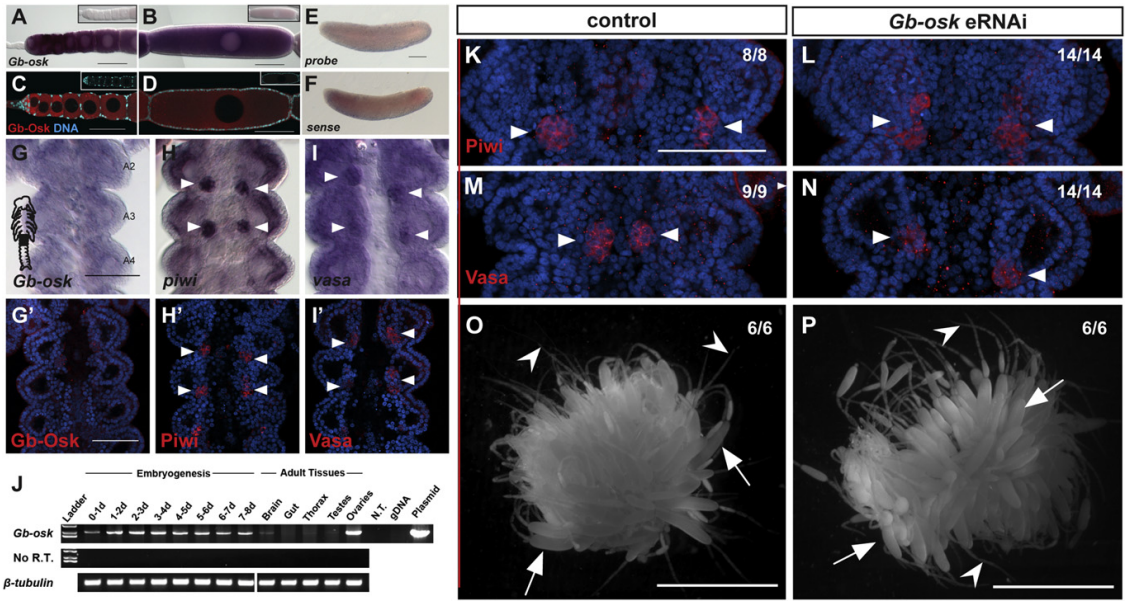


Figure 2. *Gb-oskar* mRNA and Protein Do Not Accumulate in Germ Plasm or Embryonic Germ Cells and Are Not Required for Germ Cell Formation or Development

(A–D) *Gb-oskar* mRNA (A and B) and protein (C and D) are expressed ubiquitously in oocytes and do not accumulate asymmetrically at the posterior. Insets show sense probe (A and B) and preimmune serum (C and D) controls. (E) *Gb-oskar* transcript levels are low and ubiquitous in blastoderm stage embryos. (F) Gb-oskar sense control. (G–I′) In fully segmented embryos, *Gb-oskar* mRNA (G) and protein (G′) are not enriched in embryonic germ cells, which express *piwi* and *vasa* mRNA (H and I) and protein (H′ and I′). Arrowheads in (H)–(I′) indicate germ cell clusters. A2, A3, and A4 indicate abdominal segments 2, 3, and 4. Black region of embryo schematic in (G) shows the region displayed in (G)–(I′). (J) RT-PCR analysis of *Gb-oskar* throughout embryogenesis and in different adult tissues. β-*tubulin* was used to ensure equal quantities of RNA template in the cDNA synthesis reaction and as a gel loading control. The amplified *Gb-oskar* band is 2,149 bp and was amplified with 35× PCR cycles. The highest levels of *Gb-oskar* are detected throughout embryogenesis and in adult ovaries. Lower levels are detected in the adult brain, and faint expression is detected in the adult thoracic muscles and testes. No expression is detected in the adult gut. N.T., no template control; gDNA, genomic DNA control; Plasmid, *Gb-oskar* amplified from the full-length plasmid clone; No R.T., no reverse transcriptase control. (K–N) Vasa- and Piwi-positive germ cells form in *Gb-oskar* eRNAi embryos (L and N) as in *DsRed* eRNAi controls (K and M). (P and O) *Gb-oskar* eRNAi embryos raised to adulthood form functional ovaries (P), which contain functional germaria (arrowheads) and late-stage oocytes (arrows) as in uninjected controls (O). Numbers indicate sample sizes. Left shows anterior in (A)–(E) and top shows anterior in (G)–(I′) and (K)–(N). Scale bar represents 100 μm in (A)–(M) and 5 mm in (N) and (O). Validation of all Gb-specific antibodies is shown in Figure S2; validation of *Gb-oskar* eRNAi is shown in Figure S3. Absence of a role for *Gb-oskar* in axial patterning is shown in Figure S3 and Table S1.

several other *Drosophila* genes with germline functions [4, 20], suggesting that these genes may have an ancient association with the nervous system.

Multiple *Drosophila* genes originally characterized for their role in germ cell development, including *nanos*, *pumilio*, and *Staufen*, have subsequently been shown to function in the nervous system, where they regulate translation in such processes as dendrite morphogenesis, synaptic growth, asymmetric neuroblast divisions, and neuronal specification [4, 21–23]. The co-occurrence of multiple germ plasm genes in the CNS of *D. melanogaster* raises the intriguing possibility that these genes may function within an evolutionarily conserved functional module [24], which could facilitate their co-option to a novel context such as holometabolous germ plasm. Consistent with this hypothesis, we find that *Gb-vasa* mRNA and protein are coexpressed with *Gb-oskar* in neuroblasts (Figures S2I–S2K), suggesting that a functional link between *oskar* and other germline genes predates the evolution of germ plasm. Moreover, expression of germline genes in the nervous system has also been observed in other insects belonging to both Hemimetabola (*piwi* in aphids [25]) and Holometabola (*vasa* in ants [26]).

If *oskar* were to acquire expression in germ cells due to its functional linkage with other germ plasm genes, an evolutionary change in its transcriptional, translational, or functional regulation might then have feasibly allowed its co-option to a critical function in the germline specification pathway. Consistent with this possibility, we note the presence of extremely low levels of *Gb-oskar* in *Gryllus* germ cells (Figures 2G and 2G′), although it appears to play no essential germ cell function (Figures 2K–2P). Co-option of *oskar* to assemble germ plasm probably involved molecular evolution of its regulation and function; accordingly, we find that *Gb-oskar* is not regulated by *Drosophila oskar* translational machinery in *Drosophila* oocytes or embryos (Figure S4, Table S2), suggesting that *oskar*'s translational regulation mechanisms have evolved extensively in the lineage leading to *Drosophila*. However, it is also possible that specific features of the *Gb-oskar* coding sequence, or its incompatibility with *Drosophila* UTRs, may have prevented the translation of Gb-Oskar in our transgenic experiments.

We have shown that *oskar* was present nearly 50 million years earlier in insect evolution than previously thought [3, 12] and must therefore have been lost several times in

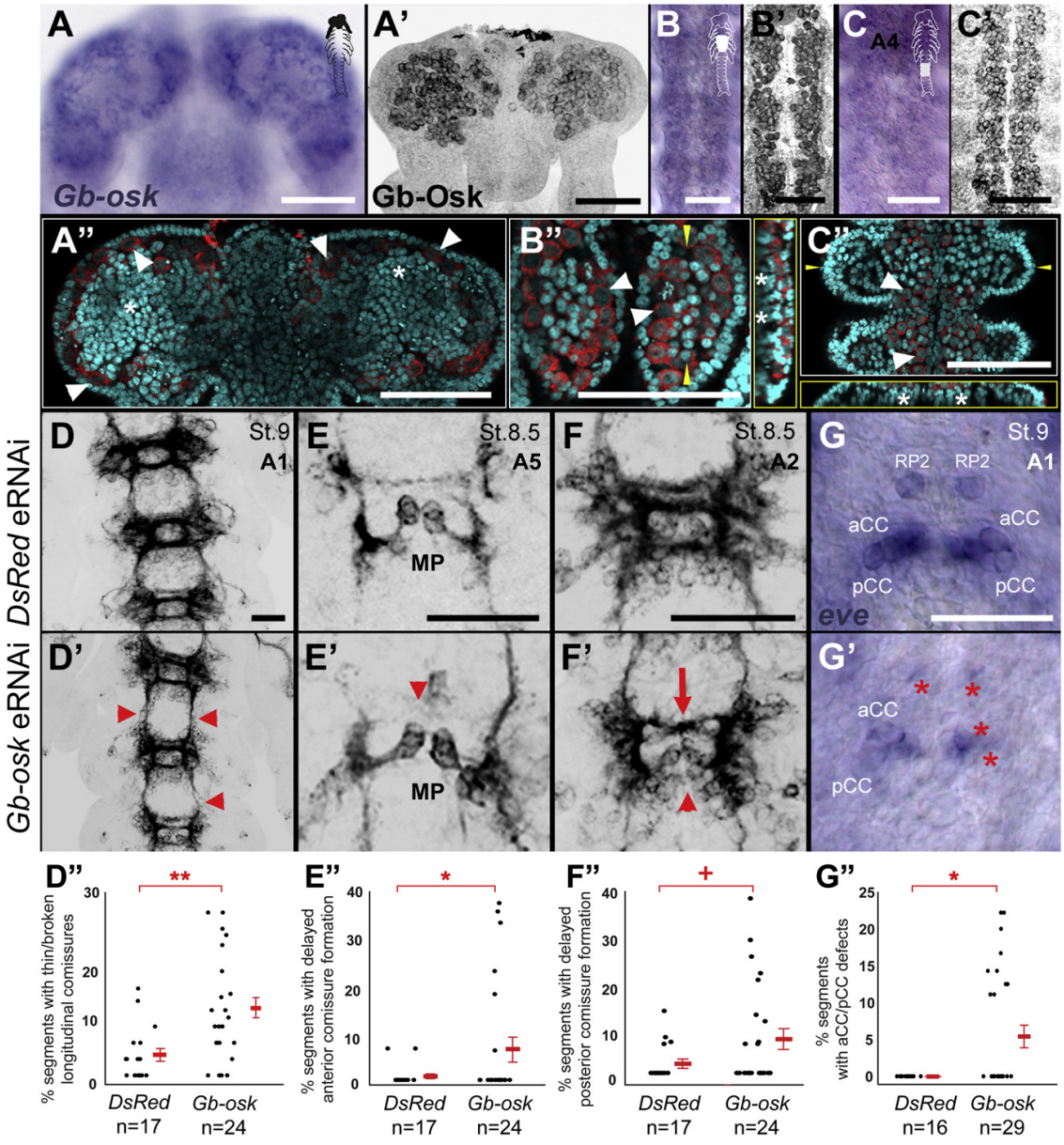


Figure 3. *Gb-oskar* Is Expressed in Neuroblasts and Is Required for Neural Development

(A– C″) *Gb-oskar* mRNA (A, B, and C) and protein (A′, B′, and C′) accumulate in neuroblasts of the brain (A and A′), thorax (T1–T3; B and B′), and abdomen (A4–A6; C and C′). (A″), (B″), and (C″) show single optical sections of Gb-Osk expression (red, Gb-Osk; cyan, nuclei), revealing the highest levels of Gb-Osk expression in neuroblasts (cells with large nuclei and diffuse chromatin; arrowheads) and absent or lower levels in neuroblast daughter cells (cells with smaller, denser nuclei; asterisks). Bright staining near midline of head in (A′) and (A″) is nonspecific staining of the extraembryonic membranes. Large panels of (B″) and (C″) show single optical sections through T3 and A5 neuromeres, respectively; anterior is on top in both central panels. Yellow framed boxes to the left and bottom of (B″) and (C″), respectively, show orthogonal sections at the plane indicated by yellow arrowheads in the large panels; in these orthogonal sections, ventral is on the right in (B″) and on top in (C″). White arrowheads indicate Gb-Osk-expressing neuroblasts immediately dorsal to the ectoderm and ventral to the underlying daughter neurons and neuronal precursors (asterisks), which show little to no expression of Gb-Osk. (D–F′) Ventral views of embryonic abdominal segments of *DsRed* eRNAi controls (D, E, and F) and *Gb-oskar* eRNAi in embryos of the same developmental stage (D′, E′, and F′) labeled with axonal marker anti-HRP. The percent of *Gb-oskar* eRNAi embryos that exhibit thin or broken longitudinal connectives (arrowheads) is 83.3% (D and D′). *Gb-oskar* eRNAi embryos show delayed formation of both anterior (E and E′) and posterior (F and F′) commissures (arrowheads) relative to the development of the midline precursors (MP). Abnormal or fused anterior commissures also appear in a greater proportion of *Gb-oskar* eRNAi embryos than in controls (arrow in F′). (G and G′) Formation of aCC and pCC neurons is impaired in *Gb-oskar* eRNAi embryos (asterisks) (G′) but never

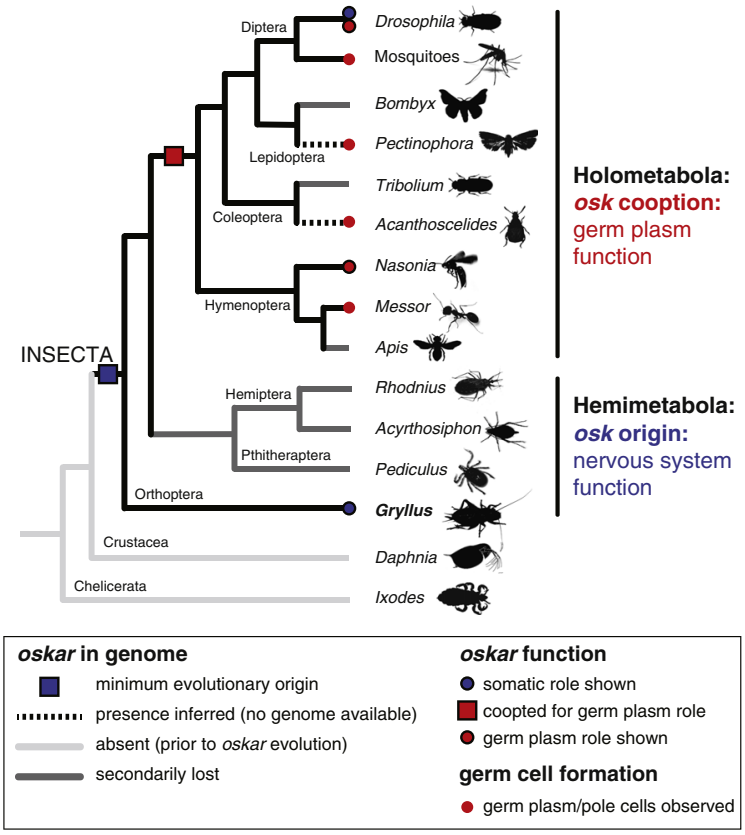


Figure 4. Phylogenetic Hypothesis of *oskar* Origin and Function across Arthropods

Species shown have sequenced genomes, allowing unambiguous determination of the presence or absence of an *oskar* ortholog; dotted lines indicate exceptions to this rule and show taxa in which pole cells and/or germ plasm have been observed but that lack available genome sequence. *oskar* is absent from the sequenced genomes of noninsect arthropods (light gray lines). Boxes indicate proposed origin of *oskar* in the last common ancestor of Holometabola and Hemimetabola (blue) and putative co-option of *oskar* to a germline role in the Holometabola (red). Circles indicate that *oskar* plays known roles in the nervous system (blue) or presumably in the germline (red); circles outlined in black indicate that there is functional evidence for the described role. Evidence for evolution of *oskar* translational regulation in the lineage leading to *D. melanogaster* is shown in Figure S4 and Table S2.

some insect lineages (Figure 4). Indeed, completely sequenced genomes of holometabolous insects lacking germ plasm or pole cells confirm that *oskar* has been lost in these lineages [3]. Germ cell specification via germ plasm is thought to have arisen independently in multiple bilaterian taxa [5], but how germ plasm evolved has remained unclear. Our results suggest a novel molecular mechanism for this process in insects: co-option of the *oskar* gene into the top of the germ plasm assembly hierarchy.

Experimental Procedures

Animal Husbandry, Gene Expression, and Functional Analysis

G. bimaculatus husbandry, gene expression analysis, RNAi experiments, egg-laying analysis, and axonal scaffold visualization were carried out as previously described [16].

Gene Cloning and Phylogenetic Analysis

Full-length *Gb-oskar* was recovered from a *G. bimaculatus* transcriptome and its identity confirmed by both Bayesian and Maximum Likelihood analysis. Details of sequence analysis are available in Supplemental Experimental Procedures.

Antibody Generation

Rabbit polyclonal antibodies were raised against an N-terminal and a C-terminal peptide from Gb-Oskar (Figures S1 and S2A) (Open Biosystems), recombinant proteins of full-length Gb-Vasa, and a 774 amino acid fragment of Gb-Piwi (McGill Biology CIAN facility). Details of antibody construction and validation are described in Supplemental Experimental Procedures.

Accession Numbers

The Genbank accession numbers for the *Gb-oskar*, *Gb-piwi-like*, and *Gb-tdrd7* sequences reported in this paper are JQ434102, JQ434103, and JQ434104, respectively.

Supplemental Information

Supplemental Information includes four figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2012.10.019.

Acknowledgments

NSF IOS-0817678 to C.G.E., DFG postdoctoral fellowship SCHW 1557/1-1 to E.E.S., and NSF predoctoral fellowships to B.E.-C. and J.R.S. supported

in controls (G). aCC/pCC are located in the corners where the longitudinal connectives meet the posterior commissure; these axonal scaffolds are visible in (G). The out-of-focus darkened spots adjacent to the in-focus aCC/pCC neurons are U/CQ neurons present ventral to the dorsally located aCC/pCC neurons. (D″, E″, F″, and G″) Quantification of neural defects illustrated in (D), (E), (F), and (G); thick red bars at the bottom of plots show mean values ± SE. Statistical significance of differences between treatments (red brackets) based on chi-square tests: $^{**}p < 0.001$, $^{*}p < 0.01$, $^{+}p < 0.025$. Anterior is shown on the top in all panels. Scale bar represents 100 μM in (A)–(C″) and 50 μM in (D)–(G″). Embryonic stage and/or the most anterior segment shown are indicated in top right corner in (C), (D), (E), (F), and (G); stages and segments shown in (D), (E), (F), and (G) apply to (D′), (E′), (F′), and (G′), respectively. Validation of *Gb-oskar* eRNAi is shown in Figures S2 and S3.

this work. We thank Elke Küster-Schöck of the McGill Biology Cell Imaging and Analysis Network (Canada) for assistance with antibody creation, Andrés Leschziner for western blot assistance, Anne Ephrussi, Satoru Kobayashi, Sam Kunes, and Akira Nakamura for reagents, and members of the Extavour laboratory for discussion.

Received: August 22, 2012
Revised: September 16, 2012
Accepted: October 9, 2012
Published online: November 1, 2012

References

1. Ephrussi, A., and Lehmann, R. (1992). Induction of germ cell formation by *oskar*. Nature *358*, 387–392.
2. Smith, J.L., Wilson, J.E., and Macdonald, P.M. (1992). Overexpression of *oskar* directs ectopic activation of *nanos* and presumptive pole cell formation in *Drosophila* embryos. Cell *70*, 849–859.
3. Lynch, J.A., Ozüak, O., Khila, A., Abouheif, E., Desplan, C., and Roth, S. (2011). The phylogenetic origin of *oskar* coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola. PLoS Genet. *7*, e1002029.
4. Dubnau, J., Chiang, A.-S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U., et al. (2003). The *staufen/pumilio* pathway is involved in *Drosophila* long-term memory. Curr. Biol. *13*, 286–296.
5. Extavour, C.G., and Akam, M.E. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. Development *130*, 5869–5884.
6. Lehmann, R., and Nüsslein-Volhard, C. (1986). Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of *oskar*, a maternal gene in *Drosophila*. Cell *47*, 141–152.
7. Mahowald, A.P. (2001). Assembly of the *Drosophila* germ plasm. Int. Rev. Cytol. *203*, 187–213.
8. Ephrussi, A., Dickinson, L.K., and Lehmann, R. (1991). Oskar organizes the germ plasm and directs localization of the posterior determinant *nanos*. Cell *66*, 37–50.
9. Anantharaman, V., Zhang, D., and Aravind, L. (2010). OST-HTH: a novel predicted RNA-binding domain. Biol. Direct *5*, 13.
10. Callebaut, I., and Mornon, J.-P. (2010). LOTUS, a new domain associated with small RNA pathways in the germline. Bioinformatics *26*, 1140–1144.
11. Webster, P.J., Suen, J., and Macdonald, P.M. (1994). *Drosophila virilis oskar* transgenes direct body patterning but not pole cell formation or maintenance of mRNA localization in *D. melanogaster*. Development *120*, 2027–2037.
12. Gaunt, M.W., and Miles, M.A. (2002). An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol. Biol. Evol. *19*, 748–761.
13. Wheeler, W.M. (1893). A contribution to insect morphology. J. Morphol. *8*, 1–160.
14. Johannsen, O.A., and Butt, F.H. (1941). Embryology of Insects and Myriapods (New York: McGraw-Hill).
15. Kim-Ha, J., Smith, J.L., and Macdonald, P.M. (1991). *oskar* mRNA is localized to the posterior pole of the *Drosophila* oocyte. Cell *66*, 23–35.
16. Kainz, F., Ewen-Campen, B., Akam, M., and Extavour, C.G. (2011). Notch/Delta signalling is not required for segment generation in the basally branching insect *Gryllus bimaculatus*. Development *138*, 5015–5026.
17. Stollewerk, A., and Simpson, P. (2005). Evolution of early development of the nervous system: a comparison between arthropods. Bioessays *27*, 874–883.
18. Ikeshima-Kataoka, H., Skeath, J.B., Nabeshima, Y., Doe, C.Q., and Matsuzaki, F. (1997). Miranda directs Prospero to a daughter cell during *Drosophila* asymmetric divisions. Nature *390*, 625–629.
19. Thomas, J.B., Bastiani, M.J., Bate, M., and Goodman, C.S. (1984). From grasshopper to *Drosophila*: a common plan for neuronal development. Nature *310*, 203–207.
20. Chintapalli, V.R., Wang, J., and Dow, J.A. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat. Genet. *39*, 715–720.
21. Li, P., Yang, X., Wasser, M., Cai, Y., and Chia, W. (1997). Inscuteable and Staufen mediate asymmetric localization and segregation of *prospero* RNA during *Drosophila* neuroblast cell divisions. Cell *90*, 437–447.
22. Menon, K.P., Sanyal, S., Habara, Y., Sanchez, R., Wharton, R.P., Ramaswami, M., and Zinn, K. (2004). The translational repressor Pumilio regulates presynaptic morphology and controls postsynaptic accumulation of translation factor eIF-4E. Neuron *44*, 663–676.
23. Ye, B., Petritsch, C., Clark, I.E., Gavis, E.R., Jan, L.Y., and Jan, Y.N. (2004). *nanos* and *pumilio* are essential for dendrite morphogenesis in *Drosophila* peripheral neurons. Curr. Biol. *14*, 314–321.
24. Wagner, G.P., Pavlicev, M., and Cheverud, J.M. (2007). The road to modularity. Nat. Rev. Genet. *8*, 921–931.
25. Lu, H.L., Tanguy, S., Rispe, C., Gauthier, J.P., Walsh, T., Gordon, K., Edwards, O., Tagu, D., Chang, C.C., and Jaubert-Possamai, S. (2011). Expansion of genes encoding piRNA-associated Argonaute proteins in the pea aphid: diversification of expression profiles in different plastic morphs. PLoS ONE *6*, e28051.
26. Khila, A., and Abouheif, E. (2010). Evaluating the role of reproductive constraints in ant social evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. *365*, 617–630.

Current Biology, Volume *22*

Supplemental Information

oskar Predates the Evolution of Germ Plasm in Insects

Ben Ewen-Campen, John R. Srouji, Evelyn E. Schwager, and Cassandra G. Extavour

Supplemental Inventory

- *Supplemental Data*
  - Figures S1-S4
  - Tables S1 and S2
- *Supplemental Experimental Procedures*
- *Supplemental References*

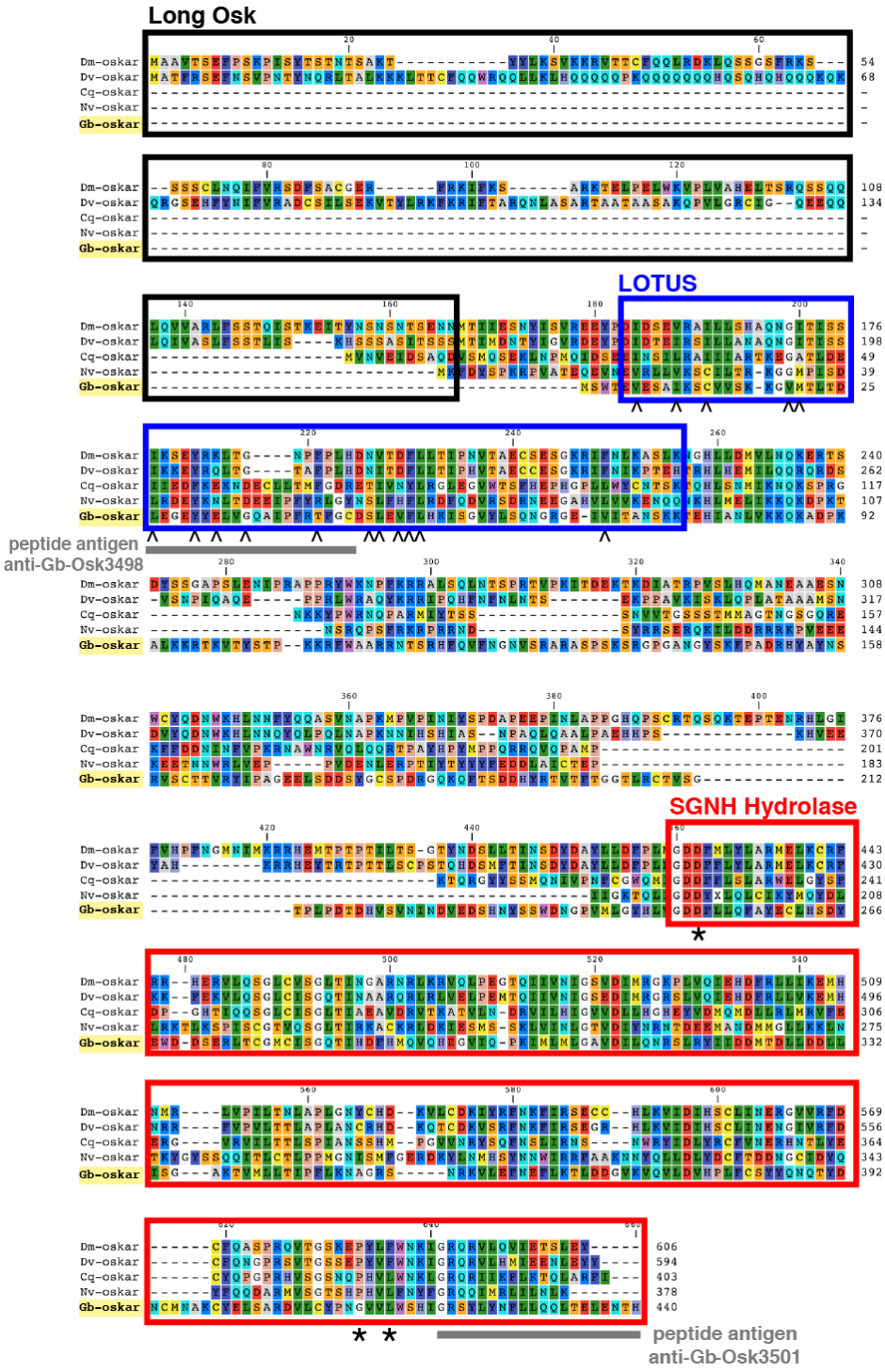
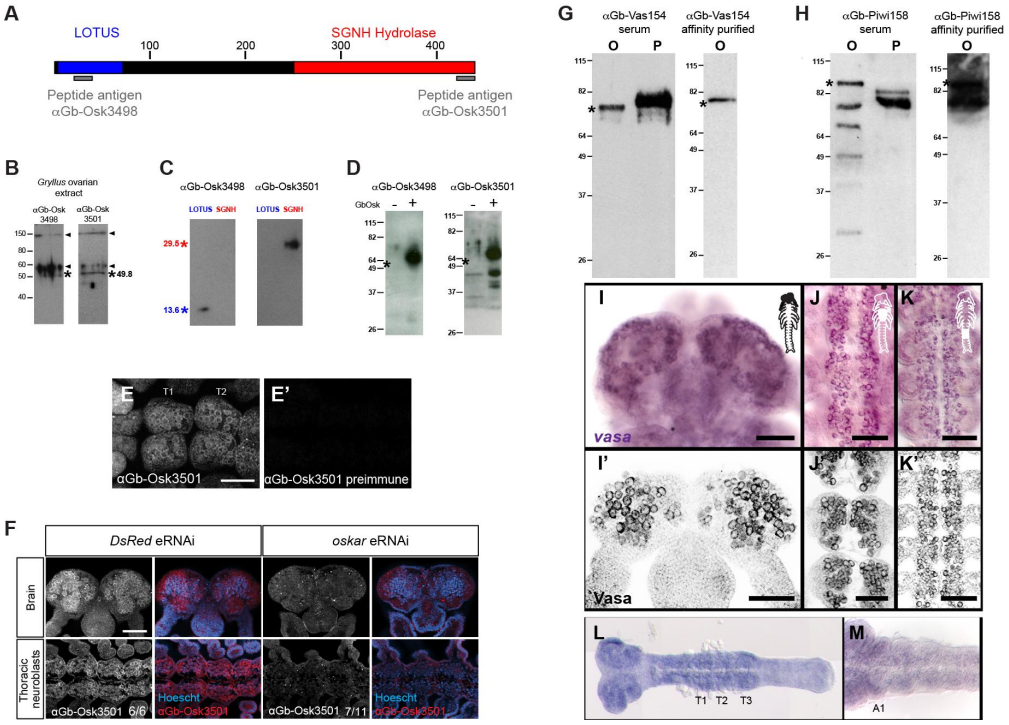
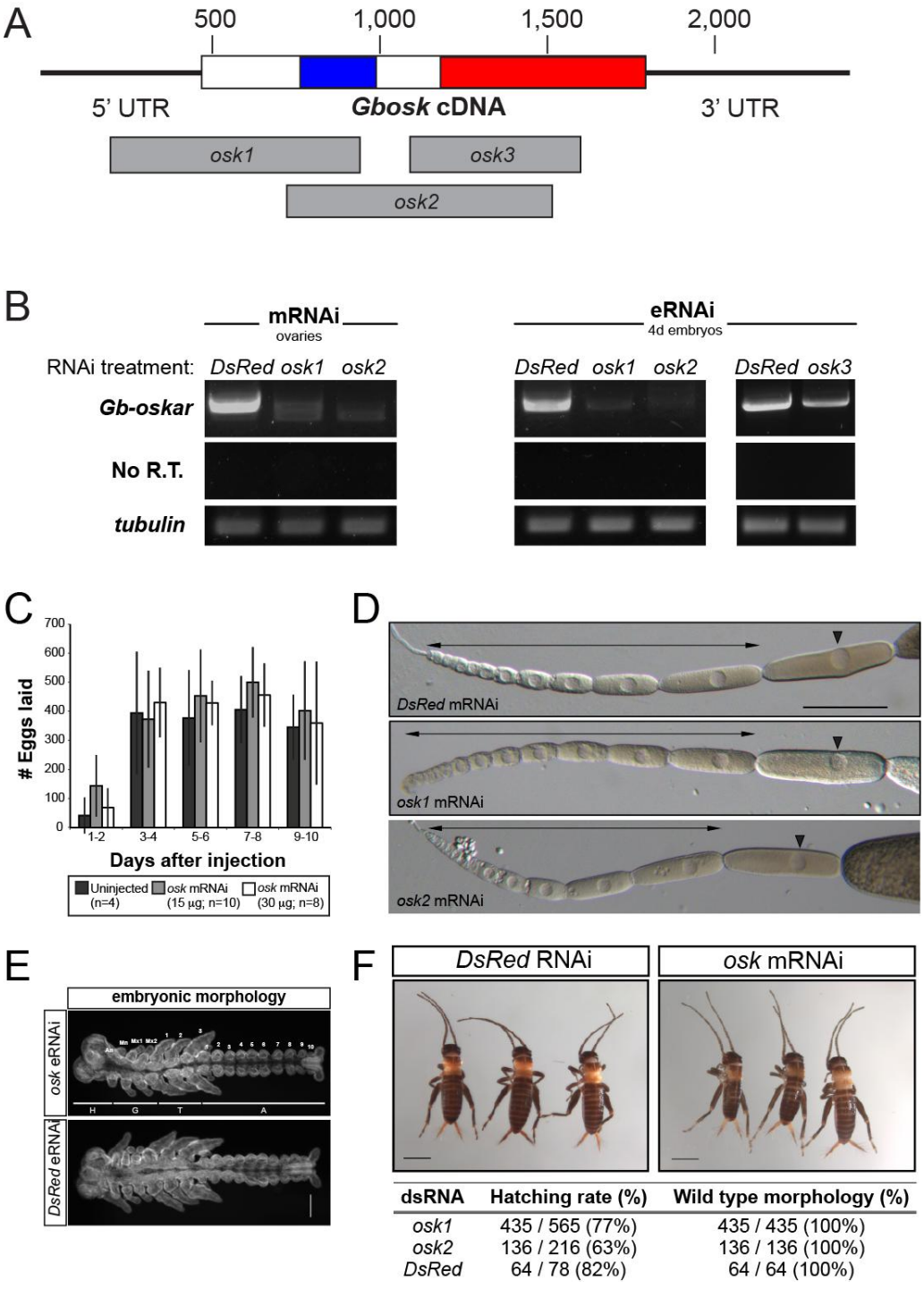

Long Osk
LOTUS
SGNH Hydrolase
peptide antigen
anti-Gb-Osk3498
peptide antigen
anti-Gb-Osk3501

Figure S1, related to Figure 1. ClustalW alignment of Gb-Oskar with known *oskar* orthologs. Gb-Oskar contains the LOTUS (*aka* OST-HTH) domain (blue box) and the SGNH hydrolase domain (red box) that are characteristic of Oskar orthologs. The Long Osk domain (black box) is only present in Oskar orthologs from Drosophilid species. Residues are colored using the RasMol color scheme to indicate physicochemical properties. Structurally conserved residues within the LOTUS domain are indicated with carets [1, 2]. Indicated with astrices are the positions of the serine, aspartate, and histidine residues (notably absent in Oskar) that constitute the catalytic triad in functional hydrolases [1]. The two peptide antigens used to generate Gb-Oskar antibodies are indicated in grey. Species (accession IDs): *Dm = Drosophila melanogaster* (AAF54306.1), *Dv = Drosophila virilis* (AAA28426.1), *Cq = Culex quinquefasciatus* (ACB20969.1)*; Nv = Nasonia vitripennis* (ADK94458.1), *Gb = Gryllus bimaculatus* (JQ434102).



Ewen-Campen et al. Supplemental Information Page 4 of 18

Figure S2, related to Figures 2, 3. Validation of rabbit αGb-Oskar, αGb-Vasa and αGb-Piwi antibodies. (A) Schematic of Gb-Oskar, showing the location of the LOTUS (blue) and SGNH hydrolase (red) domains, and the N- and C-terminal peptides used to generate αGb-Osk3498 and αGb-Osk3501 respectively (grey). **(B)**Western blot of *Gryllus* ovarian extract probed with αGb-Oskar antibodies. Bands are detected at the predicted weight of Gb-Oskar (black asterisk: 49.8 kDa). **(C)** Western blot of purified recombinant LOTUS and SGNH domains from Gb-Oskar probed with αGb-Oskar antibodies. αGb-Osk3498 specifically recognized the LOTUS domain (blue askterisk indicates expected size of 13.6 kDa) but not the SGNH hydrolase domain (red asterisk indicates expected size of 29.6 kDa). Conversely, αGb-Osk3501 specifically recognized the SGNH hydrolase domain but not the LOTUS domain. **(D)** Western blot of whole cell lysate of *E. coli* expressing full length Gb-Oskar. The predominant band is slightly larger than expected (black asterisk 49.8 Kda), as in (**A**). (-) uninduced *E. coli*, (+) induced *E. coli.***(E-E")** The expression of Gb-Oskar in neuroblasts (here shown in thoracic segments T1 and T2) is only seen using the final bleed serum (**E**), and is not present using the preimmune serum (**E'**). (**F**) Gb-Oskar immunostaining in brain (top row) and thoracic neuroblasts (bottom row) is greatly reduced following *Gb-oskar* eRNAi, validating the specificity of the antiserum and the efficacy of the RNAi knockdown. Sample sizes are given at the bottom of each panel. Anterior is to the left in (**E-E"**) and in (**F**) bottom row; anterior is up in (**F**) top row. Scale bars = 100 μM. (**G**) Western blots of unpurified αGbVas154 serum on *Gryllus* ovarian extracts (O) and purified protein (P). Serum used at 1:5,000 recognizes a band of the expected size (asterisk: 75kDa) and a smaller secondary band from ovarian extracts. Following affinity purification against recombinant Gb-Vasa protein used at 1:500, the antibody recognizes a single band of the expected size from ovarian extracts (O). (**H**) Western blots of unpurified αGbPiwi158 serum on *Gryllus* ovarian extracts (O) and purified protein (P). Serum used at 1:2,000 recognizes a band of the expected size (asterisk: 92kDa) and several smaller secondary bands from ovarian extracts. Following affinity purification agasint Protein-A, serum used at 1:500 shows reocgnition of most smaller secondary bands is abolished and the antibody strongly recognizes a band of the expected size as well as a slightly smaller secondary band from ovarian extracts (O). In situ hybridization for *Gb-vasa* (**I-K**) and antibody staining with αGbVas154 (**I'-K'**) shows expression of *Gb-vasa* transcript and protein in neuroblasts of the embryonic brain (**I, I'**), thorax (**J, J'**) and abdomen (**K, K'**) in Stage 9 embryos. *Gb-oskar* expression increases in the ectodermal neuroblast precursor territory just prior to neuroblast formation at Stage 5 (**L**) and in newly forming neuroblasts at Stage 6 (**M**). Scale bar = 100 μM.



dsRNA	Hatching rate (%)	Wild type morphology (%)
osk1	435 / 565 (77%)	435 / 435 (100%)
osk2	136 / 216 (63%)	136 / 136 (100%)
DsRed	64 / 78 (82%)	64 / 64 (100%)

Figure S3, related to Figures 2, 3. Validation and phenotypes of *Gb-oskar* RNAi. (**A**) Schematic of the three dsRNA fragments used for *Gb-oskar* RNAi experiments. Blue and red regions indicate locations of the LOTUS and SGNH hydrolase domains respectively. Grey boxes indicate fragments used for RNAi treatments. (**B**) Semi-quantitative PCR analysis of *Gb-oskar* levels in negative control (*DsRed*), *osk1* and *osk2* RNAi treatments. Maternal RNAi (mRNAi) greatly reduces *Gb-oskar* transcript levels in ovaries of injected females even 10 days after injection (left panel). Near-complete reduction of *Gb-oskar* transcript in embryos via embryonic RNAi (eRNAi) with fragments *osk1* and *osk2*, and a lesser reduction with fragment *osk3*, persists through until at least 4 days after egg laying (right panel). No R.T. = No reverse transcriptase control to ensure the absense of genomic DNA. *β-tubulin* serves as a loading and RNAi specificity control**. (C)** Average daily egg-laying rate is not significantly different between *Gb-oskar* mRNAi-injected females and controls, at two different concentrations of *Gb-oskar* dsRNA. Error bars indicate 95% confidence interval. *osk1* and *osk2* results are pooled for each concentration, as no significant differences were seen between these treatments. (**D**) *Gb-oskar* mRNAi does not disrupt oogenesis or ovarian morphology. Double-headed arrows indicate that oocytes at stages of oogenesis are present in RNAi treated ovaries as well as controls. Arrowhead indicates the posterior localization of the oocyte nucleus, which indicates normal oocyte patterning [3]. **(E)** Nuclear stain of *Gb-oskar* eRNAi embryos reveals wild type anterior-posterior patterning compared to controls. **(F)** *Gb-oskar* mRNAi hatchlings (right panel) do not display obvious morphological defects compared to controls (left panel). Hatching rate is not significantly different between *Gb-oskar* mRNAi embryos (right panel) and controls (left panel). Scale bars = 100 μM in D-E, 1 mm in F.

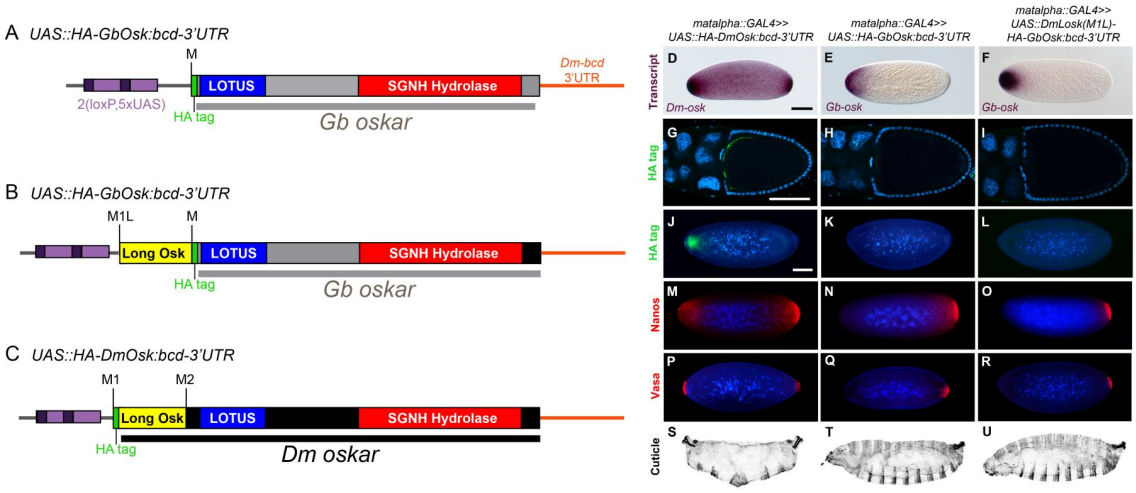

A UAS::HA-GbOsk:bcd-3'UTR
M
LOTUS
SGNH Hydrolase
Dm-bcd 3'UTR
2(loxP,5xUAS)
HA tag
Gb oskar
B UAS::HA-GbOsk:bcd-3'UTR
M1L
M
Long Osk
LOTUS
SGNH Hydrolase
HA tag
Gb oskar
C UAS::HA-DmOsk:bcd-3'UTR
M1
M2
Long Osk
LOTUS
SGNH Hydrolase
HA tag
Dm oskar
matalpha::GAL4>> UAS::HA-DmOsk:bcd-3'UTR
matalpha::GAL4>> UAS::HA-GbOsk:bcd-3'UTR
matalpha::GAL4>> UAS::DmLosk(M1L)-HA-GbOsk:bcd-3'UTR
Transcript
HA tag
HA tag
Nanos
Vasa
Cuticle
D
E
F
Dm-osk
Gb-osk
Gb-osk
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U

Figure S4, related to Figure 4. Schematic of constructs used for transgenic *D. melanogaster*, which show that *Gb-oskar* and *DmLosk(M1L)*:*GbOskar* transgenes are transcribed and localized but not translated in *Drosophila* ovaries or embryos. Full length cDNAs of *Gb-oskar* (**A**), *Gb-oskar* downstream of *Dm-Losk*(M1L) (**B**) or full length *Dm-osk* (**C**) were fused to an N-terminal HA tag fused to the *Dm-bicoid* 3'UTR sequence. Constructs were cloned into the pValium22 vector for site-directed insertion via the ΦC31 targeted transgenesis system . HA-tagged *Dm-osk* (D, G, J, M, P, S), *Gb-oskar* (E, H, K, N, Q, T), or DmLosk(M1L)-*Gb-osk* (F, I, L, O, R, U) fused to the *bicoid* 3'UTR were expressed in transgenic *D. melanogaster* with a maternal GAL4 driver (matalpha). In situ hybridization for *Dm-osk* (D) or *Gb-oskar* (E, F) shows that all transcripts are localized to the anterior of early embryos. Dm-Oskar but not Gb-Oskar protein is translated in oocytes (G, H, I) and early embryos (J, K, L). Ectopic Dm-Osk causes anterior localization of Nanos (M) and ectopic Vasa-positive pole cells (P) and a bicaudal phenotype revealed in larval cuticle preparations (S). Absence of detectable ectopic Gb-Oskar (H, I, K, L) is correlated with absence of ectopic Nanos (N, O), Vasa or pole cells (Q, R) and larval morphology is wild type (T, U). In all panels anterior is to the left, scale bar = 100 μm.

Table S1, related to Figure 2. Anterior-posterior axis phenotypes in *Gb-oskar* RNAi embryos.

RNAi method	dsRNA injected	# embryos injected[1]	# developed (%)	# surviving embryos with morphological abnormalities (%)
Maternal (mRNAi)	*osk1*	123	75 (61.0%)	0
	osk2	52	23 (44.2%)	0
	caudal	589	486 (82.5%)	321 (66.0%)
	DsRed	135	81 (60%)	0
Embryonic (eRNAi)	*osk1*	125	101 (80.8%)	0
	osk2	67	41 (61.2%)	0
	osk3	150	53 (35.3%)	0
	caudal	168	131 (78.0%)	110 (84.0%)
	DsRed	74	65 (87.8%)	0
	DsRed[2]	85	65 (76.5%)	0
	uninjected	90	64 (71.1%)	0

1. For maternal RNAi, this number refers to number of eggs dissected.

Table S2, related to Figure 4. Effect of the *Gb-oskar* transgenes in *D. melanogaster* using the bicoid 3' UTR.

Maternal genotype	bicaudal phenotype / total (%)	anterior Vasa-positive cells / total (%)
matalpha >> HA-Dm-osk:bcd3'UTR	63 / 76 (82.9%)	42 / 60 (70%)
matalpha >> HA-Gb-osk:bcd3'UTR	0 / 73 (0%)	0 / 54 (0%)
matalpha >> DmLosk(M1L):HA-Gb-osk:bcd3'UTR	0 / 91 (0%)	0 / 61 (0%)
matalpha >> DmOsk(M1L):bcd3'UTR	20 / 26 (76.9%)	14 / 28 (50%)[1]

1. Of the 14 embryos with ectopic germ cells, six (42.8%) did not form at the anterior pole, but instead formed slightly more posteriorly.

Supplemental Experimental Procedures

Gb-Oskar Identification and Phylogenetic Analysis

Gb-oskar was recovered from a de novo transcriptome generated using 454 Titanium sequencing. A 1,323 bp predicted full length *Gb-oskar* sequence was assembled from 695 raw reads (from a total of 4.2 million reads in the transcriptome), including additional 5' and 3' UTR sequence (462 bp and 607 bp, respectively). Our transcriptome assembly also contained a predicted alternate isoform of *Gb-oskar* that includes an additional 162 bp of sequence between the LOTUS and SGNH hydrolase domains. However, RT-PCR failed to amplify this larger product, and the additional sequence was not conserved in any other *oskar* orthologue. We therefore focused on the smaller of these two predicted isoforms, which we have confirmed is transcribed during *Gryllus* development by RT-PCR (Fig. 2J). Portions of *Gryllus piwi* and *tdrd7* orthologues were also found in our *Gryllus* transcriptome, and *Gryllus vasa* and *even-skipped* were cloned based on published sequence (AB378065, AB120736). All new sequence data have been submitted to NCBI (GenBank Accessions JQ434102-4).

In the course of performing reciprocal best BLAST hit analysis, we found that *Nasonia vitripennis oskar* (Nv-Osk, ADK94458.1), unlike *oskar* orthologues from *D. melanogaster* (AAF54306.1)*, Culex quinquefasciatus* (ACB20969.1)*, Aedes aegypti* (ABC41128.1)*, Anopheles gambiae* (ABC54566.1)*, and Messor pergandei* (ADM07366.1), retrieved a best hit against the *Gryllus* transcriptome sequence for *Gb-tdrd7* (e-value = 5e-11). However, this BLAST hit only extended over the LOTUS domain (amino acids 14-120), which is conserved in both *oskar* and *tdrd7* genes, while the second best hit, *Gb-oskar* (e-value = 2e-10), extended over both the LOTUS and SGNH hydrolase domain (amino acids 19-106 and 183-375 of the query). Further, *Gb-oskar* retrieves Nv-Osk as its top hit against the *N. vitripennis* proteome. We therefore concluded that across the length of the gene *Nv-oskar* is more similar to *Gb-oskar* than to *Gb-tdrd7*, a result that agrees with all other known *oskar* orthologues as well as our phylogenetic analysis (Fig. 1B).

Amino acid identity conservation (i.e. percentage of sequences containing an identical amino acid at a given alignment position) and physicochemical conservation were calculated using JalView [25].

To estimate a gene tree of *oskar* and *tdrd7* genes, *Gb-oskar* and *tdrd7* were aligned to their publically available orthologues using MUSCLE [26] (Fig. S1). Although *oskar* orthologues have been bioinformatically predicted in several recently sequenced Hymenopterans (retrievable at http://hymenopteragenome.org/ant_genomes/?q=blast), we only included those from *N. vitripennis* [NP_001234884.1] and *M. pergandei* [ADM07366.1] in our phylogenetic analysis, as these have been experimentally validated as true *oskar* orthologues [3]. Regions of uncertain alignment were removed using GBlocks [27] with the least stringent settings, which produced an alignment containing 294 amino acids. A gene tree for *oskar* and *tdrd7* was estimated under both Bayesian and maximum likelihood criteria. The "mixed" model of amino acid evolution was used in MrBayes v3.1.2 [28] to choose the optimal model for amino acid evolution for both this and the maximum likelihood tree. MrBayes selected the WAG model [29] with a probability of 1.0. Two runs of four independent MCMC chains each were executed for 2 million generations, sampling trees every 1,000 generations and the first 250,000 generations discarded as burn-in. The average standard deviation of split frequencies between the two runs fell below 0.01 after 154,000 generations, indicating that the two chains had converged. Maximum likelihood analysis

was estimated using RAxML-MPI v7.2.8. 2000 runs from independent starting trees were executed under the WAG model of protein evolution with gamma distribution of rate heterogeneity to simultaneously estimate the best scoring tree and perform rapid bootstrap analysis. All tree estimations were conducted on the Odyssey Cluster (Harvard University), supported by the FAS Sciences Division Research Computing group.

Gb-Oskar Antibody Generation

Two rabbit polyclonal antibodies were raised against an N-terminal peptide DLEGEYYELVGQAIPFRTFGC (αGb-Osk3498) and a C-terminal peptide GRSYLYNFLLQQLTELENTH (αGb-Osk3501) from Gb-Oskar (Figs. S1, S2A) (Open Biosystems, Inc., Alabama). Rabbits were boosted on days 14, 28, 42 and 82 after initial injection, and the final bleed was performed on day 138. Specificity was tested using Western blot analysis and immunostaining (Fig. S2B-F).

To prepare *Gryllus* ovaries for Western blotting, ~30-50 ovarioles were homogenized in 300 µL of 5X SDS loading buffer, boiled for 5 minutes at 95-100ºC, and stored at -20ºC until use. 1:10 and 1:100 dilutions of this homogenate, run in adjacent wells, were used in Western blots. For purified Gb-Oskar domains, 10 µg of purified protein were run. For whole-cell lysates containing full-length bacterially expressed Gb-Oskar, cells were grown (either with or without induction) overnight at 20ºC, and OD600 readings were used to calculate equal volumes of induced versus uninduced cells.

For Western blotting, samples were separated on a 12% acrylamide SDS-PAGE gel in running buffer (25 mM Tris base, 192 mM glycine, 0.1% w/v SDS) at ~180V for approximately one hour and then transferred to a nitrocellulose membrane (0.2 µm pore size, BIO-RAD cat. # 162-0112) at ~300V for 90-120 minutes at 4º C in transfer buffer (25 mM Tris base, 192 mM glycine, 0.1% w/v SDS, 20% v/v methanol). Ponceau S staining was used to verify that equal volumes were loaded in each lane and had transferred properly. Blots were blocked in 5% BSA + 5% milk powder for at least 30 minutes, and then incubated overnight with primary antibodies (1:500 dilution). Blots were washed at least 5 x 5 minutes in TBST (20 mM Tris pH 7.5, 500 mM NaCl, 0.05% (v/v) Tween-20) and then incubated with donkey anti-rabbit HRP-coupled secondary antibody (Jackson Laboratories, 1:5,000 dilution) in block for two hours. After 5 x 5 minute washes in TBST, the secondary antibody was detected with SuperSignal West Pico Chemiluminescent substrate (ThermoScientific) following manufacturer's protocols. Signal was visualized using Kodak BioMax film, and exposures were tested at 10 seconds, 1 minutes, 5 minutes, and 10 minutes to optimize exposure.

Both antibodies recognized proteins of the predicted size (asterisk in Fig. S2B) from *Gryllus* ovary extracts, as well as additional species of larger molecular weights (arrowheads in Fig. S2B), suggesting possible post-translational modification as has been reported for *Drosophila* Oskar [30]. Western blot analysis showed that as predicted, the αGb-Osk3501 serum recognized the C-terminal SGNH hydrolase domain and not the N-terminal LOTUS domain; conversely, the αGb-Osk3498 serum recognized the LOTUS domain and not the SGNH hydrolase domain (Fig. S2C). Both antibodies also recognized bacterially expressed full length Gb-Oskar at a slightly higher MW than predicted (Fig. S2D). The staining pattern seen in whole mount immunostained *Gryllus* embryos was not detected in pre-immune serum (Fig. S2E', E"), and was abolished by *oskar* RNAi (Fig. S2F).

Gb-Vasa and Gb-Piwi Antibody Generation

Rabbit polyclonal antibodies were raised against recombinant *Gryllus* Vasa and Piwi protein fragments (McGill Biology CIAN facility, Canada). Full-length *Gryllus vasa* (amino acids 2-649) and a fragment of *Gryllus piwi* (774 C-terminal amino acids ending at the stop codon) were each cloned into the pET151/dTOPO expression vector (Invitrogen Cat No. K151-01), thus introducing an N-terminal 6X-polyhistidine tag. Protein expression was induced in *E. coli* BL21 (DE3) by addition of IPTG to a final concentration of 0.25 mM and incubated at 30°C for 4 hrs. The overexpressed proteins were highly insoluble, so the cell pellets were disrupted by sonication in non-denaturing buffer, and soluble fraction were separated by centrifugation and discarded. The pellet was dissolved in 8 M urea (crude prep). Crude preps were further purified by SDS gel-purification. Protein bands were excised from the gel after reverse staining with zinc chloride and the proteins were collected by electroelution. Acetone precipitation was performed and the precipitated protein was washed and redissolved in 8M urea. Protein concentration was adjusted to 1-2 mg/ml for injection. Rabbits were injected using a standard 80-day protocol with three boosts after the initial injection; final bleed was performed after 87 days. The serum was processed by addition of NaN3 to 0.02% (w/v). Both antibodies were then affinity purified: αGb-Vasa was purified against bacterially expressed Gb-Vasa protein, and αGb-Piwi was purified with Protein A (Primm Biotech, Cambridge MA).

The CIAN facility determined specificity prior to purification using Western blot analysis (A-B, left blots), and we repeated this analysis following affinity purification using the above protocol (Fig. S2G-H, right blots).

In situ Hybridization and Antibody Staining

In situ hybridizations and antibody stainings were performed as described in [4]. Four different DIG-labeled fragments of *Gb-oskar* were used as in situ probes (ranging from 742 to 2,103 bp in size), all of which gave consistent staining results. *Gb-piwi* and *Gb-vasa* in situ probes were 781 bp and 1,953 bp, respectively. Probes were used at between 0.5-2 μg/μl during hybridization. For immunostainings, all species-specific primary antibodies used were used at a final dilution of 1:300. αGb-Oskar was preabsorbed against Oregon R *Drosophila* mixed stage embryos for 90 minutes at room temperature prior to staining to reduce background. Secondary antibodies used were goat anti-rabbit or goat anti-mouse coupled to Alexa 555, Alexa 488 or Alexa 647 (Invitrogen) at 1:1,000. Nuclei were counterstained with Hoechst 33342 at 1:5000 of a 10mg/ml stock solution. Cy3 conjugated anti-HRP (gift of Sam Kunes, Harvard University, MA, USA) was used at 1:50.

Injections for RNA Interference (RNAi)

Maternal and embryonic injections of dsRNA were carried out as described in [4]. To synchronize adult female age for maternal injections, final nymphal stage females were isolated and monitored daily for final molt. All females were injected with 15 μg of dsRNA on the third day following the final molt. Because we observed that egg laying could be reduced by the continued presence of two males, each female was housed separately with two males only for days one, two and five of a 10 day experiment.

Three different fragments of *Gb-oskar* were used as template for dsRNA synthesis (Fig. S3A): (1) *osk1*: a 742bp fragment starting 262 bp upstream of the first methionine and ending at position 480; (2) *osk2*: a 789 bp fragment from position 255-1044, and (3) *osk3*: a 503 bp fragment from position 638-1141. Results from these three fragments were consistent for all

phenotypic analyses. DsRNA was adjusted to a final concentration of 3 μg/μl (6.1 and 5.8 μM respectively) for fragments *osk1* and *osk2*, and 5 μg/μl for fragment *osk3* (15.1 μM). The *DsRed* negative control and *caudal* positive control dsRNAs have been described in [4] and were both used at the same final concentration as the corresponding *oskar* fragments used in that experiment (either 3 or 5 μg/μl).

Validating RNAi Knockdown

RNAi knockdown efficiency was estimated using both semi-quantitative RT-PCR and whole-mount antibody staining. For semi-quantitative RT-PCR (Fig. S3B), ovaries from maternal RNAi-injected females were dissected 10 days after maternal dsRNA injection, and stage 7-10 embryos laid by maternal RNAi-injected females, or resulting from zygotic RNAi injections, were dissected from 4-5 days after egg laying (AEL). All tissues were collected in Trizol (Invitrogen), and total RNA was isolated following manufacturer's protocols, followed by a 30 minute DNase treatment (Ambion) at 37°C to remove genomic DNA. Equal quantities of total RNA were used as template for first strand cDNA synthesis using SuperScript III (Invitrogen). *Gb-oskar* levels were estimated following 35 PCR cycles (98ºC for 3 minutes, 35 cycles of 98ºC for 30 seconds, 55ºC for 30 seconds, 72º C for 2 minutes 15 seconds, followed by a 10 minute final extension at 72ºC) and electrophoresis on a 1.0% agarose gel, compared to *Gb-beta-tubulin* to ensure equivalent amounts of template. Negative controls without reverse transcriptase were run in parallel and revealed no genomic DNA contamination. The *Gb-oskar* primers used (F: 5' TGGTAGTTCGAAGGGAACTTG-3'; R: 5'-CATCTTCCATTTGCCACAGA-3') amplify a band of 2,149 bp. The *Gb-beta-tubulin* primers used (F: 5'-TCAGACACCGTCGTTGAACC-3'; R: 5'-GATGGTTCAGGTCGCCGTAG-3') amplify a band of 157 bp.

Knockdown appeared to be less efficient using fragment *osk3* (Fig. S3B), but this fragment still produced CNS phenotypes consistent with those produced using fragment *osk1*.

RNAi efficiency was independently assayed by staining *osk* eRNAi and *DsRed* eRNAi embryos with the αGb-Oskar3501 antibody and imaging the embryos under identical conditions (Fig. S2F).

Scoring RNAi Phenotypes

To quantify egg laying, each injected female was individually housed with a dish of moist sand. Because we observed that *Gryllus* females typically only oviposit substantial numbers of eggs on alternate days, egg dishes were collected every 48 hours. Eggs were separated from sand using a 500 μm mesh under running tap water, and spread out in a monolayer in a 15cm petri dish. An image of the eggs was captured on a flatbed scanner and eggs were manually tallied using the ImageJ Cell Counter plug-in (Fig. S3C).

Ovarian morphology was assayed on ovaries of injected females 10 days after dsRNA injection. Wild type A-P patterning was assessed by the asymmetric localization of the oocyte nucleus in mid-stage oocytes [3]. We also tested for oogenesis phenotypes by looking for the presence of oocytes at all stages of development (Fig. S3D).

Embryonic morphology was visualized at stages 7-10 using both Nomarski optics and fluorescent nuclear staining (Fig. S3E). The presence of embryonic germ cells was assayed using in situ hybridization against *piwi* (not shown), as well as antibody staining for both Gb-Vasa and Gb-Piwi (Fig. 2G-H). The presence of functional ovarioles was also assayed in *osk* eRNAi embryos that were grown to adulthood.

Axonal scaffolds were visualized with anti-HRP immunostaining followed by confocal microscopy of the entire embryonic nervous system. Before scoring axonal defects, stage-matching of embryos was performed by comparison of morphogenetic progress in the thoracic appendages, antennae, abdominal segmentation and terminal cerci, rather than by developmental progression of the nervous system. Only embryos at stages 8.5-10 were used for this analysis, since formation of the global CNS axonal scaffold is complete by this time. Axonal patterning of *Gb-oskar* eRNAi embryos, *DsRed* eRNAi embryos, and uninjected embryos at comparable developmental stages were scored blind (embryos were grouped by matched developmental stages, but whether they were wild type, *DsRed* eRNAi or *Gb-oskar* eRNAi embryos was hidden) for consistent and symmetrical thickness and continuity of longitudinal connectives, timing/morphology of anterior commissure formation relative to central neuron mitoses, and timing/morphology of posterior commissure formation relative to anterior commissure formation. Two independent blind scoring rounds were performed and the results were consistent. After scoring was completed, the nature of the embryonic treatment was revealed, and data from embryos of the same treatment (uninjected, *DsRed* eRNAi or *Gb-oskar* eRNAi) were grouped for statistical analysis (Chi-squared test).

To test for defects in aCC/pCC neurons, we performed in situ hybridization on stage 9 embryos using a 528 bp probe against *Gryllus even-skipped.* Samples were scored blind for the absence of any of the four aCC/pCC neurons in each segment. Two independent knockdowns using two dsRNA fragments (fragments *osk1* and *osk3* described above) gave comparable results (5/11 and 5/18 embryos contained defective aCC/pCC neurons in the two experiments respectively). To reliably distinguish between aCC/pCC neurons and the nearby U/CQ neurons, we used several approaches. First, we noted that the U/CQ neurons are clearly distinguishable from aCC/pCC by their position in the dorsal-ventral axis and their relationship to the axonal scaffold. Specifically, aCC/pCC are the dorsal-most *eve*-positive cells while U/CQ are more ventral [see e.g. 3; their Figure 4a-h]. Second, aCC/pCC are reliably located in the corner where the longitudinal connectives meet the posterior commissure [see e.g. 4; their Figure 2d], and in the same focal plane. Third, in our analysis we examined optical sections through the entire dorso-ventral axis of every affected segment to ensure that apparently missing aCC/pCC cells were not visible in any other focal plane, and that U/CQ could be identified in more ventral focal planes.

Transgenic Constructs

To target HA-tagged Dm-Oskar or Gb-Oskar to the *D. melanogaster* oocyte anterior, we cloned Gb-Oskar, Dm-Losk(M1L):Gb-Oskar, or Dm-Osk-bcd3'UTR [5] into pValium22 [6], which contains attB sites for PhiC31-based targeted integration and is optimized for expression in the female germ line. We used a previously described bcd3'UTR that has a putative NRE replaced by a 10bp linker to prevent potential transcript clearing by *nanos* [5, 7]. Cloning was performed using the circular polymerase extension cloning method [8] with primers containing an N-terminal HA-tag (pValium22 was amplified with primers F: 5'-AGAACTAGAGCCGCG-3', R: 5'GCCCGAGCTTAAGACT-3'; Dm-Osk-bcd3'UTR was amplified with primers F: 5'-GCCAGTCTTAAGCTCGGGCATG
TACCCATACGATGTTCCGGATTACGCTGCCGCAGTCACAAGT-3' and R: 5'-CAATTCCGCGGCTCTAGTTCTGGATCCACCCGAGTA-3'; Gb-Oskar was amplified with primers F: 5'-GCCAGTCTTAAGCTCGGGCATG
TACCCATACGATGTTCCGGATTACGCTAGTTGGACTGAGGTT-3' and R: 5'-

CCTCTCATCCAGGCTCGAGCGCCGGCGTCAATGTGTGTTTTC-3'; to amplify the bcd3'UTR for fusion with Gb-Oskar, primers F: 5'-CTCGAGCCTGGATGA-3' and R: 5'-CAATTCCGCGGCTCTAGTTCTGGATCCACCCGAGTA-3' were used). The transgenes for GbOsk:bcd3'UTR and DmOsk:bcd3'UTR were inserted into the attp40 site on 2L, and the DmLosk(M1L):GbOsk:bcd3'UTR transgene was inserted into the attp2 site on 3L [9] both of which show high levels of specific expression. Injection to create transgenic flies was carried out by Genetic Services, Inc. (Cambridge, MA). Transgenic progeny were identified and maintained using standard genetic crosses.

Drosophila Stocks

The maternal driver line *w**; *P{w*$^{+mC}$*=matalpha4-GAL-VP16}V2H* was obtained from the Bloomington *Drosophila* Stock Center (stock number 7062). *y v; Sco/CyO* and *y sc v ; Dr e/TM3,Sb* were obtained from the Perrimon lab (Harvard University).

Imaging and Image Analysis

Images were captured with AxioVision v.4.8 driving a Zeiss Stereo Lumar equipped with an AxioCam MRc camera, or a Zeiss Axio Imager equipped with an AxioCam MRm camera, using epifluorescence either with or without an Apotome. Confocal microscopy was performed with a Zeiss LSM 710 confocal, using comparable gain, offset, and averaging parameters for all samples. Image analyses were performed with AxioVision v.4.8, Zen 2009 (Zeiss), and figures were assembled in Photoshop CS4 or Illustrator CS4 (Adobe). For the confocal images shown in Fig. 2 (G-N), a maximum-intensity projection of the antibody staining was superimposed over a single z-plane of the nuclear counterstain for visual clarity.

Author Contributions: BE-C and CGE designed research, analysed data and wrote the paper; experiments were carried out by JRS (recombinant Gb-Oskar purification), EES (cloning *Gb-vasa* and *Gb-piwi*), CGE (axonal scaffold phenotypic analysis) and BE-C (all other experiments); CGE obtained funding for the research.

Supplemental References

1. Anantharaman, V., Zhang, D., and Aravind, L. (2010). OST-HTH: a novel predicted RNA-binding domain. Biology Direct *5*, 13.

2. Callebaut, I., and Mornon, J.-P. (2010). LOTUS, a new domain associated with small RNA pathways in the germline. Bioinformatics *26*, 1140-1144.

3. Lynch, J.A., Peel, A.D., Drechsler, A., Averof, M., and Roth, S. (2010). EGF signaling and the origin of axial polarity among the insects. Curr. Biol. *20*, 1042-1047.

4. Kainz, F., Ewen-Campen, B., Akam, M., and Extavour, C.G. (2011). Delta/Notch signalling is not required for segment generation in the basally branching insect *Gryllus bimaculatus*. Development *138*, 5015-5026.

5. Tanaka, T., and Nakamura, A. (2008). The endocytic pathway acts downstream of Oskar in Drosophila germ plasm assembly. Development (Cambridge, England) *135*, 1107-1117.

6. Ni, J.-Q., Zhou, R., Czech, B., Liu, L.-P., Holderbaum, L., Yang-Zhou, D., Shim, H.-S., Tao, R., Handler, D., Karpowicz, P., et al. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nature Chemical Biology *8*, 405-407.

7. Ephrussi, A., and Lehmann, R. (1992). Induction of germ cell formation by *oskar*. Nature *358*, 387-392.

8. Quan, J., and Tian, J. (2011). Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nature protocols *6*, 242-251.

9. Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E., and Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat. Genet. *40*, 476-483.

Current Biology 23, 835–842, May 20, 2013 ©2013 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2013.03.063

CHAPTER 3

Article

Germ Cell Specification Requires Zygotic Mechanisms Rather Than Germ Plasm in a Basally Branching Insect

Ben Ewen-Campen,[1] Seth Donoughe,[1] Donald Nat Clarke,[1,2] and Cassandra G. Extavour[1,*]
[1]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Summary

Background: Primordial germ cell (PGC) specification is a universal process across animals, but the molecular mechanisms specifying PGCs are remarkably diverse. In *Drosophila*, PGCs are specified by maternally provided, asymmetrically localized cytoplasmic factors (germ plasm). In contrast, historical literature on most other arthropods reports that PGCs arise from mesoderm during midembryogenesis, suggesting that an arthropod last common ancestor may have specified PGCs via zygotic mechanisms. However, there has been no direct experimental evidence to date for germ plasm-independent arthropod PGC specification.
Results: Here we show that in a basally branching insect, the cricket *Gryllus bimaculatus*, conserved germ plasm molecules are ubiquitously, rather than asymmetrically, localized during oogenesis and early embryogenesis. Molecular and cytological analyses suggest that *Gryllus* PGCs arise from abdominal mesoderm during segmentation, and *twist* RNAi embryos that lack mesoderm fail to form PGCs. Using RNA interference we show that *vasa* and *piwi* are not required maternally or zygotically for PGC formation but rather are required for primary spermatogonial divisions in adult males.
Conclusions: These observations suggest that *Gryllus* lacks a maternally inherited germ plasm, in contrast with many holometabolous insects, including *Drosophila*. The mesodermal origin of *Gryllus* PGCs and absence of instructive roles for *vasa* and *piwi* in PGC formation are reminiscent of mouse PGC specification and suggest that zygotic cell signaling may direct PGC specification in *Gryllus* and other Hemimetabola.

Introduction

Of the many specialized cell types that comprise an animal's body, only one is capable of contributing genetic information to the next generation: the germ cells. The restriction of reproductive potential to a small subset of cells is a universal process across sexually reproducing animals and represents a profound evolutionary novelty likely required for the evolution of multicellularity [1]. The molecular mechanisms that specify these cells, however, are remarkably diverse between taxa [2–5] and only well understood in a handful of model organisms.

Primordial germ cell (PGC) specification mechanisms have been categorized into two modes: cytoplasmic inheritance and zygotic induction [3, 4, 6]. Cytoplasmic inheritance (e.g., in *Drosophila melanogaster*) involves the localization of maternal mRNAs and proteins (germ plasm) to a subcellular region of the oocyte. Germ plasm is necessary and sufficient to induce PGC fate. In zygotic induction (e.g., in *Mus musculus*), by contrast, there is no germ plasm, and PGCs instead form in response to inductive signals from neighboring somatic cells [7].

Within insects, cytoplasmic inheritance appears to be a derived character confined primarily to the holometabolous insects [8] (Figure 1A; see also Table S1 available online), where germ plasm has been demonstrated experimentally in many species (Table S1). Histological studies of insects branching basally to Holometabola (the Hemimetabola), in contrast, have reported the absence of both germ plasm and pole cells in nearly all of these taxa [3, 6] (Figure 1A; Table S1). Studies of molecular markers for PGCs in hemimetabolous insects have been limited to the highly atypical parthenogenetic embryos of the pea aphid *Acyrthosiphon pisum*, a milkweed bug, and several orthopteran species (Table S1), yet there is no conserved pattern of PGC origin across these taxa.

In this study, we use multiple conserved molecular markers and RNAi to characterize PGC formation in the cricket *Gryllus bimaculatus* (Orthoptera), a hemimetabolous model species for studying the development of basally branching insects [9]. We provide several lines of evidence that *Gryllus* PGCs form from the abdominal mesoderm via inductive signaling and discuss the implications of these results for the evolution of germ plasm and the possibility of an ancient relationship between bilaterian PGCs and mesoderm.

Results

Gryllus Germ Cells Express a Suite of Conserved Genes

Within the Orthoptera, neither germ plasm nor pole cells have been reported (Figure 1A; Table S1). Histological examinations of orthopteran embryos conducted by William Wheeler over a century ago [10] suggested that PGCs arise from or among abdominal mesoderm cells during abdominal segmentation (Figure 1C), consistent with reports of germline origin both in other Hemimetabola and in most arthropods [3, 6]. However, conserved molecular markers can reveal a cryptic germ plasm that eludes histological examinations [11–13]. We therefore examined the expression of several conserved molecular PGC markers (*vasa* [14], *piwi* [14], *tudor*, *boule*, and *germ cell-less*) and three additional PIWI family genes (Figures S1A and S1B) in *Gryllus* ovaries and embryos. Because some germ plasm components localize as proteins rather than transcripts (see for example [15, 16, 17]), we also examined the expression of Vasa and Piwi proteins [14].

In fully segmented (stage 9) *Gryllus* embryos (Figures 1B and 1C), we identified cells matching Wheeler's description [10] that express both mRNA and protein of *piwi* and *vasa*, as well as *bol* and *gcl* transcripts (Figures 1D, 1E, and S1C). These cell clusters were found in abdominal segments A2–A3 in all embryos, and in A4–A5 in 45% of embryos (Figures 1D and 1E, arrowheads). Clusters were located on the dorsal medial face of mesodermal structures termed "coelomic pouches," which are present in every gnathal, thoracic, and abdominal

[2]Present address: Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950, USA
*Correspondence: extavour@oeb.harvard.edu

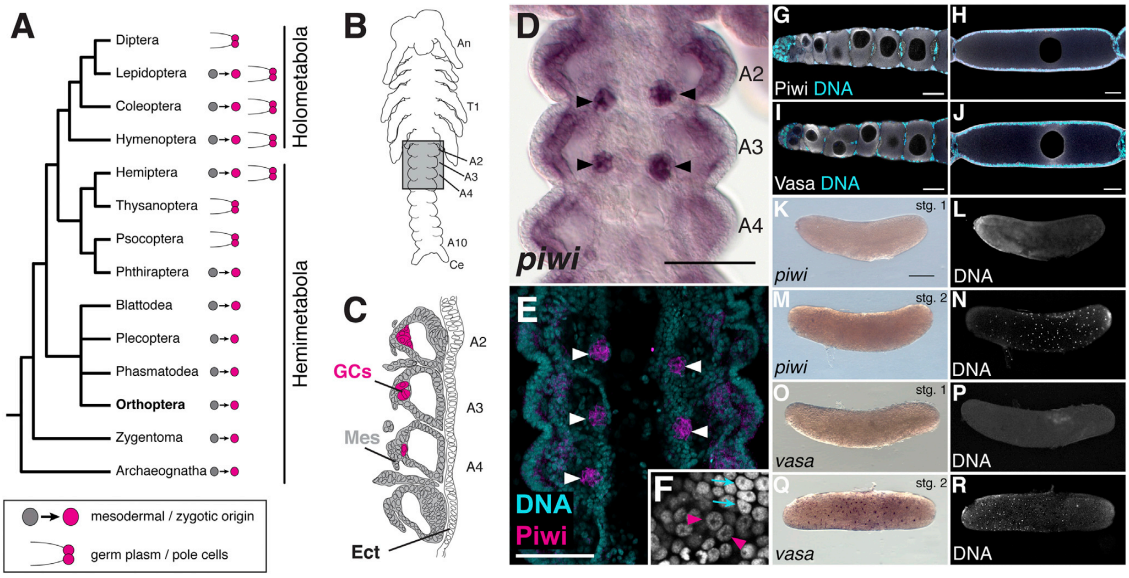


Figure 1. Molecular Markers Suggest Absence of Germ Plasm in *Gryllus*

(A) Phylogenetic distribution of reported PGC specification mechanisms across insects (see Table S1).
(B) Schematic of a stage 9 *Gryllus* embryo, highlighting the region enlarged in (D)–(E) (gray box). A2–A4, abdominal segments 2–4.
(C) Tracing of Wheeler's description of orthopteran PGCs at the earliest time point they could be identified [10]. GCs, germ cells (magenta); Mes, mesoderm (gray); Ect, ectoderm (white).
(D and E) *Gryllus* PGCs (arrowheads) express *piwi* transcripts (D) and protein (E).
(F) PGCs (arrowheads) display nuclear morphology distinct from somatic cells (arrows).
(G–J) Piwi (G and H) and Vasa (I and J) proteins do not localize asymmetrically in the ooplasm.
(K and M) *piwi* transcripts are undetectable during stages 1–2.
(O and Q) *vasa* transcripts are undetectable at stage 1 (O) and associated with all energid nuclei at stage 2 (Q).
(L, N, P, and R) Corresponding nuclear stains.
Scale bar represents 100 μm in (D) and (E); 50 μm in (G)–(J); 200 μm in (K)–(R). Anterior is up in (B)–(F), left in (G)–(R). See also Figures S1, S2, and Table S1.

segment. These cells possessed universal PGC characteristics [3] of large nuclei with diffuse chromatin and a single large nucleolus (Figure 1F). Based on these gene expression, nuclear morphology, and embryonic location data, we conclude that these cells are *Gryllus* PGCs. We also examined the expression of four additional putative candidate PGC marker genes (*tudor*, *piwi-2*, *AGO3-A*, and *AGO3-B*) but found that they were not specific PGC markers in *Gryllus* embryos (Figures S1B and S1F).

Gryllus Germline Markers Do Not Localize within Oocytes or Reveal PGCs in Early Embryos

We next examined the expression of *Gryllus* PGC markers during earlier stages of embryogenesis and oogenesis to test whether they revealed the presence of germ plasm in oocytes or PGCs in early embryos. All genes tested were consistently ubiquitous throughout oogenesis and never localized asymmetrically within the ooplasm (Figures 1G–1J, S1D, and S1F), although Vasa and Piwi proteins were enriched around the oocyte nucleus (Figures 1G–1J). In blastoderm-stage embryos (stages 1–3) and early germband-stage embryos (stage 4), *piwi* (Figures 1K–1N, S1E, and S2G–S2P′), *vasa* (Figures 1O–1R, S1E, and S2Q–S2Z′), *bol*, and *gcl* (Figures S2A–S2F) were expressed ubiquitously at low levels and showed no asymmetric localization within the embryo. These results are in stark contrast to the posterior accumulation of PGC determinants in *Drosophila* oocytes and early embryos [16, 18–20] and suggest an absence of germ-plasm-driven PGC specification in *Gryllus*.

Gryllus PGCs Arise De Novo during Midembryogenesis

To determine the embryonic origin of *Gryllus* PGCs, we examined the expression of *piwi* and *vasa* transcripts and proteins throughout abdominal elongation and segmentation. During early germband stages (stage 4), we detected low-level ubiquitous expression of both genes in all ectodermal and mesodermal cells (Figures 2A–2B′ and S3A–S3B′). It was not until thoracic limb bud enlargement began (stage 5) that *piwi* transcripts were detected at higher levels in two subsets of cells in abdominal segments A2–A4 among the lateral abdominal mesoderm (Figures 2C and 2C′). As appendage elongation began (stage 6), *piwi*-positive cells split into distinct groups along the anterior-posterior axis (Figures 2E and E′), and Piwi protein levels rose in these cells (Figures 2E″ and 2E‴). During morphological segmentation of the abdomen (stages 7–9) these cell groups coalesced into four to six distinct clusters adjacent and dorsal to the coelomic pouches in segments A2–A4 and continued to express high levels of *piwi* transcripts and protein (Figures 2F–2H‴ and S3D–S3F‴). *vasa* transcript and protein expression was similar to that of *piwi*, but *vasa* became enriched in PGCs slightly later than *piwi* and showed higher expression levels in the soma (Figure S3).

Interestingly, hallmarks of active transcription were observed in PGCs throughout all stages examined (Figures

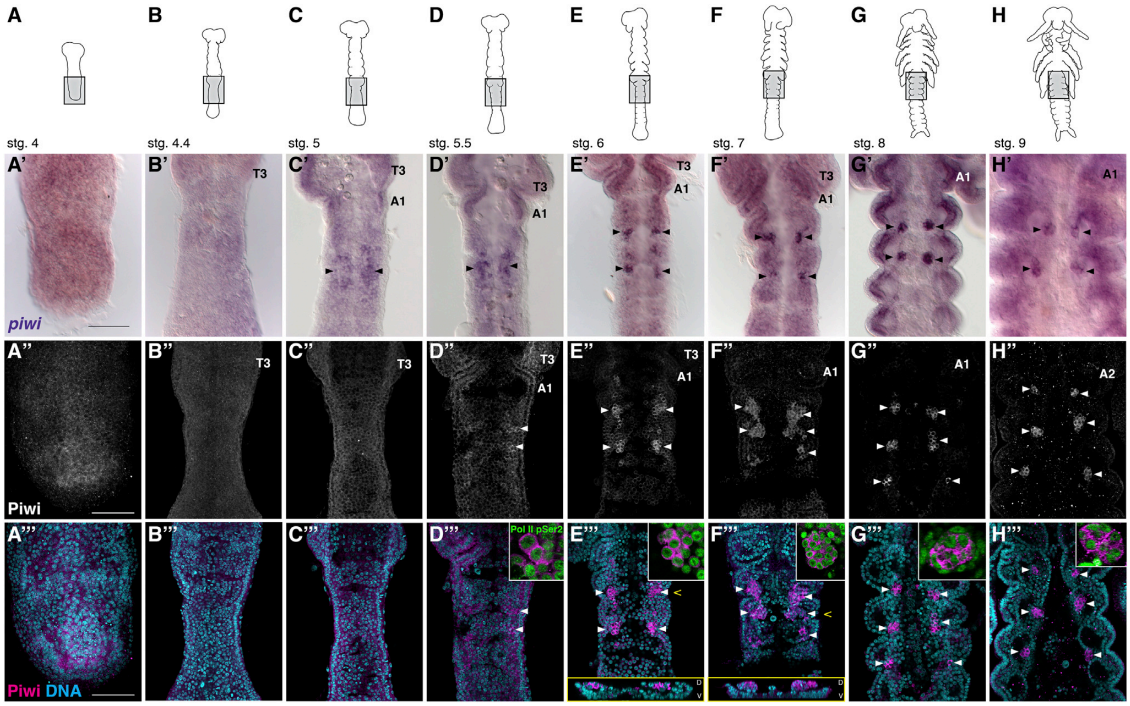


Figure 2. *Gryllus* PGCs Arise during Early Segmentation Stages

Drawings of *Gryllus* embryogenesis (A–H), highlighting the region depicted below (gray boxes). *piwi* transcripts are ubiquitous in stages 4 and 4.4 embryos (A′ and B′), but beginning at stage 5 (C′), two bilateral groups of *piwi*-positive cells arise on the dorsal surface of the embryo, then resolve into clusters during later stages in abdominal segments A2–A4 (D′–H′). Piwi protein shows a similar expression pattern to that of *piwi* transcript but is enriched slightly later in development (A″–H″). Piwi (magenta) overlaid on nuclear stain (cyan) reveals that PGCs arise prior to coelomic pouch formation and ultimately reside medial to these mesodermal structures (A‴–H‴). Yellow-framed insets in (E‴) and (F‴) show orthogonal projections at the position of the caret, illustrating the dorsal location of PGCs. Insets in (D‴)–(H‴) show expression of RNA polymerase II pSer2 (green) in PGCs. T3, thoracic segment 3; A1 and A2, abdominal segments 1 and 2.

Scale bar represents 100 μm. Anterior is up in all panels. See also Figures S2 and S3.

2D‴–2H‴, insets). This is consistent with *Gryllus* PGC formation via active transcriptional response to inductive signaling between cells rather than PGCs being a transcriptionally quiescent subpopulation of early-segregated cells as seen in *Drosophila* and other species with germ plasm [21].

Consistent with a conversion of presumptive mesoderm cells to PGCs beginning at stage 5, the nuclear morphology of mesodermal cells correlated with the relative levels of Piwi expression throughout development. At stage 4, all mesoderm cells had uniform Piwi expression and nuclear morphology, relatively compact chromatin, and multiple nucleoli (Figures 3A and 3A′). As Piwi expression increased in presumptive PGCs, their nuclei became larger with increasingly diffuse chromatin, whereas nuclei of neighboring Piwi-poor cells decreased in size, and their chromatin became compact as they progressed through mesoderm differentiation (Figures 3B–3D′). By stages 8–9, PGCs were clearly distinguished by their high nuclear-cytoplasmic ratio, diffuse chromatin, and single large nucleolus (Figures 3E–3F′), criteria used to identify PGCs in historical studies of Orthoptera and other animals [3]. Following stage 10, PGC clusters merged via short-range cell migration (Figure 3G) and coalesced into two bilateral gonad primordia (Figure 3H) located in segments A3–A4. Thus, *Gryllus* PGCs do not undergo long-range migration, as they do in many other species including *Drosophila* [22], but rather arise near the location of the embryonic gonad.

Knockdown of *Gryllus piwi* or *vasa* Does Not Disrupt PGC Formation or Maintenance

We knocked down *vasa* and *piwi* function using both maternal and zygotic RNAi (mRNAi and eRNAi, respectively) and confirmed knockdown using qPCR and immunostaining (Figures 4A, 4B, 4E, and 4H). In contrast to *Drosophila*, in which *vasa* and *piwi* are required maternally for embryonic PGC formation [15, 18], mRNAi against *vasa* and *piwi* did not disrupt PGC formation in *Gryllus* embryos (Figures 4C–4H), and there was no significant difference in the number of PGCs in either *vasa* or *piwi* mRNAi or eRNAi embryos relative to controls (Figures 4I and 4J). Furthermore, female embryos laid by mothers injected with *vasa* or *piwi* double-stranded (ds)RNA ultimately grew into fertile adults with fully functioning ovaries (Figure 4K–4M). In contrast to the *Drosophila* requirement for *vasa* and *piwi* in oogenesis and axial patterning [15, 23], *Gryllus* females injected with *vasa* or *piwi* dsRNA displayed no defects in egg laying, oogenesis, or axial patterning (Figures S4A–S4C). Moreover, double knockdown of *vasa* + *piwi* maternally or zygotically did not

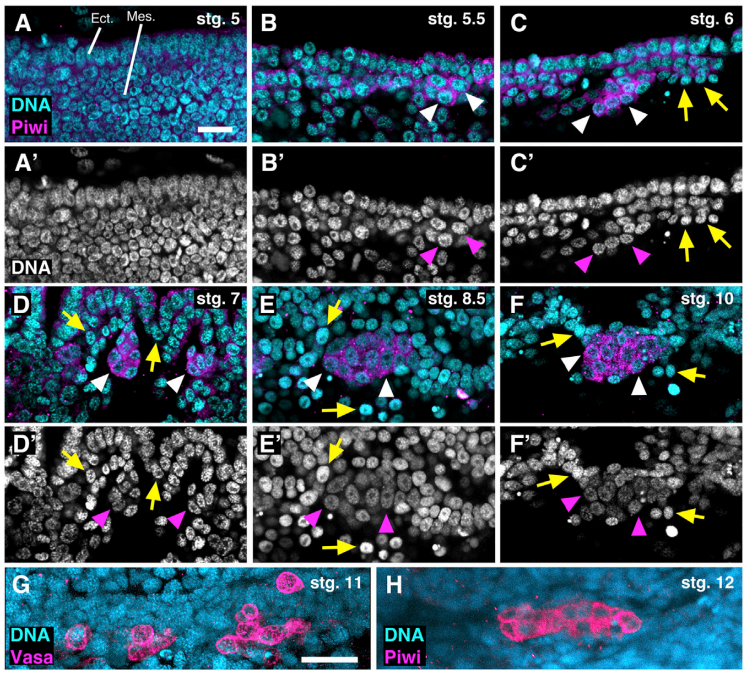


Figure 3. Piwi Expression Correlates with Acquisition of PGC Nuclear Morphology in Abdominal Mesoderm Cells

(A–F′) Between stages 5 and 10, Piwi expression increases in PGCs (A–F, arrowheads), and nuclear morphology of Piwi-enriched cells changes accordingly (A′–F′, arrowheads). Late stage 5 PGCs have chromatin compaction and multiple nucleoli similar to neighboring mesodermal cells (B and B′). As Piwi enrichment in PGCs increases (arrowheads), their chromatin becomes more diffuse and nuclear size increases (yellow arrows) (C and C′). In subsequent stages chromatin morphology differences become more pronounced (D–F′).

(G and H) At stage 11 (G), PGCs commence short-range migration along the anterior-posterior axis toward the intersegmental region of A3–A4 to form a single gonad primordium on each side of the embryo by stage 12 (H).

Scale bars represent 50 μm in (A) (applies to A′–F′) and (G) (applies to H). Anterior is to the left.

disrupt PGC formation or axial patterning (Figures S4C–S4E), indicating that these genes do not act redundantly to direct PGC specification.

vasa and *piwi* Play Roles in *Gryllus* Spermatogenesis

In mice, which lack germ plasm and specify PGCs from presumptive mesoderm via signaling, *vasa* and *piwi* are not required for PGC specification but do mark established PGCs of both sexes and play roles in adult spermatogenesis [24, 25]. We tested whether these genes were required for adult spermatogenesis in *Gryllus* by injecting adult males with dsRNA for *vasa* or *piwi* to achieve paternal RNAi (pRNAi). *Gryllus* testes comprise 200–300 testioles (sperm tubules) [26], within which spermatogenesis proceeds from anterior to posterior (Figures 5A and S5A–S5G). The anterior region of each testiole expresses Vasa and Piwi proteins (Figures S5T and S5U) and contains primary and secondary spermatogonia (Figure 5A). Knockdown of *vasa* or *piwi* via pRNAi severely reduced spermatogonial region length (Figure 5H). In both *vasa* and *piwi* pRNAi testes, meiotic spermatocytes were found in the anterior region of testioles, in some cases almost abutting primary spermatogonia (Figures 5C, 5D, 5F, and 5G, yellow arrowheads), and secondary spermatonial cysts were reduced (Figures 5C, 5D, and 5G, red arrows) or absent (Figure 5F), suggesting that the mitotic divisions of primary spermatogonia were affected. The misregulation of primary spermatocyte divisions was not due to absence of the germline stem cell niche (apical cell), which was present in *piwi* and *vasa* pRNAi testes (Figures 5E–5G, asterisks, and S5B, S5H, and S5N). Postspermatogonial stages of spermatogenesis appeared unaffected (Figures S5I–S5M and S5O–S5S). These data indicate that, as in mice and other animals (see Discussion), *piwi* and *vasa* play a role in *Gryllus* gametogenesis in adult males.

Mesoderm Is Required for *Gryllus* PGCs

Our observations thus far suggested that PGCs arise from among mesodermal cells during abdominal segmentation. To test this hypothesis, we took advantage of the conserved role of the *twist* gene in mesoderm development [27] to ask whether PGCs could form if mesoderm development was compromised. *Gryllus twist* is expressed in the abdominal mesoderm beginning during axial elongation, including in cells of the region where PGCs arise (Figures S6A–S6D2′). In *Drosophila*, *twist* mutants display gastrulation defects [28], yet PGCs form normally because PGC specification occurs via germ plasm well before gastrulation (Figures 6A and 6E). In *Gryllus*, *twist* eRNAi similarly causes disorganization or loss of major mesodermal structures within all body segments (Figures 6F and 6G, compare to 6B and 6C). In contrast to *Drosophila*, however, 49% of *Gryllus twist* eRNAi embryos lack PGCs, compared to 0% of controls ($p < 0.01$, Figures 6D and 6G′), and those *twist* eRNAi embryos that specify PGCs have fewer than controls ($p = 0.05$, Figure 6H). These results are consistent with the hypothesis that PGCs form from a subset of abdominal mesoderm. Alternatively, PGCs may be formed normally at stage 5 but fail to be maintained due to absent or compromised mesodermal surroundings.

Discussion

We have shown that neither *vasa* nor *piwi* are required maternally or zygotically for the formation of functional PGCs (Figures 4 and S4) but instead play a role in spermatogonial divisions in adult males. Our results differ from those of analogous experiments in *D. melanogaster* [15, 18], indicating that the functions of these genes have diverged between *Gryllus* and *Drosophila*. Although these genes are not required for *Gryllus* PGC formation, we propose that, together with *gcl* and *boule* expression (Figure S2) and the transition from mesodermal to PGC-like morphology in situ (Figure 3), *vasa* and *piwi* are nevertheless informative *Gryllus* PGC markers, despite their pleiotropic roles in other developmental

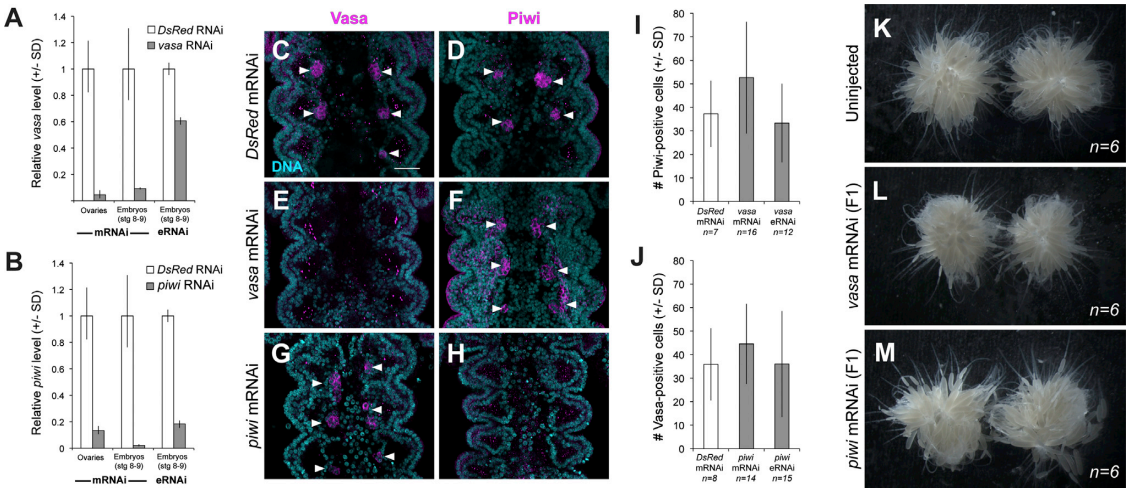


Figure 4. *vasa* and *piwi* Are Not Required for *Gryllus* PGC Specification

(A and B) qPCR validation of *vasa* and *piwi* knockdown following mRNAi and eRNAi.

(C–H) Piwi-positive PGCs (arrowheads) form in *vasa* RNAi embryos, and Vasa-positive PGCs form in *piwi* RNAi embryos. Consistent with qPCR results, *vasa* mRNAi (E; 100%, n = 9) and *piwi* mRNAi (H; 60%, n = 10) abolished respective protein expression. eRNAi produced similar results (not shown).

(I and J) PGC quantification confirms that PGC formation is not reduced (Student's t test: *vasa* mRNAi p = 0.07; *vasa* eRNAi p = 0.57; *piwi* mRNAi p = 0.24; *piwi* eRNAi p = 0.77).

(K–M) Ovaries from adult offspring of *vasa* and *piwi* pRNAi mothers (L–M) are indistinguishable from uninjected controls (K). Scale bar represents 50 μm in (C)–(H).

See also Figure S4.

processes. We cannot eliminate the possibility that untested marker genes might show an earlier PGC specification event than the one we identify in stage 5 (Figure 2C). However, given the conserved coexpression of the tested genes in PGCs of multiple metazoans, we believe it unlikely that all four would be absent from *Gryllus* PGCs at the time of their specification.

Evidence from multiple systems suggests that functional divergence of *vasa* and *piwi* is widespread. In

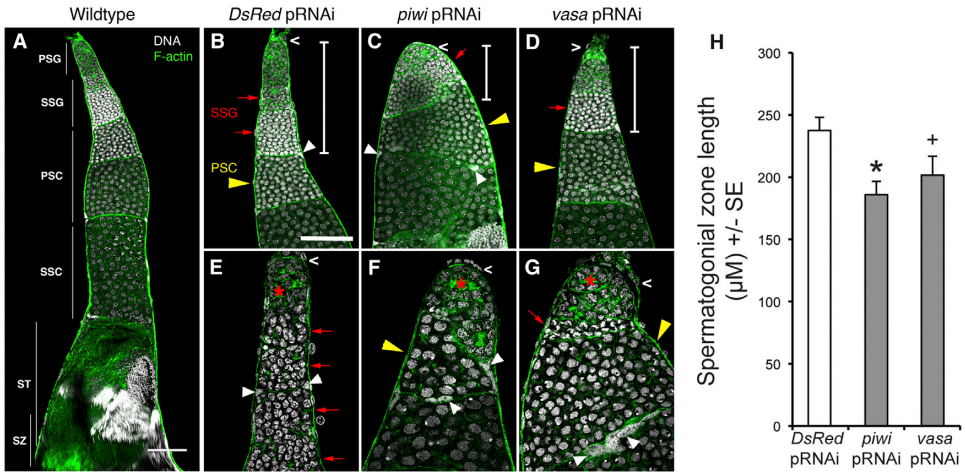


Figure 5. *piwi* and *vasa* pRNAi Causes Defects in Spermatogonial Proliferation

(A) Wild-type *Gryllus* testiole showing the stages of spermatogenesis.

(B–G) White bars in (B)–(D) indicate the spermatogonial zone containing secondary spermatogonia (SSG, red arrows). The zone of primary spermatocytes (PSC, yellow arrowheads) nearly abuts the primary spermatogonial zone in *piwi* (C and F) and *vasa* (D and G) pRNAi testes because of the shortened SSG zone but is absent from the anterior region of control testioles that have extensive SSG populations (B and E). Higher magnification (E–G) is shown of anterior testiole regions in control (E), *piwi* RNAi (F), and *vasa* RNAi (G) testes.

(H) *vasa* or *piwi* paternal RNAi results in a shortened spermatogonial zone compared to controls (Student's t test: *p < 0.01, +p = 0.06).

Scale bar represents 100 μm in (A) and 50 μm in (B) (applies also to C–G). Anterior is up in (A)–(G). See also Figure S5.

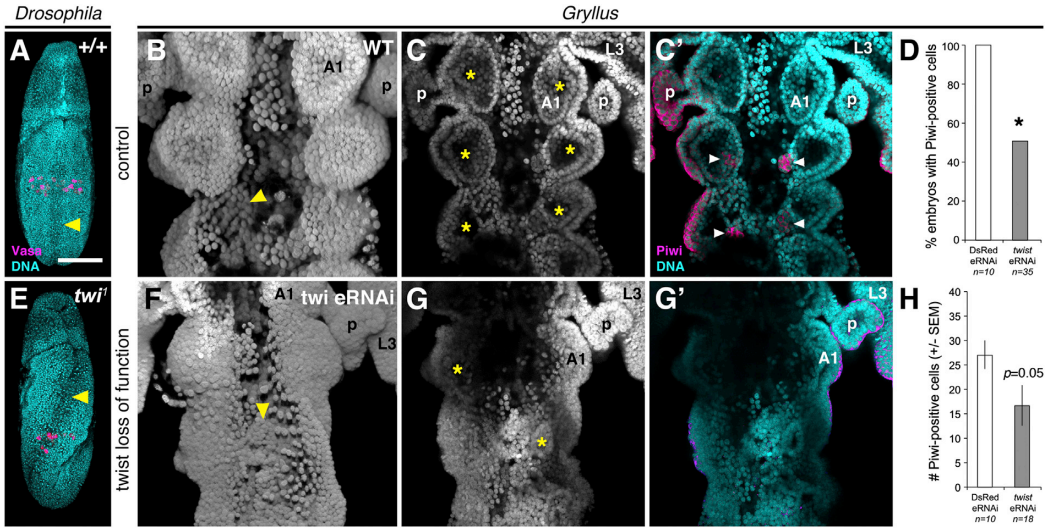


Figure 6. *twist* eRNAi Disrupts *Gryllus* PGC Formation

Drosophila twist^{1} embryos display gastrulation defects (A and E; arrowhead indicates ventral furrow) but specify PGCs (magenta) properly. In comparison to wild-type (B–C′), *Gryllus twist* eRNAi embryos (F–G′) have disorganized mesoderm (F, arrowhead), show absent or rudimentary coelomic pouches (G, asterisks), and fail to specify PGCs (G′) significantly more often than controls (D; Fisher's exact test, *p < 0.01). When they do specify PGCs, *Gryllus twist* eRNAi embryos have fewer PGCs than controls (H; Student's t test, p = 0.05).

Scale bar represents 50 μm and applies to all panels. L3, third thoracic leg; A1, first abdominal segment; p, pleuropodia. Anterior is up. See also Figure S6.

D. melanogaster, where both genes were first discovered, mutations in *vasa* or either of the two *piwi* orthologs (*piwi* and *aubergine*) cause defects in germ plasm formation, oogenesis, PGC specification, and posterior patterning [15, 18, 19, 29]. Similarly, *vasa* and *piwi* orthologs are required for PGC specification, development, and oogenesis in *C. elegans*, *D. rerio* [see 2], and medaka [30]. In mice, however, *vasa* is expressed in embryonic PGCs of both sexes, but *vasa*$^{-/-}$ homozygotes display no discernable defects in PGC specification or oogenesis and instead show a male-specific defect in spermatogenesis [25]. Similarly, knockout mice for any of the three PIWI family homologs display spermatogenic defects only, with no defects in females [24, 31, 32]. Our data therefore suggest that the roles of *Gryllus vasa* and *piwi* are similar to those of their mouse homologs. Functional genetic and gene expression data from insects (Table S1) suggest that, in this clade, an instructive role for these genes in PGC formation may be restricted to the Holometabola, perhaps concomitant with the co-option of *oskar* to the top of the PGC specification pathway [14]. Consistent with this hypothesis, *vasa* is dispensable for PGC formation in another hemimetabolous insect, the milkweed bug, *Oncopeltus fasciatus* [33].

Our data indicate that a zygotic mode of PGC specification is likely present in *Gryllus*, whereby PGCs appear to arise from presumptive mesoderm. Because *twist* is expressed broadly in mesodermal cells (Figure S6), our *twist* RNAi results could indicate either that mesoderm gives rise to PGCs directly or that mesoderm is required to maintain PGCs (we note that these interpretations are not mutually exclusive). However, our morphological (Figure 3) and gene expression (Figures 2, S3, and S6) analyses strongly suggest that cells convert from mesodermal to PGC fate in situ. Alternatively, an undifferentiated population of PGC precursors could exist that do not express any of the tested PGC marker genes but are induced to adopt PGC fate by adjacent mesodermal cells. If this is the case, however, we note that such pluripotent precursors cannot require maternal provision of *vasa* or *piwi* and would most likely be specified by zygotic mechanisms.

Several lines of evidence suggest that a cell lineage relationship between mesoderm and the germline may be a cell type association predating the emergence of Bilateria. Bilaterian germ cells are strikingly similar in gene expression and cytological characteristics to endomesodermally derived stem cells in bilaterian outgroups. Whereas nonbilaterians do not have a dedicated germline per se, their pluripotent stem cell populations serve the function of the germline (reviewed in [34]), and cnidarian pluripotent stem cells are derived from endomesoderm during embryogenesis [35–39]. Within bilaterians, gametogenic cells are consistently described as arising from gonadal epithelia of mesodermal origin in most arthropods and many marine invertebrates (reviewed in [3]). In many spiralians, cytological, cell lineage, and molecular data indicate that PGCs originate from a multipotent mesodermal precursor or precursors (see also [3, 40–47]). Recent studies suggest that mouse PGCs may default to a mesodermal specification program if germline induction signals are absent [48, 49]. The work presented here illuminates broad similarities between PGC specification and *vasa* function in *Gryllus* and in the mouse. Future work will be required to explore this apparent similarity in greater depth and to determine the extent of conservation in the developmental and molecular processes involved in specifying the germline across Bilateria.

Experimental Procedures

Gryllus husbandry, gene expression analysis, mRNAi, eRNAi, and phenotypic analysis were carried out as previously described [50]. For pRNAi, 5 μl of 3 μg/ml dsRNA was injected into the coelomic cavity of adult males

1–3 days after the final molt to sexual maturity, and testes of injected males were dissected for analysis 7 days after injection (details in Supplemental Experimental Procedures).

Accession Numbers

Sequences have been deposited in GenBank (accession numbers KC242803–KC242808).

Supplemental Information

Supplemental Information includes six figures, Supplemental Experimental Procedures, and one table and and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.03.063.

Acknowledgments

National Science Foundation (NSF) IOS-0817678 and IOS-1257217 to C.G.E. and NSF predoctoral fellowships to B.E.-C. and S.D. supported this work. We thank Paul Lasko and Siegfried Roth for reagents, Tony De Tomaso for use of the Olympus compound microscope, Franz Kainz for performing *twist* in situ hybridization, and members of the Extavour laboratory for discussion.

Received: January 4, 2013
Revised: March 3, 2013
Accepted: March 22, 2013
Published: April 25, 2013

References

1. Buss, L.W. (1987). The Evolution of Individuality (Princeton, NJ, USA: Princeton University Press).
2. Ewen-Campen, B., Schwager, E.E., and Extavour, C.G. (2010). The molecular machinery of germ line specification. Mol. Reprod. Dev. *77*, 3–18.
3. Extavour, C.G., and Akam, M.E. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. Development *130*, 5869–5884.
4. Extavour, C.G. (2007). Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. Integr. Comp. Biol. *47*, 770–785.
5. Juliano, C.E., Swartz, S.Z., and Wessel, G.M. (2010). A conserved germline multipotency program. Development *137*, 4113–4126.
6. Nieuwkoop, P.D., and Sutasurya, L.A. (1981). Primordial Germ Cells in the Invertebrates: From Epigenesis to Preformation (Cambridge: Cambridge University Press).
7. Hayashi, K., de Sousa Lopes, S.M., and Surani, M.A. (2007). Germ cell specification in mice. Science *316*, 394–396.
8. Lynch, J.A., Ozüak, O., Khila, A., Abouheif, E., Desplan, C., and Roth, S. (2011). The phylogenetic origin of *oskar* coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola. PLoS Genet. *7*, e1002029.
9. Mito, T., and Noji, S. (2009). The Two-spotted Cricket *Gryllus bimaculatus*: An Emerging Model for Developmental and Regeneration Studies. Emerging Model Organisms: A Laboratory Manual, *Volume 1* (Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press), pp. 331–346.
10. Wheeler, W.M. (1893). A contribution to Insect Embryology. J. Morphol. *8*, 1–160.
11. Yoon, C., Kawakami, K., and Hopkins, N. (1997). Zebrafish *vasa* homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells. Development *124*, 3157–3165.
12. Wu, H.-R., Chen, Y.-T., Su, Y.-H., Luo, Y.-J., Holland, L.Z., and Yu, J.-K. (2011). Asymmetric localization of germline markers Vasa and Nanos during early development in the amphioxus *Branchiostoma floridae*. Dev. Biol. *353*, 147–159.
13. Tsunekawa, N., Naito, M., Sakai, Y., Nishida, T., and Noce, T. (2000). Isolation of chicken *vasa* homolog gene and tracing the origin of primordial germ cells. Development *127*, 2741–2750.
14. Ewen-Campen, B., Srouji, J.R., Schwager, E.E., and Extavour, C.G. (2012). *Oskar* predates the evolution of germ plasm in insects. Curr. Biol. *22*, 2278–2283.
15. Schüpbach, T., and Wieschaus, E. (1986). Maternal-effect mutations altering the anterior-posterior patterns of the *Drosophila* embryo. Wilhelm Roux's Arch. Dev. Biol. *195*, 302–317.
16. Lasko, P.F., and Ashburner, M. (1990). Posterior localization of *vasa* protein correlates with, but is not sufficient for, pole cell development. Genes Dev. *4*, 905–921.
17. Lasko, P.F., and Ashburner, M. (1988). The product of the *Drosophila* gene *vasa* is very similar to eukaryotic initiation factor-4A. Nature *335*, 611–617.
18. Megosh, H.B., Cox, D.N., Campbell, C., and Lin, H. (2006). The role of PIWI and the miRNA machinery in *Drosophila* germline determination. Curr. Biol. *16*, 1884–1894.
19. Harris, A.N., and Macdonald, P.M. (2001). *Aubergine* encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. Development *128*, 2823–2832.
20. Jongens, T.A., Hay, B., Jan, L.Y., and Jan, Y.N. (1992). The *germ cell-less* gene product: a posteriorly localized component necessary for germ cell development in *Drosophila*. Cell *70*, 569–584.
21. Seydoux, G., and Dunn, M.A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of *Caenorhabditis elegans* and *Drosophila melanogaster*. Development *124*, 2191–2201.
22. Kunwar, P.S., Siekhaus, D.E., and Lehmann, R. (2006). In vivo migration: a germ cell perspective. Annu. Rev. Cell Dev. Biol. *22*, 237–265.
23. Lin, H., and Spradling, A.C. (1997). A novel group of *pumilio* mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. Development *124*, 2463–2476.
24. Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development *131*, 839–849.
25. Tanaka, S.S., Toyooka, Y., Akasu, R., Katoh-Fukui, Y., Nakahara, Y., Suzuki, R., Yokoyama, M., and Noce, T. (2000). The mouse homolog of *Drosophila Vasa* is required for the development of male germ cells. Genes Dev. *14*, 841–853.
26. Snodgrass, R.E. (1937). The male genitalia of orthopteroid insects. Smithsonian Misc. Coll. *96*, 1–107.
27. Castanon, I., and Baylies, M.K. (2002). A Twist in fate: evolutionary comparison of Twist structure and function. Gene *287*, 11–22.
28. Simpson, P. (1983). Maternal-Zygotic Gene Interactions during Formation of the Dorsoventral Pattern in *Drosophila* Embryos. Genetics *105*, 615–632.
29. Hay, B., Ackerman, L., Barbel, S., Jan, L.Y., and Jan, Y.N. (1988). Identification of a component of *Drosophila* polar granules. Development *103*, 625–640.
30. Li, M., Hong, N., Xu, H., Yi, M., Li, C., Gui, J., and Hong, Y. (2009). Medaka *vasa* is required for migration but not survival of primordial germ cells. Mech. Dev. *126*, 366–381.
31. Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev. Cell *12*, 503–514.
32. Deng, W., and Lin, H. (2002). *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. Dev. Cell *2*, 819–830.
33. Ewen-Campen, B., Jones, T., and Extavour, C. (2013). Evidence against a germ plasm in the milkweed bug *Oncopeltus fasciatus*, a hemimetabolous insect. Biol. Open. Published online April 19, 2013. http://dx.doi.org/10.1242/bio.20134390.
34. Agata, K., Nakajima, E., Funayama, N., Shibata, N., Saito, Y., and Umesono, Y. (2006). Two different evolutionary origins of stem cell systems and their molecular basis. Semin. Cell Dev. Biol. *17*, 503–509.
35. Extavour, C.G., Pang, K., Matus, D.Q., and Martindale, M.Q. (2005). *vasa* and *nanos* expression patterns in a sea anemone and the evolution of bilaterian germ cell specification mechanisms. Evol. Dev. *7*, 201–215.
36. Martindale, M.Q., Pang, K., and Finnerty, J.R. (2004). Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone *Nematostella vectensis* (phylum, Cnidaria; class, Anthozoa). Development *131*, 2463–2474.
37. Kumé, M., and Dan, K. (1968). Invertebrate Embryology (Belgrade: Prosveta).
38. Pilato, G. (2000). The ontogenetic origin of germ cells in Porifera and Cnidaria and the "theory of the endoderm as secondary layer". Zool. Anz. *239*, 289–295.

39. Martin, V.J., and Archer, W.E. (1997). Stages of larval development and stem cell population changes during metamorphosis of a hydrozoan planula. Biol. Bull. *192*, 41–52.
40. Swartz, S.Z., Chan, X.Y., and Lambert, J.D. (2008). Localization of *Vasa* mRNA during early cleavage of the snail *Ilyanassa*. Dev. Genes Evol. *218*, 107–113.
41. Meyer, N.P., Boyle, M.J., Martindale, M.Q., and Seaver, E.C. (2010). A comprehensive fate map by intracellular injection of identified blastomeres in the marine polychaete *Capitella teleta*. Evodevo *1*, 8.
42. Agee, S.J., Lyons, D.C., and Weisblat, D.A. (2006). Maternal expression of a NANOS homolog is required for early development of the leech *Helobdella robusta*. Dev. Biol. *298*, 1–11.
43. Rebscher, N., Zelada-González, F., Banisch, T.U., Raible, F., and Arendt, D. (2007). Vasa unveils a common origin of germ cells and of somatic stem cells from the posterior growth zone in the polychaete *Platynereis dumerilii*. Dev. Biol. *306*, 599–611.
44. Dill, K.K., and Seaver, E.C. (2008). *Vasa* and *nanos* are coexpressed in somatic and germ line tissue from early embryonic cleavage stages through adulthood in the polychaete *Capitella sp. I*. Dev. Genes Evol. *218*, 453–463.
45. Lyons, D.C., Perry, K.J., Lesoway, M.P., and Henry, J.Q. (2012). Cleavage pattern and fate map of the mesentoblast, 4d, in the gastropod *Crepidula*: a hallmark of spiralian development. Evodevo *3*, 21.
46. Giani, V.C., Jr., Yamaguchi, E., Boyle, M.J., and Seaver, E.C. (2011). Somatic and germline expression of *piwi* during development and regeneration in the marine polychaete annelid *Capitella teleta*. Evodevo *2*, 10.
47. Rabinowitz, J.S., Chan, X.Y., Kingsley, E.P., Duan, Y., and Lambert, J.D. (2008). Nanos is required in somatic blast cell lineages in the posterior of a mollusk embryo. Curr. Biol. *18*, 331–336.
48. Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. *22*, 1617–1635.
49. Yabuta, Y., Kurimoto, K., Ohinata, Y., Seki, Y., and Saitou, M. (2006). Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. Biol. Reprod. *75*, 705–716.
50. Kainz, F., Ewen-Campen, B., Akam, M., and Extavour, C.G. (2011). Notch/Delta signalling is not required for segment generation in the basally branching insect *Gryllus bimaculatus*. Development *138*, 5015–5026.

Current Biology, Volume *23*
Supplemental Information

Germ Cell Specification Requires Zygotic Mechanisms Rather Than Germ Plasm in a Basally Branching Insect

Ben Ewen-Campen, Seth Donoughe, Donald Nat Clarke, and Cassandra G. Extavour

Author Contributions

B.E.-C. and C.G.E. designed research, analyzed data, and wrote the paper; experiments were carried out by S.D. (*twist* eRNAi experiment), D.N.C. (pRNAi egg-laying and embryonic survival scoring, PGC migration analysis), C.G.E. (nuclear morphology analysis, *twist* in situ hybridization analysis, spermatogenesis analysis of pRNAi experiments), and B.E.-C. (all other experiments); C.G.E. obtained funding for the research.

Supplemental Inventory

- Supplemental Figures S1–S6
- Table S1, Related to Figure 1. Data on Insect PGC Origin during Embryogenesis
  Please see accompanying Excel file. Only studies directly addressing the mechanism and/or description of the first embryonic appearance of PGCs are referenced. Numbered references in the Excel file are listed at the end of this Supplemental Information document.
- Supplemental Experimental Procedures
- Supplemental References

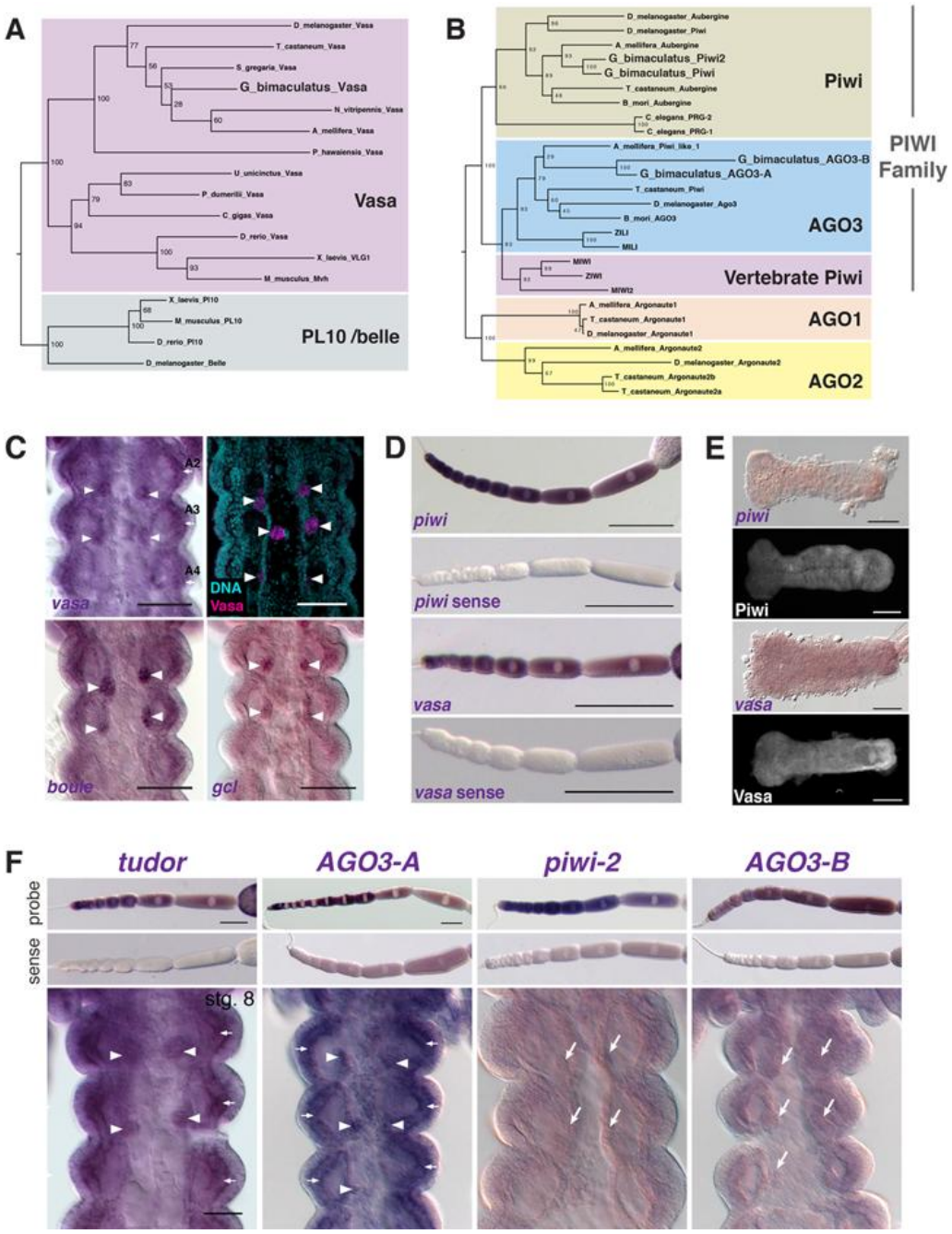


Figure S1, Related to Figure 1. Phylogenetic Analysis of *Gryllus* Germ Line Genes and Expression Patterns of Additional Germ Line Marker Candidates

(A and B) Maximum-likelihood phylogenetic reconstruction of Vasa and Piwi amino acid sequences. As previously reported [1], *Gryllus* Vasa falls clearly within other insect *vasa* genes, not the PL10/Belle class of RNA helicases (A). *Gryllus* possesses two *piwi*-like genes and two AGO3-like genes, both of which represent species-specific duplications (B). As only the first identified *piwi*-like gene [2] was enriched in *Gryllus* PGCs, we focus the present analyses on this orthologue, which we refer to here simply as *piwi* as it is clearly orthologous to other animal *piwi* genes. Note that *aubergine* is a *Drosophila*-specific duplication of *piwi*.
(C) *Gryllus* PGCs (arrowheads) express high levels of Vasa protein and transcripts of *vasa*, *boule* and *germ cell less*. All genes are also expressed at lower levels throughout the somatic tissues of the embryo.
(D) *piwi* and *vasa* transcripts are expressed ubiquitously during all stages of oogenesis and do not localise to the posterior ooplasm.
(E) *piwi* and *vasa* transcripts and protein products are expressed ubiquitously in stage 4 embryos. The apparent increased expression levels at the germ band posterior are an artifact of tissue thickness.
(F) Top row: *tudor*, *AGO3-A, piwi-2*, and *AGO3-B* transcripts are not localized asymmetrically in oocytes. Bottom row: these genes do not specifically label *Gryllus* PGCs. In the PGC-containing region (Figure 1B-E) at stage 9, *tudor* and *AGO3-A* are detectable in PGCs (arrowheads), but are also expressed throughout the somatic tissues of the embryo (arrows). *piwi-2* and *AGO3-B* are not detected above background levels in stage 8-9 embryos. Arrows mark PGCs recognisable based on morphology and anatomical position independent of gene expression. Anterior is to the left in D and top two rows row of E, and up in C and bottom row of E. Scale bars = 100 μM in C, 500 μM in D; 200 μM in E and top two rows of F; 50 μM bottom row of F.

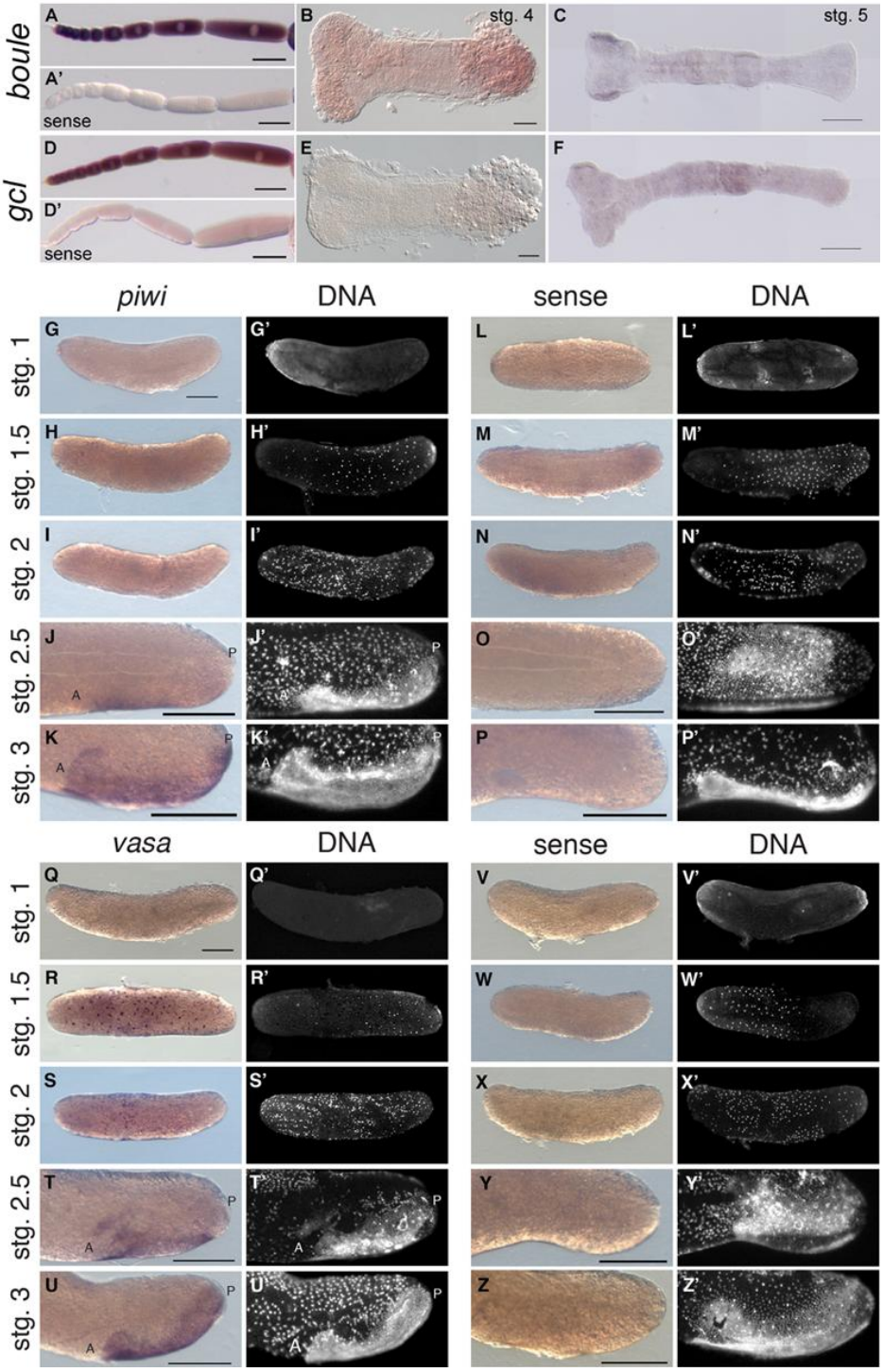

boule
A
A'
sense
B
stg. 4
C
stg. 5
gcl
D
D'
sense
E
F
piwi
DNA
sense
DNA
stg. 1
stg. 1.5
stg. 2
stg. 2.5
stg. 3
G
G'
L
L'
H
H'
M
M'
I
I'
N
N'
J
J'
O
O'
K
K'
P
P'
P
A
vasa
DNA
sense
DNA
Q
Q'
V
V'
R
R'
W
W'
S
S'
X
X'
T
T'
Y
Y'
U
U'
Z
Z'

Figure S2, Related to Figures 1 and 2. Additional Gene Expression Data in Support of the Absence of Germ Plasm in *Gryllus*

(A–F) Expression of *boule* and *gcl*, which mark PGCs in stage 9 embryos (Figure 1) during oogenesis and early embryogenesis. Neither gene is asymmetrically localized in oocytes (A and D; A' and D' show sense controls). Both genes are expressed ubiquitously during stages 4 (B and E) and 5 (C and F), and do not reveal any segregated PGCs during these stages.

(G–Z) *piwi* (G–K) and *vasa* (Q–U) expression during blastoderm stages. (L-P and V-Z) Corresponding sense controls. (G'–Z') Nuclear stains of adjacent panels. *piwi* transcripts are undetectable during blastoderm stages (G–I), and are found ubiquitously at low levels as the germ band condenses (J and K). *vasa* transcripts are undetectable in just-laid eggs (Q), and energids are associated with all nuclei along the A–P axis as they populate the blastoderm surface (R). During subsequent blastoderm divisions (S), *vasa* expression is not localized to any specific subset of nuclei. As the germ band condenses at the posterior of the egg (T, U), *vasa* expression is detected at similar, low levels throughout the germ band but not enriched at the posterior or in any other specific region. A = germ band anterior, P = germ band posterior. Scale bars = 200 μM in (A)–(A') and (D)–(D'); 100 μM in (B) (applies also to E); 200 μM in (C) (applies also to F), (G), and (Q) (applies also to H–I and R–S, respectively), and (J)–(K), (O)–(P), (T)–(U), and (Y)–(Z). Anterior is to the left in all panels.

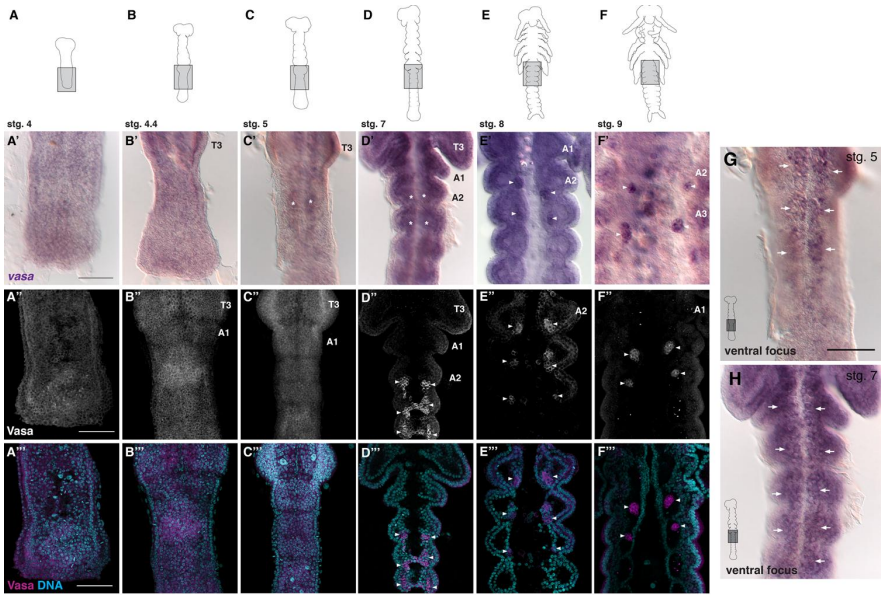


Figure S3, Related to Figure 2. Expression of *vasa* Transcript and Protein throughout Abdominal Segmentation

(A–F) Schematic drawings of *Gryllus* mid-staged embryos of the stages shown here; boxed grey areas indicate regions shown in panels below. *vasa* transcripts are expressed ubiquitously during stages 4–4.4 (A and B), and do not reveal the presence of PGCs at this stage, consistent with *piwi* expression. (C and D) During stages 5 and 7 *vasa* transcripts do not reveal the presence of PGCs. Astrices in (C' and D') denote out-of-focus staining in the ventrally located nervous system, which is shown in focus in (G–H). *vasa* transcripts are detected in PGCs during stage 8 and 9 (E and F). Vasa protein is ubiquitously expressed during stages 4-5 (A'–C', A"–C"). In stage 7 embryos (D' and D"), Vasa protein is strongly enriched in PGCs, and this expression continues in stage 8 and 9 (E'–F' and E"–F"). Arrowheads indicate PGC clusters. T3 = thoracic segment 3; A1, A2 = abdominal segments 1 and 2. Scale bar = 100 μM. Anterior is up in all panels.

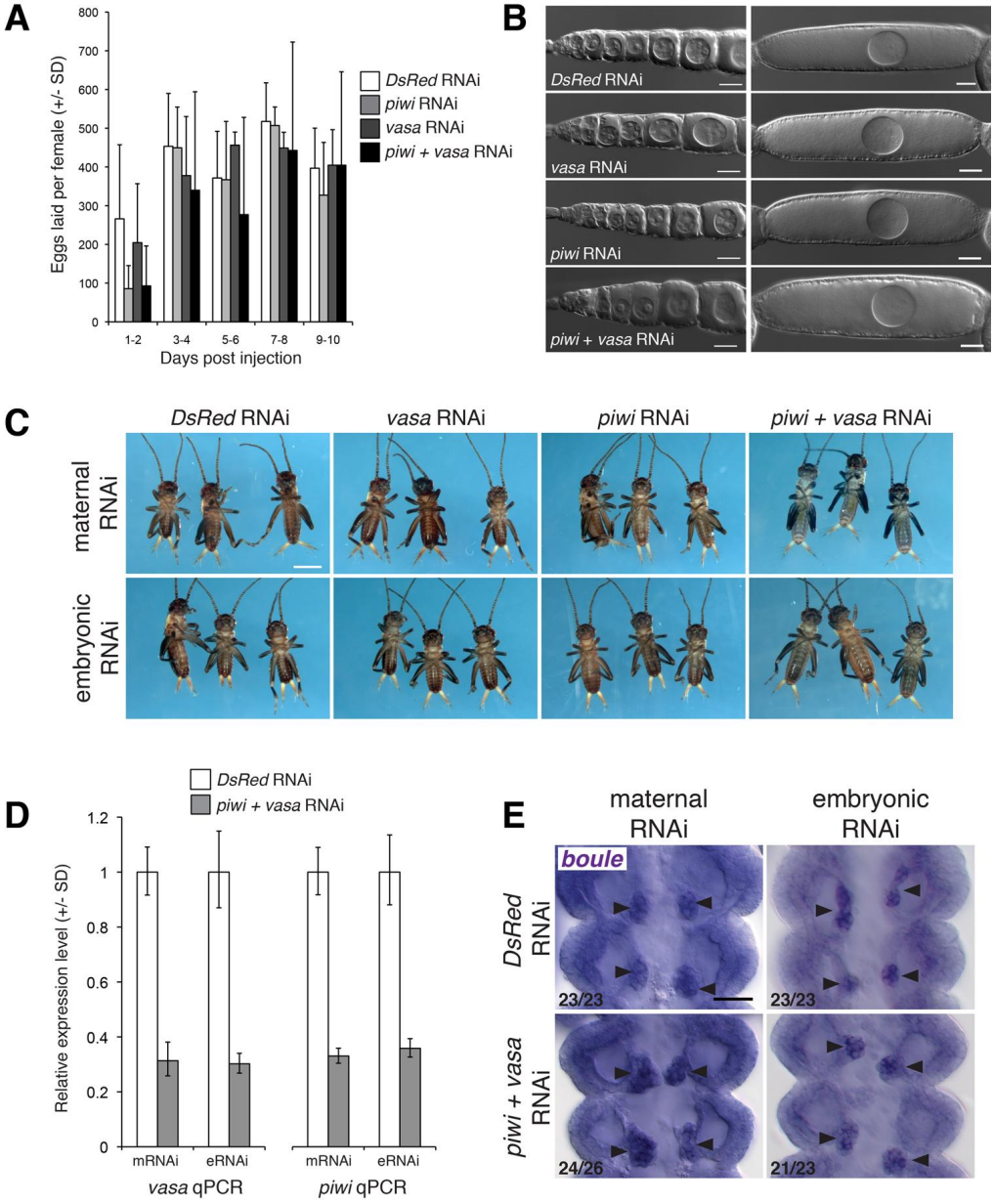


Figure S4, Related to Figure 4. Phenotypic Analysis of *vasa* and *piwi* mRNAi and *vasa* + *piwi* Double eRNAi and mRNAi, in Ovaries and Embryos

(A) Females injected with dsRNA against *vasa* or *piwi* lay numbers of eggs that do not differ significantly from controls (student's t-test, $p>0.05$ in every pairwise comparison of vasa or piwi RNAi with DsRed RNAi on the indicated days post-injection).

(B) Ovaries dissected from *vasa* or *piwi* mRNAi females 10 days after injection are morphologically

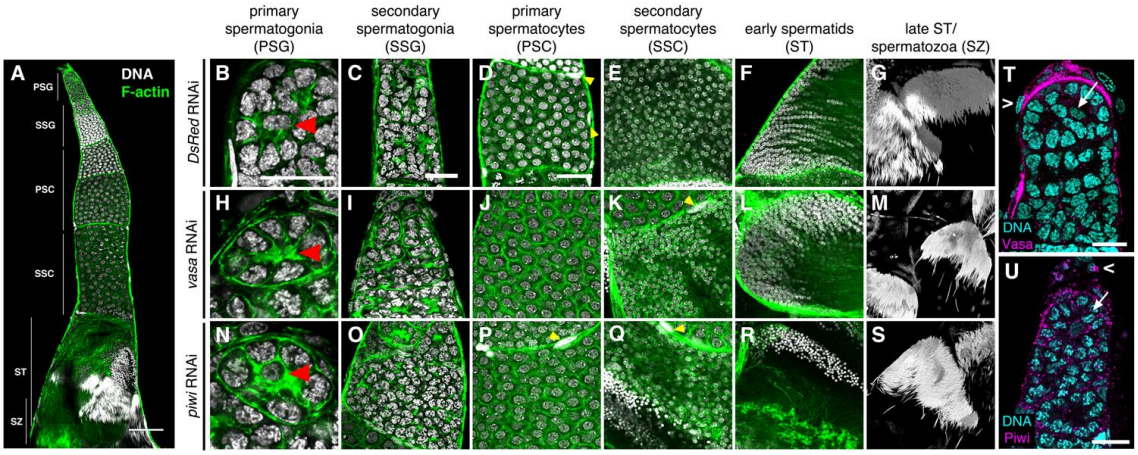


Figure S5, Related to Figure 5. *vasa* or *piwi* pRNAi Does Not Disrupt Postspermatogonial Stages of Spermatogenesis in *Gryllus*
Wild type *Gryllus* testiole showing the stages of spermatogenesis (A). Primary spermatogonia (PSC) undergo self-renewing divisions, which are thought to occur under the influence of a single apical cell (red arrowheads in B, H, N) that provides a "stem cell niche" analogous to the hub of *Drosophila* testes [3, 4]. Following seven mitotic divisions by cysts of secondary spermatogonia (SSG) enclosed by somatic cell sheaths (yellow arrowheads in D, K, P–Q), the resulting 128 primary spermatocytes (PSC) undergo meiosis (secondary spermatocytes: SSC) to produce 512 clonally related spermatids (ST), which undergo synchronous spermeiogenesis to produce bundles of mature spermatozoa (SZ) [5]. Although *vasa* and *piwi* pRNAi testes display a reduction in the number of secondary spermatogonial cysts (Figure 7), they possess normal apical cells (B, H, N, arrowheads), and cysts proceeding normally through all stages of spermatogenesis (D–G, I–M, O–S), which are surrounded by somatic sheath cells (arrowheads) as in controls. Vasa (T) and Piwi (U) proteins are expressed in the anterior region of testioles, which contains primary spermatogonia (arrows) and somatic sheath cells (arrowheads), and is enclosed by a cellular peritoneal sheath (carets). Scale bars = 100 μm in A; 50 μm in (D) (applies also to E–G, J–M, P–S); 25 μm in (B) (applies also to H and N) and (C) (applies also to I and O); 20 μm in (T)–(U).

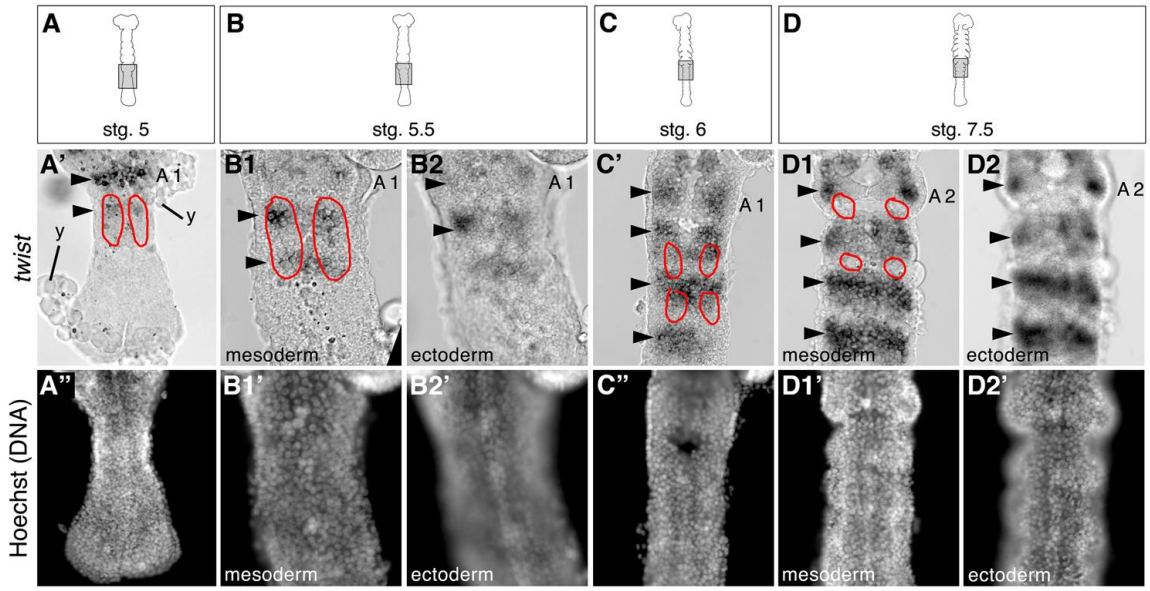


Figure S6, Related to Figure 6. *twist* Expression in the Abdominal Mesoderm where *Gryllus* Germ Cells Arise

(A–D) Schematic drawings of progressive stages of *Gryllus* embryogenesis; grey box indicates region shown in panels below. *twist* transcripts accumulate in an anterior to posterior progression in abdominal segments, indicated by black arrowheads in (A'), (B1)–(B2), (C'), and (D1)–(D2). Red outlines in (A'), (B1), (C'), and (D1) indicate the regions that become enriched for *piwi* expression at these stages, suggesting that these are the sites of PGC origin and showing that these regions express *twist* at the proposed onset of PGC specification (stage 5). Bottom row shows nuclear staining of corresponding bright field images in the row above. Anteriormost abdominal segment is labeled in each panel. y = yolk. *twist* expression is confined to the mesoderm and absent from the ectoderm, as shown in micrographs of mesodermal focal planes in (B1 and D1), and ectodermal focal planes in (B2 and D2).

Supplemental Experimental Procedures

Insect Cultures and Embryonic Staging

Gryllus bimaculatus cultures were maintained as previously described [139] and embryos were staged according to [140]. *Drosophila Oregon R* and *twist*^{1} stocks were obtained from the Bloomington Drosophila Stock Center (#5, #2381).

Cloning and Phylogenetic Analysis

Orthologues of *boule*, *tudor*, *germ cell-less,* an additional *piwi*-like gene, and two *AGO3*-related genes were identified in a *Gryllus* developmental transcriptome via reciprocal best BLAST hit analysis against the *Drosophila melanogaster* proteome. *Gryllus twist* was a gift of S. Roth (University of Cologne, Germany).

To resolve the orthology of the four *Gryllus* PIWI family proteins, we used maximum-likelihood based phylogenetic reconstruction as implemented by RAxML v 7.2.8 [141, 142] on the Odyssey Cluster, maintained by the FAS Sciences Division Research Computing Group (Harvard University). The alignment was produced using Muscle [143] and trimmed using Gblocks [144] under the least stringent settings. The best tree and rapid bootstrap analysis were conducted from 2000 independent runs under the WAG model of protein evolution with a gamma distribution of rate heterogeneity.

In Situ Hybridization

DIG-labeled probes were hybridised at 68° C following standard protocols [139], with 50% polyvinyl alcohol included in the NBT/BCIP development step. Probe lengths were as follows: *vasa*: 1,953 bp*; piwi-1*: 781 bp*; piwi-2*: 821 bp*; AGO3-A*: 760 bp*; AGO3-B*: 832 bp*; germ cell-less*: 1,691 bp; *boule:* 995 bp*; tudor*: 1,707 bp. Our results for *vasa* expression (both mRNA and protein) in *Gryllus* differ from those reported by Mito *et al*. [1], who failed to identify the germ cell clusters that we observed beginning at stages 6/7 (Figures 1, 4, S1, S4). This discrepancy may be due to the strong nervous system expression of *vasa* that can obscure the relatively weaker PGC expression (their Figures 3I, 3J, 4A-D), and to our use of a species-specific Vasa antibody [2] as opposed to the cross-reactive antibody [145] used by Mito *et al*.. Mito *et al*. also reported detection of transient *vasa* mRNA staining at the posterior of stage 4 embryos (their Figure 3C-D), and interpreted it as consistent with Heymons' 1895 claim that germ cell precursors arose among the posterior germ band mesoderm shortly after gastrulation [125]. However, we found that this apparently stronger expression is due to the thickness of the posterior germ band tissue at these stages. In three *in situ* hybridization replicates and ≥30 early stage embryos, *vasa* did not show consistent enrichment in any specific embryonic region before stage 5. We therefore conclude that germ cells are not specified until this stage, in agreement with the majority of previous authors on orthopteran germ cell origin [131-133].

Immunohistochemistry

Primary antibodies used were rabbit anti-Gb-Vasa and anti-Gb-Piwi [2] at 1:300, mouse anti-RNA polymerase II pSer 6 Mab H5 (Covance MMS-129R) 1:100, FITC-conjugated anti-alpha Tubulin (Sigma F2168) 1:100 and rabbit anti-Drosophila Vasa 1:500 (gift of P. Lasko) following standard procedures. Goat anti-rabbit secondary antibodies conjugated to Alexa 488, Alexa 555 or Alexa 568 (Invitrogen) were used at 1:500 or 1:1000. Counterstains were Hoechst 33342 (Sigma B2261) 0.1 to 0.05 mg/ml and FITC-conjugated phalloidin (Sigma P5282) 1 U/ml.

RNA Interference

dsRNA injection into adult females (maternal RNAi = mRNAi) and newly laid embryos (embryonic RNAi = eRNAi) was conducted as previously described [2]. dsRNA fragments for *vasa* and *piwi* were 541 bp and 646 bp, respectively. For eRNAi double knockdown experiments, equal volumes of *vasa* and *piwi* dsRNA were mixed prior to injection. For mRNAi double knockdown experiments, twice the volume of dsRNA as that used for single RNAi experiments (15 µg each of *vasa* and *piwi* dsRNA, or 30 µg of the DsRed control dsRNA) was injected into adult females. dsRNA was used at a concentration of 3 µM (mRNAi) and 5 µM (eRNAi and pRNAi).

qPCR Analysis of Knockdown

qPCR was used to verify RNAi efficacy as follows: total RNA was extracted from RNAi-treated ovaries or stage 8-9 (day 4) embryos using TRIzol (Invitrogen) and including a 30-minute DNase digestion at 37° C to remove genomic contamination. Equal volumes of RNA were used as template for first strand cDNA synthesis using SuperScript III (Invitrogen) including a no reverse transcriptase control. cDNA was diluted 1:5 prior to qPCR. qPCR was conducted using PerfeCta SYBR Green SuperMix (Low ROX, Quanta Biosciences) in a Stratagene MxP3005 machine. Primers amplifying single amplicons of *piwi* (129 bp; F: TTCGGCCAACTACTTCAAGC; R: AGAGTTTCCCGATGAACACG), *vasa* (150 bp; F: GAACATTGTGAGCCTCATGC ; R: TTGCTGAGCCTGGTGGTAT) and beta-tubulin (166 bp; F: TGGACTCCGTCCGGTCAGGC; R: TCGCAGCTCTCGGCCTCCTT) were used. Each reaction was conducted in triplicate, and fluorescence measurements were normalised and background-subtracted using the ROX dye present in the PCR reactions.

C_t values were used to calculate fold change compared to *DsRed*-injected controls using the $2^{-\Delta\Delta Ct}$ method [146]. Triplicate C_t values were averaged and the standard deviation was propagated using standard methods.

Imaging and Image Analysis

Micrographs were captured with AxioVision v.4.8 driving a Zeiss Stereo Lumar equipped with an AxioCam MRc camera, Zen Blue 2011 driving a Zeiss Stereo Zoom equipped with an AxioCam HRc camera, a Zeiss Axio Imager equipped with an AxioCam MRm camera using epifluorescence either with or without an Apotome, or an Olympus IX71 equipped with a Hamamatsu C10600-108 camera. Confocal microscopy was performed with a Zeiss LSM 710 or 780 confocal, using comparable gain, offset, and averaging parameters for all samples. Image analyses were performed with AxioVision v.4.8, Zen 2009 or Zen 2011 (Zeiss), and figures were assembled in Photoshop CS4, InDesign CS4, or Illustrator CS4 (Adobe). For confocal images shown in Figures 1E; 2A”-H”; 4C-H; 6C’ and G’; S3A”’-F”’, a maximum-intensity projection of multiple optical sections of the antibody staining was superimposed over a single optical section of the nuclear counterstain for visual clarity. All other confocal micrographs are maximum intensity projections (Figures 2A”-H’; 3G and H; S3A”-F”), three-dimensional projections (Figure 6B, F) or single optical sections (all other confocal micrographs).

Supplemental References

1. Mito, T., Nakamura, T., Sarashina, I., Chang, C.C., Ogawa, S., Ohuchi, H., and Noji, S. (2008). Dynamic expression patterns of *vasa* during embryogenesis in the cricket *Gryllus bimaculatus*. Dev. Genes Evol. *218*, 381-387.
2. Ewen-Campen, B., Srouji, J.R., Schwager, E.E., and Extavour, C.G. (2012). *oskar* Predates the Evolution of Germ Plasm in Insects. Curr. Biol. *22*, 2278-2283.
3. Carson, H.L. (1945). A comparative study of the apical cell of the insect testis. J. Morphol. *77*, 141-161.
4. Hardy, R.W., Tokuyasu, K.T., Lindsley, D.L., and Garavito, M. (1979). The germinal proliferation center in the testis of *Drosophila melanogaster*. Journal of Ultrastructure Research *69*, 180-190.
5. Dumser, J.B. (1980). The regulation of spermatogenesis in insects. Annu. Rev. Entomol. *25*, 341-369.
6. Staiber, W. (2000). Immunocytological and FISH analysis of pole cell formation and soma elimination of germ line-limited chromosomes in the chironomid *Acricotopus lucidus*. Cell Tissue Res. *302*, 189-197.
7. Christophers (1960). *Aedes aegypti* L. The yellow fever mosquito, its life history, bionomics and structure. (Cambridge: Cambridge University Press).
8. Juhn, J., and James, A.A. (2006). *oskar* gene expression in the vector mosquitoes, *Anopheles gambiae* and *Aedes aegypti*. Insect Mol. Biol. *15*, 363-372.
9. Leloup, A.M. (1974). Morphogenèse des Gonades chez *Calliphora erythrocephala*: étude descriptive. Cellule *1*, 51-93.
10. Noack, W. (1901). Beitrage zur Entwicklungsgeschichet der Musciden. Zeit. f. wiss Zool. *70*, 1-55.
11. Riparbelli, M.G., Callaini, G., and Dallai, R. (1996). Primordial germ cell migration in the *Ceratitis capitata* embryo. Tissue Cell *28*, 99-105.
12. Ritter, R. (1890). Die Entwicklung der Geschlechtsorgane und des Darmes bei *Chironomous*. Zeit. f. wiss Zool. *50*, 408-425.
13. Curtis, D., Apfelt, J., and Lehmann, R. (1995). *nanos* is an evolutionarily conserved organizer of anterior-posterior polarity. Development, 1899-1910.
14. Rohr, K.B., Tautz, D., and Sander, K. (1999). Segmentation gene expression in the mothmidge *Clogmia albipunctata* (Diptera, Psychodidae) and other primitive dipterans. Dev. Genes Evol. *209*, 145-154.
15. Hegner, R.W. (1914). Studies on Germ Cells. I. The history of the germ cells in insects with special reference to the Keimbahn-determinants. II. The origin and significance of the Keimbahn-determinants in animals. J. Morphol. *25*, 375-509.
16. Davis, C.W.C. (1967). A comparative study of larval embryogenesis in the mosquito *Culex fatigans* Wiedemann (Diptera: Culicidae) and the sheep-fly *Lucilia sericata* Meigen (Diptera: Calliphoridae). I. Description of Embryonic Development. Aust. J. Zool. *15*, 547-579.
17. Oelhafen, F. (1961). Zur embryogenese von *Culex pipiens*: Markierungen und exstirpationen mit UV-strahlenstich. Roux Arch. Dev. Biol. *153*, 120-157.
18. Juhn, J., Marinotti, O., Calvo, E., and James, A.A. (2008). Gene structure and expression of *nanos* (*nos*) and *oskar* (*osk*) orthologues of the vector mosquito, *Culex quinquefasciatus*. Insect Mol. Biol. *17*, 545-552.

19. Anderson, D.T. (1962). The Embryology of *Dacus tyroni* (Frogg.) (Diptera, Trypetidae (= Tephritidae)), the Queensland Fruit-Fly. J. Emb. Exp. Morph. *3*, 248-292.
20. Huettner, A.F. (1923). The origin of the germ cells in *Drosophila melanogaster*. J. Morphol. *2*, 385-422.
21. Warn, R. (1975). Restroation of the capacity to form pole cells in u.v.-irradiated *Drosophila* embryos. J. Embryol. Exp. Morphol. *33*, 1003-1011.
22. Hathaway, D.S., and Selman, G.S. (1961). Certain aspects of cell lineage and morphogenesis studied in *Drosophila melanoagster* with an ultraviolet microbeam. J. Embryol. Exp. Morphol. *9*, 310-325.
23. Geigy, R. (1931). Action de l'ultra-violet sur le pôle germinal dans l'oef de *Drosophila melanogaster*. Rev. Suisse Zool. *38*, 187-288.
24. Poulson, D.F., and Waterhouse, D.F. (1960). Experimental studies on pole cells and midgut differentiation in Diptera. Aust. J. Biol. *13*, 541-566.
25. Poulson, D.F. (1950). Diagram of cell lineage in the embryo of *D. melanogaster*. The Biology of Drosophila, 243 (from 168-274).
26. Illmensee, K., and Mahowald, A.P. (1974). Transplantation of Posterior Polar Plasm in *Drosophila*. Induction of Germ Cells at the Anterior Pole of the Egg. Proc. Natl. Acad. Sci. USA *4*, 1016-1020.
27. Illmensee, K., and Mahowald, A.P. (1976). The autonomous function of germ plasm in a somatic region of the *Drosophila* egg. Exp. Cell Res. *97*, 127-140.
28. Underwood, E.M., Caulton, J.H., Allis, C.D., and Mahowald, A.P. (1980). Developmental Fate of Pole Cells in *Drosophila melanogaster*. Dev. Biol., 303-314.
29. Callaini, G., Riparbelli, M. G., and Dallai, R. (1994). Pole cells migration through the midgut wall of the *Drosophila* embryo. Cell Biol. Int. *5*, 514.
30. Junquera, P. (1983). Polar plasm and pole cell formation in an insect egg developing with or without follicular epithelium: an ultrastructural study. J. Exp. Zool., 441-452.
31. Lassmann, G.W.P. (1936). The early embryological development of *Melophagus ovinus* L., with special reference to the development of the germ cells. Ann. Entomol. Soc. Am. *29*, 397-413.
32. Hegner, R.W. (1914). The Germ-Cell Cycle in Animals, (New York: The MacMillan Company).
33. Metschnikoff, E. (1866). Embryologische Studien an Insekten. Zeit. f. wiss Zool. *16*, 389-500.
34. Mahowald, A.P. (1975). Ultrastructural changes in the germ plasm during the life cycle of *Miastor* (Cecidomyidae, Diptera). Roux Arch. Dev. Biol. *176*, 223-240.
35. West, J.A., Cantwell, G.E., and Shorino, T.J. (1968). Embryology of the house fly, *Musca domestica* (Diptera: Muscidae), to the blastoderm stage. Ann. Entomol. Soc. Am. *61*, 13-17.
36. Kowalevsky, A. (1886). Zur Embryonalen Entwicklung der Musciden. Biol. Zentralbl. *6*, 49-54.
37. Escherich, K. (1900). Über die Bilding der Keimblatter bei den Musciden. Nova Acta Abh. der Kaiserl. Leop-Carol. Deutschen Akad. Nat. *77*, 303-364.
38. Schüpbach, P.M., and Camenzind, R. (1983). Germ-Cell Lineage and Follicle Formation in Pedogenetic Development of *Mycophila speyeri* Barnes (Diptera, Cecidomyiidae). Int. J. Ins. Morph. Emb. *12*, 211-223.

39. Fish, W.A. (1952). Embryology of *Phaenica sericata* (Meigen) (Diptera: Calliphoridae). Part IV. The inner layer and mesenteron rudiments. Ann. Entomol. Soc. Am. *45*, 1-22.
40. Auten, M. (1934). The early embryological development of *Phormia regina*: Diptera (Calliphoridae). Ann. Entomol. Soc. Am. *27*, 481-506.
41. Rezende-Teixeira, P., Palomino, N.B., and Machado-Santelli, G.M. (2012). Rananos expression pattern during oogenesis and early embryonic development in Rhynchosciara americana. Dev. Genes Evol. *222*, 153-164.
42. Du Bois, M. (1932). A contribution to the embryology of *Sciara coprophilia* (Diptera). J. Morphol. *54*, 161-192.
43. Gambrell, F.L. (1933). The embryology of the black fly, *Simulium pictipes* Hagen. Ann. Entomol. Soc. Am. *26*, 641-671.
44. Zissler, D., and Sander, K. (1973). The cytoplasmic architecture of the egg cell of *Smittia spec.* (Diptera, Chironomidae). Roux Arch. Dev. Biol., 175-186.
45. Kessel, E.L. (1939). The embryology of fleas. Smithsonian Miscellaneous Collections *98*, 1-78.
46. Saito (1937). On the development of the Tusser, *Antheraea pernyi* Guerin-Meneville, with special reference to the comparative embryology of insects. J. Fac. Agr. Hokkaido Imp. Univ. *40*, 35-109.
47. Toshiki, T., Chantal, C., R., Toshio, K., Eappen, A., Mari, K., Natuo, K., Jean-Luc, T., Bernard, M., Gérard, C., Paul, S., et al. (2000). Germline transformation of the silkworm *Bombyx mori* L. using a piggyBac transposon-derived vector. Nature Biotech., 81-84.
48. Miya, K. (1953). The presumptive genital region at the blastoderm stage of the silkworm egg. J. Fac. Agr. Iwate Univ. *1*, 223-227.
49. Nakao, H., Hatakeyama, M., Lee, J.M., Shimoda, M., and Kanda, T. (2006). Expression pattern of Bombyx vasa-like (BmVLG) protein and its implications in germ cell development. Dev Genes Evol *216*, 94-99.
50. Zhao, G., Chen, K., Yao, Q., Wang, W., Wang, Y., Mu, R., Chen, H., Yang, H., and Zhou, H. (2008). The nanos gene of *Bombyx mori* and its expression patterns in developmental embryos and larvae tissues. Gene Expr. Patterns *8*, 254-260.
51. Miya, K. (1975). Ultrastructural changes of embryonic cells during organogenesis in the silkworm, *Bombyx mori*. I. The Gonad. J. Fac. Agr. Iwate Univ. *12*, 329-338.
52. Tomaya, K. (1902). On the embryology of the silkworm. Bull. Coll. Agriculture, Tokyo *5*, 73-111.
53. Miya, K. (1958). Studies on the embryonic development of the gonad in the silkworm, *Bombyx mori* L. Part I. Differentiation of germ cells. J. Fac. Agr. Iwate Univ. *3*, 436-467.
54. Nakao, H. (1999). Isolation and characterization of a *Bombyx vasa*-like gene. Dev. Genes Evol. *209*, 312-316.
55. Tanaka, M. (1987). Differentiation and behaviour of Primordial Germ Cells during the Early Embryonic Development of *Parnassius glacialis* Butler, *Luehdorfia japonica* Leech and *Byasa (Atrophaneura) alcinous alcinous* Klug (Lepidoptera: Papilionidae). In Recent Advances in Insect Embryology in Japan and Poland, H. Ando and C. Jura, eds. (Tsukuba: Arthropod. Embryol. Soc. Jpn., ISEBU Co. Ltd.), pp. 255-266.
56. Johannsen, O.A. (1929). Some phases in the embryonic development of *Diacrisia virginica* Fabr. (Lepidoptera). J. Morphol. Physiol. *2*, 493-541.

57. Ando, H., and Tanaka, M. (1979). Early embryonic devekopment of the primitive moths, *Enduclyta signifer* Walker and *E. excrescens* Butler (Lepidoptera: Hepialidae). Int. J. Ins. Morph. Emb. *9*, 67-77.
58. Schwangart, F. (1905). Zur Entwickslungsgeschichte der Lepidopteren. Biol. Centralbl. *25*, 777-789.
59. Guelin, M. (1994). [Activity of W-sex heterochromatin and accumulation of the nuage in nurse cells of the lepidopteran *Ephestia*]. C. R. Acad. Sci. Paris. Ser. III *317*, 54-61.
60. Sehl, A. (1931). Furchung und Bildung der Keimanlage bei der Mehlmotte *Ephestia kuehniella*. Zell. Zeit. Morph. U. Okol. *1*, 429-506.
61. Anderson, D.T., and Wood, E.C. (1968). The morphological basis of embryonic movements in the light brown apple moth, *Epiphyas postvittana* (Walk.) (Lepidoptera, Tortricidae). Aust. J. Zool. *16*, 763-793.
62. Woodworth, C.W. (1889). Studies on the embryological development of *Euvanessa antiopa*. In Butterflies of Eastern United States and Canada, Volume 1, Scudder, ed., p. 102.
63. Presser, B.D., and Rutschky, C.W. (1957). The embryonic development of the corn earworm, *Heliothis zea* (Boddie) (Lepidoptea, Phalaenidae). Ann. Entomol. Soc. Am. *50*, 133-164.
64. Kobayashi, Y., and Ando, H. (1984). Mesodermal Organogenesis in the Embryo of the Primitive Moth, *Neomicropteryx nipponensis* Issiki (Lepidoptera, Micropterygidae). J. Morphol. *181*, 29-47.
65. Berg, G.J., and Gassner, G. (1978). Fine structure of the blastoderm embryo of the pink bollworm, *Pectinophora gossypiella* (Saunders) (Lepidoptera: gelechiidae). Int. J. Ins. Morph. Emb. *1*, 81-105.
66. Eastham, L.E.S. (1930). The embryology of *Pieris rapae* - Organogeny. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *219*, 1-50.
67. Lautenschlager, F. (1932). Die Embryonalentwicklung der weiblichen Keimdruse bei der Psychide *Solenobia triquetella*. Zool. Jarh. *56*, 121-162.
68. Inkman, F. (1933). Beitrage zur Entrwicklungsgeschichte des Kornkafers (*Calandra granaria* L.). Zool. Jarh. *56*, 521-557.
69. Tiegs, O.W., and Murray, F.V. (1938). The embryonic development of *Calandra oryzae*. Quart. J. Microscop. Sci. *80*, 159-284.
70. Wray, D.L. (1937). The embryology of *Calendra callosa* Olivier; the southern corn billbug (Coleoptera, Rhynchophoridae). Ann. Entomol. Soc. Am. *30*, 361-409.
71. Hegner, R.W. (1908). Effects of removing the germ-cell determinants from the eggs of some chrysomelid beetles. Preliminary report. Biol. Bull. *16*, 19-26.
72. Paterson, N.F. (1935). Observations on the embryology of *Corynodus pusis* (Coleoptera, Chrysomelidae). Quart. J. Microscop. Sci. *78*, 91-132.
73. Wheeler, W.M. (1889). The embryology of *Blatta germanica* and *Doryphora decemlineata*. J. Morphol. *3*, 291-386.
74. Heider, K. (1889). Die Embryonalentwicklung von *Hydrophilus piceus* L., (Jena: Verlag von G. Fischer).
75. Kobayashi, Y., Suzuki, H., and Ohba, N. (2002). Embryogenesis of the glowworm *Rhagophthalmus ohbai* Wittmer (Insecta: Coleoptera, Rhagophthalmidae), with emphasis on the germ rudiment formation. J. Morphol. *253*, 1-9.

76. Saling, R. (1907). Zur Kenntnis der Entwicklung der Keimdrüsen von *Tenebrio molitor* L. Zeit. f. wiss Zool. *86*, 238-300.
77. Ullmann, S.L. (1964). The origin and structure of the mesoderm and the formation of the coelomic sacs in *Tenebrio molitor* L. (Insecta, Coleoptera). Philos. Trans. R. Soc. Lond. B. Biol. Sci. *248*, 245-277.
78. Zissler, D. (1992). From egg to pole cells: ultrastructural aspects of early cleavage and germ cell determination in insects. Micr. Res. and Tech., 49-74.
79. Schröder, R. (2006). *vasa* mRNA accumulates at the posterior pole during blastoderm formation in the flour beetle *Tribolium castaneum*. Dev Genes Evol *216*, 277-283.
80. von Levetzow, C. (2008). Konservierte und divergente Aspekte der *twist*-, *snail*- und *concertina*-Funktion im Käfer *Tribolium castaneum*. In Mathematisch-Naturwissenschaftlichen Fakultät, Volume PhD. (Köln: Universität zu Köln), p. 100.
81. Handel, K., Grünfeld, C.G., Roth, S., and Sander, K. (2000). *Tribolium* embryogenesis: a SEM study of cell shapes and movements from blastoderm to serosal closure. Dev. Genes Evol., 167-179.
82. Beeman, S.L., and Norris, D.M. (1977). Embryogenesis of *Xyleborus ferrugineus* (Fabr.) (Coleoptera, Scolytidae). I. External morphogenesis of male and female embryos. J. Morphol. *152*, 177-215.
83. Gatenby, J.B. (1920). The Cytoplasmic Inclusions of the Germ Cells. Part VI. On the origin and probable constitution of the germ-cell determinant of *Apanteles glomeratus*, with a note on the secondary nuclei. Q. J. Microsc. Sci. *64*, 133-153.
84. Tawfik, M.F.S. (1957). Alkaline phosphatase in the germ-cell determinant of the egg of *Apanteles*. Journal of Insect Physiology *1*, 286-291.
85. Fleig, R., and Sander, K. (1985). Blastoderm development in honey bee embryogenesis as seen in the scanning electron microscope. Int. J. Inver. Rep. Dev. *8*, 279-286.
86. Gutzeit, H.O., Zissler, D., and Fleig, R. (1993). Oogenesis in the Honeybee *Apis mellifera* - Cytological Observations on the Formation and Differentiation of Previtellogenic Ovarian Follicles. Roux Arch. Dev. Biol. *202*, 181-191.
87. Nelson, J.A. (1915). The embryology of the honey bee., (Princeton: Princeton University Press).
88. Dearden, P.K. (2006). Germ cell development in the Honeybee (*Apis mellifera*); *vasa* and *nanos* expression. BMC Dev. Biol. *6*, 6.
89. Bütschli, O. (1870). Zur Entwicklungsgeschichte der Biene. Z. Wiss. Zool. *20*, 519-564.
90. Fleig, R., and Sander, K. (1986). Embryogenesis of the Honeybee *Apis mellifera* L (Hymenoptera, Apidae) - an Sem Study. Int. J. Ins. Morph. Emb. *15*, 449-462.
91. Hegner, R.W. (1914). Studies on germ cells. III. The origin of the Keimbahn-determinants in a parasitic Hymenopteran, *Copidosoma*. Anat. Anz. *3-4*, 51-69.
92. Zhurov, V., Terzin, T., and Grbic, M. (2004). Early blastomere determines embryo proliferation and caste fate in a polyembryonic wasp. Nature *432*, 764-769.
93. Grbic', M., Nagy, L.M., Carroll, S.B., and Strand, M. (1996). Polyembryonic development: insect pattern formation in a cellularised environment. Development, 795-804.
94. Grbic', M. (2000). "Alien" wasps and evolution of development. BioEssays *22*, 920-932.
95. Strand, M.R., and Grbic', M. (1997). The Development and Evolution of Polyembryonic Insects. Curr. Top. Dev. Biol. *35*, 121-159.

96. Silvestri, F. (1906). Contribuzioni alla conoscenza biologica degli Imenotteri parassiti. I. Biologia del *Litomastix truncellatus* Dalm. Ann. R. Scuola Sup. Agric. Portici *6*, 3-51.
97. Donnell, D.M., Corley, L.S., Chen, G., and Strand, M.R. (2004). Caste determination in a polyembryonic wasp involves inheritance of germ cells. Proc. Natl. Acad. Sci. USA *101*, 10095-10100.
98. Grbic', M. (2003). Polyembryony in parasitic wasps: evolution of a novel mode of development. Int. J. Dev. Biol. *47*, 633-642.
99. Amy, R.L. (1961). The embryology of *Habobracon juglandis* (Ashmead). J. Morphol. *109*, 199-217.
100. Bull, A.L. (1982). Stages of living embryos in the jewel wasp *Mormoniella (Nasonia) vitripennis* (Walker) (Hymenoptera: Pteromalidae). Int. J. Ins. Morph. Emb. *1*, 1-23.
101. Lynch, J.A., Özüak, O., Khila, A., Abouheif, E., Desplan, C., and Roth, S. (2011). The Phylogenetic Origin of *oskar* Coincided with the Origin of Maternally Provisioned Germ Plasm and Pole Cells at the Base of the Holometabola. PLoS Genetics *7*, e1002029.
102. Lynch, J.A., and Desplan, C. (2010). Novel modes of localization and function of *nanos* in the wasp *Nasonia*. Development *137*, 3813-3821.
103. Bronskill, J.F. (1959). Embryology of *Pimpla turionellae* (L.) (Hymenoptera: Ichneumonidae). Can. J. Zool. *37*, 655-688.
104. Fleischmann, V.G. (1975). Origin and embryonic development of fertile gonads with and without pole cells of *Pimpla turionellae* L. (Hymenoptera, Ichneumonidae). Zool. Jb. Anat. Bd. *94*, 375-411.
105. Shafiq, S.A. (1954). A study of the embryonic development of the Gooseberry Sawfly, *Pteronidea ribesii*. Q. J. Microsc. Sci. *95*, 93-114.
106. Gatenby, J.B. (1917). The segregation of the germ-cells in *Trichogramma evanescens*. Q. J. Microsc. Sci. *62*, 149-187.
107. Gatenby, J.B. (1918). The segregation of germ cells in *Trichogramma evanescens*. Quart. J. Microscop. Sci. *63*, 161-173.
108. Koscielska, M.K., and Koscielski, B. (1987). Early embryonic development of *Tritneptis diprionis* (Chalcidoidea, Hymenoptera). In Recent Advances in Insect Embryology in Japan and Poland, H. Ando and C. Jura, eds. (Tsukuba: Arthropod. Embryol. Soc. Jpn. ISEBU Co. Ltd.), pp. 207-214.
109. Lu, H.L., Tanguy, S., Rispe, C., Gauthier, J.P., Walsh, T., Gordon, K., Edwards, O., Tagu, D., Chang, C.C., and Jaubert-Possamai, S. (2011). Expansion of genes encoding piRNA-associated Argonaute proteins in the pea aphid: diversification of expression profiles in different plastic morphs. PLoS ONE *6*, e28051.
110. Miura, T., Braendle, C., Shingleton, A., Sisk, G., Kambhampati, S., and Stern, D.L. (2003). A comparison of parthenogenetic and sexual embryogenesis of the pea aphid *Acyrthosiphon pisum* (Hemiptera : Aphidoidea). J. Exp. Zool. Part B *295B*, 59-81.
111. Chang, C.C., Lee, W.C., Cook, C.E., Lin, G.W., and Chang, T. (2006). Germ-plasm specification and germline development in the parthenogenetic pea aphid *Acyrthosiphon pisum*: Vasa and Nanos as markers. Int. J. Dev. Biol. *50*, 413-421.
112. Chang, C.C., Lin, G.W., Cook, C.E., Horng, S.B., Lee, H.J., and Huang, T.Y. (2007). *Apvasa* marks germ-cell migration in the parthenogenetic pea aphid *Acyrthosiphon pisum* (Hemiptera: Aphidoidea). Dev. Genes Evol. *217*, 275-287.

113. Chang, C.C., Huang, T.Y., Cook, C.E., Lin, G.W., Shih, C.L., and Chen, R.P. (2009). Developmental expression of *Apnanos* during oogenesis and embryogenesis in the parthenogenetic pea aphid *Acyrthosiphon pisum*. Int. J. Dev. Biol. *53*, 169-176.
114. Will, L. (1888). Entwicklungsgescshichte der viviparen Aphiden. Zool. Jarh. *3*, 201-280.
115. Witlaczil, E. (1884). Entwicklungsgeschichte der Aphiden. Zeit. f. wiss Zool. *40*, 559-690.
116. Butt, F.H. (1949). Embryology of the Milkweed Bug, *Oncopeltus fasciatus* (Hemiptera). Cornell Exp. Sta. Mem. *283*, 2-43.
117. Ewen-Campen, B., Jones, T., and Extavour, C. (2013). Evidence against a germ plasm in the milkweed bug *Oncopeltus fasciatus*, a hemimetabolous insect. Biology Open (Company of Biologists) *in press*.
118. Kelly, G.M., and Huebner, E. (1989). Embryonic development of the hemipteran insect *Rhodnius prolixus*. J. Morphol. *199*, 175-196.
119. Seidel, F. (1924). Die Geschlechtsorgane in der Embryonalentwicklung von *Pyrrhocoris apterus*. Zeit. Morph. U. Okol. *1*, 429-506.
120. Mellanby, H. (1935). The early embryonic development of *Rhodnius prolixus* (Hemiptera, Heteroptera). Q. J. Microsc. Sci. *78*, 71-90.
121. Heming, B.S., and Huebner, E. (1994). Development of the Germ Cells and Reproductive Primordia in Male and Female Embryos of *Rhodnius prolixus* Stal (Hemiptera, Reduviidae). Can. J. Zool. *72*, 1100-1119.
122. Heming, B.S. (1979). Origin and Fate of Germ Cells in Male and Female Embryos of *Haplothrips verbasci* (Osborn) (Insecta, Thysanoptera, Phlaeothripidae). J. Morphol. *160*, 323-344.
123. Goss, R.J. (1952). The early embryology of the book louse, *Liposcelis Divergens* Badonnel (Psocoptera; Liposcelidae). J. Morphol. *91*, 135-167.
124. Matsuda, R. (1976). Morphology and evolution of the insect abdomen: with special reference to developmental patterns and their bearings upon systematics, (Oxford: Pergamon Press).
125. Heymons, R. (1895). Die Embryonalentwicklung von Dermapteren und Orthopteren, (Jena: Verlag von G. Fischer).
126. Miller, A. (1939). The egg and early development of the stonefly, *Pteronarcys proteus* Newman (Plecoptera). J. Morphol. *64*, 555-608.
127. Cavallin, M. (1971). La "polyembryonie substitutive" et le probleme de l'origine de la lignée germinale chez le Phasme *Carausius morosus* Br. C. R. Acad. Sci. Paris. Ser. III *272*, 462-465.
128. Cavallin, M., and Hajji, K. (1979). Marqueurs endogènes des gonocytes primordiaux et ségrégation germinale chez l'embryon de *Clitumnus extradenticus* Br. (Phasmida: Lonchodidae). Int. J. Ins. Morphol. Embryol. *8*, 85-94.
129. Louvet, J.P. (1963). Contribution a l'étude histologique et histochimique de l'embryogenèse précoce chez le Phasme *Carausius morosus* Br.: ségrégation du mésoderme et segmentation de l'embryon. Ph.D. Thesis *Bordeaux*.
130. Cavallin, M. (1976). La ségrégation de la lignée germinale chez le Phasme Carausius morosus Br. Bull. Soc. Zool. Fr. Evol. Zool. *Supplement 4*, 15-.
131. Roonwal, M.L. (1937). Studies on the Embryology of the African Migratory Locust, *Locusta migratoria migratoroides* Reiche and Frm. (Orthoptera, Acrididae). II. Organogeny. Phil. Trans. R. Soc. Lond. B. *227*, 175-244.

132. Nelsen, O.E. (1934). The segregation of the germ cells in the grasshopper, *Melanoplus differentialis* (Acrididae; Orthoptera). J. Morphol., 545-575.
133. Wheeler, W.M. (1893). A contribution to Insect Morphology. J. Morphol. *8*, 1-160.
134. Woodland, J.T. (1957). A contribution to our knowledge of Lepismatid development. J. Morphol. *101*, 523-577.
135. Heymons, R. (1897). Entwicklungsgeschichtliche Untersuchungen an Lepisma Saccharina L. Z. Wiss. Zool. *62*, 583-631.
136. Klag, J. (1977). Differentiation of primordial germ cells in the embryonic development of *Thermobia domestica*, Pack. (Thysanura): an ultrastructural study. J. Emb. Exp. Morph. *38*, 93-114.
137. Machida, R., Nagashima, T., and Ando, H. (1990). The Early Embryonic Development of the Jumping Bristletail Pedetontus-Unimaculatus Machida (Hexapoda, Microcoryphia, Machilidae). J. Morphol. *206*, 181-195.
138. Larink, O. (1983). Embryonic and postembryonic development of Machilidae and Lepismatidae (Insecta: Archaeognatha et Zygentomata). Entomol. Gen. *8*, 119-133.
139. Kainz, F., Ewen-Campen, B., Akam, M., and Extavour, C.G. (2011). Delta/Notch signalling is not required for segment generation in the basally branching insect *Gryllus bimaculatus*. Development *138*, 5015-5026.
140. Mito, T., and Noji, S. (2009). The Two-spotted Cricket *Gryllus bimaculatus*: An Emerging Model for Developmental and Regeneration Studies. In Emerging Model Organisms: A Laboratory Manual, Volume 1. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 331-346.
141. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688-2690.
142. Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. *57*, 758-771.
143. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.
144. Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. *56*, 564-577.
145. Chang, C., Dearden, P., and Akam, M. (2002). Germ Line Development in the Grasshopper *Schistocerca gregaria*: *vasa* As a Marker. Dev. Biol. *252*, 100-118.
146. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Supplementary Table 1

Insect Order	Species[1]	n.d.	Presumptive PGC Origin: Reported Location/Derivation[6]	Mode of PGC Specification[2]	PGC Identification Criteria[3]	Functional or Experimental Evidence[4]	References
Diptera							
	Acricotopus lucidus	Blastoderm	Pole cell[7]	GP	LM, LSM	N	[6]
	Aedes aegypti	Blastoderm	Pole cells	GP	LM, MM	N	[7, 8]
	Anopheles gambiae	Blastoderm	Pole cells	GP	MM	N	[8]
	Anopheles maculipennis	Blastoderm	Pole cells	GP	LM	N	[7]
	Calliphora crythrocephala	Blastoderm	Pole cells	GP	LM	N	[9]
	Calliphora erythrocephala	Blastoderm	Pole cells	GP	LM	N	[10]
	Ceratitis capitata	Blastoderm	Pole cells	GP	LM, TEM, LSM	N	[11]
	Chironomous sp.	Blastoderm	Pole cells	GP	LM, MM	N	[12, 13]
	Clogmia albipunctata	nd[8]	nd	nd	LM	N	[14]
	Compisura concinnata	Blastoderm	Pole cells	GP	LM	N	[15]
	Culex fatigans	Blastoderm	Pole cells	GP	LM	N	[16]
	Culex pipiens	Blastoderm	Pole cells	GP	LM	Y	[7, 17]
	Culex quinquefasciatus	Blastoderm	Pole cells	GP	MM	N	[18]
	Dacus tyroni	Blastoderm	Pole cells	GP	LM	N	[19]
	Drosophila melanogaster	Blastoderm	Pole cells	GP	LM	Y	[20-29]
	Heteropeza pygmaea	Blastoderm	Pole cells	GP	LM	N	[30]
	Lucilia cuprina	Blastoderm	Pole cells	GP	LM	Y	[24]
	Lucilia sericata	Blastoderm	Pole cells	GP	LM	N	[16]
	Melophagus ovirus	Blastoderm	Pole cells	GP	LM	N	[31]
	Miastor americana	Blastoderm	Pole cell	GP	LM	N	[32]
	Miastor metraloas	Blastoderm	Pole cell	GP	LM	N	[33, 34]
	Musca domestica	Blastoderm	Pole cells	GP	LM, MM	N	[13, 35]
	Musca vomitora	Blastoderm	Pole cells	GP	LM	N	[36, 37]
	Mycophila speyeri	Blastoderm	Pole cells	GP	LM	N	[38]
	Phoenicia sericata	Blastoderm	Pole cells	GP	LM	N	[39]
	Phormia regina	Blastoderm	Pole cells	GP	LM	N	[40]
	Rhynchosciara americana	Blastoderm	Pole cells	GP	MM	N	[41]
	Sciara coprophila	Blastoderm	Pole cells	GP	LM	N	[42]
	Simulium pictipes	Blastoderm	Pole cells	GP	LM	N	[43]
	Smittia sp.	Blastoderm	Pole cells	GP	LM	N	[44]
Siphonaptera							
	Ctenocephalides felis	Blastoderm	Pole cells	GP	LM	N	[45]
	Hystrichopsylla dippie	Blastoderm	Pole cells	GP	LM	N	[45]
	Neosopsyllus fasciatus	Blastoderm	Pole cells	GP	LM	N	[45]
Lepidoptera							
	Antheraea pernyi	Germ band	CP Mesoderm	MZ	LM	N	[46]
	Bombyx mori	Blastoderm	Lateral ventral blastoderm	GP	LM, MM	Y	[47-54]
	Byasa (Atrophaneura) alcinous alcinous	Germ rudiment	Ventral germ rudiment	MZ	LM	N	[55]
	Diacrisia virginica	Germ band	CP Mesoderm	MZ	L	N	[56]
	Endoclyta execrescens	Germ band	PGB	MZ	LM	N	[57]
	Endoclyta signifer	Germ band	PGB	MZ	LM	N	[57]
	Endromis versicolora	Germ band	PGB primary ectoderm	MZ	LM	N	[58]
	Ephestia kuehniella	Germ band	PGB primary ectoderm	MZ	LM, MM	N	[59, 60]
	Epiphyas postvittana	Germ band	Germ band midline	MZ	LM	N	[61]
	Euvanessa antiopa	Blastoderm	Ventral cellular blastoderm	MZ	LM	N	[62]
	Heliothis zea	Germ band	Germ band midline	MZ	LM	N	[63]
	Luehdorfia japonica	Germ rudiment	Ventral germ rudiment	MZ	LM	N	[55]
	Neomicropteryx nipponensis	Germ band	CP Mesoderm	MZ	LM	N	[64]
	Parnassius glacialis	Germ rudiment	Ventral germ rudiment	MZ	LM	N	[55]
	Pectinophora gossypiella	Blastoderm	Pole cells	GP	LM, TEM	N	[65]
	Pieris rapae	Germ band	Mesoderm	MZ	LM	N	[66]
	Solenobia triquetrella	Germ band	PGB primary ectoderm	MZ	LM	N	[67]
Coleoptera							
	Calandra granaria	Blastoderm	Pole cells	GP	LM	N	[68]
	Calandra oryzae	Blastoderm	Pole cells	GP	LM	N	[69]
	Calendra callosa	Blastoderm	Pole cells	GP	LM	N	[70]
	Calligrapha multipunctata	Blastoderm	Pole cells	GP	LM	N	[71]
	Corynodes pusis	Blastoderm	Posterior blastoderm	GP	LM	N	[72]
	Doryphora decemlineata	Germ band	PGB	MZ	LM	N	[73]
	Hydrophilus piceus	Germ band	CP mesoderm	MZ	LM	N	[74]
	Leptinotarsa decemlineata	Blastoderm	Pole cells	GP	LM	N	[71]
	Rhagophthalmus ohbai	Germ disc	Posterior germ disc	MZ	LM, SEM	N	[75]
	Tenebrio molitor	Germ band	PGB mesoderm	MZ	LM	N	[76, 77]
	Tribolium castaneum	Germ band	PGB	MZ	TEM, LM, SEM, MM	N	[78-81]
	Xyleborus terrugineus	Blastoderm	Pole cells	GP	LM	N	[82]
Hymenoptera							
	Apanteles glomeratus	Blastoderm	Pole cells	GP	LM	N	[83, 84]
	Apis mellifera	Germ band	Mesoderm	MZ	SEM, MM	N	[78, 85-90]
	Copidosoma sp.	2nd cleavage	B4 at four cell stage[9]	GP	LM, MM	Y	[91-98]
	Habobracon juglandis	Blastoderm	Pole cells	GP	LM	N	[99]
	Mormoniella (Nasonia) vitripennis	Blastoderm	Pole cells	GP	LM	N	[100-102]
	Pimpla turionellae	Blastoderm	Pole cells	GP	LM	"+"	[103, 104]
	Pteronidea ribesii	Germ band	PGB	MZ	LM	N	[105]
	Trichogramma evanescens	Blastoderm	Pole cells	GP	LM	N	[106, 107]
	Tritneptis dipronis	Blastoderm	Pole cells	GP	LM	N	[108]
Hemiptera							
	Acyrthosiphon pisum	Blastoderm	Pole cells	GP	LM, MM	N	[109-113]

Supplementary Table 1 (Continued)

	Aphis pelargonii	Blastoderm	PGB	MZ	LM	N	[114]
	Aphis plantoides	Blastoderm	PGB	MZ	LM	N	[115]
	Aphis rosea	Blastoderm	PGB	MZ	LM	N	[33]
	Oncopeltus fasciatus	Blastoderm	PGB	MZ	LM, TEM, MM	Y	[116-118]
	Pyrrhocoris apterus	Blastoderm	PGB	MZ	LM	N	[119]
	Rhodnius prolixus	Blastoderm	PGB	GP	LM	N	[120, 121]
Thysanoptera							
	Haplothrips verbasci	Blastoderm	Pole cells	GP	LM	N	[122]
Psocoptera							
	Liposcelis divergens	Blastoderm	PV blastoderm	MZ	LM	N	[123]
Phthiraptera							
		Blastoderm	Pole cells	GP	LM	N	[cited in 122, 124]
Blattodea							
	Periplaneta orientalis	Germ band	Mesoderm	MZ	LM	N	[125]
	Phyllodromia (Blatta) germanica	Germ band	Mesoderm	MZ	LM	N	[125]
Plecoptera							
	Pteronarcys proteus	Germ band	Mesoderm	MZ	LM	N	[126]
Phasmatodea							
	Carausius morosus	Germ band	Mesoderm	MZ	LM, TEM	Y	[127-130]
Orthoptera							
	Gryllus bimaculatus	Germ band	Mesoderm	MZ	LM, MM	Y	This study
	Gryllus campestris	Germ band	PGB	MZ	LM	N	[125]
	Gryllus domesticus	Germ band	PGB	MZ	LM	N	[125]
	Locusta migratoria	Germ band	CP Mesoderm	MZ	LM	N	[131]
	Melanoplus differentialis	Germ band	Lateral abdominal ectoderm	MZ	LM	N	[132]
	Xiphidium enisterum	Germ band	Mesoderm	MZ	LM	N	[133]
Zygentoma							
	Ctenolepisma lineata	Germ band	CP Mesoderm	MZ	LM	N	[134]
	Lepisma saccharina	Germ band	PGB	MZ	LM	N	[135]
	Thermobia domestica	Germ band	CP Mesoderm	MZ	LM, TEM	N	[134, 136]
Archaeognatha							
	Pedetontus unimaculatus	Blastoderm	PGB	MZ	LM	N	[137]
	Petrobius brevistylis	Germ band	CP Mesoderm	MZ	TEM	N	[138]

[1] Species name is shown as reported in the primary data reference, listed alphabetically within an insect order.

[2] MZ = mesodermal/zygotic origin; GP = germ plasm/pole cells

[3] LM = light microscopic histological analysis, of either whole mounts or sections; TEM = transmission electron microscopy; SEM = scanning electron microscopy; EM = enzymatic markers; MM = molecular markers; LI = cell lineage studies.

[4] Y = yes; N = no. Includes data derived from functional genetic evidence and/or physical perturbation/ablation.

[5] Since direct comparison of the duration of stages of development in different species is not appropriate due to differences in culture conditions, we describe here developmental stages rather than absolute time.

[6] Pole cell(s) = Drosophila-like PGC formation at the blastoderm posterior; CP = coelomic pouch; PGB = posterior germ band (germ layer indicated if specified by author of primary reference); PV = posterior ventral

[7] In embryos of Acricotopus lucidus, Miastor americana and Miastor metraloas, a single pole cell (PGC) gives rise to all germ line cells.

[8] nd = no data; for Clogmia albipunctata pole cells have been specifically noted as absent, but no hypotheses on PGC origin have been proposed in the literature.

[9] Copidosoma floridanum is unusual among insects in that it displays holoblastic cleavage rather than syncytial cleavage in early development. B4 is the name given to the small blastomere formed at second cleavage, which inherits the oosome (germ plasm) and is the primordial germ cell.

Evidence against a germ plasm in the milkweed bug *Oncopeltus fasciatus*, a hemimetabolous insect

Ben Ewen-Campen*, Tamsin E. M. Jones* and Cassandra G. Extavour‡

Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

*These authors contributed equally to this work

‡Author for correspondence (extavour@oeb.harvard.edu)

Biology Open 0, 1–13
doi: 10.1242/bio.20134390
Received 11th February 2013
Accepted 1st March 2013

Summary

Primordial germ cell (PGC) formation in holometabolous insects like *Drosophila melanogaster* relies on maternally synthesised germ cell determinants that are asymmetrically localised to the oocyte posterior cortex. Embryonic nuclei that inherit this "germ plasm" acquire PGC fate. In contrast, historical studies of basally branching insects (Hemimetabola) suggest that a maternal requirement for germ line genes in PGC specification may be a derived character confined principally to Holometabola. However, there have been remarkably few investigations of germ line gene expression and function in hemimetabolous insects. Here we characterise PGC formation in the milkweed bug *Oncopeltus fasciatus*, a member of the sister group to Holometabola, thus providing an important evolutionary comparison to members of this clade. We examine the transcript distribution of orthologues of 19 *Drosophila* germ cell and/or germ plasm marker genes, and show that none of them localise asymmetrically within *Oncopeltus* oocytes or early embryos. Using multiple molecular and cytological criteria, we provide evidence that PGCs form after cellularisation at the site of gastrulation. Functional studies of *vasa* and *tudor* reveal that these genes are not required for germ cell formation, but that *vasa* is required in adult males for spermatogenesis. Taken together, our results provide evidence that *Oncopeltus* germ cells may form in the absence of germ plasm, consistent with the hypothesis that germ plasm is a derived strategy of germ cell specification in insects.

Key words: *vasa*, *tudor*, *boule*, Germ line, RNA interference, Spermatogenesis

Introduction

In sexually reproducing animals, only germ cells contribute genetic information to future generations. The germ line/soma separation is a cell fate decision shared across Metazoa (Buss, 1987). Despite the fundamental commonality of germ cell function in animals, the molecular mechanisms underlying germ cell specification are remarkably diverse across different taxa (Extavour and Akam, 2003; Extavour, 2007; Ewen-Campen et al., 2010; Juliano et al., 2010).

Primordial germ cells (PGCs) can be specified via different developmental mechanisms; here we call these "cytoplasmic inheritance" and "zygotic induction." (We and others have previously referred to these mechanisms as "preformation" and "epigenesis" respectively (Nieuwkoop and Sutasurya, 1981; Extavour and Akam, 2003; Extavour, 2007). However, these terms can hold different meanings in other contexts of the history and philosophy of biology (e.g. Callebaut, 2008). We therefore avoid them here in favour of more mechanistically descriptive terms.) Cytoplasmic inheritance is characterised by the asymmetric formation of a specialised cytoplasmic region within the oocyte or early embryo, termed "germ plasm." Germ plasm contains maternally provided mRNAs and proteins that are individually necessary and collectively sufficient for PGC formation. Cells that inherit germ plasm during embryogenesis acquire germ line fate. The best understood example of cytoplasmic inheritance occurs in *Drosophila melanogaster*, where germ plasm is maternally synthesised, localised to the posterior of the oocyte during oogenesis, and subsequently incorporated into PGCs (pole cells) during cellularisation. Removing pole cells after their formation, or compromising the molecular components of germ plasm, leads to loss of PGCs and sterility in adulthood (reviewed by Mahowald, 2001). In contrast, zygotic induction of PGCs takes place later in development and requires signalling from neighbouring somatic cells to induce germ line fate. This mode of PGC development is exemplified by *Mus musculus*, wherein PGCs develop from a subset of presumptive mesodermal cells after the segregation of embryonic and extraembryonic tissues in response to local signalling (reviewed by Magnúsdóttir et al., 2012).

Across Insecta, germ plasm has been almost exclusively reported in taxa nested within Holometabola ("higher" insects, which undergo complete metamorphosis) including *D. melanogaster* (reviewed by Kumé and Dan, 1968; Anderson, 1973; Nieuwkoop and Sutasurya, 1981), and in only three species belonging to the sister assemblage to the Holometabola (see below). Thus, although the vast majority of our knowledge of insect germ cell development comes from studies of germ plasm in *D. melanogaster*, this mode of germ cell specification is likely a derived feature of Holometabolous insects and their close sister taxa.

Our present knowledge of PGC specification in basally branching insects (Hemimetabola) is based almost entirely on classical histological studies of insect development conducted over the past 150 years. Nearly all of these report that PGCs arise late in embryogenesis, raising the possibility that they may be specified through inductive mechanisms (Wheeler, 1893; Heymons, 1895; Hegner, 1914; Nelsen, 1934; Roonwal, 1937). Experimental approaches to discovering germ plasm in Hemimetabola are limited, but a study involving destruction of the germ rudiment via irradiation in the cricket *Gryllus domesticus* (Schwalm, 1965) showed that no specific region of early embryos in this species contains a germ line determinant. Functional tests of genes that may specify germ cells in Hemimetabola have been performed in only one insect, the cricket *Gryllus bimaculatus*. In this cricket, the conserved germ line markers *vasa* and *piwi* are dispensable maternally and zygotically for PGC formation (Ewen-Campen et al., 2013). Most evidence available for the Hemimetabola therefore suggests the absence of germ plasm and the operation of zygotic PGC specification mechanisms.

Exceptions have been reported, however, in some members of the Paraneoptera, an assemblage of insect orders (including Hemiptera [true bugs], Psocoptera [book lice], and Thysanoptera [thrips]) that collectively form the sister group to Holometabola (Yeates et al., 2012). Cytological studies of three paraneopteran species, a book louse (Psocoptera (Goss, 1952)), a thrip (Thysanoptera (Heming, 1979)) and an aphid (Hemiptera (Chang et al., 2009)) suggested the presence of germ plasm in oocytes or early embryos, as did expression studies of *vasa*, *piwi* and *nanos* expression during asexual development of the pea aphid *Acyrthosiphon pisum* (Chang et al., 2006; Chang et al., 2007; Chang et al., 2009; Lu et al., 2011). However, *A. pisum* embryogenesis is highly modified relative to that of other hemimetabolous insects and even relative to other members of the same order (Miura et al., 2003). Studies of embryogenesis in most other hemipterans describe absence of germ plasm and PGC origin after cellularisation from the blastopore region at gastrulation stages (Metschnikoff, 1866; Witlaczil, 1884; Will, 1888; Seidel, 1924; Mellanby, 1935; Butt, 1949; Sander, 1956; Kelly and Huebner, 1989; Heming and Huebner, 1994). We therefore wished to examine the expression and function of germ line genes in a hemipteran displaying embryological characteristics more representative of the order.

Here we characterise germ cell formation and migration in the milkweed bug *Oncopeltus fasciatus* (Hemiptera). We examine the expression of 19 molecular markers including *vasa*, *nanos*, and *piwi*, and test the germ cell function of three of these using RNA interference. We show that in striking contrast to *Drosophila*, transcripts of none of these genes localise asymmetrically within *Oncopeltus* oocytes or early embryos. We identify PGCs using multiple criteria, and show that neither *vasa* nor *tudor* are required for PGC specification or oogenesis in this species, but that *vasa* is required for spermatogenesis in adult males. These data show that the PGC specification role of *vasa* has diverged between *Oncopeltus* and the Holometabola, and suggest that *Oncopeltus* PGCs may form in the absence of maternally supplied germ plasm.

Results

Putative germ cells are first detectable in the late blastoderm stage

In contrast to *D. melanogaster*, classical studies of *Oncopeltus fasciatus* embryogenesis have not revealed a germ plasm in oocytes or early embryos, and instead first identify cells with cytological characteristics of PGCs at the posterior of the embryo at the end of the cellular blastoderm stage (Butt, 1949). We used semi-thin plastic sectioning and fluorescence microscopy to confirm these observations, and traced the development of these putative PGCs throughout gastrulation and germ band elongation (supplementary material Fig. S1). Our observations of putative PGC formation in *Oncopeltus* were consistent with historical studies (Butt, 1949), showing that these cells first arise at the blastoderm posterior immediately prior to gastrulation (supplementary material Fig. S1). Unlike pole cells in *D. melanogaster*, presumptive PGCs in *Oncopeltus* arise on the basal side of the blastoderm surface, adjacent to the yolk (supplementary material Fig. S1G–H). In order to obtain further evidence that these cells were PGCs and test for the presence of a maternally supplied germ plasm, we examined the expression of conserved germ line markers.

Cloning *Oncopeltus* germ line markers

We cloned fragments of *vasa*, *nanos*, and *piwi* (Ewen-Campen et al., 2010) and confirmed that each was the best reciprocal BLAST hit to its respective orthologue in *D. melanogaster*. *vasa* was cloned using degenerate primers (supplementary material Table S1). *nanos* and a single *piwi* gene were recovered from the *Oncopeltus* transcriptome, in addition to a single AGO-3 orthologue (an additional PIWI family protein belonging to a separate sub-family; not shown). We believe it is unlikely that *Oncopeltus* possesses additional orthologues of these genes because (1) the *Oncopeltus* ovarian and embryonic transcriptome, which is nearly saturated for gene discovery and has an average coverage of 23× (Ewen-Campen et al., 2011), contained only one orthologue of each gene; and (2) degenerate PCR for *vasa* using primers flanking the conserved DEAD box helicase domain (Rocak and Linder, 2004) recovered only a single *vasa* orthologue.

Phylogenetic reconstruction confirmed that *Oncopeltus vasa* is nested within other insect *vasa* genes (supplementary material Fig. S2A), and that *Oncopeltus piwi* belongs to the PIWI sub-family containing the *Drosophila* genes *piwi* and *aubergine* (which are *Drosophila*-specific duplications) (supplementary material Fig. S2B). The portion of animal Nanos proteins with conservation sufficient for confident alignment (48 amino acids) is too short to yield significant phylogenetic signal (supplementary material Fig. S2C, note low support values), but *Oncopeltus* Nanos does contain the diagnostic 2×(CCHC) zinc finger domain found in all Nanos orthologs (supplementary material Fig. S2D).

Our analysis of the *Oncopeltus nanos* sequence produced an unexpected result: we found that a stop codon is present 771 bp upstream of the first CCHC zinc finger domain, although no methionine is found anywhere in this region. This is unlikely to be a sequencing error, as it was identified with high coverage (22 reads/bp at this position) in the transcriptome (Ewen-Campen et al., 2011) and confirmed using Sanger sequencing of independent clones generated from a different cDNA pool than that used to generate the transcriptome. Furthermore, repeated attempts at 5′ RACE using a third independent DNA pool failed to amplify a start codon. Several lines of evidence confirm that this *Oncopeltus nanos* sequence represents a highly expressed mRNA and is therefore unlikely to be a pseudogene: it was recovered from a transcriptome made solely from poly(A)-RNA, and is detected via both RT-PCR (Ewen-Campen et al., 2011) and in situ hybridisation (see below). We hypothesise that a large, unspliced intron downstream of the start codon may have been

present in our mRNA preparations. Alternatively, given that the length of the predicted translated region upstream of the first CCHC zinc finger domain (266 amino acids) is within the range of known arthropod Nanos orthologues (95 to 332 amino acids) (Wang and Lehmann, 1991; Curtis et al., 1995; Calvo et al., 2005; Lynch and Desplan, 2010), it may be that the *Oncopeltus* Nanos N terminus has a non-methionine start codon. Although rare, eukaryotic non-AUG translation initiation can occur in nuclear-encoded genes, including developmentally relevant genes (Hellen and Sarnow, 2001), and can be recognized by insect ribosomes (Sasaki and Nakashima, 2000; Jan et al., 2001). In the absence of a complete genome sequence we cannot distinguish between these hypotheses. Despite this uncertainty, we report *nanos* transcript expression here for the sake of completeness.

vasa, *nanos*, and *piwi* transcripts do not localise asymmetrically in ovaries

The distinct cytoplasm inherited by early-specified PGCs in multiple organisms, including *D. melanogaster*, *Caenorhabditis elegans*, *Xenopus laevis* and *Danio rerio*, contains transcripts of the highly conserved *piwi*, *vasa* and *nanos* gene families. Of these, only *nanos* mRNA is asymmetrically localised to *D. melanogaster* germ plasm, while *piwi* and *vasa* transcripts are ubiquitous throughout the fly oocyte and embryo. However, in several other organisms *vasa* orthologue transcripts are asymmetrically localised germ plasm components (reviewed by Ewen-Campen et al., 2010).

To test whether any of these transcripts were asymmetrically localised to putative germ plasm in *Oncopeltus* oocytes, we conducted in situ hybridisation on adult ovaries. The structure of *Oncopeltus* ovaries is typical of Hemiptera and several other insect orders but differs remarkably from that of *Drosophila* (Fig. 1A) (Büning, 1994). Rather than each oocyte developing together with its own complement of 15 nurse cells as in *Drosophila*, all oocytes in *Oncopeltus* ovarioles share a common pool of syncytial nurse cells located at the anterior of each ovariole in a region termed the “tropharium” (Fig. 1A1). The nurse cell syncytium connects to all oocytes via elongated, microtubule-rich tubes called “nutritive tubes” (Hyams and Stebbings, 1979; Harrison et al., 1991) through which maternal factors, including mRNA, proteins and mitochondria, are transported to developing oocytes (Fig. 1A2,A3) (Stebbings et al., 1985; Stebbings and Hunt, 1987; Anastas et al., 1991; Hurst et al., 1999; Stephen et al., 1999).

vasa, nanos, and *piwi* were expressed at high levels in *Oncopeltus* nurse cells and oocytes of all stages, but at no stage of oogenesis did any of these three transcripts localise asymmetrically within oocytes (Fig. 1B–D). Expression was detected in nurse cells, resting oocytes, nutritive tubes, and developing oocytes, suggesting that these transcripts are synthesised in the nurse cells and subsequently transported to oocytes via nutritive tubes (Fig. 1B–D). *nanos* and *piwi* were expressed throughout the tropharium (Fig. 1C,D), in contrast to *vasa*, whose expression was primarily in nurse cells of the posterior tropharium, resting and developing oocytes (Fig. 1B). In late stage oocytes, expression remained ubiquitous (not shown), similar to the expression in just-laid eggs (see below).

In situ screen of conserved *Drosophila* germ plasm markers fails to reveal a germ plasm in *Oncopeltus*

The expression of *piwi*, *vasa* and *nanos* suggests that a maternally localised germ plasm containing transcripts of these genes is not

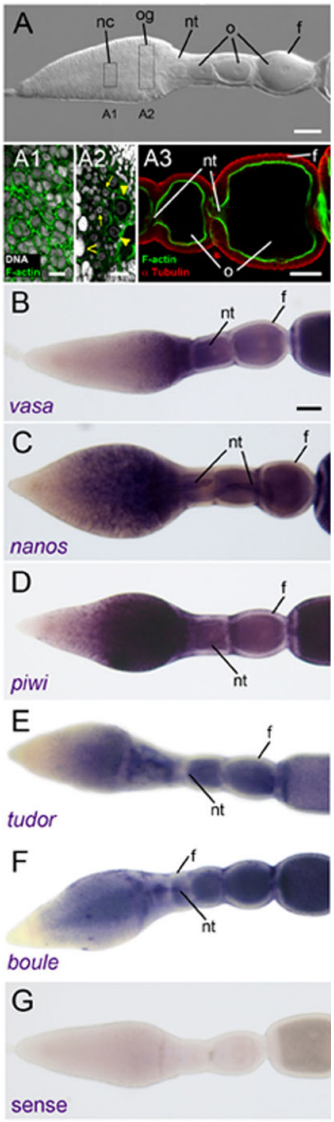


Fig. 1. Germ cell markers do not localise asymmetrically during oogenesis. (**A**) Overview of a single *Oncopeltus* ovariole. *nc*: nurse cells; *o*: oocytes; *nt*: nutritive tubes; *f*: follicle cells. Boxed regions are enlarged in (**A1–A3**). (A1) Nurse cell syncytium containing polyploid nurse cell nuclei (white) connected by cytoplasmic bridges (green). (A2) Posterior tropharium containing oogonia (arrows) and resting oocytes (arrowheads). Caret indicates polyploid nurse cells in the anterior of this region. (A3) Nutritive tubes (*nt*) are actin-rich at the end that enters the anterior of each oocyte. Transcripts of *vasa* (**B**), *nanos* (**C**), *piwi* (**D**), *tudor* (**E**) and *boule* (**F**) are detected in nurse cells, nutritive tubes, and uniformly in oocytes. (**G**) A representative sense control (for *vasa*) is shown; sense controls for other genes were similar. Scale bars: 100 µm in A,A3,B (applies to C–G); 25 µm in A1,A2. Anterior to the left in all panels.

present in *Oncopeltus* oocytes. However, a functional germ plasm that contains gene products other than those encoded by these three genes could be present in oocytes or early embryos. To

explore this possibility, we examined the expression of 14 additional genes whose transcripts are enriched in the germ plasm and germ cells of *Drosophila* (supplementary material Table S2) (Tomancak et al., 2002; Lécuyer et al., 2007; Tomancak et al., 2007) that were also recovered from the *Oncopeltus* ovarian and embryonic transcriptome (Ewen-Campen et al., 2011) based on best reciprocal BLAST hit analysis with the *Drosophila* proteome (Zeng and Extavour, 2012). Although several of these genes do not have documented mutant phenotypes for germ cell formation in *Drosophila* (supplementary material Table S2), all are expressed at high levels in germ plasm and/or pole cells and are therefore molecular markers for germ plasm in *Drosophila*. We reasoned that if *Oncopeltus* possessed germ plasm it would likely be revealed by at least one of these genes.

In addition, we examined the expression of *boule* and *tudor*, which have widely conserved functions in germ cells across Metazoa (Eberhart et al., 1996; Ewen-Campen et al., 2010; Shah et al., 2010). *tudor* is one of 23 Tudor domain-containing proteins in *Drosophila* (Ying and Chen, 2012), but there is no evidence that loss of function of other Tudor domain-containing genes have grandchildless phenotypes in *Drosophila* (Handler et al., 2011; Pek et al., 2012). We therefore focus only on the expression and function of the orthologue of *Drosophila tudor* (CG9450). We examined *boule* and *tudor* transcript expression throughout oogenesis and embryogenesis through mid-germ band stages.

None of these 16 transcripts localised asymmetrically in ovaries (Fig. 1E,F; supplementary material Fig. S3). Instead, like *vasa*, *piwi* and *nanos* (Fig. 1B–D), all of these genes were expressed ubiquitously throughout oogenesis. Half of the genes examined (*sra*, *CycB*, *Bsg25D*, *Uev1A*, *CG16817*, *Unr*, *mael* and *tud*) were expressed, like *vasa* (Fig. 1B), in nurse cells adjacent to resting oocytes, as well as in the resting and early oocytes themselves (Fig. 2E; supplementary material Fig. S3B–H). Five genes (*Gap1*, *eIF5*, *bel*, *orb* and *bol*) were, like *piwi* and *nanos* (Fig. 1C,D), strongly expressed in all nurse cells of the

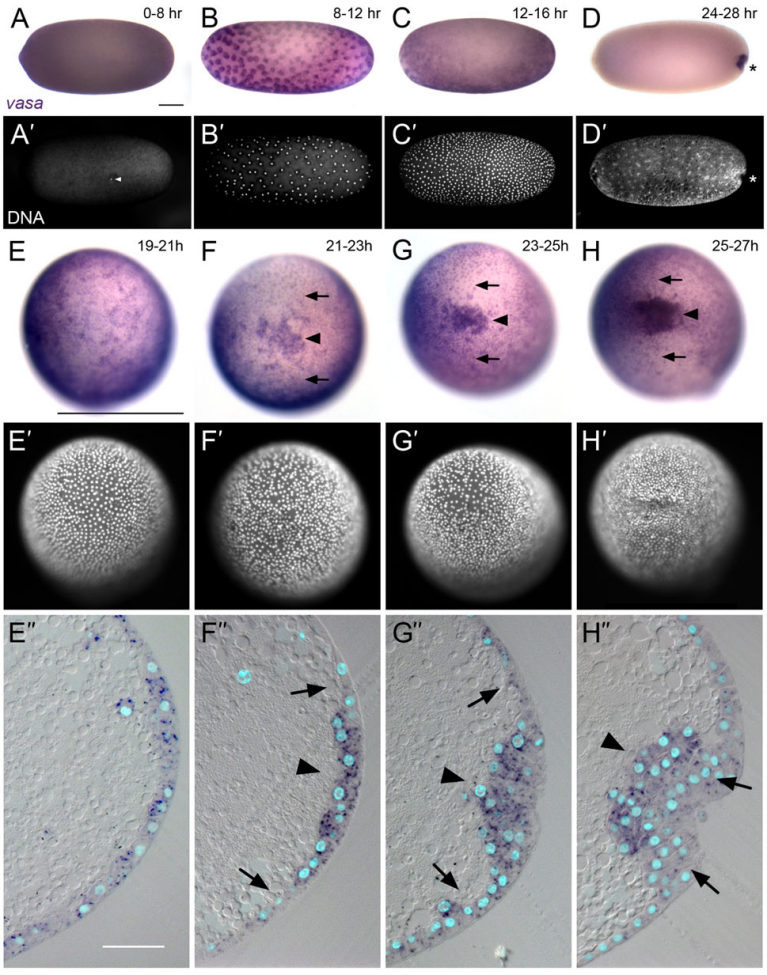


Fig. 2. *vasa* transcript expression first labels PGCs at late blastoderm stages. (**A–D**) *vasa* transcript expression. (**A′–D′**) Corresponding images of nuclear stains. (A,A′) Immediately following fertilisation *vasa* is detected ubiquitously. Arrowhead: polar body. (B,B′) In early cleavage stages *vasa* transcripts are associated with all energid nuclei. (C,C′) During early syncytial blastoderm stages, *vasa* expression remains ubiquitous. (D,D′) At cellular blastoderm stages (24–28 h AEL), *vasa* marks putative PGCs at the posterior pit (asterisk). (**E–H**) End-on perspective of the posterior of *Oncopeltus* embryos showing *vasa* expression during PGC formation. (**E′–H′**) Corresponding images of nuclear stains. (**E″–H″**) Medial sections of *vasa*- (purple) and nuclear- (cyan) stained embryos at corresponding time points. (E–E″) In late syncytial blastoderm stages, *vasa* is expressed ubiquitously. (F–F″) In early cellular blastoderm embryos, *vasa* expression increases in some posterior cells (arrowheads in F,F″) while levels in the remainder of the blastoderm decrease (arrows in F,F″). (G–G″) At posterior germ band invagination *vasa*-positive cells (arrowheads) are the first cells to enter the yolk; *vasa* transcripts continue to be cleared from somatic tissue (arrows). (H–H″) As invagination proceeds *vasa* expression is largely restricted to PGCs (arrowhead) and nearly cleared from the soma (arrows). Scale bars: 100 µm in A (applies to B–D,A′–D′); 500 µm in E (applies also to F–H′); 50 µm in E″ (applies also to F″–H″). Anterior is to the left in A–D′ and E″–H″.

tropharium (Fig. 2F; supplementary material Fig. S3I–L). Two genes (*cta* and *Tao*) were expressed in resting and early oocytes but barely at all in the tropharium (supplementary material Fig. S3M,N), suggesting that these genes may be transcribed by resting oocyte nuclei rather than by nurse cells. Finally, *aret* (aka *bruno*), which is a translational regulator of Oskar in *Drosophila* (Kim-Ha et al., 1995; Webster et al., 1997), was expressed in nurse cells of the posterior tropharium and in early stages of oogenesis but excluded from resting oocytes (supplementary material Fig. S3O), suggesting that it is transcribed by oocyte nuclei after the onset of oogenesis. In summary, although transcripts of most of these genes are likely to be supplied maternally to oocytes, they are not asymmetrically localised within oocytes of any stage.

vasa, *boule* and *tudor* transcripts mark PGCs throughout embryogenesis but are not asymmetrically localised in early embryos

Although none of the genes examined showed asymmetric localisation during oogenesis or early embryogenesis, at late blastoderm stages many of the genes appeared enriched at the posterior pit, where PGCs had previously been identified based on cytological criteria (supplementary material Fig. S1) (Butt, 1949). However, because at this stage of development gastrulation begins at the posterior, this region of the blastoderm is multilayered. Upon close examination, we found that the apparent transcript enrichment was an artifact of tissue thickness for all genes except *vasa*, *tudor* and *boule*, whose transcripts appeared truly enriched in putative PGCs at late blastoderm/early gastrulation stages (Fig. 2D–H″, Fig. 4N,S).

Strikingly, we found that *vasa*, *tudor* and *boule* marked PGCs from the time of their formation at cellular blastoderm stages, but that none of these genes' transcripts were asymmetrically localised prior to PGC formation. Immediately after egg laying, *vasa* transcripts were not localised asymmetrically but rather were ubiquitously distributed throughout the embryo (Fig. 2A). As energid nuclei reached the embryonic surface (Fig. 2B), cytoplasmic islands enriched with these transcripts were distributed evenly across the embryonic surface, remaining there as these energids divided to form the uniform blastoderm (Fig. 2C). Prior to posterior pit formation, *vasa* expression became restricted to putative PGCs at the embryonic posterior (Fig. 2D).

To visualise *vasa* expression in the developing PGCs in greater detail, we collected staged embryos in two-hour intervals over the period during which PGCs arise (19 to 27 hours after egg laying (AEL)), performed in situ hybridisation for *vasa* (Fig. 2E–H), and sectioned the embryos in plastic resin (Fig. 2E″–H″). During this eight-hour period, the blastoderm nuclei undergo two concurrent, dynamic processes: continuing cell divisions increase the nuclear density throughout the blastoderm, and the blastoderm nuclei move towards the posterior pole and ultimately into the yolk (Butt, 1949; Liu and Kaufman, 2004) (Fig. 2E′–H′). From 19–21 hours AEL, the ubiquitous *vasa* expression seen in early embryos remained unchanged (Fig. 2E–E″). However, from 21–23 hours AEL *vasa* expression became enriched in a subset of cells at the blastoderm posterior (Fig. 2F–F″). From 23–25 hours AEL, *vasa*-positive cells increased in density at the blastoderm posterior and began to move into the yolk (Fig. 2G–G″). This movement appeared passive, due to the formation of the posterior pit by invagination of the germ rudiment. However, in the absence of time-lapse data we cannot rule out the possibility of active PGC movement out of the blastoderm epithelium and

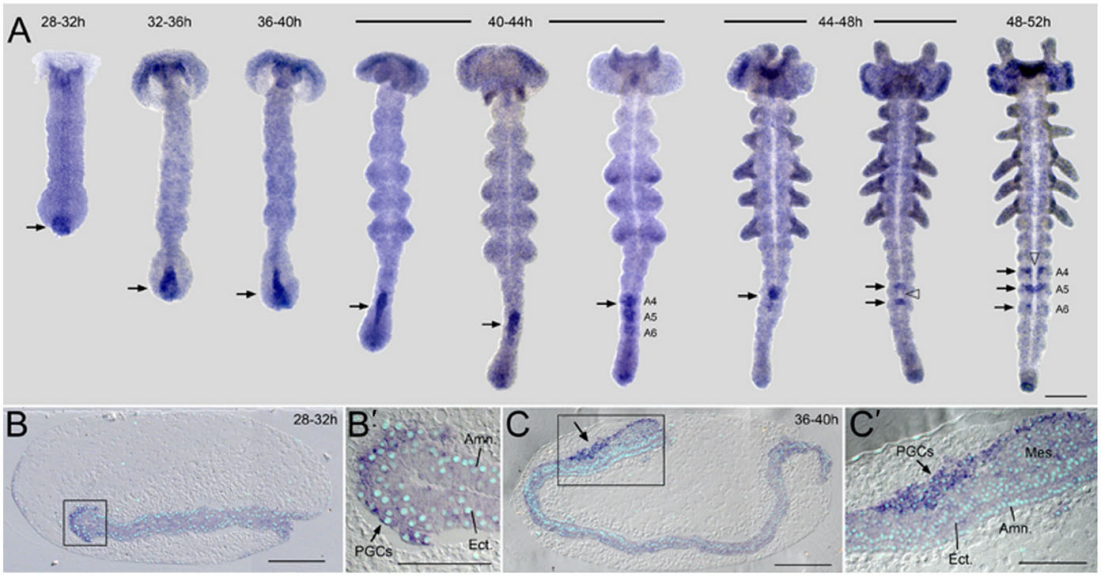


Fig. 3. *vasa* marks PGCs throughout migration. (**A**) *vasa* transcript expression during progressive stages of germ band development. Arrows indicate *vasa*-positive PGCs. (**B**) Medial section of an embryo at 28–32 h AEL, showing *vasa* in situ hybridisation (purple) and nuclear stain (cyan). Boxed region enlarged in (**B′**) shows PGCs in contact with ectoderm (*Ect.*) and the amnion (*Amn.*). (**C**) Medial section of an embryo at 36–40 h AEL, when PGCs (arrow) initiate migration along the mesoderm (*Mes.*). Boxed region enlarged in (**C′**). Scale bars: 200 µm in A–C; 100 µm in B′,C′. Anterior is up in A, left in B–C′.

towards the yolk. From 25–27 hours AEL, as the germ rudiment began its invagination into the yolk, *vasa*-positive cells formed a distinct mesenchymal clump within the yolk at the posterior of the embryo (Fig. 2H–H″). During this and all following stages, in addition to the marked enrichment in PGCs, *vasa* transcripts were additionally observed ubiquitously at low levels throughout somatic tissue (Figs 2, 3).

Throughout all subsequent stages of germ band elongation and patterning, *vasa* continued to mark PGCs (Fig. 3). During early stages of germ band elongation prior to limb bud formation (~28–32 hours AEL) *vasa*-positive PGCs remained at the embryonic posterior on the dorsal surface of the newly forming mesoderm (Fig. 3A,B,B′). The PGC cluster then became pear-shaped from 32–42 hours AEL, as the anteriormost PGCs began to move towards the anterior of the embryo (Fig. 3A,C,C′). As the head lobes enlarged (36–40 h AEL), PGCs began to migrate anteriorly on the dorsal surface of the embryo and continued their migration during limb bud stages (40–44 h AEL) (Fig. 3A). During appendage elongation stages (44–48 h AEL) PGCs split into distinct clusters spanning the midline in abdominal segments A4–A6, one cluster per segment. As appendage segmentation became morphologically distinct (48–52 h AEL), the segmental clusters split along the ventral midline into bilateral clusters in A4–A6.

tudor and *boule* were also expressed in PGCs at all stages in a pattern indistinguishable from that of *vasa* (Fig. 4), providing further evidence that the *vasa*-positive cells are *Oncopeltus* PGCs. None of the other genes we examined (supplementary material Fig. S3), including *nanos* and *piwi* (Fig. 4), were enriched in PGCs at any stage of embryogenesis.

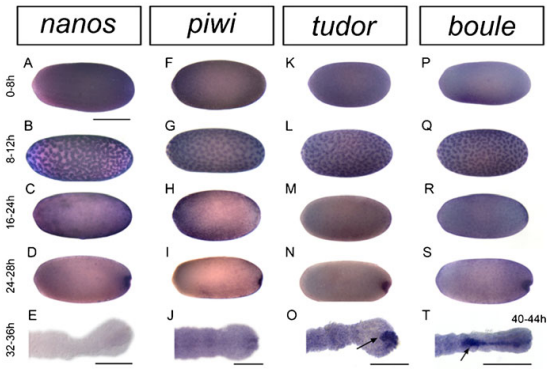


Fig. 4. *Oncopeltus* PGCs express *tudor* and *boule*, but not *nanos* or *piwi*. In early embryos, expression of all four genes remains ubiquitous during energid proliferation (**A**,**F**,**K**,**P**) and blastoderm formation (**B**,**C**,**G**,**H**,**L**,**M**,**Q**,**R**). During posterior pit formation *nanos* is expressed throughout the length of the embryo (**D**), whereas *piwi* expression is reduced in the presumptive extraembryonic serosal tissue in the anterior of the embryo (**I**). Apparent posterior staining in (D) and (I) is the result of tissue thickness in that location, and is not specific to PGCs. *tudor* (**N**) and *boule* (**S**) transcripts become restricted to presumptive PGCs at the time of their specification. In germ band stage embryos, while *tudor* (**O**) and *boule* (**T**) mark presumptive PGC clusters, *nanos* is not detected (**E**), while *piwi* expression is ubiquitous (**J**). Scale bars: 500 µm in A (applies also to B–D,F–I,K–N,P–S); 100 µm in E,J; 200 µm in O,T. Anterior to the left.

Neither *vasa* nor *tudor* are required for PGC formation

Our gene expression analysis demonstrates that *vasa*, *tudor*, and *boule* are specifically expressed in PGCs beginning at the putative time of their specification at the embryonic posterior just prior to gastrulation. To determine whether these genes were required for PGC formation or development in *Oncopeltus*, we performed maternal RNAi (mRNAi) for each gene. We confirmed that mRNAi effectively reduced zygotic transcript levels in our experiments using RT-PCR (Fig. 5E). PGC presence or absence was determined with in situ hybridisation against PGC markers at ~40–54 hours AEL, when germ cells are visible on the dorsal mesoderm.

RNAi knockdown of *vasa* or *tudor* did not disrupt embryonic patterning or germ band development (supplementary material Table S3), despite the widespread expression of these genes at early blastoderm stages (Figs 2, 4), and their persistent low levels of expression in somatic cells even after PGC formation (Figs 3, 4). Strikingly, germ cells were clearly present in both *vasa* (93.8%, $n=16$) and *tudor* (100%, $n=20$) knockdowns, suggesting that neither of these genes is required for PGC specification (Fig. 5A–C′). It is formally possible that residual *vasa* or *tudor* transcripts that may have escaped destruction by mRNAi could be sufficient to play an instructive role in PGC formation. However, we note that transcript levels of both genes in the progeny of injected mothers were barely detectable in the case of *vasa*, and undetectable in the case of *tudor*, when assessed with RT-PCR even as late as 4 days AEL (Fig. 5E). Moreover, even hypomorphic alleles of *tudor* (Schüpbach and Wieschaus, 1986) and *vasa* (Lasko and Ashburner, 1990; Schüpbach and Wieschaus, 1991; Liang et al., 1994) lead to loss of PGCs in *Drosophila*. We therefore hypothesise that in *Oncopeltus*, *vasa* and *tudor* are required neither maternally nor zygotically for germ cell specification, although they are expressed in the cells specified as PGCs.

To address the possibility of redundancy between these two genes, we performed double knockdown of *vasa* and *tudor*, which reduced transcripts of both genes to undetectable levels (Fig. 5E). Eggs laid by *vasa* + *tudor* double RNAi females had an increased rate of embryonic lethality relative to controls (supplementary material Table S3; 47.4%, $n=19$ vs 10.3%, $n=39$), which may mean that these genes work together to play roles in somatic development. However, embryos that escaped this lethality still had PGCs (100%, $n=10$) (Fig. 5D,D′).

None of the knockdowns caused any qualitative or quantitative change in egg laying by injected females compared to controls, and ovaries of injected females showed neither morphological abnormalities nor signs of disrupted oogenesis (not shown). This indicates that, in contrast to *Drosophila* (Schüpbach and Wieschaus, 1991; Styhler et al., 1998; Tomancak et al., 1998; Johnstone and Lasko, 2004), *vasa* is not required individually or redundantly with *tudor* for *Oncopeltus* oogenesis or egg laying.

boule is necessary for *Oncopeltus* oogenesis and embryonic survival

boule mRNAi caused a complete cessation of egg laying by injected females after four to five clutches (one clutch is laid every one to two days). In contrast, *vasa*, *tudor* and control mRNAi females continued to lay up to 12 clutches. Ovaries of *boule* dsRNA-injected females possessed only a few oocytes at early stages of oogenesis, and few or no mature oocytes (not shown), indicating a requirement for *boule* in the progression of

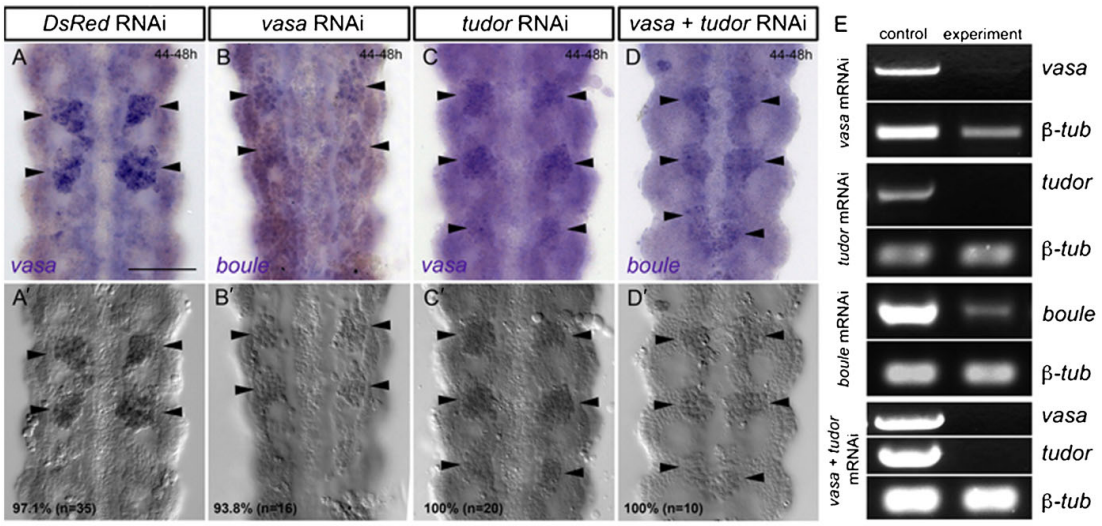


Fig. 5. *vasa* and *tudor* are not required for PGC specification in *Oncopeltus*. (**A–D**) Bright field images of in situ hybridisations for PGC markers in different RNAi conditions; numbers indicate sample sizes and % of embryos with PGCs. Arrowheads indicate PGC clusters in abdominal segments A4–A6. (**A′–D′**) DIC images of the same embryos shown in (A–D) showing distinct PGC cluster morphology. (A,A′) In control embryos *vasa*-positive PGCs are visible on the dorsal surface of abdominal segments 4–5. PGCs are present in *vasa* RNAi (B,B′), *tudor* RNAi (C,C′), and double *vasa* + *tudor* RNAi (D,D′) embryos. (**E**) RT-PCR validation of RNAi knockdown. Controls are animals injected with *DsRed* dsRNA. Expression of β-tubulin was analysed to confirm cDNA integrity and allow comparison of amounts of template per lane. Scale bar: 100 µm. Anterior is up in A–D.

oogenesis. Eggs laid by *boule* RNAi females displayed nearly complete embryonic lethality (81.8%, $n=22$) in all but the first clutch laid. (The first clutch of *Oncopeltus* eggs laid following mRNAi typically displays no abnormalities, as these eggs have developed their chorion by the time of injection and are therefore impervious to dsRNA (Liu and Kaufman, 2004).) This was a striking increase in embryonic lethality compared to *DsRed* controls (26.8%, $n=190$), *vasa* knockdowns (23.2%, $n=198$) and *tudor* knockdowns (5.6%, $n=54$). The oogenesis requirement for *boule* and resulting embryonic lethality thus prevented us from determining whether *boule* is required for germ cell specification in *Oncopeltus*, and we do not further report on the role of *boule* on oogenesis in the present study.

vasa is required for *Oncopeltus* spermatogenesis

Given that in contrast to *Drosophila*, *vasa* is not required for germ cell specification or oogenesis in *Oncopeltus*, we wished to test for other possible functions of this gene. In mice, despite its expression in the embryonic PGCs of both sexes once they reach the genital ridge (Fujiwara et al., 1994; Diez-Roux et al., 2011). *vasa* is required not for PGC specification, but rather for gametogenesis in males (Tanaka et al., 2000). Similarly, we recently showed that *vasa* plays a role in spermatogenesis in the cricket *G. bimaculatus* (Ewen-Campen et al., 2013). We therefore asked whether *vasa* also functions during spermatogenesis in *Oncopeltus*.

The testes of *Oncopeltus* show an organisation typical of insect testes (Dumser, 1980), with stages of spermatogenesis located in an anterior–posterior progression (supplementary material Fig. S4). Unlike *Drosophila*, which has a single sperm tubule (testiole) per testis (Hardy et al., 1979), each *Oncopeltus* testis comprises seven testioles (Bonhag and Wick, 1953). In situ hybridisation for *vasa* showed that it is strongly expressed in secondary spermatogonia of each testiole, and at lower levels in early primary spermatocytes and post-spermatocyte stages, but not in primary spermatogonia or somatic cells (Fig. 6A). Adult males injected with dsRNA against *vasa* displayed multiple abnormalities in spermatogenesis. Testioles of *vasa* RNAi males lacked clearly defined cysts and contained large numbers of small, dense nuclei in the anterior region (Fig. 6H,I,I′), which in controls contained only spermatocytes with large, pale nuclei (Fig. 6C,D,D′; supplementary material Fig. S4E). The primary spermatogonial region of *vasa* RNAi testioles contained cysts of irregular size (Fig. 6I, arrowheads) with poorly defined cytoplasmic bridges (Fig. 6I′, arrows). In the spermatocyte region *vasa* RNAi testioles contained large, poorly defined clusters of several hundred cells (Fig. 6J, arrowheads) at varying stages of spermatogenesis (Fig. 6J′). The nuclear morphology of cells in these cysts corresponded to spermatocyte (Fig. 6E; supplementary material Fig. S4E) or early spermatid (Fig. 6F) stages, as well as shell stage-like nuclei (Fig. 6K,K′) typical of the mid-stage spermatids of controls (Fig. 6F′). Cysts of wild type shell stage spermatids are no longer syncytial as the actin-rich cytoplasmic bridges disappear during spermatocyte stages (supplementary material Fig. S4G). In contrast, the anterior shell stage-like nuclei in *vasa* RNAi testioles remained connected by cytoplasmic bridges (Fig. 6K′, red arrows), consistent with precocious spermatid differentiation. Moreover, although they displayed clear shell stage nuclear morphology (Fig. 6K,K′, red arrowheads), they were larger than wild type shell stage nuclei (Fig. 6F′, red arrowheads), suggesting that they had begun spermatid differentiation as syncytial diploid cells without first

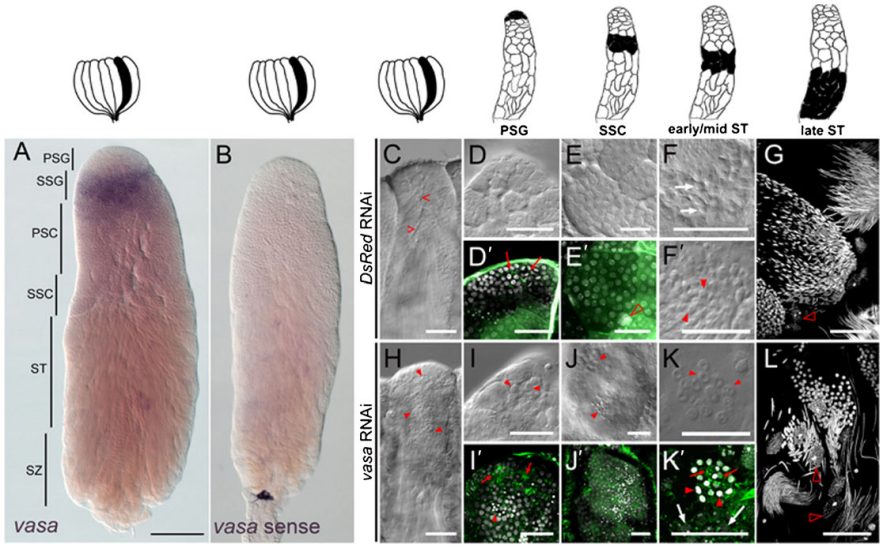


Fig. 6. *vasa* is expressed in adult testes and required for spermatogenesis in *Oncopeltus*. Schematics indicate the region of the testis (**A–C**) or testiole (**D–L**) shown in each column. (A) *vasa* in situ of an adult wild-type testiole showing expression in the secondary spermatogonia. PSG: primary spermatogonia; SSG: secondary spermatogonia; PSC: primary spermatocytes; SSC: secondary spermatocytes; ST: spermatids undergoing spermiogenesis; SZ: spermatozoa. (B) *vasa* sense control probe. DIC optics (C–F,**F′**) and F-actin (green) and nuclear staining (white) (**D′**,**E′**) of control testioles reveals distinct, synchronized spermatogenic cysts separated by clear cyst boundaries (carets) (C), small cysts of spermatogonia (PSG) at the apex (D,D′), larger cysts of secondary spermatocytes (SSC) posterior to the apex (E,E′; arrowhead in (E′) indicates somatic sheath cells associated with cysts of germ cells), early spermatids with round prominent nuclei (early ST) (F) and mid-stage spermatids with smaller, compact round nuclei (mid ST) (F′). (G) Late spermatid cysts in controls are synchronized in spermiogenesis; hollow arrowheads indicate somatic sheath cells. *vasa* RNAi testioles contain large masses of cells with heterogeneous nuclear morphologies (H; arrowheads). (I) PSG cysts are abnormal in shape and size, contain nuclei of multiple sizes (arrowheads), and (**I′**) have filamentous actin masses interspersed between nuclei (arrows) rather than clearly defined cytoplasmic bridges (compare with D′, arrows). (J,**J′**) Abnormal cysts contain clusters of small dense nuclei (arrowheads). (K,**K′**). Aberrant cysts retain cytoplasmic bridges at spermatid stages (red arrows), and contain nuclei with morphologies corresponding to different spermatogenic stages, including both early (white arrows; compare with F) and mid ST (red arrowheads; compare with F′) stages. (L) *vasa* RNAi late spermatid cysts are asynchronous, comprising multiple late spermatid and spermatozoon differentiation stages within a single cyst; cysts remain associated with sheath cells (hollow arrowheads). Scale bars: 200 µm in A (applies to B); 100 µm in C,H; 50 µm in D–E′, G, I–J′, L; 25 µm in F,F′,K,K′. Anterior is up in all panels.

proceeding through meiosis as in wild type. Finally, the posteriormost region of *vasa* RNAi testioles contained irregular groups of cells at mixed stages of late spermatid and spermatozoon differentiation (Fig. 6L), rather than the perfectly synchronised cysts of late spermiogenic stages seen in controls (Fig. 6G). These defects were observed in testes examined 28–29 days following injection of adult males, but are not artefacts of age, as testes of 10 week old wild type adult males showed normal progression through all stages of spermatogenesis (supplementary material Fig. S4B).

Taken together, these data suggest that *vasa* is required for the maintenance of synchrony within cysts at multiple stages of spermatogenesis. In addition, *vasa* may be required for secondary spermatogonia to enter correct meiotic progression as spermatocytes, in the absence of which germ cells are nonetheless able to continue with subsequent stages of spermatogenesis.

Discussion

Oncopeltus germ cell formation

In several cases, analyses of molecular markers such as *vasa* mRNA have revealed the presence of a cryptic germ plasm that had eluded prior histological studies (Yoon et al., 1997; Tsunekawa et al., 2000; Wu et al., 2011). In *Oncopeltus*, we have shown that none of the transcripts of an extensive suite of conserved germ cell markers localise asymmetrically within oocytes or in early embryos (Figs 1, 2, 4; supplementary material Fig. S3), including transcripts of genes that localise to and are required for the function of germ plasm in *Drosophila*. Gene products of at least one of these conserved germ line markers have been found in the germ plasm of every species where a germ plasm is known to exist (Ewen-Campen et al., 2010), although we note the important caveat that in *Drosophila*, several of these genes (*vasa*, *piwi*, and *tudor*) are localized as proteins rather than mRNAs. Thus, the lack of localisation of transcripts of any of these 19 genes during oogenesis or early embryogenesis suggests that *Oncopeltus* lacks germ plasm. Instead, our data support the hypothesis that *Oncopeltus* germ cells form in the absence of germ plasm, and are not present prior to the onset of posterior invagination at the end of the cellular blastoderm stage. We cannot, however, rule out the possibility that untested molecular markers, including protein products of the genes examined here, could be asymmetrically localised to a putative germ plasm in *Oncopeltus*.

While we provide multiple markers of PGCs, further experiments could be useful to confirm the identity of these cells. However, demonstration that these cells are functional PGCs via ablation experiments is complicated by the fact

that they arise at the inner face of the blastoderm at the gastrulation center, so that their physical disruption would likely also compromise mesoderm formation and subsequent embryogenesis. Moreover, we note that while pole cell removal experiments in *Drosophila* result in sterility, pole cell removal in another insect with germ plasm, the wasp *Pimpla turionellae*, yields fertile adults despite the fact that these pole cells are bona fide PGGs in wild type embryos (Bronskill, 1959; Achtelig and Krause, 1971; Fleischmann, 1975). Further, we are currently unable to genetically ablate these cells and determine their effect on fertility, as our *vasa*, *tudor* and *vasa* + *tudor* RNAi double RNAi experiments do not disrupt their formation (Fig. 5). Lineage tracing techniques that would permit tracking of the putative PGCs over the six-week period between PGC formation and sexual maturity are not currently available for *Oncopeltus*. These caveats notwithstanding, the molecular and morphological evidence that the cells we identify in this report are bona fide *Oncopeltus* PGCs is comparable to that available for PGC identification in most studied animal species: (1) three conserved germ line genes, *vasa*, *tudor*, and *boule*, are specific germ cell markers in *Oncopeltus* (Figs 1–4); (2) transcripts of these genes first become enriched in germ cells specifically at the time that these cells were previously reported to arise based on morphological and cytological criteria (Figs 2–4) (Butt, 1949); and (3) cells with these molecular markers undergo migration and primordial gonad occupation (supplementary material Fig. S1; Figs 3, 4) consistent with the well-documented behavior of PGCs in many other hemipterans (Seidel, 1924; Mellanby, 1935; Butt, 1949; Sander, 1956; Kelly and Huebner, 1989; Heming and Huebner, 1994).

Although the posterior location of germ cells at the time of their specification is superficially similar to that of pole cells in *Drosophila* and other Diptera, PGC specification and development in *Oncopeltus* differs in several important ways. First, while *Drosophila* pole cells form on the exterior of the posterior syncytial blastoderm before somatic cellularisation, *Oncopeltus* germ cells appear on the yolk side of the cellular blastoderm. Second, while *Drosophila* pole cells are the first cells in the embryo to cellularise (Huettner, 1923), *Oncopeltus* germ cells arise after blastoderm cellularisation is complete (Butt, 1949). Third, because *Oncopeltus* is an intermediate-germ insect, only the gnathal and thoracic segments have been specified at the time that germ cells arise (Liu and Kaufman, 2004), whereas in the long-germ insect *Drosophila*, pole cells form posterior to the abdominal embryonic segments. Lastly, *Oncopeltus* germ cells form on the dorsal surface of the embryo, and remain on the yolk-facing surface of the mesoderm during their migration to the gonad primordium in anterior abdominal segments (Fig. 3). As a result, they do not undergo a transepithelial migration through the hindgut epithelium as in *Drosophila* (reviewed by Richardson and Lehmann, 2010).

The function of "germ line genes" in *Oncopeltus*

Our functional analysis led to the surprising discovery that neither *vasa* nor *tudor* play instructive roles in germ cell specification in *Oncopeltus*. Both of these genes are required for germ cell specification in *Drosophila* (Boswell and Mahowald, 1985; Schüpbach and Wieschaus, 1986) and other species (Sunanaga et al., 2007; Spike et al., 2008). However, *vasa* has widely divergent roles across Metazoa (reviewed by Yajima and Wessel, 2011), and in many cases is dispensable for PGC specification (Tanaka et al., 2000; Braat et al., 2001; Li et al., 2009; Özhan-Kizil et al., 2009). In several organisms it plays a role in adult gametogenesis (Tanaka et al., 2000; Ohashi et al., 2007; Salinas et al., 2007; Fabioux et al., 2009; Salinas et al., 2012; Ewen-Campen et al., 2013).

Intriguingly, we find that similar to the mouse and the cricket, *vasa* is required for spermatogenesis in adult *Oncopeltus* (Fig. 6), but not for oogenesis. This sex-specific function may relate to a putative role in stem cell function. As in other hemimetabolous insects (Büning, 1994), in *Oncopeltus* germ line stem cells are likely active in the apex of the testes (Schmidt and Dorn, 2004) but are not thought to be present in adult ovaries. One caveat to this hypothesis is that *vasa* transcript was not detected by in situ hybridisation in the primary spermatogonia (Fig. 6A), although it may be present at very low levels in those stem cells. Alternatively, given its strong expression in secondary spermatocytes and the defects in cyst integrity and synchrony caused by *vasa* RNAi (Fig. 6), *Oncopeltus vasa* may play a male-specific role in the onset or synchrony of meiosis. Consistent with a conserved role for *vasa* in bilaterian meiosis, male germ cells in *vasa* knockout mice arrest just prior to meiosis onset (Tanaka et al., 2000), and in human stem cell-derived germ cells *vasa* overexpression enhances meiotic progression (Medrano et al., 2012). *Oncopeltus vasa* RNAi leads to premature spermatid differentiation by some diploid secondary spermatocytes within a cyst, resulting in cyst asynchrony. In *Drosophila*, mutations are known that disrupt meiosis but do not prevent sperm formation (Davis, 1971), consistent with the hypothesis that spermiogenesis can be decoupled from meiotic status.

The evolutionary origins of germ plasm in insects

Together with recent molecular and classical histological data on germ cell specification in other insects, our results are consistent with the hypothesis that germ plasm is a derived mode of germ cell specification that arose in the ancestor to holometabolous insects (Fig. 7) (Lynch et al., 2011; Ewen-Campen et al., 2012). The only other functional genetic analysis of germ line specification in a hemimetabolous insect to date (Ewen-Campen et al., 2013) has also provided evidence that maternally supplied posterior germ plasm is absent, and that *vasa* is dispensable maternally and zygotically for germ cell formation. Our data thus support the notion that germ plasm-driven germ cell specification mechanisms operative in *Drosophila melanogaster* and *Nasonia vitripennis* are derived relative to the Hemimetabola (Fig. 7).

The ubiquitous distribution of germ cell markers in early *Oncopeltus* embryos and their subsequent enrichment in presumptive germ cells at the blastoderm posterior is reminiscent of *vasa* expression in the beetle *Tribolium* (Fig. 7) (Schröder, 2006; C. von Levetzow, Konservierte und divergente Aspekte der *twist*-, *snail*- und *concertina*-Funktion im Käfer *Tribolium castaneum*, PhD thesis, Universität zu Köln, 2008). Further taxonomic sampling, and functional studies in *Tribolium*, will be needed to determine whether the PGC specification mechanisms in these two species may be the result of common ancestry (Fig. 7).

A large number of transcripts that localise to germ plasm in *Drosophila* are expressed ubiquitously in *Oncopeltus* oocytes and early embryos. This suggests that the evolution of germ plasm in Holometabolous insects involved a large-scale change in the localisation of many transcripts in the oocyte. We propose that this likely resulted from a change in the localisation of an upstream component capable of recruiting many downstream

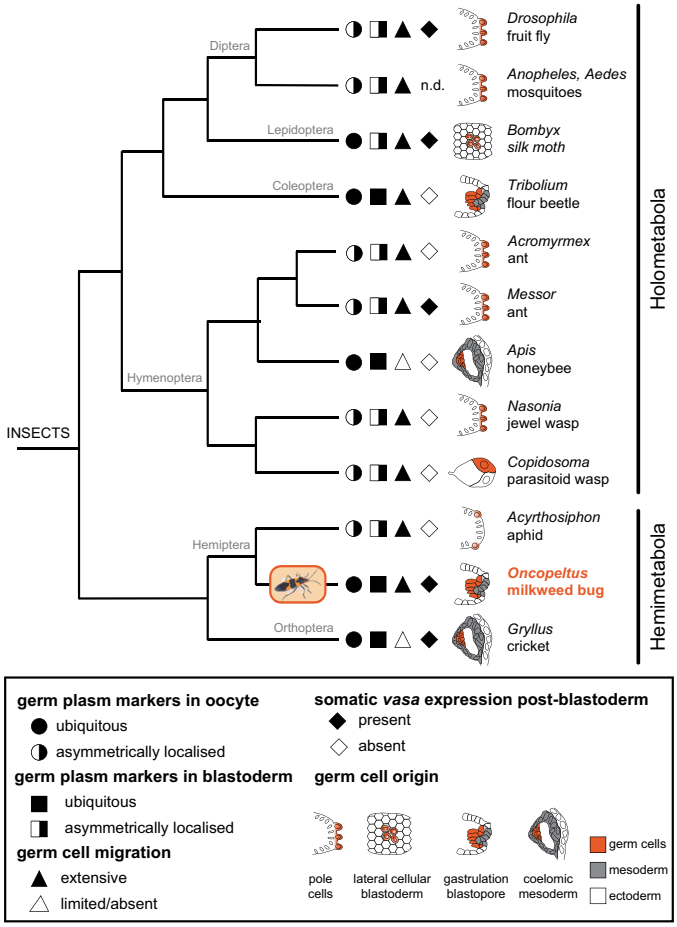


Fig. 7. Phylogenetic distribution of germ cell specification mechanisms and migration patterns across insects. Species shown are those for which data on the expression and/or function of molecular markers for germ cells during oogenesis and embryogenesis are available. Molecular data suggest absence of germ plasm in oocytes (circles) and early embryos (squares) of some Holometabola (*Tribolium*, *Apis*) and Hemimetabola (*Oncopeltus* and *Gryllus*), and somatic expression of *vasa* at post-blastoderm stages of development (diamonds) is not uncommon. In most species, PGCs undergo extensive migration from the site of specification to the gonad primordia (triangles). Data from this study (also Nakao, 1999; Mahowald, 2001; Donnell et al., 2004; Zhurov et al., 2004; Chang et al., 2006; Dearden, 2006; Juhn and James, 2006; Nakao et al., 2006; Schröder, 2006; Chang et al., 2007; Juhn et al., 2008; Zhao et al., 2008; Chang et al., 2009; Khila and Abouheif, 2010; Lynch et al., 2011; Ewen-Campen et al., 2013; C. von Levetzow, Konservierte und divergente Aspekte der *twist*-, *snail*- und *concertina*-Funktion im Käfer *Tribolium castaneum*, PhD thesis, Universität zu Köln, 2008). Phylogenetic relationships from Yeates et al. (Yeates et al., 2012).

transcripts, rather than via the sequential evolution of distinct localisation mechanisms for individual transcripts. Studies on the genetic mechanism of evolutionary redeployments of multiple downstream genes have largely focused on transcription factors, as individual transcription factors are capable of regulating large numbers of target genes (Hoekstra and Coyne, 2007; Moczek, 2008; Stern and Orgogozo, 2008; Craig, 2009). Interestingly, in the case of germ plasm, transcription factors are unlikely to have been key players in the mechanisms of evolutionary change for a number of reasons. First, regulation of germ line determinants is largely post-transcriptional (Arkov and Ramos, 2010; Richter and Lasko, 2011; Sengupta and Boag, 2012; Nousch and Eckmann, 2013). Second, germ plasm transcript function relies on their subcellular localisation (often mediated via signals in their 3′UTRs) rather than their presence or absence (Rangan et al., 2009). Finally, unlike key transcription factors identified as largely sufficient to induce specific somatic cell fates (e.g. Akam, 1998; Kozmik, 2005; Baena-Lopez and García-Bellido, 2006), there is no single conserved gene that is sufficient to confer germ cell fate across metazoans. The evolution of germ plasm may therefore serve as an example of how a novelty (asymmetrically localized germ plasm in the oocyte) arose via changes in RNA localisation rather than transcriptional regulation.

Materials and Methods

Animal culture

Oncopeltus fasciatus were cultured at 28°C as previously described (Ewen-Campen et al., 2011). Timing of embryonic events reported here may differ from that reported in other studies using lower rearing temperatures (e.g. Liu and Kaufman, 2004).

Cloning and phylogenetic analysis

Total RNA was extracted from mixed-stage embryos and ovaries using TRIzol (Invitrogen) and used for first strand cDNA synthesis with qScript cDNA SuperMix (Quanta BioSciences). An *Oncopeltus vasa* fragment was cloned using degenerate primers (supplementary material Table S1). *nanos* and *piwi* fragments were obtained from the *Oncopeltus* transcriptome (Ewen-Campen et al., 2011). Fragments were extended using RACE PCR (SMART RACE cDNA kit, Clontech), and used as templates for DIG-labeled in situ probes and dsRNA fragments following sequence verification (supplementary material Table S1). Genes for the in situ hybridisation screen (supplementary material Tables S1, S2) were obtained from the *Oncopeltus* transcriptome (Ewen-Campen et al., 2011; Zeng and Extavour, 2012) and amplified using primers containing linker sequence (5′-CCCGGGGC-3′) enabling direct addition of a T7 site to the 3′ end in a subsequent PCR reaction. Extended sequences are available from ASGARD (http://asgard.rc.fas.harvard.edu) (Zeng and Extavour, 2012). All coding sequences

reported in this study have been submitted to GenBank [accession numbers KC261571–KC261587] except for *orb* and *Uev1A*, for which we obtained only 3′ UTR sequence.

Maximum-likelihood phylogenetic analysis was performed for *vasa*, *piwi*, and *nanos* as previously described (Ewen-Campen et al., 2012).

Tissue fixation and gene expression analysis

Embryos were fixed and stained as previously described (Liu and Kaufman, 2004; Erezyilmaz et al., 2009; Kainz et al., 2011). Adult gonads were dissected in 1× PBS and fixed in 4% formaldehyde in 1× PBS for at least one hour. Antibodies used were mouse anti-alpha tubulin DM1A (Sigma) 1:50 and goat anti-mouse Alexa Fluor 568 (Invitrogen) 1:500–1:1000, and counterstains were FITC-phalloidin (Invitrogen) 0.5–1 µl and Hoechst 33342 (Sigma) 0.1–0.5 µg/ml.

Plastic sectioning

In situ hybridisation and/or Sytox Green (Invitrogen) staining were performed prior to embedding embryos in Durcupan ACM Fluka (Sigma), mixed at a ratio of 32:27:1:0.6 = components A:B:C:D. Embryos were dehydrated through 10-minute washes in each of 50%, 70%, 90%, 2× 100% ethanol and 100% acetone, transferred to a 1:1 mixture of acetone: catalysed Durcupan, and left uncovered in a fume hood overnight. Embryos were individually transferred to fresh Durcupan in silicone molds (Electron Microscope Sciences NO. 70903) and oriented following a 30-minute initial hardening at 65℃. Resin blocks were baked for 24 hours at 65℃.

Block fronts were trimmed with a razor blade and sectioned at 5–6 µm on a Leica RM2255 microtome with a high-profile knife holder using High-Profile disposable “diamond-edge” steel knives (C.L. Sturkey NO. D554D50). Sections were collected on water droplets on charged slides, dehydrated on a heat block, and mounted in Permount (Fisher Scientific).

Parental RNAi

dsRNA for all genes (supplementary material Table S1) was prepared as previously described (Kainz et al., 2011) and resuspended in injection buffer (5 mM KCl, 0.1 mM NaH_2PO_4) to a concentration of 2 µg/uL. Male and female adults were injected three days after final molt with 5 µL of 2 µg/uL dsRNA using a Hamilton syringe and size 26 needles. Testes were collected from injected males 27–29 days after injection.

Reverse-transcription PCR

Half of each clutch laid by injected females was fixed for in situ hybridisation, and the other half was homogenised in TRIzol (Invitrogen) and stored at −80℃ before isolation of total RNA. RNA was isolated separately from late blastoderm (24–29 hours AEL), early germ band (24–48 hours AEL) and late germ band (72–96 hours AEL) embryos laid by injected mothers. Genomic DNA was treated with Turbo DNase (Ambion) at 37℃ for 30 minutes, followed by DNase heat-inactivation and phenol/chloroform extraction. cDNA was synthesised from 120 ng of each RNA sample using Superscript III Supermix (Invitrogen). PCR was performed using Advantage 2 DNA Polymerase from 1 µL of cDNA template and primers indicated in supplementary material Table S1 at 60℃ annealing temperature with 35 PCR cycles. RT-PCR results for samples of all three embryonic ages tested yielded indistinguishable results, indicating that maternal RNAi was effective at reducing zygotic transcripts in embryos at least up to four days AEL.

Acknowledgements

We thank Omar Delannoy Bruno for conducting additional in situ hybridisations, Ian Dunn for assistance collecting embryos, and Jeremy Lynch for providing the T7-linker PCR probe synthesis protocol. This work was partially supported by National Science Foundation (NSF) awards [grant numbers IOS-0817678 and IOS-1257217] to C.G.E., an NSF Predoctoral Fellowship to B.E.-C., and a Frank Knox Memorial Fellowship to T.E.M.J.

Author Contributions

B.E.-C. and C.G.E. designed the research; B.E.-C., T.E.M.J. and C.G.E. performed experiments, collected and analysed data and wrote the manuscript; C.G.E. obtained funding for the research.

Competing Interests

The authors have no competing interests to declare.

References

Achtelig, M. and Krause, G. (1971). Experimente am ungefurchten ei von *Pimpla turionellae* L. (Hymenoptera) zur funktionsanalyse des oosombereichs. *Wilhelm Roux Arch. Entwickl. Mech. Org.* **167**, 164-182.

Akam, M. (1998). Hox genes: from master genes to micromanagers. *Curr. Biol.* **8**, R676-R678.

Anastas, A., Hunt, C. and Stebbings, H. (1991). Characterization of a nucleotide-sensitive high molecular weight microtubule-associated protein in the ovary of a hemipteran insect. *Cell Motil. Cytoskeleton* **19**, 37-48.

Anderson, D. T. (1973). *Embryology And Phylogeny In Annelids And Arthropods*. Oxford: Pergamon Press.

Arkov, A. L. and Ramos, A. (2010). Building RNA-protein granules: insight from the germline. *Trends Cell Biol.* **20**, 482-490.

Baena-Lopez, L. A. and García-Bellido, A. (2006). Control of growth and positional information by the graded *vestigial* expression pattern in the wing of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **103**, 13734-13739.

Bardsley, A., McDonald, K. and Boswell, R. E. (1993). Distribution of tudor protein in the *Drosophila* embryo suggests separation of functions based on site of localization. *Development* **119**, 207-219.

Bonhag, P. F. and Wick, J. R. (1953). The functional anatomy of the male and female reproductive systems of the milkweed bug, *Oncopeltus fasciatus* (Dallas) (Heteroptera: Lygaeidae). *J. Morphol.* **93**, 177-283.

Boswell, R. E. and Mahowald, A. P. (1985). *tudor*, a gene required for assembly of the germ plasm in *Drosophila melanogaster*. *Cell* **43**, 97-104.

Braat, A. K., van de Water, S., Korving, J. and Zivkovic, D. (2001). A zebrafish vasa morphant abolishes vasa protein but does not affect the establishment of the germline. *Genesis* **30**, 183-185.

Bronskill, J. F. (1959). Embryology of *Pimpla turionellae* (L.) (Hymenoptera: Ichneumonidae). *Can. J. Zool.* **37**, 655-688.

Büning, J. (1994). *The Insect Ovary: Ultrastructure, Previtellogenic Growth And Evolution*. London: Chapman and Hall.

Buss, L. W. (1987). *The Evolution Of Individuality*. Princeton, NJ: Princeton University Press.

Butt, F. H. (1949). *Embryology Of The Milkweed Bug: Oncopeltus Fasciatus (Hemiptera)*, pp. 2-43. Ithaca, NY: Cornell University Agricultural Experiment Station.

Callebaut, M. (2008). A review: Historical evolution of preformistic versus neoformistic (epigenetic) thinking in embryology. *Belg. J. Zool.* **138**, 20-35.

Calvo, E., Walter, M., Adelman, Z. N., Jimenez, A., Onal, S., Marinotti, O. and James, A. A. (2005). *Nanos* (*nos*) genes of the vector mosquitoes, *Anopheles gambiae*, *Anopheles stephensi* and *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **35**, 789-798.

Chang, C. C., Lee, W. C., Cook, C. E., Lin, G. W. and Chang, T. (2006). Germ-plasm specification and germline development in the parthenogenetic pea aphid *Acyrthosiphon pisum*: Vasa and Nanos as markers. *Int. J. Dev. Biol.* **50**, 413-421.

Chang, C. C., Lin, G. W., Cook, C. E., Horng, S. B., Lee, H. J. and Huang, T. Y. (2007). *Apvasa* marks germ-cell migration in the parthenogenetic pea aphid *Acyrthosiphon pisum* (Hemiptera: Aphidoidea). *Dev. Genes Evol.* **217**, 275-287.

Chang, C. C., Huang, T. Y., Cook, C. E., Lin, G. W., Shih, C. L. and Chen, R. P. (2009). Developmental expression of *Apnanos* during oogenesis and embryogenesis in the parthenogenetic pea aphid *Acyrthosiphon pisum*. *Int. J. Dev. Biol.* **53**, 169-176.

Craig, L. R. (2009). Defending evo-devo: a response to Hoekstra and Coyne. *Philos. Sci.* **76**, 335-344.

Curtis, D., Apfeld, J. and Lehmann, R. (1995). *nanos* is an evolutionarily conserved organizer of anterior-posterior polarity. *Development* **121**, 1899-1910.

Davis, B. K. (1971). Genetic analysis of a meiotic mutant resulting in precocious sister-centromere separation in *Drosophila melanogaster*. *Mol. Gen. Genet.* **113**, 251-272.

Dearden, P. K. (2006). Germ cell development in the honeybee (*Apis mellifera*); *vasa* and *nanos* expression. *BMC Dev. Biol.* **6**, 6.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I. et al. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* **9**, e1000582.

Donnell, D. M., Corley, L. S., Chen, G. and Strand, M. R. (2004). Caste determination in a polyembryonic wasp involves inheritance of germ cells. *Proc. Natl. Acad. Sci. USA* **101**, 10095-10100.

Dumser, J. B. (1980). The regulation of spermatogenesis in insects. *Annu. Rev. Entomol.* **25**, 341-369.

Eberhart, C. G., Maines, J. Z. and Wasserman, S. A. (1996). Meiotic cell cycle requirement for a fly homologue of human *Deleted in Azoospermia*. *Nature* **381**, 783-785.

Economopoulos, A. P. and Gordon, H. T. (1971). Growth and differentiation of the testes in the large milkweed bug, *Oncopeltus fasciatus* (Dallas). *J. Exp. Zool. B* **177**, 391-405.

Erezyilmaz, D. F., Rynerson, M. R., Truman, J. W. and Riddiford, L. M. (2009). The role of the pupal determinant *broad* during embryonic development of a direct-developing insect. *Dev. Genes Evol.* **219**, 535-544.

Ewen-Campen, B., Schwager, E. E. and Extavour, C. G. (2010). The molecular machinery of germ line specification. *Mol. Reprod. Dev.* **77**, 3-18.

Ewen-Campen, B., Shaner, N., Panfilio, K. A., Suzuki, Y., Roth, S. and Extavour, C. G. (2011). The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. *BMC Genomics* **12**, 61.

Ewen-Campen, B., Srouji, J. R., Schwager, E. E. and Extavour, C. G. (2012). *oskar* predates the evolution of germ plasm in insects. *Curr. Biol.* **22**, 2278-2283.

Ewen-Campen, B., Donoughe, S., Clarke, D. N. and Extavour, C. G. (2013). Germ cell specification requires zygotic mechanisms rather than germ plasm in a basally branching insect. *Curr. Biol.* (in press).

Extavour, C. G. (2007). Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. *Integr. Comp. Biol.* **47**, 770-785.
Extavour, C. G. and Akam, M. E. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* **130**, 5869-5884.
Fabioux, C., Corporeau, C., Quillien, V., Favrel, P. and Huvet, A. (2009). *In vivo* RNA interference in oyster – *vasa* silencing inhibits germ cell development. *FEBS J.* **276**, 2566-2573.
Fleischmann, V. G. (1975). Origin and embryonic development of fertile gonads with and without pole cells of *Pimpla turionellae* L. (Hymenoptera, Ichneumonidae). *Zoologische Jahrbücher. Abteilung für Anatomie und Ontogenie der Tiere.* **94**, 375-411.
Fujiwara, Y., Komiya, T., Kawabata, H., Sato, M., Fujimoto, H., Furusawa, M. and Noce, T. (1994). Isolation of a DEAD-family protein gene that encodes a murine homolog of Drosophila vasa and its specific expression in germ cell lineage. *Proc. Natl. Acad. Sci. USA* **91**, 12258-12262.
Golumbeski, G. S., Bardsley, A., Tax, F. and Boswell, R. E. (1991). tudor, a posterior-group gene of Drosophila melanogaster, encodes a novel protein and an mRNA localized during mid-oogenesis. *Genes Dev.* **5**, 2060-2070.
Goss, R. J. (1952). The early embryology of the book louse, *Liposcelis divergens* Badonnel (Psocoptera; liposcelidae). *J. Morphol.* **91**, 135-167.
Handler, D., Olivieri, D., Novatchkova, M., Gruber, F. S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R. and Brennecke, J. (2011). A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* **30**, 3977-3993.
Hardy, R. W., Tokuyasu, K. T., Lindsley, D. L. and Garavito, M. (1979). The germinal proliferation center in the testis of *Drosophila melanogaster*. *J. Ultrastruct. Res.* **69**, 180-190.
Harrison, A., Stebbings, H. and Hyams, J. S. (1991). Different patterns of α-tubulin post-translational modification in ovarian nutritive tubes of two hemipteran insects. *J. Cell Sci.* **100**, 501-507.
Hegner, R. W. (1914). Studies on germ cells. I. The history of the germ cells in insects with special reference to the Keimbahn-determinants. II. The origin and significance of the Keimbahn-determinants in animals. *J. Morphol.* **25**, 375-509.
Hellen, C. U. and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* **15**, 1593-1612.
Heming, B. S. (1979). Origin and fate of germ cells in male and female embryos of *Haplothrips verbasci* (Osborn) (Insecta, Thysanoptera, Phlaeothripidae). *J. Morphol.* **160**, 323-343.
Heming, B. S. and Huebner, E. (1994). Development of the germ cells and reproductive primordia in male and female embryos of *Rhodnius prolixus* Stål (Hemiptera, Reduviidae). *Can. J. Zool.* **72**, 1100-1119.
Heymons, R. (1895). *Die Embryonalentwicklung Von Dermapteren Und Orthopteren*. Jena: G. Fischer.
Hoekstra, H. E. and Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995-1016.
Huettner, A. F. (1923). The origin of the germ cells in *Drosophila melanogaster*. *J. Morphol.* **37**, 385-423.
Hurst, S., Talbot, N. J. and Stebbings, H. (1999). A staufen-like RNA-binding protein in translocation channels linking nurse cells to oocytes in Notonecta shows nucleotide-dependent attachment to microtubules. *J. Cell Sci.* **112**, 2947-2955.
Hyams, J. S. and Stebbings, H. (1979). The formation and breakdown of nutritive tubes – massive microtubular organelles associated with cytoplasmic transport. *J. Ultrastruct. Res.* **68**, 46-57.
Jan, E., Thompson, S. R., Wilson, J. E., Pestova, T. V., Hellen, C. U. and Sarnow, P. (2001). Initiator Met-tRNA-independent translation mediated by an internal ribosome entry site element in cricket paralysis virus-like insect viruses. *Cold Spring Harb. Symp. Quant. Biol.* **66**, 285-292.
Johnstone, O. and Lasko, P. (2004). Interaction with eIF5B is essential for Vasa function during development. *Development* **131**, 4167-4178.
Juhn, J. and James, A. A. (2006). *oskar* gene expression in the vector mosquitoes, *Anopheles gambiae* and *Aedes aegypti*. *Insect Mol. Biol.***15**, 363-372.
Juhn, J., Marinotti, O., Calvo, E. and James, A. A. (2008). Gene structure and expression of *nanos* (*nos*) and *oskar* (*osk*) orthologues of the vector mosquito, *Culex quinquefasciatus*. *Insect Mol. Biol.* **17**, 545-552.
Juliano, C. E., Swartz, S. Z. and Wessel, G. M. (2010). A conserved germline multipotency program. *Development* **137**, 4113-4126.
Kainz, F., Ewen-Campen, B., Akam, M. and Extavour, C. G. (2011). Notch/Delta signalling is not required for segment generation in the basally branching insect *Gryllus bimaculatus*. *Development* **138**, 5015-5026.
Kaye, J. S. and McMaster-Kaye, R. (1966). The fine structure and chemical composition of nuclei during spermiogenesis in the house cricket. I. Initial stages of differentiation and the loss of nonhistone protein. *J. Cell Biol.* **31**, 159-179.
Kelly, G. M. and Huebner, E. (1989). Embryonic development of the hemipteran insect *Rhodnius prolixus*. *J. Morphol.* **199**, 175-196.
Khila, A. and Abouheif, E. (2010). Evaluating the role of reproductive constraints in ant social evolution. *Philos. Trans. R. Soc. B* **365**, 617-630.
Kim-Ha, J., Kerr, K. and Macdonald, P. M. (1995). Translational regulation of *oskar* mRNA by Bruno, an ovarian RNA-binding protein, is essential. *Cell* **81**, 403-412.
Kozmik, Z. (2005). Pax genes in eye development and evolution. *Curr. Opin. Genet. Dev.* **15**, 430-438.
Kumé, M. and Dan, K. (1968). *Invertebrate Embryology*. Belgrade: Prosveta.
Lasko, P. F. and Ashburner, M. (1990). Posterior localization of vasa protein correlates with, but is not sufficient for, pole cell development. *Genes Dev.* **4**, 905-921.
Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T. R., Tomancak, P. and Krause, H. M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174-187.
Li, M., Hong, N., Xu, H., Yi, M., Li, C., Gui, J. and Hong, Y. (2009). Medaka *vasa* is required for migration but not survival of primordial germ cells. *Mech. Dev.* **126**, 366-381.
Liang, L., Diehl-Jones, W. and Lasko, P. (1994). Localization of vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. *Development* **120**, 1201-1211.
Liu, P. Z. and Kaufman, T. C. (2004). *hunchback* is required for suppression of abdominal identity, and for proper germband growth and segmentation in the intermediate germband insect *Oncopeltus fasciatus*. *Development* **131**, 1515-1527.
Lu, H. L., Tanguy, S., Rispe, C., Gauthier, J. P., Walsh, T., Gordon, K., Edwards, O., Tagu, D., Chang, C. C. and Jaubert-Possamai, S. (2011). Expansion of genes encoding piRNA-associated argonaute proteins in the pea aphid: diversification of expression profiles in different plastic morphs. *PLoS ONE* **6**, e28051.
Lynch, J. A. and Desplan, C. (2010). Novel modes of localization and function of *nanos* in the wasp *Nasonia*. *Development* **137**, 3813-3821.
Lynch, J. A., Ozüak, O., Khila, A., Abouheif, E., Desplan, C. and Roth, S. (2011). The phylogenetic origin of *oskar* coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola. *PLoS Genet.* **7**, e1002029.
Magnúsdóttir, E., Gillich, A., Grabole, N. and Surani, M. A. (2012). Combinatorial control of cell fate and reprogramming in the mammalian germline. *Curr. Opin. Genet. Dev.* **22**, 466-474.
Mahowald, A. P. (2001). Assembly of the *Drosophila* germ plasm. *Int. Rev. Cytol.* **203**, 187-213.
Medrano, J. V., Ramathal, C., Nguyen, H. N., Simon, C. and Reijo Pera, R. A. (2012). Divergent RNA-binding proteins, DAZL and VASA, induce meiotic progression in human germ cells derived *in vitro*. *Stem Cells* **30**, 441-451.
Mellanby, H. (1935). The early embryonic development of Rhodnius prolixus (Hemiptera, Heteroptera). *Q. J. Microsc. Sci.* **78**, 71-90.
Metschnikoff, E. (1866). Embryologische studien an insekten. *Z. Wiss. Zool.* **16**, 389-500.
Miura, T., Braendle, C., Shingleton, A., Sisk, G., Kambhampati, S. and Stern, D. L. (2003). A comparison of parthenogenetic and sexual embryogenesis of the pea aphid *Acyrthosiphon pisum* (Hemiptera: Aphidoidea). *J. Exp. Zool. B Mol. Dev. Evol.* **295**, 59-81.
Moczek, A. P. (2008). On the origins of novelty in development and evolution. *Bioessays* **30**, 432-447.
Nakao, H. (1999). Isolation and characterization of a *Bombyx vasa*-like gene. *Dev. Genes Evol.* **209**, 312-316.
Nakao, H., Hatakeyama, M., Lee, J. M., Shimoda, M. and Kanda, T. (2006). Expression pattern of *Bombyx vasa*-like (BmVLG) protein and its implications in germ cell development. *Dev. Genes Evol.* **216**, 94-99.
Nelsen, O. E. (1934). The segregation of the germ cells in the grasshopper, *Melanoplus differentialis* (Acrididae; Orthoptera). *J. Morphol.* **55**, 545-575.
Nieuwkoop, P. D. and Sutasurya, L. A. (1981). *Primordial Germ Cells In The Invertebrates: From Epigenesis To Preformation*. Cambridge: Cambridge University Press.
Nousch, M. and Eckmann, C. R. (2013). Translational control in the *Caenorhabditis elegans* germ line. *Adv. Exp. Med. Biol.* **757**, 205-247.
Ohashi, H., Umeda, N., Hirazawa, N., Ozaki, Y., Miura, C. and Miura, T. (2007). Expression of *vasa* (*vas*)-related genes in germ cells and specific interference with gene functions by double-stranded RNA in the monogenean, *Neobenedenia girellae*. *Int. J. Parasitol.* **37**, 515-523.
Özhan-Kizil, G., Havemann, J. and Gerberding, M. (2009). Germ cells in the crustacean *Parhyale hawaiensis* depend on Vasa protein for their maintenance but not for their formation. *Dev. Biol.* **327**, 230-239.
Panfilio, K. A. (2008). Extraembryonic development in insects and the acrobatics of blastokinesis. *Dev. Biol.* **313**, 471-491.
Pek, J. W., Anand, A. and Kai, T. (2012). Tudor domain proteins in development. *Development* **139**, 2255-2266.
Rangan, P., DeGennaro, M., Jaime-Bustamante, K., Coux, R.-X., Martinho, R. G. and Lehmann, R. (2009). Temporal and spatial control of germ-plasm RNAs. *Curr. Biol.* **19**, 72-77.
Richardson, B. E. and Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nat. Rev. Mol. Cell Biol.* **11**, 37-49.
Richter, J. D. and Lasko, P. (2011). Translational control in oocyte development. *Cold Spring Harb. Perspect. Biol.* **3**, a002758.
Rocak, S. and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. *Nat. Rev. Mol. Cell Biol.* **5**, 232-241.
Roonwal, M. L. (1937). Studies on the embryology of the african migratory locust, *Locusta migratoria migratoides* Reiche and Frm. (Orthoptera, Acrididae). II. Organogeny. *Philos. Trans. R. Soc. B* **227**, 175-244.
Salinas, L. S., Maldonado, E., Macías-Silva, M., Blackwell, T. K. and Navarro, R. E. (2007). The DEAD box RNA helicase VBH-1 is required for germ cell function in *C. elegans*. *Genesis* **45**, 533-546.
Salinas, L. S., Franco-Cea, A., Láscarez-Lagunas, L. I., Villanueva-Chimal, E., Maldonado, E. and Navarro, R. E. (2012). Germ cell survival in *C. elegans* and *C.*

remanei is affected when the DEAD box RNA helicases VBH-1 or Cre-VBH-1 are silenced. *Genesis* **50**, 801-818.
Sander, K. and Mirza, M. B. (1956). *The Early Embryology Of Pyrilla Perpusilla Walker (Homoptera), Including Some Observations On The Later Development (On Indian Insect Types, no. 4)*, p. 61. Aligarh: Aligarh Muslim University Publications.
Sasaki, J. and Nakashima, N. (2000). Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc. Natl. Acad. Sci. USA* **97**, 1512-1515.
Schmidt, E. D. and Dorn, A. (2004). Structural polarity and dynamics of male germline stem cells in the milkweed bug (*Oncopeltus fasciatus*). *Cell Tissue Res.* **318**, 383-394.
Schröder, R. (2006). *vasa* mRNA accumulates at the posterior pole during blastoderm formation in the flour beetle *Tribolium castaneum*. *Dev. Genes Evol.* **216**, 277-283.
Schüpbach, T. and Wieschaus, E. (1986). Maternal-effect mutations altering the anterior-posterior pattern of the *Drosophila* embryo. *Roux. Arch. Dev. Biol.* **195**, 302-317.
Schüpbach, T. and Wieschaus, E. (1991). Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129**, 1119-1136.
Schwalm, F. E. (1965). Zell – und mitosenmuster der normalen und nach rötgenbestrahlung regulierenden keimanlage von *Gryllus domesticus*. *Zeitschrift für Morphologie und Ökologie der Tiere* **55**, 915-1023.
Seidel, F. (1924). Die geschlechtsorgane in der embryonalentwicklung von pyrrhocoris apterus L. *Zeitschrift für Morphologie und Ökologie der Tiere* **1**, 429-506.
Sengupta, M. S. and Boag, P. R. (2012). Germ granules and the control of mRNA translation. *IUBMB Life* **64**, 586-594.
Shah, C., Vangompel, M. J. W., Naeem, V., Chen, Y., Lee, T., Angeloni, N., Wang, Y. and Xu, E. Y. (2010). Widespread presence of human *BOULE* homologs among animals and conservation of their ancient reproductive function. *PLoS Genet.* **6**, e1001022.
Spike, C., Meyer, N., Racen, E., Orsborn, A., Kirchner, J., Kuznicki, K., Yee, C., Bennett, K. and Strome, S. (2008). Genetic analysis of the *Caenorhabditis elegans* GLH family of P-granule proteins. *Genetics* **178**, 1973-1987.
Stebbings, H. and Hunt, C. (1987). The translocation of mitochondria along insect ovarian microtubules from isolated nutritive tubes: a simple reactivated model. *J. Cell Sci.* **88**, 641-648.
Stebbings, H., Sharma, K. and Hunt, C. (1985). Protein turnover in the cytoplasmic transport system within an insect ovary – a clue to the mechanism of microtubule-associated transport. *FEBS Lett.* **193**, 22-26.
Stephen, S., Talbot, N. J. and Stebbings, H. (1999). Poly(A) mRNA is attached to insect ovarian microtubules *in vivo* in a nucleotide-sensitive manner. *Cell Motil. Cytoskeleton* **43**, 159-166.
Stern, D. L. and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155-2177.
Styhler, S., Nakamura, A., Swan, A., Suter, B. and Lasko, P. (1998). *vasa* is required for GURKEN accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. *Development* **125**, 1569-1578.
Sunanaga, T., Watanabe, A. and Kawamura, K. (2007). Involvement of *vasa* homolog in germline recruitment from coelomic stem cells in budding tunicates. *Dev. Genes Evol.* **217**, 1-11.
Tanaka, S. S., Toyooka, Y., Akasu, R., Katoh-Fukui, Y., Nakahara, Y., Suzuki, R., Yokoyama, M. and Noce, T. (2000). The mouse homolog of *Drosophila Vasa* is required for the development of male germ cells. *Genes Dev.* **14**, 841-853.
Tomancak, P., Guichet, A., Zavorszky, P. and Ephrussi, A. (1998). Oocyte polarity depends on regulation of *gurken* by Vasa. *Development* **125**, 1723-1732.
Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E. et al. (2002). Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **3**, research0088.1-research0088.14.
Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S. E. and Rubin, G. M. (2007). Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **8**, R145.
Tsunekawa, N., Naito, M., Sakai, Y., Nishida, T. and Noce, T. (2000). Isolation of chicken *vasa* homolog gene and tracing the origin of primordial germ cells. *Development* **127**, 2741-2750.
Wang, C. and Lehmann, R. (1991). Nanos is the localized posterior determinant in Drosophila. *Cell* **66**, 637-647.
Webster, P. J., Liang, L., Berg, C. A., Lasko, P. and Macdonald, P. M. (1997). Translational repressor *bruno* plays multiple roles in development and is widely conserved. *Genes Dev.* **11**, 2510-2521.
Wheeler, W. M. (1893). A contribution to insect embryology. *J. Morphol.* **8**, 1-161.
Will, L. (1888). Entwicklungsgeschichte der viviparen aphiden. *Zoologische Jahrbücher* **3**, 201-280.
Witlaczil, E. (1884). Entwicklungsgeschichte der aphiden. *Z. Wiss. Zool.* **40**, 559-696.
Wu, H.-R., Chen, Y.-T., Su, Y.-H., Luo, Y.-J., Holland, L. Z. and Yu, J.-K. (2011). Asymmetric localization of germline markers *Vasa* and *Nanos* during early development in the amphioxus *Branchiostoma floridae*. *Dev. Biol.* **353**, 147-159.
Yajima, M. and Wessel, G. M. (2011). The multiple hats of Vasa: its functions in the germline and in cell cycle progression. *Mol. Reprod. Dev.* **78**, 861-867.
Yeates, D. K., Cameron, S. L. and Trautwein, M. (2012). A view from the edge of the forest: recent progress in understanding the relationships of the insect orders. *Aust. J. Entomol.* **51**, 79-87.
Ying, M. and Chen, D. (2012). Tudor domain-containing proteins of *Drosophila melanogaster*. *Dev. Growth Differ.* **54**, 32-43.
Yoon, C., Kawakami, K. and Hopkins, N. (1997). Zebrafish *vasa* homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells. *Development* **124**, 3157-3165.
Zeng, V. and Extavour, C. G. (2012). ASGARD: an open-access database of annotated transcriptomes for emerging model arthropod species. *Database* **2012**, bas048.
Zhao, G., Chen, K., Yao, Q., Wang, W., Wang, Y., Mu, R., Chen, H., Yang, H. and Zhou, H. (2008). The *nanos* gene of *Bombyx mori* and its expression patterns in developmental embryos and larvae tissues. *Gene Expr. Patterns* **8**, 254-260.
Zhurov, V., Terzin, T. and Grbić, M. (2004). Early blastomere determines embryo proliferation and caste fate in a polyembryonic wasp. *Nature* **432**, 764-769.

Biology Open

Supplementary Material

Ben Ewen-Campen et al. doi: 10.1242/bio.20134390

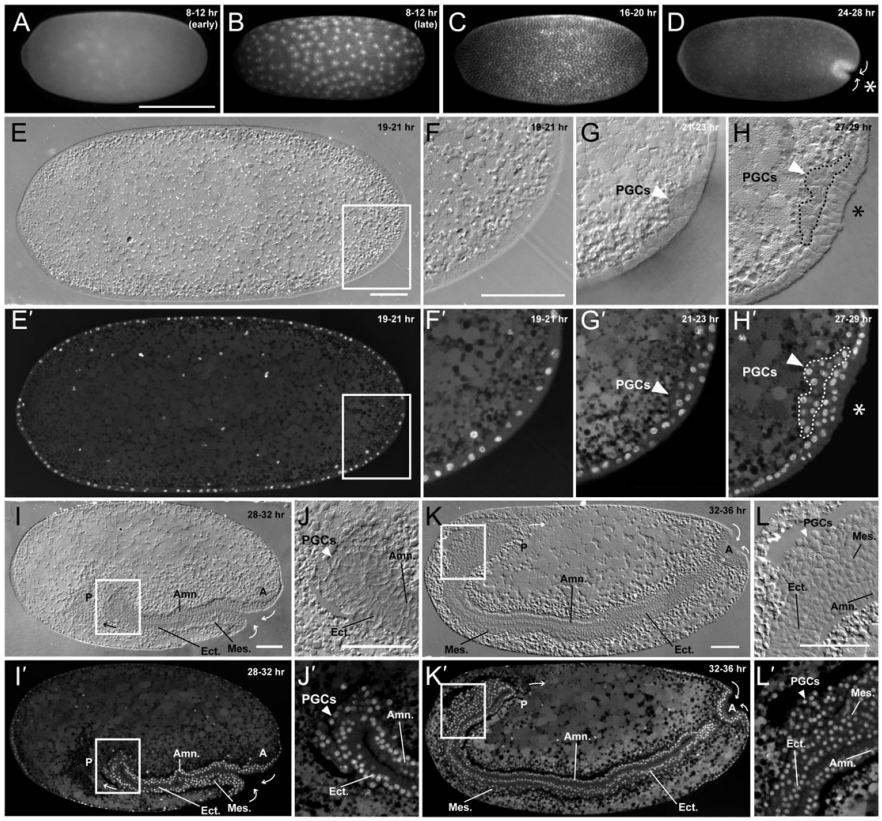


Fig. S1. Early embryogenesis of *Oncopeltus fasciatus* and morphological identification of putative PGCs. (**A–C**) The *Oncopeltus* syncytial blastoderm forms as a single layer of nuclei spread evenly across the surface of the yolk. (A) 8–12 hours after egg laying (h AEL) syncytially dividing nuclei are visible beneath the yolk surface. (B) Energid nuclei populate the yolk surface within 12 hours, and (C) undergo repeated mitosis and cellularisation to form a uniform cellular blastoderm by approximately 20 hours. Consistent with previous reports (Butt, 1949), we did not detect pole cell-like cells at any time during syncytial blastoderm or early cellular stages. (**D**) By 24–28 h AEL the posterior of the embryo has begun to invaginate into the yolk (arrows indicate direction of embryonic movements), forming the posterior pit (asterisk) where gastrulation takes place. This embryonic invagination is the beginning of the axial elongation process that will create the abdominal segments (Liu and Kaufman, 2004; reviewed by Panfilio, 2008). Immediately before posterior pit formation (~21 hours AEL), we observed putative PGCs on the inner surface of the blastoderm adjacent to the yolk (**E–H′**). (E) Medial section of a 19–21 h AEL embryo viewed with DIC optics and (E′) stained with Sytox Green to reveal nuclei. Boxed region is enlarged in (F–H) and (F′–H′). (F,F′) In 19–21 h AEL embryos, the early blastoderm is single-layered. (G,G′) Between 21–23 h AEL, the embryonic posterior becomes multilayered, and the first cells visible within the yolk mass are the presumptive PGCs (arrowheads). (H,H′) By 27–29 h AEL the putative PGCs (arrowhead) have fully entered the yolk. These putative PGCs are visible as a mesenchymal cluster with large round, centrally located nuclei, directly adjacent to the epithelialized somatic cells of the posterior blastoderm (asterisk), which are columnar in shape with smaller, basally located nuclei. (**I,I′**) As the germ band elongates and its posterior end invaginates into the yolk at 28–32 h AEL (arrows indicate direction of movement), the putative PGCs remain in a mesenchymal cluster at the germ band posterior. During early stages of germ band elongation (28–32 hours AEL), ongoing gastrulation produces mesodermal cells on the dorsal surface of the ectoderm (**I–L′**). The single-layered amnion (*Amn.*) is ventral to the ectoderm (*Ect.*); anterior mesoderm (*Mes.*) is on the dorsal surface of the ectoderm. Boxed region is enlarged in (J,J′). (J,J′) Putative PGCs (arrowhead) form a cluster of cells at the posterior of the germ band, distinct from the adjacent ectoderm and amnion. (K,K′) By 32–36 h AEL the embryo has nearly completed germ band elongation and its posterior end begins to curl towards the anterior of the egg within the yolk (arrows indicate direction of movement). Mesoderm now extends along its entire anteriorposterior extent. Boxed region is enlarged in (L,L′). (L,L′) Putative PGCs (arrowhead) remain in a distinct cluster dorsal to the mesoderm and begin to migrate anteriorly along the dorsal surface of the mesoderm (see Fig. 3). Scale bars: 500 µm in A–D, 100 µm in E–L′. Egg anterior is to the left in all panels.

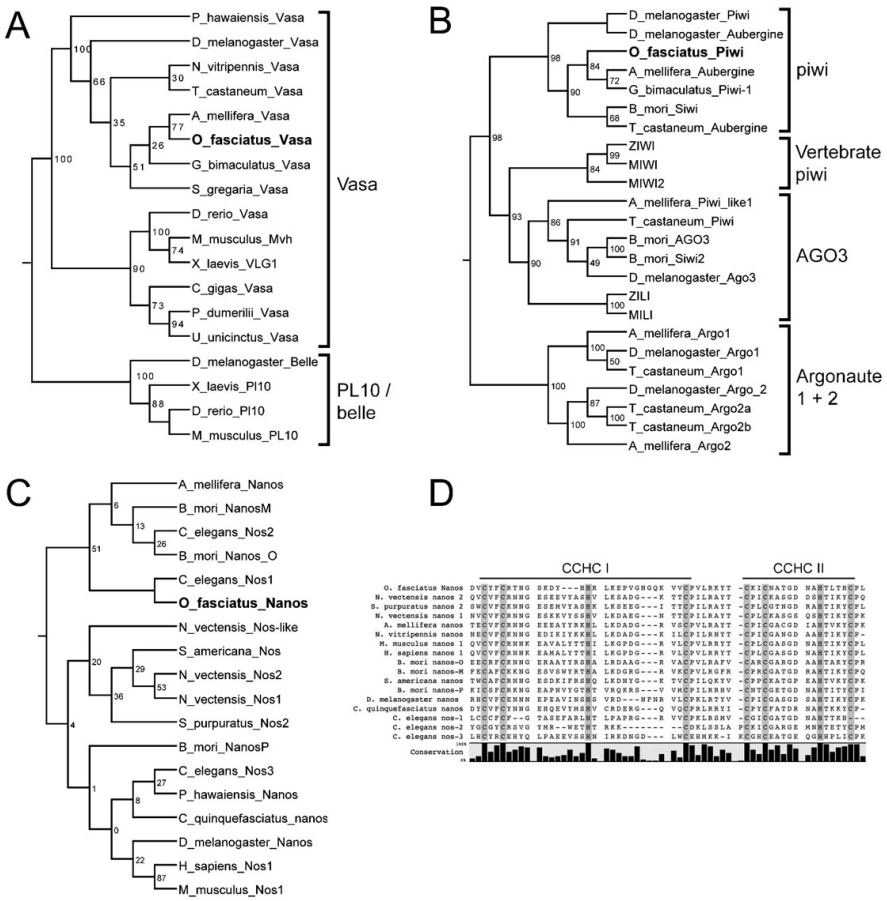


Fig. S2. Phylogenetic analysis of *vasa*, *piwi*, and *nanos*. Best-scoring maximum likelihood cladograms are shown with bootstrap values from 2000 replicates at nodes. (**A**) *Oncopeltus* Vasa is a member of the Vasa family of RNA helicases, not the closely related PL10/belle proteins. (**B**) *Oncopeltus* Piwi is a member of the *piwi* clade of PIWI proteins, closely related to *Drosophila* Piwi and Aubergine. (**C**) Phylogenetic reconstruction fails to resolve the internal relationships of *nanos* genes, because the conserved region of these proteins suitable for alignment (48 amino acids) is too short to provide sufficient phylogenetic signal. (**D**) A protein alignment of *Oncopeltus* Nanos protein with known orthologues demonstrates the presence of the diagnostic 2×(CCHC) zinc finger domain.

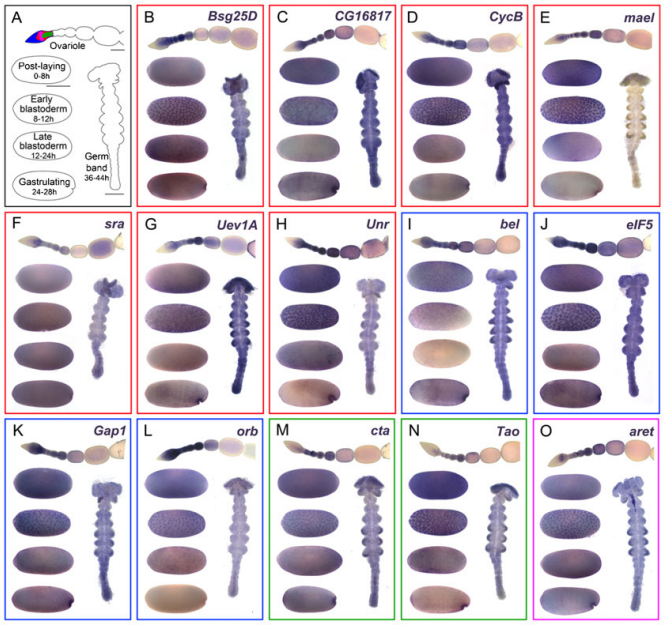


Fig. S3. Expression of additional germ plasm candidate markers in *Oncopeltus* ovaries and embryos. Transcripts were chosen for analysis based on their expression in the germ plasm and PGCs of *Drosophila melanogaster* (supplementary material Table S2). (**A**) Schematic figure showing the tissues depicted in subsequent panels. Embryonic ages shown in hours AEL. Coloured shading in ovariole schematic indicates spatial expression pattern of genes shown in boxes outlined in the corresponding colours. Blue = throughout entire tropharium in all nurse cells, as well as oogonia and resting oocytes; red = posterior nurse cells, oogonia and resting oocytes; green = oogonia and resting oocytes but absent from or very low in only posterior nurse cells of tropharium; magenta = posterior nurse cells of tropharium but not oogonia or resting oocytes. (**B–O**) Expression patterns of genes studied in ovaries (top of each panel), blastoderm stages from 0–28 hours AEL (arranged vertically along the left of each panel), and in mid-germ band stages (to right of each panel), when PGCs are easily discernable in embryos stained for *vasa*, *tudor* or *boule* (Figs 3, 4). None of the 14 genes shown here were asymmetrically localised within oocytes, or to PGCs in later stages of development. (O) *aret* was strongly expressed in a population of cells located at the posterior, dorsal surface of the head at germ band stages, perhaps implicating this gene in foregut development. Scale bars: 500 µm for ovarioles and non-germ band embryos, 200 µm for germ band embryos.

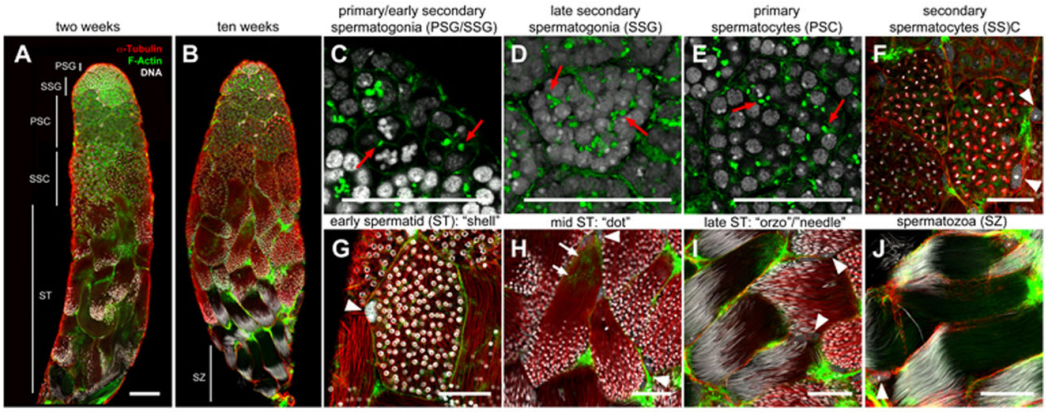


Fig. S4. Spermatogenesis in wild type *Oncopeltus* adult testioles. (**A**) Sperm tubule (testiole) from a two-week old male. Germ line stem cells (primary spermatogonia) and their putative niche are located at the anterior apex of each testiole. Cysts of clonally related secondary spermatogonia, spermatocytes, spermatids and spermatozoa are arranged in order posterior to the niche. All cells of a given cyst proceed synchronously through all stages of spermatogenesis, and all cysts at the same position along the anterior–posterior axis of the testiole are also roughly synchronised with each other (Economopoulos and Gordon, 1971). PSG: primary spermatogonia; SSG: secondary spermatogonia; PSC: primary spermatocytes; SSC: secondary spermatocytes; ST: spermatids undergoing spermiogenesis; SZ: spermatozoa. (**B**) Testiole from a ten week-old male. All stages of spermatogenesis continue to progress normally, although a greater number of mature spermatodesms are present. (**C**) Primary spermatogonia divide mitotically to form cysts of two to eight cells, and remain connected by actin-rich cytoplasmic bridges (red arrows). (**D**) Secondary spermatogonia undergo six synchronous mitotic transit amplifying divisions to produce cysts of 64 nuclei that retain cytoplasmic bridges (green; red arrows) (Economopoulos and Gordon, 1971). (**E**) Primary spermatocytes undergo the first meiotic division to produce 128 clonally related diploid cells, still connected by cytoplasmic bridges (red arrows). (**F**) Secondary spermatocytes undergo the second meiotic division synchronously; two cysts at anaphase (left) and metaphase (right) are shown. Each cyst is accompanied by a single large somatic sheath cell (arrowheads). (**G**) Nuclei of early "shell stage" spermatids appear hollow or shell-shaped (Kaye and McMaster-Kaye, 1966) and begin to develop elongated tubulin-rich tails (red); cytoplasmic bridges are no longer present. (**H**) Mid-stage ("dot stage") spermatid nuclei condense (top cyst) and begin to elongate; tubulin-rich tails continue to elongate and actin-rich elongation complexes proceed posteriorly along the growing sperm tail (white arrows). (**I**) Late stage ("orzo" and "needle" stage) spermatid nuclei are further elongated. (**J**) Mature spermatozoa remain associated in spermatodesms containing all clonally related products of a single primary spermatogonium. Arrowheads in (F–J) indicate somatic sheath cells that are associated with each cyst. White = nuclei (Hoechst 33342), green = F-actin (FITC-phalloidin), red = anti-alpha Tubulin. Scale bars: 100 µm in A (applies also to B); 50 µm in C–J. Anterior is to the top in all panels.

Table S1. Primers used for degenerate PCR, in situ hybridisation probe preparation, dsRNA preparation for RNAi, and RT-PCR. Sequence lengths are in nucleotides. F = forward; R = reverse.

Gene Name	degenerate primers		In situ probe length	Primers for in situ probe fragment (5′ to 3′)		dsRNA length	dsRNA primers (5′ to 3′)		RT-PCR amplicon length	RT-PCR primers (5′ to 3′)	
vasa	F	CCGATCGCATGCTGGAYATGGGNTT	1403	F	AAAAGGACTGGCAATGATGG	601	F	TGAGAGTATGACGAC	1392	F	GGAAGAGAAAGGGGACAAGG
	R	GGTGCGGCCGATGCKRTGNACRTA		R	AAACCTGGATCCCCAAATTC		R	TCCCGTCTGTTCAAGAATCC		R	TCCCGTCTGTTCAAGAATCC
boule			1287	F	ATTGAGGCACCAACTTCGAT	716	F	AGCCTCACCACCAGTATTCG	614	F	ATTGAGATGAAACCTCCGGCCT
				R	AGGGTGCCTAGGATTGGACT		R	AGGGTGCCTAGGATTGGACT		R	AGTTCAGTGCCTCAGGGAAA
tudor			1215	F	GGTTAGCAAGCCTTGGAGTG	802	F	CCGAGAGTGCTCAAGTTTCC	1876	F	TGCTTCCAGTTGGTTCTTCC
				R	CACACCTGTTGCCATAATCG		R	AACTTGGTACGCCTGTGGTC		R	CCACCAAAATCGCTTCTCATT
nanos			843	F	GAAGGAACCCGTAGGGAA						
				R	ATAATCCCTGAAGTAGTGC						
piwi			1023	F	TGAAGAAGTCAGAGCCAG						
				R	GATTGAGAGACGAGAAAGAA						
arrest (aka *bruno*)			703	F	ATGTTCACTGCCCCTGGTAG						
				R	TACAGTGCCATACGGTTGGA						
belle			725	F	GGGTTGAGGAGCAAGACAAA						
				R	GCTCTGGtTTGCGGAGTAAG						
beta-tubulin									199	F	TGATCCCTTACCTCGACACA
										R	CCCGCAAGGAAATCACTCT
Blastoderm-specific - gene 25D (aka *Bsg25D*)			821	F	AGCTGGTGGAACTCCAGAGA						
				R	TCAGTTCCTCCGAGTGCTTT						
CG16817			717	F	GCCATAGCTGGTTTCTCCAG						
				R	GGGTCGAGGTCATCAAAAGA						
concertina			795	F	GGGAAGTCAACGTTCCTCAA						
				R	TTGAAGAAGAAAGGCCTGGA						
Cyclin B			913	F	GTAAGGGAGAAGTTTTGG						
				R	TGTATATGAGGTGTAGTGG						

Table S1. Continued.

Gene Name	degenerate primers	In situ probe length	Primers for in situ probe fragment (5′ to 3′)		dsRNA length	dsRNA primers (5′ to 3′)	RT-PCR amplicon length	RT-PCR primers (5′ to 3′)
elongation initiation factor 5 (aka *eIF5*)		790	F	GGTTCTCCCAAATCTGACGA				
			R	TGTCCAGGTCATCTCCTTCC				
GTPase-activating protein 1 (aka *Gap1*)		748	F	AAAGTGGCCATTAAGCGAGA				
			R	AGCCATGGTGAAGAAACACC				
maelstrom		543	F	GCAGTACGTTGTGCTTTGGA				
			R	AACAGGATCGgCATCAAATC				
oo18 RNA-binding protein (aka *orb*)		772	F	TCTCGGTGATGGTGTTTTCA				
			R	TGCTGTGAGGTGTAGCCTTG				
sarah		351	F	TAAACCGGAATTGGACGTTT				
			R	GATGTTGGCAAATCTTCATCA				
Tao		746	F	TCAGGCTCTCCTCCGTTTTA				
			R	AGCTCCCATGGCCTATTTCT				
Uev1A		792	F	TTTCTGCCATAGCCCATTTT				
			R	GCAACAAGTTTCCAGATGTCC				
Upstream of N-ras		848	F	AAACCCATGAGCCTTCACAG				
			R	GCCCTCTCAATGAATCCAAA				

Table S2. Genes included in *Oncopeltus* germ plasm in situ screen.

Drosophila Gene Name	*Drosophila* gene symbol	*Drosophila* CG #	Transcript Expression in *Drosophila* Germ Plasm and PGCs: Germ Plasm	Pole Cells	Stage 9 PGCs	*Drosophila* germ line mutant phenotype (Molecular function)	Functional conservation outside *Drosophila*	References
Boule	*bol*	CG4760	No	No	No	Spermatogenesis defects (RNA binding)	Germ line expression/function across bilateria, often specific to spermatogenesis	[Lécuyer et al., 2007; Shah et al., 2010]
tudor	*tud*	CG9450	No*	No*	N.D.	Pole cell formation defects (tudor domain protein)	Germ line expression/function across bilateria	[Golumbeski et al., 1991; Bardsley et al., 1993; Ewen-Campen et al., 2010]
orb	*orb*	CG10868	Yes	Yes	No	Oogenesis defects (RNA binding)	N.D.	[Lécuyer et al., 2007]
sarah	*sra*	CG6072	Yes	Yes	No	Oogenesis defects (Calcineurin regulation)	N.D.	[Lécuyer et al., 2007]
Cyclin B	*CycB*	CG3510	Yes	Yes	N.D.	Fertility defects in both sexes (Cyclin protein)	N.D.	[Lécuyer et al., 2007]
arrest (bruno)	*aret*	CG31762	Yes	Yes	Yes	Oogenesis defects (RNA binding)	N.D.	[Lécuyer et al., 2007]
concertina	*cta*	CG40010	Yes	Yes	Yes	No germ line phenotype reported (G-protein alpha subunit)	N.D.	[Lécuyer et al., 2007]
Gap1	*Gap1*	CG6721	Yes	Yes	Yes	No germ line phenotype reported (PH & C2-domain, Ras GTPase activation)	N.D.	[Lécuyer et al., 2007]
eIF5	*eIF5*	CG9177	Yes	Yes	Yes	No germ line phenotype reported (translation initiation)	N.D.	[Lécuyer et al., 2007]
Blastoderm-specific gene 25D	*Bsg25D*	CG14025	Yes	Yes	Yes	No germ line phenotype reported (Unknown)	N.D.	[Lécuyer et al., 2007]
Uev1A	*Uev1A*	CG10640	Yes	Yes	Yes	No germ line phenotype reported (ubiquitin-protein ligase)	N.D.	[Lécuyer et al., 2007]
CG16817	–	CG16817	Yes	Yes	Yes	No germ line phenotype reported (Unknown)	N.D.	[Lécuyer et al., 2007]
Tao	*Tao*	CG14217	Yes	Yes	Yes	No germ line phenotype reported (Protein S/T kinase)	N.D.	[Lécuyer et al., 2007]
Upstream of N-ras	*Unr*	CG7015	Yes	Yes	N.D.	No germ line phenotype reported (RNA and protein binding)	N.D.	[Lécuyer et al., 2007]
Belle	*bel*	CG9748	N.D.	Yes	No	Oogenesis and spermatogenesis defects (ATP-dependent RNA helicase)	N.D.	[Tomancak et al., 2007]
maelstrom	*mael*	CG11254	No	Yes	Yes	Oogenesis and spermatogenesis defects (HMG-box DNA binding)	Germ line function in mouse	[Lécuyer et al., 2007]

*Tudor protein is localised to both pole plasm and pole cells in *Drosophila*. N.D. = no data available.

Table S3. Effects of RNAi on *Oncopeltus* PGC formation.

RNAi	Total # Scored	# embryos with non-specific defects* (%)	# surviving embryos with PGCs (%)
DsRed	39	4 (10.3%)	34 (97.1%)
vasa	16	0	15 (93.8%)
tudor	20	0	20 (100%)
vasa and *tudor*	19	9 (47.4%)	10 (100%)
boule	22	18 (81.8%)	4 (100%)

*"Non-specific defects" includes failure to develop a germ band as well as the formation of grossly defective germ bands, both of which ultimately resulted in embryonic lethality before 40–54 hours AEL, the time when we scored for PGCs. These embryos were not scored for PGC presence/absence.

CHAPTER 5

oskar functions in adult neural stem cells to influence long-term memory formation in the cricket *Gryllus bimaculatus*

Ben Ewen-Campen[1], Ryo Wakuda[2], Kanta Terao[2], Yukihisa Matsumoto[2,3], Makoto Mizunami[4], and Cassandra Extavour[1]

Author Affiliations:

1. Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA
2. Graduate School of Life Science, Hokkaido University, Japan
3. Faculty of Liberal Arts, Tokyo Medical and Dental University, Japan
4. Faculty of Science, Hokkaido University, Japan

ABSTRACT

Although *oskar* was first described for its role in the *Drosophila* germ line, recent studies have shown that this gene has additional functions in the nervous system of larval *Drosophila* and in the embryos of a basally-branching insect, the cricket *Gryllus bimaculatus*. However, the specific molecular functions of *oskar* in the nervous system of either species remain unclear. In this study, we show that RNAi against *Gb-oskar* impairs long-term olfactory memory, but not short-term memory, in the cricket. We then show that *Gb-oskar* is expressed in the adult brain in a cluster of neuroblasts that are responsible for adult neurogenesis in the mushroom body, the anatomical substrate of olfactory memory in insects. Previous studies have shown that killing these mushroom body neuroblasts specifically impairs olfactory learning. Thus, our results are consistent with the hypothesis that *Gb-oskar* is involved in adult neuroblast function (i.e. proliferation, maintenance, or survival), and that its role in long-term olfactory memory is mediated through these cells. We discuss these results in both a functional and an evolutionary

context, and propose necessary additional experiments to directly test the role of *Gb-oskar* in adult neuroblasts.

INTRODUCTION

The *oskar* gene was first discovered in *Drosophila melanogaster*, where it is both necessary and sufficient to specify embryonic germ cells and recruit the posterior determinant *nanos* (Ephrussi et al., 1991; Ephrussi and Lehmann, 1992; Lehmann and Nüsslein-Volhard, 1986). Interestingly, *oskar* and several other genes originally identified as *Drosophila* germ line genes, including *nanos*, *pumilio*, *staufen, orb*, and the PIWI proteins *aubergine* and *AGO3*, have since been shown to have a variety of roles in the nervous system (Dubnau et al., 2003; Pai et al., 2013; Perrat et al., 2013; Wharton et al., 1998; Xu et al., 2013; Ye et al., 2004). Furthermore, a role for *oskar* in the nervous system, but not in the germ line, is conserved between *Drosophila* and a basally-branching hemimetabolous insect, the cricket *Gryllus bimaculatus*, suggesting that a neural function of *oskar* may in fact be ancestral (Ewen-Campen et al., 2012). However, the precise role(s) of *oskar* in the nervous system remain largely unknown.

The biochemical functions of some of the *Drosophila* "germ line genes" are well characterized, allowing for relatively detailed models of their function in neurons. For example, studies of Nanos (Nos) and Pumilio (Pum) during *Drosophila* embryogenesis have revealed that Pum binds to specific sequences in the 3'UTR of target mRNAs via the highly conserved PUF domain, and subsequently recruits Nos to form a translational repression complex (Murata and Wharton, 1995; Sonoda and Wharton, 1999; Zamore et al., 1997). In the *Drosophila* nervous system, Pum has been shown to function at neuromuscular junctions as a translational regulator (Mee et al., 2004; Menon et al., 2004) and in long-term memory formation (Dubnau et al., 2003), and to act together with Nanos in larval dendrite morphogenesis (Xu et al., 2013; Ye et al., 2004). A variety of relevant target neuronal mRNAs regulated by Pum have been identified at the neuromuscular junction, including the voltage-gated sodium channel *paralytic* (Mee et al., 2004) and the translation factory *eIF-4E* (Menon et al., 2004), and in the adult brain, including a membrane-associated guanylate kinase, *dlg1* (G. Chen et al., 2008). There is also evidence that

the neural roles of Nos and Pum are phylogenetically widespread: a Pum ortholog, Pumilio-2, has been shown to function mammalian neurons (Driscoll et al., 2013), and a role for Nos in the development of neurons has been functionally demonstrated in a cnidarian (Kanska and Frank, 2013).

In contrast to Nos and Pum, the biochemical function of Oskar remains largely unknown, despite decades of research (reviewed in Ewen-Campen et al., 2010). It is thus unclear how this protein may function in neurons. *oskar* is a novel, insect-specific gene, and contains two predicted protein domains, both of unknown function: a LOTUS (aka OST-HTH) domain and a SGNH hydrolase domain (Anantharaman et al., 2010; Callebaut and Mornon, 2010; Ewen-Campen et al., 2012; Lynch et al., 2011). The LOTUS/OST-HTH domain has been predicted to bind RNA (Anantharaman et al., 2010; Callebaut and Mornon, 2010), although Oskar has never been directly shown to bind RNA. The SGNH hydrolase domain belongs to a large family of lipid-processing enzymes, but, enigmatically, this domain in Oskar is predicted to be catalytically inactive (Anantharaman et al., 2010). A third domain, termed Long Osk, is found only in *Drosophilid* insects, and also has unknown biochemical function (Ewen-Campen et al., 2012; Lynch et al., 2011). The current model is that Oskar serves as a scaffolding protein, facilitating the assembly of ribonucleoprotein complexes that include a variety RNA-processing components (Jones and Macdonald, 2007; Suyama et al., 2008).

Despite the unknown molecular function of Oskar, there is evidence that this gene functions in the nervous system of *Drosophila* as well as in a basally branching hemimetabolous insect, the cricket *Gryllus bimaculatus*. In *Drosophila*, larvae that are mutant for *oskar* or express neuron-specific *oskar* RNAi display defects in *nanos* localization, ultimately leading to a defect in dendrite morphogenesis and an associated defect in motor response to mechanical stimulation (Xu et al., 2013). In addition, Dubnau *et al*. (2003) have reported a role for *oskar* in long-term memory formation based on an insertional pGal4 mutant upstream of *oskar* (the *norka* mutant), but it is unclear if *norka* is a *bona fide* allele of *oskar* (Xu et al., 2013) (See also Appendix 2 of

this dissertation). In the cricket, we have previously shown that *Gb-oskar* is expressed in embryonic neuroblasts, and that RNAi against *oskar* disrupts the development of the central nervous system (Ewen-Campen et al., 2012). Thus, although *oskar* functions in the nervous system of both species, it is unclear if it plays similar or divergent roles in each.

In the present study, we address whether *Gb-oskar* has an additional neural function in the adult brain of the cricket. We first show that RNAi against *Gb-oskar* in adult crickets impairs long-term memory formation in an olfactory associative learning assay. Next, we show that *Gb-oskar* is expressed in a well-characterized group of mushroom body neuroblasts responsible for adult neurogenesis. Given that these mushroom body neuroblasts have previously been implicated in olfactory learning in related species (Scotto-Lomassese et al., 2003), we hypothesize that *Gb-oskar* is involved in the survival and/or proliferation of neuroblasts in the adult brain. We propose future experiments to directly test whether the role of *Gb-oskar* in olfactory learning involves adult neurogenesis.

METHODS

Gryllus husbandry

For behavior experiments, *Gryllus bimaculatus* were maintained in the Mizunami laboratory at 27ºC on a 12:12 light cycle, with a diet of insect food pellets, as previously described (Matsumoto and Mizunami, 2000). For gene expression analysis, qPCR, and cell proliferation experiments, crickets were maintained in the Extavour laboratory at 28ºC on a 12:12 light cycle, with a diet of grain and cat food, as previously described (Kainz et al., 2011).

RNAi

Adult male crickets within one week of their final moult were injected 2 µL of 10µM dsRNA through a hole pierced in the median ocellus using a 10 µL syringe fitted with 26S gauge tip (Hamilton Inc., Nevada, USA). Behavioral tests were repeated using two non-overlapping fragments of *Gb-oskar* (742 bp and 503 bp) with a 678 bp fragment of DsRed, as a negative control (Ewen-Campen et al., 2012).

qPCR

Two days after dsRNA injection, brains were dissected in ice-cold PBS, then immediately homogenized in TRIzol (Life Technologies). Total RNA was extracted from a total of six brains per treatment, following the manufacturer's instructions, including a 30 minute DNAse treatment. 1 µg of total RNA was used as template for cDNA synthesis using SuperScript III (Life Technologies) with oligo-dT primers. cDNA was diluted 1:10 prior to qPCR, and 6 µL of template was used per 25 µL qPCR reaction. (PerfeCta SYBR Green SuperMix, Low ROX, Quanta Biosciences). qPCR reactions were conducted in triplicate, and fold change was calculated using the delta-delta Ct method (Livak and Schmittgen 2001), with standard deviation propagated following standard methods. Beta-tubulin was used as a reference gene (Kainz et al., 2011). Primers amplifying a 234 fragment of *Gb-oskar* (F: TTGTTGACCATTCCCTTCCT, R: ACTCCACAACACCACTCC) and a 166 bp fragment of Beta-tubulin (F: TGGACTCCGTCCGGTCAGGC, R: TCGCAGCTCTCGGCCTCCTT) (Kainz et al., 2011) were used.

Behavioral analysis

Adult male crickets at one week after final moult were used in the experiments. Three days before conditioning, individual crickets were separated into 100-mL beakers and deprived of

drinking water to enhance their motivation to search for water. Two days before a conditioning, each cricket was injected with dsRNA as described above.

Two days after dsRNA injection, each cricket was subjected to an odor preference test, in which the animal was allowed to freely visit peppermint and vanilla odors (Matsumoto and Mizunami, 2002). The time spent at each of the peppermint and vanilla source was measured cumulatively to evaluate relative odor preference (Matsumoto and Mizunami, 2002).

Crickets were subjected to 4-trial conditioning, in which an odor was paired with water reward, with inter-trial interval of 5 min (Takahashi et al., 2009). For conditioning, a small filter paper was attached to the needle of a hypodermic syringe. The syringe was filled with water reward (unconditioned stimulus: US^{+}), and the filter paper was soaked with peppermint essence (conditioned stimulus: CS).

At one hour and one day after the end of the conditioning, each cricket was subjected to an odor preference test. Relative odor preference of each animal was measured using the preference index (PI) for rewarded odor (peppermint), defined as $t_P/(t_P+t_V)$ *100 (%), where t_P is the time spent exploring the peppermint source and t_V is the time spent exploring the vanilla source. Wilcoxon's (WCX) test was used to compare odor preferences before and after training. For multiple comparisons, Holm's method was used to adjust the significance level.

Gene expression analysis

Previously, we generated an antibody against *Gb-oskar*, and showed that this antibody recognizes a band of the correct size in *Gryllus* ovaries, and that the neuroblast-specific signal observed in embryos is abolished following *Gb-oskar* RNAi (Ewen-Campen et al., 2012). However, when we tested this antibody on *Gryllus* brains via Western blot, we observed that this antibody strongly cross-reacted with a ~23-24 kDa protein that is not present in ovaries or embryos. RNAi against *Gb-oskar* in adult brains did not reduce the intensity of this band on

Western blot, and Mass spectrometry of an excised band of 23-24 kDa size from an SDS-PAGE separation of *Gryllus* brain lysate did not identify any peptides corresponding to *Gb-oskar*. Thus, we conclude that, unlike in ovaries and embryos, this antibody recognizes something other than GbOsk protein in *Gryllus* brains. For this reason, we used *in situ* hybridization to detect *Gb-oskar* expression in brains.

For *in situ* hybridization, brains were dissected in ice-cold PBS, and de-sheathed following one hour in 4% paraformaldehyde, followed by an additional overnight fixation at 4ºC or 3-4 hours at room temperature. Brains were then stored in methanol overnight at room temperature, due to the surprising observation that storage at -20ºC greatly increased background tracheal staining. *Gb-oskar* was detected using a 788 bp probe, following standard protocols (Kainz et al., 2011), with the following modifications to reduce background: a 20 minute Proteinase K treatment followed by a 20-30 minute fix in 0.8% glutaraldehyde and 4% paraformaldehyde. The *Gb-oskar* probe was used at 1.0 ng/µl, and hybridized at 69-70ºC.

EdU assay

Cell proliferation was assayed using the Click-iT EdU Alexa 488 kit (Life Technologies). Crickets were injected with 15 µl of EdU either into the abdomen or into the head capsule through the median ocellus (both methods successfully labeled dividing neuroblasts), and cell proliferation was detected following manufacturer's instructions. For tissue double-stained by *in situ* hybridization and EdU, the *in situ* hybridization was conducted before the visualization of incorporated EdU.

Vibratome sectioning

Gryllus brains were embedded in 4% low-melt agarose, and sectioned at 50-90 µm using a Leica VT1000S vibratome. For antibody staining, brains were sectioned prior to incubation

with the primary antibody. For *in situ* hybridization, brains were sectioned after staining had been completed.

RESULTS

Gb-oskar *RNAi impairs long-term olfactory memory*

Crickets have robust olfactory learning capabilities that can be experimentally investigated by training crickets to associate an novel odor with a water reward (Matsumoto and Mizunami, 2000). A behavioral preference for the rewarded odor can be detected one hour after training, and as few as three training sessions are sufficient to form a memory that is that does not significantly decay between 1-7 days (Matsumoto and Mizunami, 2000). In addition, the injection of dsRNA into the hemolymph of adult crickets has been shown to trigger a systemic reduction in target gene levels, and has been used successfully to interfere with the function of genes known to be involved in learning and memory in *Gryllus bimaculatus* (Takahashi et al., 2009). Importantly, because crickets are not injected with dsRNA until adulthood, this approach bypasses any potential developmental requirements for a given gene.

In order to test whether *Gb-oskar* is required in adults for olfactory memory in *Gryllus*, we assayed the olfactory memory of *Gb-oskar* RNAi adult male crickets at one hour and one day after training, compared to control crickets injected with an equal amount of DsRed dsRNA. We used qPCR to quantify RNAi efficiency, and found that *Gb-oskar* levels were reduced to 66.4% of control levels (**Figure 5.1A**). Although this reduction is modest compared to the reductions seen for many maternal or zygotic RNAi experiments (e.g. Ewen-Campen et al., 2013), it is comparable to reported knockdown levels of other genes in the *Gryllus* brain in adult crickets (Takahashi et al., 2009).

In DsRed RNAi control crickets, four training sessions led to a significant short-term preference for the rewarded hour at one hour after training, and this preference was retained after

one day (**Figure 5.1B**). Strikingly, although *Gb-oskar* RNAi crickets formed a short-term preference for the rewarded odor at one hour after training, this memory was lost by one day after training (**Figure 5.1C**). To ensure the specificity of this knockdown, we repeated these experiments using a non-overlapping fragment of *Gb-oskar* dsRNA, and observed a similar result (**Figure 5.1D**). Thus, *Gb-oskar* RNAi impairs long-term, but not short-term, olfactory memory formation. The fact that short-term olfactory learning remains intact in *Gb-oskar* suggests that the effect of *Gb-oskar* RNAi does not globally disrupt such processes as olfaction or locomotion, but is instead specific to long-term memory.

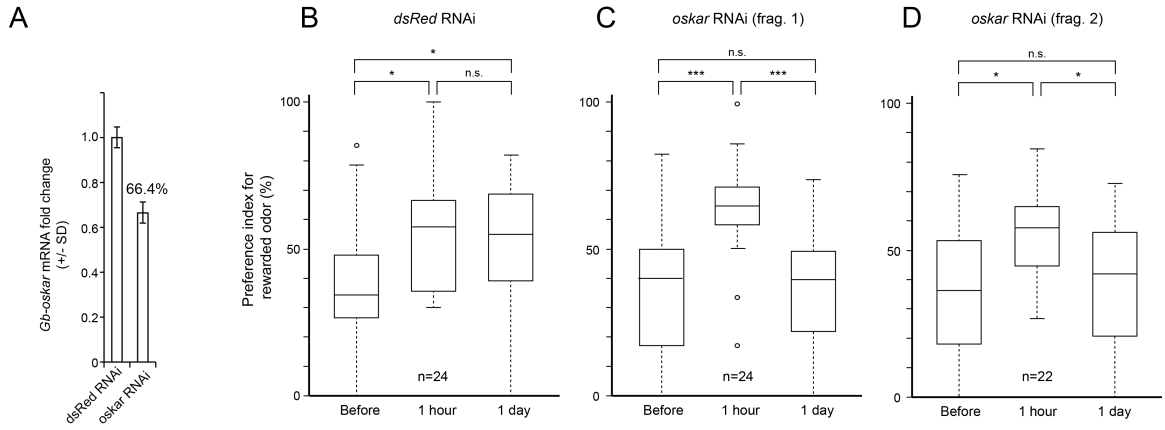


Figure 5.1. *Gb-oskar* RNAi impairs long-term olfactory memory formation in crickets. (A) qPCR validation of *Gb-oskar* RNAi. A modest reduction to 66.4% of control levels is observed. Graph represents *Gb-oskar* fragment #1. (B-D) Effects of *Gb-oskar* RNAi on olfactory learning. Relative preference between the rewarded odor (peppermint) and control odor (vanilla) was tested before training, 1 hour post-training, and 1 day post training for dsRed controls (B), *Gb-oskar* fragment #1 (C) and *Gb-oskar* fragment #2 (D). Boxes represent the 1st-3rd quartiles surrounding the median (middle line). Whiskers extend to extreme values within 1.5x of interquartile range. Wilcoxon's test was used for comparison of preference before and after conditioning. For multiple comparisons, the Holm method was used to adjust the significance level. (* $p < 0.05$, *** $p<0.001$, n.s. = not significantly different). Behavior experiments and the associated statistical analysis were performed by Ryo Wakuda, Kanta Terao, Yukihisa Matsumoto, and Makoto Mizunami in the Mizunami lab.

Gb-oskar *is expressed in adult neuroblasts of the mushroom body*

To address the cellular mechanism by which *Gb-oskar* may influence olfactory memory, we examined the expression of *Gb-oskar* in the cricket brain. *Gb-oskar* expression was restricted to a cluster of cells located at the apex of each mushroom body (**Figure 5.2A**), and expression was not detected elsewhere in the brain. These *Gb-oskar*-positive cells matched descriptions of

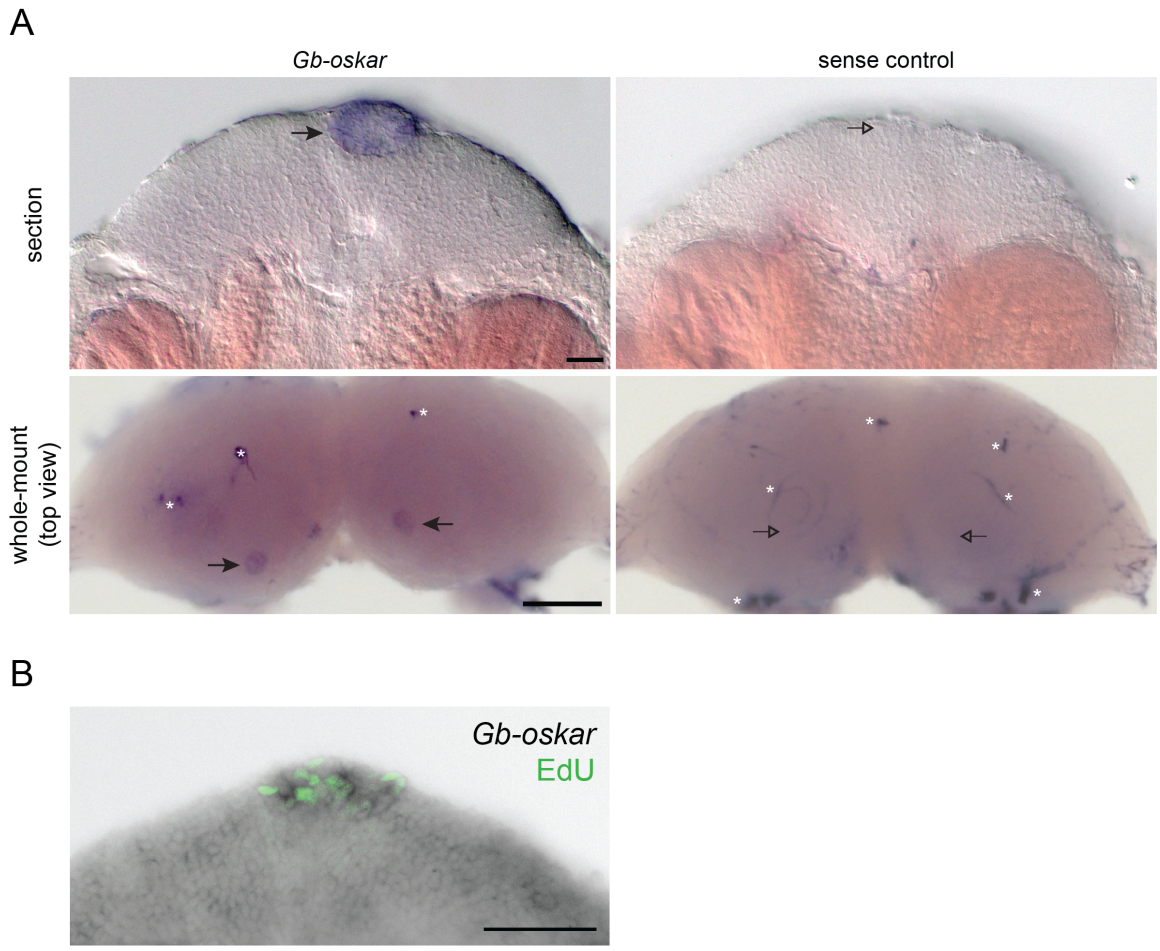


Figure 5.2. *Gb-oskar* is expressed in adult neuroblasts in the mushroom body. (A) *Gb-oskar in situ* hybridization (left panels) and sense controls (right panels), shown in 90 µm vibratome section (top row; dorsal is up) and in a top-down view (looking down onto the dorsal surface) in whole-mount preparations (bottom row). *Gb-oskar* is detected in a cluster of cells at the apex of the Kenyon cells (black arrows). (B) The cells that stain for *Gb-oskar* are EdU-positive, confirming that these are mitotically active adult neuroblasts. Crickets were injected with EdU six hours prior to dissection. Scale bars = 50 µm in top row of (A) and (B), 200 µm in bottom row of (A).

mushroom body neuroblasts, the only proliferative cells in the adult cricket brain, which divide continuously during adult life to produce new Kenyon cells (the cells that make up the mushroom

body) (Cayre et al., 1996). To test whether *Gb-oskar*-positive cells were in fact the neuroblasts, we injected crickets with EdU six hours prior to dissection. Following *in situ* hybridization, we detected EdU specifically in the cells expressing *Gb-oskar* (**Figure 5.2B**), demonstrating that the *Gb-oskar*-expressing cells are indeed the adult neuroblast of the mushroom body.

DISCUSSION

We have shown that *Gb-oskar* RNAi impairs olfactory learning in the cricket *Gryllus bimaculatus* (**Figure 5.1**), and that *Gb-oskar* is specifically expressed in the neuroblasts of the adult mushroom body (**Figure 5.2**). Intriguingly, proliferation of these mushroom body neuroblasts has previously been linked to olfactory learning in a related cricket species: when these neuroblasts are killed using radiation, olfactory learning is impaired (Scotto-Lomassese et al., 2003). Thus, our results support the hypothesis that *Gb-oskar* is required for neuroblast function (i.e. proliferation, survival, and/or maintenance in an undifferentiated state), and that the role of *Gb-oskar* in olfactory memory is mediated through a cellular role in these neuroblasts. Directly testing this hypothesis will require assaying the effects of *Gb-oskar* RNAi on the proliferation and/or survival of these neuroblasts. Such experiments are currently underway.

The relationship between adult neurogenesis and long-term olfactory memory formation is intriguing, because it is well-established that the formation of long-term memory, including olfactory memory in insects, does not strictly require the birth of new neurons (Heisenberg, 2003; Kandel, 2012). Accordingly, we emphasize that killing adult neuroblasts impairs, but does not abolish, olfactory learning in the cricket (Scotto-Lomassese et al., 2002). In fact, adult neurogenesis does not occur in the mushroom bodies of *Drosophila* (or several other insect species)(reviewed in Cayre et al., 2007), yet olfactory memory is robust in these species (Heisenberg, 2003). Classical experiments in a wide range of animals have demonstrated that

long-term memory involves changes in synaptic strength between existing neurons, regulated via cyclic adenosine monophosphate (cAMP) response element-binding protein (CREB)-dependent *de novo* protein synthesis (reviewed in Kandel, 2012), including in *Gryllus bimaculatus* (Matsumoto et al., 2006). In *Drosophila*, long-term olfactory memory requires the ~2,500 Kenyon cells of the mushroom body (Heisenberg, 2003), each of which receives synaptic inputs from a small and random subset of the antennal lobe projection neurons (which themselves receive direct synaptic input from the odorant receptor neurons of the antennae) (Caron et al., 2013). A given Kenyon cell thus responds with high selectivity to a small number of odors (or other stimuli), allowing the mushroom body to house an "explicit" representation of a large number of olfactory cues (Caron et al., 2013; Heisenberg, 2003). Specific olfactory stimuli are then associated with learned behavioral responses via specific sets of neurons connecting the mushroom body to other brain regions in a protein synthesis-dependent fashion, to form long-term memories (C.-C. Chen et al., 2012; Pai et al., 2013). Thus, it seems possible that adult-born Kenyon cells in *Gryllus* (and other species which display adult neurogenesis in the mushroom body) are recruited into an existing circuit, and allow for a constantly increasing repertoire of olfactory associations, and it could be this process that is disrupted by *Gb-oskar* RNAi. It is also interesting to note that of the two mammalian brain regions known to undergo adult neurogenesis, one (the subventricular zone) contributes to the olfactory bulb, and neurogenesis in this region is involved in olfactory memory (Lazarini and Lledo, 2011).

It remains unclear whether *oskar* is involved in long-term memory formation in *Drosophila* (Dubnau et al., 2003). Given that adult *Drosophila* lack the mushroom body neuroblasts seen in *Gryllus* (Cayre et al., 2007), a straight-forward test for a directly comparable *oskar* function is not possible. However, although *Drosophila* mushroom body stem cells are absent in adults, analogous mushroom body neuroblasts remain mitotically active late into pupal development (Ito and Hotta, 1992). Thus, it will be interesting to test whether *oskar* functions in

these mushroom body neuroblasts during larval and/or pupal stages, which would suggest a conserved function.

In addition, there is good reason to hypothesize that *oskar* could function in *Drosophila* olfactory long-term memory independent of a possible role in neuroblasts. Specifically, a recent study of the mushroom body output neurons has suggested that long-term memory involves the activity-dependent de-repression of mRNAs localized to granules containing Pum, Staufen, and Orb (Pai et al., 2013). Given that Oskar is thought to nucleate similar granules containing these proteins in the *Drosophila* oocyte (Breitwieser et al., 1996; Chang et al., 1999), it would be very interesting to test whether Oskar is involved in the formation and/or activity of these granules in the brain. Although I have thus far been unable to detect Oskar protein in the adult brain via antibody staining (see Appendix 2 of this dissertation), direct functional tests are necessary to establish whether or not this is the case. In addition, it should be noted that such a function for *Gb-oskar* in the formation of RNP in neurons is formally possible in *Gryllus*; although *Gb-oskar* is detected at highest levels in the mushroom body neuroblasts, we cannot rule out that it is expressed at lower levels in differentiated Kenyon cells (see Figure 2B).

We have previously shown that *Gb-oskar* functions in embryonic neuroblasts (Ewen-Campen et al., 2012), and have now extended these observations to show that *Gb-oskar* is also present in neuroblasts of the adult brain. It will be interesting to test whether *Gb-oskar* functions in neuroblasts throughout the entirety of nymphal development, or whether its activity is limited to the adult neuroblasts. In addition, future studies could test whether additional "germ line" genes also function in these adult neuroblasts, which would suggest that *oskar* may cooperate with conserved molecular partners in different cellular contexts. Indeed, we have previously shown that *Gb-vasa* is co-expressed with *Gb-oskar* in the embryonic neuroblasts (Ewen-Campen et al., 2012), although *Gb-vasa*'s function in these cells was not addressed.

Both germ cells and neuroblasts are stem cells that give rise to highly specialized, post-mitotic daughter cells while they themselves remain proliferative for long periods of time. Thus,

the role of *oskar* in both cell types could conceivably be related to stem cell maintenance and/or asymmetric division. Indeed, it has been noted that a variety of highly conserved "germ line genes" including *vasa*, *nanos*, and *piwi* are found in a variety of multipotent cells in diverse animals (Juliano and Wessel, 2010), raising the possibility that such genes are involved in establishing multipotency rather than specifically germ cells *per se*. A broader understanding of the function(s) of *oskar* will thus require additional studies of phylogenetically diverse insects, as well as further detailed biochemical analysis in *Drosophila* germ cells and neurons.

ACKNOWLEDGEMENTS

We thank Dr. Sumihare Noji and the members of his laboratory for hosting the 2nd International Conference on the Cricket, where this collaboration began. We thank Yue Meng for *Gryllus* colony maintenance, and members of the Extavour and Mizunami labs for feedback.

REFERENCES

Anantharaman, V., Zhang, D., Aravind, L., 2010. OST-HTH: a novel predicted RNA-binding domain 1–8.

Breitwieser, W., Markussen, F.H., Horstmann, H., Ephrussi, A., 1996. Oskar protein interaction with Vasa represents an essential step in polar granule assembly. Genes Dev 10, 2179–2188.

Callebaut, I., Mornon, J.P., 2010. LOTUS, a new domain associated with small RNA pathways in the germ line. Bioinformatics 26, 1140–1144.

Caron, S.J.C., Ruta, V., Abbott, L.F., Axel, R., 2013. Random convergence of olfactory inputs in the Drosophila mushroom body. Nature 497, 113–117.

Cayre, M., Scotto-Lomassese, S., Malaterre, J., Strambi, C., Strambi, A., 2007. Understanding the Regulation and Function of Adult Neurogenesis: Contribution from an Insect Model, the House Cricket. Chemical Senses 32, 385–395.

Cayre, M., Strambi, C., Charpin, P., Augier, R., Meyer, M.R., Edwards, J.S., Strambi, A., 1996. Neurogenesis in adult insect mushroom bodies. J. Comp. Neurol. 371, 300–310.

Chang, J.S., Tan, L., Schedl, P., 1999. The Drosophila CPEB homolog, orb, is required for oskar protein expression in oocytes. Dev Biol 215, 91–106.

Chen, C.-C., Wu, J.-K., Lin, H.-W., Pai, T.-P., Fu, T.-F., Wu, C.-L., Tully, T., Chiang, A.-S., 2012. Visualizing long-term memory formation in two neurons of the Drosophila brain. Science 335, 678–685.

Chen, G., Li, W., Zhang, Q.-S., Regulski, M., Sinha, N., Barditch, J., Tully, T., Krainer, A.R., Zhang, M.Q., Dubnau, J., 2008. Identification of synaptic targets of Drosophila pumilio. PLoS Comput. Biol. 4, e1000026.

Driscoll, H.E., Muraro, N.I., He, M., Baines, R.A., 2013. Pumilio-2 regulates translation of Nav1.6 to mediate homeostasis of membrane excitability. J. Neurosci. 33, 9644–9654.

Dubnau, J., Chiang, A.-S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U., Broger, C., Tully, T., 2003. The staufen/pumilio pathway is involved in Drosophila long-term memory. Curr Biol 13, 286–296.

Ephrussi, A., Dickinson, L.K., Lehmann, R., 1991. Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. Cell 66, 37–50.

Ephrussi, A., Lehmann, R., 1992. Induction of germ cell formation by oskar. Nature 358, 387–392.

Ewen-Campen, B., Donoughe, S., Clarke, D.N., Extavour, C.G., 2013. Germ Cell Specification Requires Zygotic Mechanisms Rather Than Germ Plasm in a Basally Branching Insect. Current Biology 23, 835–842.

Ewen-Campen, B., Schwager, E.E., Extavour, C.G.M., 2010. The molecular machinery of germ line specification. Mol. Reprod. Dev. 77, 3–18.

Ewen-Campen, B., Srouji, J.R., Schwager, E.E., Extavour, C.G., 2012. oskar Predates the Evolution of Germ Plasm in Insects. Curr Biol 22, 2278–2283.

Heisenberg, M., 2003. Mushroom body memoir: from maps to models. Nat. Rev. Neurosci. 4, 266–275.

Ito, K., Hotta, Y., 1992. Proliferation pattern of postembryonic neuroblasts in the brain of Drosophila melanogaster. Dev Biol 149, 134–148.

Jones, J.R., Macdonald, P.M., 2007. Oskar controls morphology of polar granules and nuclear bodies in Drosophila. Development 134, 233–236.

Juliano, C., Wessel, G., 2010. Developmental biology. Versatile germ line genes. Science 329, 640–641.

Kainz, F., Ewen-Campen, B., Akam, M., Extavour, C.G., 2011. Notch/Delta signalling is not required for segment generation in the basally branching insect Gryllus bimaculatus. Development 138, 5015–5026.

Kandel, E.R., 2012. The molecular biology of memory: cAMP, PKA, CRE, CREB-1, CREB-2,

and CPEB. Mol Brain 5, 14.

Kanska, J., Frank, U., 2013. New roles for Nanos in neural cell fate determination revealed by studies in a cnidarian. J Cell Sci 126, 3192–3203.

Lazarini, F., Lledo, P.-M., 2011. Is adult neurogenesis essential for olfaction? Trends Neurosci. 34, 20–30.

Lehmann, R., Nüsslein-Volhard, C., 1986. Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila. Cell 47, 141–152.

Lynch, J.A., Ozüak, O., Khila, A., Abouheif, E., Desplan, C., Roth, S., 2011. The phylogenetic origin of oskar coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola. PLoS Genet. 7, e1002029.

Matsumoto, Y., Mizunami, M., 2000. Olfactory learning in the cricket Gryllus bimaculatus. J. Exp. Biol. 203, 2581–2588.

Matsumoto, Y., Mizunami, M., 2002. Temporal determinants of long-term retention of olfactory memory in the cricket Gryllus bimaculatus. J. Exp. Biol. 205, 1429–1437.

Matsumoto, Y., Unoki, S., Aonuma, H., Mizunami, M., 2006. Critical role of nitric oxide-cGMP cascade in the formation of cAMP-dependent long-term memory. Learn. Mem. 13, 35–44.

Mee, C.J., Pym, E.C.G., Moffat, K.G., Baines, R.A., 2004. Regulation of neuronal excitability through pumilio-dependent control of a sodium channel gene. J. Neurosci. 24, 8695–8703.

Menon, K.P., Sanyal, S., Habara, Y., Sanchez, R., Wharton, R.P., Ramaswami, M., Zinn, K., 2004. The translational repressor Pumilio regulates presynaptic morphology and controls postsynaptic accumulation of translation factor eIF-4E. Neuron 44, 663–676.

Murata, Y., Wharton, R.P., 1995. Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in Drosophila embryos. Cell 80, 747–756.

Pai, T.-P., Chen, C.-C., Lin, H.-H., Chin, A.-L., Lai, J.S.-Y., Lee, P.-T., Tully, T., Chiang, A.-S., 2013. Drosophila ORB protein in two mushroom body output neurons is necessary for long-term memory formation. Proc Natl Acad Sci USA 110, 7898–7903.

Perrat, P.N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., Waddell, S., 2013. Transposition-driven genomic heterogeneity in the Drosophila brain. Science 340, 91–95.

Scotto-Lomassese, S., Strambi, C., Aouane, A., Strambi, A., Cayre, M., 2002. Sensory inputs stimulate progenitor cell proliferation in an adult insect brain. Curr Biol 12, 1001–1005.

Scotto-Lomassese, S., Strambi, C., Strambi, A., Aouane, A., Augier, R., Rougon, G., Cayre, M., 2003. Suppression of adult neurogenesis impairs olfactory learning and memory in an adult insect. J. Neurosci. 23, 9289–9296.

Sonoda, J., Wharton, R.P., 1999. Recruitment of Nanos to hunchback mRNA by Pumilio. Genes Dev 13, 2704–2712.

Suyama, R., Jenny, A., Curado, S., Pellis-van Berkel, W., Ephrussi, A., 2008. The actin-binding protein Lasp promotes Oskar accumulation at the posterior pole of the Drosophila embryo. Development 136, 95–105.

Takahashi, T., Hamada, A., Miyawaki, K., Matsumoto, Y., Mito, T., Noji, S., Mizunami, M., 2009. Systemic RNA interference for the study of learning and memory in an insect. Journal of Neuroscience Methods 179, 9–15.

Wharton, R.P., Sonoda, J., Lee, T., Patterson, M., Murata, Y., 1998. The Pumilio RNA-binding domain is also a translational regulator. Mol Cell 1, 863–872.

Xu, X., Brechbiel, J.L., Gavis, E.R., 2013. Dynein-dependent transport of nanos RNA in Drosophila sensory neurons requires Rumpelstiltskin and the germ plasm organizer Oskar. J. Neurosci. 33, 14791–14800.

Ye, B., Petritsch, C., Clark, I.E., Gavis, E.R., Jan, L.Y., Jan, Y.N., 2004. Nanos and Pumilio are essential for dendrite morphogenesis in Drosophila peripheral neurons. Curr Biol 14, 314–321.

Zamore, P.D., Williamson, J.R., Lehmann, R., 1997. The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. RNA 3, 1421–1433.

DISCUSSION AND OUTLOOK

Divergent mechanisms for specifying a conserved cell type

In this thesis, I provide evidence that a highly conserved cell type – the germ cells – can arise via highly un-conserved developmental mechanisms across species. Specifically, whereas *Drosophila* PGCs acquire their identity via maternal factors localized within the oocyte cytoplasm, my dissertation research demonstrates that such maternal determinants do not specify PGCs in two different basally branching insect species. Instead, my data suggest that PGCs arise later during development in both *Gryllus* and *Oncopeltus*, after cellularization has occurred and the zygotic genome has been activated, and therefore this process likely requires cell signaling. In other words, although PGCs are undoubtedly a homologous cell type across all animals, the mechanisms that specify these cells are diverse.

The observation that homologous structures can develop via divergent mechanisms is certainly not new (reviewed in Scholtz, 2005). Structures as obviously homologous and highly specialized as the wings of insects develop via remarkably different mechanisms across species: in holometabolous insects, wings arise from imaginal discs that survive metamorphosis, whereas imaginal discs (and metamorphosis itself) are entirely lacking in hemimetabolous insects (reviewed in Giorgianni and Patel, 2007). Similarly, while all adult insects (and arthropods) share a segmented body plan, the molecular mechanisms of segmentation during embryogenesis are entirely distinct between species (reviewed in Davis and Patel, 2002). On a broader phylogenetic scale, such cell types as neurons or embryonic mesoderm are each believed to be homologous across Bilateria, based on conserved gene expression patterns and analogy of function, yet both cell types arise via distinct developmental mechanisms even between relatively closely related taxa (reviewed in Roth, 2004; Stollewerk, 2008). Altogether, there is a growing body of modern genomic evidence supporting Karl von Baer's famous 1828 observation that the conserved

"phylotypic stage" of a given phylum (that embryonic stage during which all members of a phylum most closely resemble one another) is arrived at via highly diverse embryonic avenues (Domazet-Lošo and Tautz, 2010; Kalinka et al., 2010).

However, while there are plentiful examples of homologous structures arising via divergent developmental mechanisms, there is much less empirical data on *how*, mechanistically, development can evolve while consistently maintaining a conserved outcome. The data presented in this dissertation suggest that, in the case of PGC specification, gene co-option played an important role. Whereas previous studies proposed that the divergent mechanisms of PGC specification involved the *de novo* evolution of a novel gene (i.e. *oskar*; Lynch et al., 2011), the data presented here instead suggest that this process involved the redeployment of a pre-existing gene into a novel function. Below, I discuss how the modularity of the "germ line genes" may have influenced this evolutionary process.

The roles of gene co-option and modularity in developmental evolution

To understand how PGC specification has evolved, it may be helpful to imagine the molecular machinery of germ cells as a functional module. By module, I mean that the "germ line genes" are characterized by relatively strong interactions between one another, and that their expression in a cell is largely co-regulated by common upstream factors. Indeed, across nearly all animals that have been examined, germ cells express some or all of a conserved suite of genes, including Vasa, Nanos, Pumilio, Tudor genes, and PIWI family genes (reviewed in Ewen-Campen et al., 2010). These proteins often co-localize to the same subcellular structures within germ cells, a granular ribonucleoprotein structure lining the cytoplasmic face of the nuclear membrane, forming a germ cell-specific organelle sometimes referred to as the "nuage"(reviewed in Ewen-Campen et al., 2010; Voronina et al., 2011). Furthermore, there is evidence for physical interactions between PIWI proteins and Tudor proteins (Chen et al., 2009), and it has been

proposed that the ensemble of proteins in the nuage collectively act as a "hub" for post-transcriptional regulation of various classes of RNA as they exit the nucleus, a fundamental aspect of germ cell function (Voronina et al., 2011). Thus, we can imagine a germ line module, comprised of a group of largely conserved proteins with essentially conserved functions and interactions, which collectively perform the cellular activities required of germ cells.

To specify a germ cell, therefore, is a matter of correctly co-expressing the proteins in germ line module. Seen this way, a germ line module is conceptually analogous to a "gene regulatory network," a familiar concept in evolutionary developmental biology (reviewed in Peter and Davidson, 2011). Gene regulatory networks are sets of co-regulated genes that respond to common upstream transcription factors; thus, a single so-called "master regulator" (a highly upstream transcription factor) can simultaneously deploy the expression of hundreds of downstream genes, which collectively effect a cellular phenotype. It is widely accepted that, due to the modular organization of gene regulatory networks, relatively major evolutionary change can be achieved through relatively minor changes in the expression of upstream transcription factors (reviewed in Stern and Orgogozo, 2008). For example, such novel structures as beetle horns and the pigmentation patterns on butterfly wings are believed to have evolved via novel expression domains of transcription factors that in turn deploy largely intact downstream gene regulatory networks (Moczek and Rose, 2009; Monteiro, 2012).

This leads to a simple conceptual model for how PGC specification mechanisms could evolve: while the PGC module itself remains largely intact across species, the upstream signal that assembles these factors changes over time. In the case of insects, my data supports the hypothesis that the PGCs of ancestral insects activated this module in response to secreted cell signals. (Indeed, a recent paper from the Extavour lab has identified BMP signaling as one of the pathways implicated in PGC specification in *Gryllus* [Donoughe et al., 2014]). During the course of insect evolution, likely near the base of Holometabola, the *oskar* gene acquired a novel expression domain in the oocyte, where it was became able to physically assemble the members

of the PGC module in a subcellular region of the oocyte, thus nucleating a germ plasm and establishing a continuity of PGC module protein expression between the generations[1].

On a broader phylogenetic scale, the fact that germ plasm has independently arisen so many times in widely divergent taxa may be the result of the simple fact that embryos all physically begin as germ cells (gametes) of the previous generation, and are thus likely to express at least some genes of the PGC module during the earliest stages of development. Generating a germ plasm may simply require evolving a mechanism to maintain the expression of these genes into embryogenesis and localize them to a subset of embryonic cells (Extavour, 2007). For insects, my data implies that the evolution of this nucleating factor involved the co-option of *oskar* from a somatic role (perhaps in the nervous system) into a novel role in the germ line of holometabolous insects. As *oskar* and additional nucleating factors (such as *e.g. buckyball* in zebrafish and/or the PGL proteins of *C. elegans*) are characterized on a structural and biochemical level, it will be interesting to compare the common properties of these proteins.

Why has oskar been lost in so many lineages?

Given that *oskar* plays an indispensable role in the PGCs of *Drosophila* and *Nasonia*, why has it been repeatedly lost from so many other insect lineages? One explanation for this seeming paradox has been suggested by Lynch *et al*. (2011). In order for germ plasm to have evolved from an ancestral signaling-based PGC specification mechanism, these two processes must have co-existed within an individual for at least some portion of evolutionary history (an example of the so-called "transition model" *sensu* Extavour, 2007). That is, it is difficult to imagine the simultaneous evolution of germ plasm precisely coincident with the loss of signaling-based PGC

[1] It is unknown whether the ability of *oskar* to physically recruit other germ line genes predates its expression in the germ line, or whether this represents an additional evolutionary change in the molecular function of *oskar* and/or it's interaction partners. It will be interesting, in the future, to test whether *oskar* has similar interaction partners in the neuroblasts of *Gryllus*.

specification. In those species where both mechanisms are present, there would be a level of redundancy in the specification of PGCs, which could then allow for evolutionary decay of either one or the other mechanisms in various lineages (and also perhaps for the existence of both mechanisms in some extant lineages, including the crustacean *Parhyale hawaiiensis* [MS Modrell, AL Price, J Havemann, CG Extavour, M Gerberding, and NH Patel, *in revision*], the wasp *Pimpla turionellae* [summarized in Lynch et al., 2011] and the sea urchin *Strongylocentrotus purpuratus* [Yajima and Wessel, 2011], all of which show some evidence for germ plasm yet can regenerate germ cells when PGC precursors are embryonically ablated). This would explain why, in some lineages of holometabolous insects, *oskar*-mediated germ plasm has become the exclusive mode of PGC specification, whereas in others *oskar* appears to have been lost altogether and PGCs are specified via other means (Lynch et al., 2011). We note that this scenario would also require that the functions of *oskar* in the nervous system would also have become dispensable in these lineages, perhaps also through the evolution of compensatory mechanisms.

The idea that redundancy can allow for divergence and eventual loss of an ancestral state, while maintaining the functional output of the system, has been explored both theoretically and experimentally (reviewed in Wagner et al., 2007). For example, an analysis of the *cis*-regulatory elements regulating ribosomal gene expression across the yeast phylogeny revealed that a novel regulatory site arose and eventually replaced an ancestral site in derived taxa, but that both sites remain present in several intermediate taxa (Tanay et al., 2005). Future studies examining the function of *oskar* in a wider variety of intermediate taxa between *Gryllus* and Holometabola will be very interesting to explore these ideas.

Outlook

Given the fundamental differences we observe in PGC specification between *Drosophila* and

the species studied in this dissertation, many new questions arise. Here, I note three questions that I find particularly interesting for future research:

1. **What is the mechanism for inductive PGC specification in *Gryllus*?** My dissertation suggests that a maternally-supplied germ plasm is absent from *Gryllus* ovaries, but does not suggest possible pathways or mechanisms which could induce PGC fate. An important first step has recently been taken by others in the Extavour lab to show that *dpp* signaling is involved in PGC specification (Donoughe et al., 2014). Interestingly, the mouse orthologs of *dpp* (BMP-family ligands) specify PGC fate in the mouse embryo, raising the possibility that this pathway may in fact have an ancestral role in animal PGC specification. Further work to understand the mechanism of inductive PGC specification in *Gryllus* would be very interesting, as many questions remain. Given the extreme pleiotropy of the *dpp* pathway in embryonic development, how is its role in PGCs accomplished specifically? How is PGC fate restricted to abdominal segments 2-5? Are other pathways involved?
2. **How are PGCs specified in additional insect species?** The immense diversity of insects provides investigators with a wealth of opportunities to study evolution, yet this immense diversity also makes it perpetually difficult to generalize from findings from one species to other insects. Historical descriptions of insect embryology suggest PGCs arise in a wide variety of times and places during development (see Introduction), and modern studies using molecular markers and functional manipulations would greatly expand on these classical descriptions. Specifically, focusing on those hemimetabolous species described to have germ plasm and/or pole cells (see Introduction Figure 1.3) would be very interesting. Studying PGC specification in a wide diversity of insects would greatly improve our understanding of how this process evolves.
3. **What new experimental tools would be most useful for future studies of germ**

cells? Although non-model organisms provide important phylogenetic contrasts with such models as *Drosophila*, there are major experimental limitations to working with such organisms. However, with the recent emergence of (nearly) affordable technologies to sequence and annotate genomes, make stable transgenic lines, and knock-in/knock-out genes using such genome-editing techniques as CRISPR, entire new avenues of experimentation are becoming available. A fluorescent reporter of PGCs in *Gryllus*, such as a Gb-Piwi or Gb-Vasa reporter, would be an extremely valuable tool, and is being developed by Seth Donoughe and Taro Nakamura in the Extavour lab. Currently, screening for PGC phenotypes is time-consuming, as embryos must be dissected, fixed, stained via antibody staining or *in situ* hybridization, and imaged via confocal microscopy. Being able to visualize PGCs in a faster and easier way would allow future researchers to greatly expand the scope of manipulations they test for effects on PGC formation. In mouse PGC research, such tools as *Blimp1* reporter lines have allowed for an explosion in mechanistic studies of inductive PGC specification. In addition, such a tool may allow for a more detailed description of PGC specification, as this process has thus far only been examined in fixed tissue, at various snap-shots of development, as levels of Gb-Piwi and Gb-Vasa become differentially higher in PGCs during stages 5-7 of *Gryllus* development. Live imaging of PGC formation is likely to be quite technically challenging due to the fact that the embryo is deep within yolk and is undergoing significant physical movements during this time. However, such an approach could yield a far more detailed picture of PGC development than is currently known.

This dissertation has also provided evidence that *oskar*, once thought to be a Dipteran-specific novel gene functioning solely in the germline, in fact evolved quite early in insect

evolution and has an additional function in the nervous system of *Gryllus*. In light of recently published work demonstrating a neural role for *Drosophila oskar*, it seems likely that this neural role may in fact be its ancestral function. However, many questions still remain regarding the enigmatic *oskar* gene.

1. **Is *oskar* present in additional insect genomes?** As additional insect genomes are sequenced, it will be very interesting to know where *oskar* is found in the insect tree (and possibly other arthropods). This study is currently being initiated by Tamsin Jones, a graduate student in the Extavour lab. These data will provide far greater resolution to say when *oskar* first arose, and where it has been lost. Furthermore, this study will identify a number of additional insect species that should be studied in detail, specifically to know whether *oskar* functions in the nervous system, germline, or both.
2. **What is the molecular function of *oskar* in the nervous system?** Given the ambiguity of the *norka* "allele" of *oskar* (see Appendix 2 of this dissertation), a first priority should be to test for behavioral phenotypes in *bona fide* mutants of *oskar*, to know whether this gene does indeed have a function in olfactory learning and/or other neural roles in *Drosophila*. If so, this would provide a new system in which to probe *oskar* function, which has thus far proved remarkably recalcitrant. Does *oskar* function during neural stem cell divisions in *Drosophila*? Does it function in the formation of ribonucleoprotein granules within specific neurons (which are known to contain such Oskar-associated proteins as Staufen, Pumilio, and Orb; see Chapter 5)? Or, is the neural function of *oskar* in *Gryllus* simply distinct from that of its roles in *Drosophila*?

REFERENCES:

Chen, C., Jin, J., James, D.A., Adams-Cioaba, M.A., Park, J.G., Guo, Y., Tenaglia, E., Xu, C., Gish, G., Min, J., Pawson, T., 2009. Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. Proc Natl Acad Sci USA 106, 20336–

20341.

Davis, G.K., Patel, N.H., 2002. Short, long, and beyond: molecular and embryological approaches to insect segmentation. Annu. Rev. Entomol. 47, 669–699.

Domazet-Lošo, T., Tautz, D., 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature 468, 815–818.

Donoughe, S., Nakamura, T., Ewen-Campen, B., Green, D.A., Henderson, L., Extavour, C.G., 2014. BMP signaling is required for the generation of primordial germ cells in an insect. Proc Natl Acad Sci USA 111, 4133–4138.

Ewen-Campen, B., Schwager, E.E., Extavour, C.G.M., 2010. The molecular machinery of germ line specification. Mol. Reprod. Dev. 77, 3–18.

Extavour, C.G.M., 2007. Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. Integr Comp Biol 47, 770–785.

Giorgianni, M.W., Patel, N.H., 2007. Conquering Land, Air and Water: The Evolution and Development of Arthropod Appendages. In "Evolving Form and Function: Fossils and Development" (Briggs, Derek .E.G. ed.) Peabody Museum of Natural History, New Haven, CT. pp.159-180. 1–24.

Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M., Tomancak, P., 2010. Gene expression divergence recapitulates the developmental hourglass model. Nature 468, 811–814.

Lynch, J.A., Ozüak, O., Khila, A., Abouheif, E., Desplan, C., Roth, S., 2011. The phylogenetic origin of oskar coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola. PLoS Genet. 7, e1002029.

Moczek, A.P., Rose, D.J., 2009. Differential recruitment of limb patterning genes during development and diversification of beetle horns. Proc Natl Acad Sci USA 106, 8992–8997.

Monteiro, A., 2012. Gene regulatory networks reused to build novel traits: co-option of an eye-related gene regulatory network in eye-like organs and red wing patches on insect wings is suggested by optix expression. Bioessays 34, 181–186.

Peter, I.S., Davidson, E.H., 2011. Evolution of gene regulatory networks controlling body plan development. Cell 144, 970–985.

Roth, S., 2004. Gastrulation in Other Insects, in: Stern, C. (Ed.), Gastrulation: From Cells to Embryos. Cold Spring Harbor Laboratory Press, pp. 105–122.

Scholtz, G., 2005. Homology and ontogeny: Pattern and process in comparative developmental biology. Theory in Biosciences 124, 121–143.

Stern, D.L., Orgogozo, V., 2008. The loci of evolution: how predictable is genetic evolution? Evolution 62, 2155–2177.

Stollewerk, A., 2008. Evolution of neurogenesis in arthropods, in: Fusco, G., Minelli, A. (Eds.),

Evolving Pathways : Key Themes in Evolutionary Developmental Biology. Cambridge University Press, pp. 359–380.

Tanay, A., Regev, A., Shamir, R., 2005. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. Proc Natl Acad Sci USA 102, 7203–7208.

Voronina, E., Seydoux, G., Sassone-Corsi, P., Nagamori, I., 2011. RNA granules in germ cells. Cold Spring Harb Perspect Biol 3.

Wagner, G.P., Pavlicev, M., Cheverud, J.M., 2007. The road to modularity. Nat Rev Genet 8, 921–931.

Yajima, M., Wessel, G.M., 2011. Small micromeres contribute to the germline in the sea urchin. Development 138, 237–243.

APPENDIX A

REVIEW ARTICLE

Molecular Reproduction & Development 77:3–18 (2010)

The Molecular Machinery of Germ Line Specification

BEN EWEN-CAMPEN, EVELYN E. SCHWAGER, AND CASSANDRA G.M. EXTAVOUR*

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts

SUMMARY

Germ cells occupy a unique position in animal reproduction, development, and evolution. In sexually reproducing animals, only they can produce gametes and contribute genetically to subsequent generations. Nonetheless, germ line specification during embryogenesis is conceptually the same as the specification of any somatic cell type: germ cells must activate a specific gene regulatory network in order to differentiate and go through gametogenesis. While many genes with critical roles in the germ line have been characterized with respect to expression pattern and genetic interactions, it is the molecular interactions of the relevant gene products that are ultimately responsible for germ cell differentiation. This review summarizes the current state of knowledge on the molecular functions and biochemical connections between germ line gene products. We find that homologous genes often interact physically with the same conserved molecular partners across the metazoans. We also point out cases of nonhomologous genes from different species whose gene products play analogous biological roles in the germ line. We suggest a preliminary molecular definition of an ancestral ''pluripotency module'' that could have been modified during metazoan evolution to become specific to the germ line.

Mol. Reprod. Dev. 77: 3–18, 2010. © *2009 Wiley-Liss, Inc.*

Sexually reproducing animals must ensure that one particularly important cell type is determined: the germ cells

* Corresponding author:
Department of Organismic and Evolutionary Biology
Harvard University, 16 Divinity Avenue
BioLabs Building Room 4103,
Cambridge, MA 02138.
E-mail: extavour@oeb.harvard.edu

Ben Ewen-Campen and Evelyn E. Schwager contributed equally to this work.

Published online 29 September 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/mrd.21091

Received 1 June 2009; Accepted 7 July 2009

INTRODUCTION

Across the plant and animal kingdoms, embryogenesis is that crucial developmental phase during which a single pluripotent cell, the fertilized ovum, must divide and differentiate to produce a plethora of differentiated, unipotent cell types. Sexually reproducing animals must ensure that one particularly important cell type is determined: the germ cells. These cells will be the sole progenitors of eggs and sperm in the sexually mature adult, and as such, their correct specification during embryonic development is critical for reproductive success and species survival. Germ cells and their embryonic origin have fascinated biologists for centuries, resulting in an enormous amount of primary literature on the subject (last comprehensively reviewed by Nieuwkoop and Sutasurya, 1979, 1981). The time and place when germ cells are first observed in embryogenesis, their histological and cytological characteristics, and the results of experimental manipulation of embryos on germ cell formation have been described in great detail for dozens of different species across the metazoa. All studies coincide in their observation of germ cell-specific cytoplasmic inclusions, visible under transmitted light and electron microscopy alike. Molecular studies from the last three decades have shown that this special cytoplasm, often called germ plasm, houses germ cell-specific gene products. Several excellent reviews have examined germ cell formation in specific animals (Saffman and Lasko, 1999; Raz, 2003; Strome, 2005; Hayashi et al., 2007; Saitou, 2009), the genetic mechanisms of specific germ line specification modes (Houston and King, 2000a; Strome and Lehmann, 2007), general molecular

characteristics of germ cells (Seydoux and Braun, 2006; Cinalli et al., 2008; Nakamura and Seydoux, 2008), or the function of germ cell-specific genes (Raz, 2000; Noce et al., 2001). As we strive to put biological processes into an evolutionary perspective, however, we now need to begin to consider the ancestral histories of not just germ cell-specific genes themselves, but also their molecular interactions and collective functions in the germ plasm. While it is clear that many of these genes are conserved across metazoa, it is less clear to what extent the specific molecular interactions of these mRNAs, proteins, and cellular organelles have changed or remained the same throughout evolution.

The recent molecular revisitation of classical comparative embryology, otherwise known as evolutionary developmental biology or ''evo-devo,'' has clarified a key paradigm that is relevant to the germ cell problem in this context. It is now possible, and moreover useful, to speak of molecular modules comprising gene regulatory networks (GRNs) (see e.g. Davidson et al., 2002). Such modules consist of a group of genes whose genetic interactions, or physical interactions of their gene products, are highly biochemically stable and thus highly conserved. The result of this genetic and molecular interaction stability is that the same batteries of genes, or modules, are found to operate in similar ways both in different organisms, and in different places and/or times during the development of a single organism (discussed in Wagner et al., 2007; Monteiro and Podlaha, 2009). The Notch–Delta pathway, for example, is a ligand/receptor-activated signal transduction pathway that ultimately regulates gene expression at the level of transcription. All members of this pathway are both highly conserved and operate together in all metazoans (reviewed by Kopan and Ilagan, 2009). Over the course of animal development, this module participates in a wide variety of developmental processes, including segmentation, neuroblast specification, and stem cell maintenance (reviewed by Artavanis-Tsakonas et al., 1999; Lai, 2004).

We can therefore ask, in the case of germ line-specific molecules, if it is possible to identify a group of genes that are not only highly conserved, but whose products also display conserved molecular interactions. If so, does this putative ''module'' also participate, like the Notch pathway, in a variety of developmental decisions, or is it confined to germ line specification? In this review, we will establish a framework for answering these questions by reviewing and summarizing recent data on the molecular functions and interactions of several genes that are critical for germ cell specification. Because examining all known genes involved in the process is beyond the scope of this review, several genes that are conserved in animal genomes, but whose role in germ line specification is either poorly understood or likely to be indirect, are indicated in Table 1, but not discussed further. Instead, we have focused on a subset of genes whose germ line specificity and critical roles in specification are well established. Some of these genes are highly conserved across the Metazoa, while for others, either their presence in the genome or their role in germ line specification, are lineage-specific.

CONSERVED MOLECULAR COMPONENTS OF GERM LINE SPECIFICATION AND DIFFERENTIATION

Vasa

Products of the *vasa* gene family are the most widely used molecular germ cell markers for the Metazoa (discussed in Raz, 2000; Noce et al., 2001; Extavour and Akam, 2003). Vasa proteins are ATP-dependent RNA helicases of the DEAD box class, which was originally identified as a helicase family based on conservation of eight functional domains (Linder et al., 1989). DEAD box helicases are generally involved in RNA metabolism and can mediate both RNA–RNA and RNA–protein interactions (Rocak and Linder, 2004). A significant body of functional data for these helicases exists, based largely on studies of yeast DEAD box proteins (reviewed by Rocak and Linder, 2004). However, much less is known about the specific molecular function of Vasa, the germ cell-specific member of this class.

The *vasa* (*vas*) locus was first identified in *Drosophila* in a screen for maternal effect genes involved in anterior–posterior axis formation (Schüpbach and Wieschaus, 1986). *Drosophila* Vasa protein localizes to cytoplasmic granules within pole plasm (Lasko and Ashburner, 1988; Hay et al., 1988a,b), and localization of the mRNA, protein, or both to germ plasm and germ cells at some stage of development is a universal characteristic of the *vasa* gene family (see e.g. Lasko and Ashburner, 1988, 1990; Hay et al., 1988a,b; Fujiwara et al., 1994; Komiya et al., 1994; Ikenishi and Tanaka, 1997; Yoon et al., 1997; Braat et al., 2000; Knaut et al., 2000; Tanaka et al., 2000; Toyooka et al., 2000; Özhan-Kizil et al., 2009).

Localization of *vasa* gene products to germ plasm is consistent with its loss-of-function phenotypes in *Drosophila*, which are loss of or defective primordial germ cells (PGCs; also called pole cells in *Drosophila*) (Lasko and Ashburner, 1990), with additional oogenesis defects seen for null mutations (Styhler et al., 1998). Similarly, nematode (Gruidl et al., 1996; Kuznicki et al., 2000; Spike et al., 2008), frog (Ikenishi and Tanaka, 1997), flatworm (Ohashi et al., 2007), crustacean (Özhan-Kizil et al., 2009), tunicate (Sunanaga et al., 2007), and mouse (Tanaka et al., 2000) *vasa* knockdowns or mutants show germ line defects at various stages of germ cell development, including gametogenesis. In zebrafish, however, morpholino-mediated protein knockdowns of Vasa affect neither germ cell number nor fertility (Braat et al., 2001). While *vasa* is almost always required for some stage of germ cell development, in no animal has it been shown to be sufficient (see, e.g., Ikenishi and Yamakita, 2003). However, a recent study (Lavial et al., 2009) has shown that experimentally induced *vasa* expression can reprogram chicken embryonic stem cells and direct them toward a germ cell fate. This suggests that *vasa* might be able to function as a germ cell determinant for cells that are already pluripotent.

Genetic interactions between *vasa* and other germ line genes have suggested a complex network of positive and negative regulation at multiple levels, including transcription, translation, and post-translational modification. In *Caenorhabditis elegans* and mice, various components of germ

cell-specific cytoplasmic aggregations such as P granules, chromatoid bodies, and nuage lose their localization in *vasa* mutants (Chuma et al., 2006; Hosokawa et al., 2007; Spike et al., 2008). Vasa's identity as an RNA helicase suggests a role in translational regulation, and indeed, higher levels of some proteins in *vasa* mutants (Johnstone and Lasko, 2004) and a physical and genetic interaction with a translation initiation factor (Carrera et al., 2000) are both consistent with this hypothesis. However, very few direct molecular interactors have been identified for Vasa to date, and most of them effect or stabilize its localization (but see Carrera et al., 2000; Johnstone and Lasko, 2004). The SOCS-box/SPRY-domain gene *gustavus* was identified in a yeast two-hybrid screen that used *Drosophila* Vasa protein as bait (Styhler et al., 2002). Gustavus is a highly conserved protein whose zebrafish homolog localizes to germ plasm (Li et al., 2009b), suggesting an ancient origin for this protein interaction. *Drosophila gustavus* mutants fail to localize Vasa protein to the germ plasm, and other identified binding partners of Vasa protein also appear to play a role in localization, rather than function, of Vasa. The novel protein Oskar (discussed below) and the ubiquitin-specific protease Fat facets (Liu et al., 2003) interact physically with Vasa, and both are required for Vasa's correct localization to germ plasm. Detailed studies of multiple *vasa* mutant alleles have shown that the RNA-binding domains of the Vasa protein are not necessary for its localization to the pole plasm, but are necessary for its germ cell function (Liang et al., 1994). While further work will be needed to identify the molecular partners and direct targets of Vasa's regulatory function, it is clear that Vasa co-localizes to the germ plasm together with several other highly conserved germ cell gene products. Those for which the most functional data are available are discussed in the following sections.

Nanos and *Pumilio*

Orthologs of Nanos localize to germ cells of nearly all taxa studied (Extavour and Akam, 2003). The specific functions played by Nanos vary, but the phylogenetically widespread expression of these proteins in germ cells suggests that a germ line function of Nanos may have evolved very early in animals (Extavour and Akam, 2003; Extavour, 2007). Pumilio, which has orthologs in organisms as diverse as yeast and plants (Zamore et al., 1997), has been shown to physically interact with Nanos proteins in flies (Sonoda and Wharton, 1999), nematodes (Kraemer et al., 1999), frogs (Nakahata et al., 2001), and humans (Jaruzelska et al., 2003), implying that this interaction is ancestral in bilaterians.

Like *vasa*, *nanos* and *pumilio* were first discovered in *Drosophila* (Nüsslein-Volhard et al., 1987) where both genes play essential roles in abdominal patterning and germ cell survival (Nüsslein-Volhard et al., 1987; Lehmann and Nüsslein-Volhard, 1991; Wang and Lehmann, 1991; Kobayashi et al., 1996). The molecular functions of Nanos and Pumilio were first investigated in studies of their role in repressing anterior identity in *Drosophila* embryos (Irish et al., 1989; Tautz and Pfeifle, 1989), and subsequent biochemical studies have suggested mechanisms by which these two proteins form a complex that binds target RNAs and regulates their translation.

The Nanos protein contains a highly conserved C-terminal domain encoding two CCHC zinc-finger domains that bind RNA with high affinity but low sequence specificity (Curtis et al., 1997). Specificity is provided through complexing with the conserved Puf domain of Pumilio (named for *Pu*m and *F*BF, its ortholog in *C. elegans*), which binds specific sequences in the 3′-UTRs of target RNAs (Zamore et al., 1997; Zhang et al., 1997; Sonoda and Wharton, 1999). Structural analyses of Puf domains reveal a ''rainbow-shaped molecule'' formed of eight tandem helical repeats (Edwards et al., 2001), each of which usually binds a single RNA nucleotide (Wang et al., 2002; Miller et al., 2008; Gupta et al., 2009). Both Nanos and Pumilio proteins thus bind RNA and each other, and can conditionally recruit additional protein factors to regulate target RNAs (see, e.g., Sonoda and Wharton, 2001).

The mechanisms by which the Nanos/Pumilio complex regulate translation likely involve recruitment of the deadenylation machinery to target RNAs. In both flies and worms, binding of Nanos and Pumilio orthologs to target RNAs correlates with translational repression (Wreden et al., 1997; Zhang et al., 1997). In flies, such binding drives RNA deadenylation (Wreden et al., 1997), and Nanos itself has been shown to physically interact with the Ccr4p-Pop2p-Not deadenylase complex member Not4 (Kadyrova et al., 2007). Additionally, Puf proteins in yeast are able to bind Pop2, another member of this deadenylation complex (Olivas and Parker, 2000; Goldstrohm et al., 2006), suggesting that both Nanos and Pumilio have active roles in regulating translation.

Importantly, the ultimate roles played by Nanos and Pumilio orthologs vary in the germ cells of different organisms. In *Drosophila*, Nanos and Pumilio directly regulate many RNAs in migrating PGCs to repress somatic identity (Kobayashi et al., 1996; Deshpande et al., 1999; Hayashi et al., 2004), halt the cell cycle (Asaoka-Taguchi et al., 1999), and prevent apoptosis (Hayashi et al., 2004; Sato et al., 2007). Similarly, in *C. elegans*, *nos-1* and *nos-2* are not necessary for the initial formation of PGCs, but rather for their maintenance and survival during embryogenesis (Subramaniam and Seydoux, 1999). In contrast, the *C. elegans* NOS-3/FBF complex is not involved in germ cell development, but rather in the sperm-to-oocyte transition in hermaphrodites (Kraemer et al., 1999). In zebrafish (Koprunner et al., 2001) and mice (Tsuda et al., 2003) *nanos*-related genes are required for PGC survival in both sexes, but specific targets of Nanos are largely unreported in vertebrates.

In both *Drosophila* and *C. elegans*, Nanos is also genetically implicated in the maintenance of a specific ''chromatin architecture'' that is associated with general transcriptional repression (Schaner et al., 2003). However, it has been pointed out that the cytoplasmic localization of Nanos protein, as well as our knowledge of its molecular function, implies that this function may be indirect (discussed in Seydoux and Braun, 2006).

Finally, there is also evidence for Nanos proteins functioning in the absence of known interactions with Pumilio

proteins. As mentioned above, *C. elegans* NOS-3 physically interacts with the Puf protein FBF, but the two other nematode *nanos* orthologs, *nos-1* and *nos-2*, do not do so in a yeast two-hybrid assay (Kraemer et al., 1999). However, the *C. elegans* genome encodes eight Puf proteins, and knock down of several of these proteins produces PGC defects indistinguishable from those observed in *nos-1* and *nos-2* knock downs (Subramaniam and Seydoux, 1999), suggesting that these two *nanos* orthologs could interact with other Puf proteins. In *Drosophila*, protein expression and detailed mutant analysis suggest that Nanos and Pumilio may have nonoverlapping roles in early oocyte development (Forbes and Lehmann, 1998). Therefore, while most of the studied roles of Nanos involve Pumilio-related proteins, it remains to be seen how the two may function in each other's absence.

Tudor

The "grandchildless" phenotype of *tudor* mutants was first described in *Drosophila* by Boswell and Mahowald (1985). *tud*$^{-}$ mutants do not maintain expression of germ granule components Oskar and Vasa (Thomson and Lasko, 2004), form abnormal germ granules, and ultimately fail to produce pole cells (Boswell and Mahowald, 1985; Thomson and Lasko, 2004). Proteins containing the so-called Tudor domains have since been found in organisms ranging from yeast to humans (Ponting, 1997), and Tudor proteins localize to germ granules of flies (Arkov et al., 2006), zebrafish (Mishima et al., 2006; Strasser et al., 2008), and male mice (Chuma et al., 2006; Hosokawa et al., 2007). In flies, Tudor protein also localizes between mitochondria and germ granules, and is required for transferring ribosomal RNAs from the mitochondria (Table 1) to germ granules, an essential process in germ cell specification (Amikura et al., 2001).

Insight into the molecular basis for Tudor function has come from studies of the protein and its interactors. Studies of Tudor domain proteins in humans revealed that these domains interact with methylated arginine and lysine residues of diverse protein partners, including Sm proteins of the spliceosome (Friesen and Dreyfuss, 2000; Brahms et al., 2001). Recent studies suggest that Tudor's interactions with methylated proteins, as well as with proteins of the methylosome itself, may be required for the formation of germ granules. Localization of *Drosophila* Tudor to germ granules genetically requires the activity of the methylosome components *capsuleen* (also called *dart5*, the fly ortholog of human *dPRMT5*) and *Valois* (the fly ortholog of human *MEP50*) (Anne and Mechler, 2005; Gonsalvez et al., 2006). Further, Tudor can bind both Capsuleen and Valois in vitro, and these latter two proteins methylate Sm proteins, with which Tudor also physically interacts (Anne and Mechler, 2005). As Seydoux and Braun (2006) have pointed out, Sm proteins are common components of germ granules from vertebrates to *C. elegans*, and have been shown to be required for P granule localization and function in *C. elegans* (Barbee et al., 2002; Barbee and Evans, 2006), suggesting that the role of *Tudor* in assembling germ granules may involve its association with methylosome components and Sm proteins.

A Link Between Tudor and PIWI-Family Proteins

An additional role for Tudor was recently suggested by the finding that PIWI-family proteins in mice, frogs, and flies contain symmetrically methylated arginine residues of the type recognized by Tudor proteins (Kirino et al., 2009; Vagin et al., 2009). In the *Drosophila* ovary, PIWI-family proteins (discussed below) require *capsuleen*-dependent methylation in order to maintain their own expression and to maintain wild-type levels of piRNAs. In addition, *capsuleen*$^{-}$ mutant ovaries accumulate abnormally high levels of retrotransposons that are normally silenced by PIWI-family proteins (Kirino et al., 2009). In mice, the three PIWI proteins were recently shown to directly interact with the methylosome complex of PRMT5 and WDR77/MEP50. This complex methylates arginine residues of Mili, Miwi and Miwi2, which in turn interact with Tudor domain-containing proteins. Additionally, specific PIWI and Tudor proteins also colocalize to nuage components (Vagin et al., 2009, Wang et al., 2009). Together, these results suggest a previously unrecognized connection between the interacting networks of proteins and RNAs in germ granules.

The PIWI family of proteins (called Piwi, Aubergine, and Ago3 in *Drosophila*; Ziwi and Zili in zebrafish; Miwi, Miwi2, and Mili in mice; and Xiwi, Xili, and Xiwi2 in frogs) were named for a founding member (*P*-element-*i*nduced *wi*mpy testis) uncovered in a screen for genes involved in maintaining germ line stem cells in the *Drosophila* ovary (Lin and Spradling, 1997). This protein family has since been intensively studied for its role in silencing retrotransposons in the germ line through interactions with a special class of small RNAs called piRNAs or rasiRNAs (reviewed by Hartig et al., 2007).

Specific roles for PIWI proteins in the specification and/or maintenance of germ cells have been suggested by mutant analyses. *Piwi*$^{-}$ mutant flies have reduced numbers of pole cells (Megosh et al., 2006). Both Piwi and Aubergine localize to germ granules in nurse cells and pole cells, and Piwi physically interacts with Vasa until Piwi translocates to the nucleus, where it remains throughout germ cell migration and gametogenesis (Megosh et al., 2006). In zebrafish, *ziwi* RNA (Tan et al., 2002) and protein (Houwing et al., 2007) co-localize with Vasa to germ line-specific ribonucleoprotein complexes (RNPs), and *ziwi* mutants are agametic owing to progressive apoptosis of germ cells (Houwing et al., 2007). In mice, mutants for *miwi*, *mili*, or *miwi2* fail to complete spermatogenesis, although these genes are not required for female germ line development (Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004; Carmell et al., 2007). Recently it has been shown that Miwi and Miwi2 also form a complex with Mvh, the mouse vasa homologue (Vagin et al., 2009).

Importantly, although the PIWI-family proteins are best known for their germ cell function, some members of the related Argonaute family localize not only to germ plasm, but also to somatic RNA processing organelles such as P bodies (see, e.g., Lin et al., 2006). P bodies and germ plasm granules may thus contain organelles with closely related roles in RNA processing in both germ line and soma. While it is clear that the germ line and somatic organelles are

TABLE 1. Additional Genes With Roles in Germ Cell Specification and/or Function

Gene name	Functional gene product(s)	Molecular/genetic germline role demonstrated	Molecular function	Descriptive summary	References
Maelstrom	Protein	Dm, Mm	Putative protein binding	Nuage localization; piRNA pathway	Costa et al. (2006) Findley et al. (2003), Lim and Kai (2007), Soper et al. (2008), Zhang et al. (2008)
Par-1	Protein	Dm, Mm, Ce	Kinase	Cytoskeletal polarization; asymmetric cell division; Osk/PIE-1 stabilization	Cheeks et al. (2004), Doerflinger et al. (2003, 2006), Guo and Kemphues (1995), Reese et al. (2000), Riechmann et al. (2002), Shulman et al. (2000), Tian and Deng (2009), Vaccari and Ephrussi (2002), Zimyanin et al. (2007)
Mitochondrial rRNA	RNA	Dm, Xl	Ribosomal component	Polar granule localization; rescues UV-ablated PGCs	Amikura et al. (2001, 2005), Kashikawa et al. (2001), Kloc et al. (2001, 2002), Kobayashi et al. (1993), Kobayashi and Okada (1989)
Germ cell less	Protein	Dm, Mm	Protein binding	Cell cycle and transcriptional regulation; PGC nuclear envelope localization	de la Luna et al. (1999), Jongens et al. (1992), Kimura et al. (1999, 2003), Leatherman et al. (2000), Li et al. (2006), Nili et al. (2001), Scholz et al. (2004)
Dead end	Protein	Mm, Dr, Xl, Gg	RNA binding	Protects germline RNAs from miRNA-based degradation; vertebrate-specific	Kedde et al. (2007)
Staufen	Protein	Dm, Mm, Xl	dsRNA binding	*osk* localization; germ plasm component maintenance	Ephrussi et al. (1991), Irion et al. (2006), Ramasamy et al. (2006), St Johnston et al. (1991, 1992), Thomas et al. (2009), Yoon and Mowry (2004)
Dazl	Protein	Dm, Mm, Dr, Xl, Ce, Ol, Hs, Am	RNA binding	Germ plasm component	Anderson et al. (2007), Hashimoto et al. (2004), Houston and King (2000b), Houston et al. (1998), Johnson et al. (2001), Karashima et al. (2000), Kosaka et al. (2007), Moore et al. (2003), Reynolds et al. 2005, Saunders et al. (2003), Tung et al. (2006), Venables et al. (2001), Xu et al. (2007)

Am, *Ambystoma mexicanum*; Dm, *Drosophila melanogaster*; Mm, *Mus musculus*; Hm, *Homo sapiens*; Ol, *Oryzias latipes*; Xl, *Xenopus laevis*; Ce, *Caenorhabditis elegans*; Gg, *Gallus gallus*; Dr, *Danio rerio*. Due to space limitations, we have focused here on those primary references that focus on the germline role of these gene products.

not identical, there is some overlap in the proteins and RNAs that they contain (Megosh et al., 2006; Gallo et al., 2008; Lykke-Andersen et al., 2008). These observations highlight the molecular similarities underpinning the functional analogies between RNA processing organelles in both germ cells and somatic cells, and are consistent with the hypothesis that RNPs are repressive regulatory organelles with an ancient eukaryotic history, predating the origin of the dedicated metazoan germline (see Eulalio et al., 2007 for a detailed review).

Silencing of transposable elements in the germ line is the most well-established role of PIWI-family proteins. In the above-listed PIWI-family mutations, germ cell failure is correlated with reduced piRNA levels and abnormal accumulation of transposable elements. This function is mediated through the interaction of PIWI-family proteins and additional factors with a special class of small RNAs that provide sequence specificity to a transcript-silencing complex (see Klattenhoff et al., 2007 for additional details). Two exciting studies have recently demonstrated a role for the piRNA pathway in silencing transposable elements in somatic cells of the gonad rather than in the germ cells themselves (Malone et al., 2009; Li et al., 2009a).

Among many other defects, PIWI family mutants also exhibit defects in maintaining the localization of essential germ granule components in *Drosophila*. For example, although *piwi* is not required for the initial expression of Oskar, Vasa, or Nanos, ectopic expression of Piwi protein is able to recruit these maternal factors and increase their expression levels (Megosh et al., 2006), suggesting that Piwi acts in a positive feedback loop with these factors. Recent studies have also shown that Vasa localization genetically requires *aubergine* and *ago3*, and that these two proteins require one another for their own localization (Li et al., 2009a). Additionally, *aubergine*$^{-}$ mutants fail to properly localize the RNAi pathway members Krimper and Maelstrom (Lim and Kai, 2007). The mechanism by which PIWI-family proteins act to recruit and/or maintain localized expression of other factors to germ granules is unknown.

SYSTEM-SPECIFIC MOLECULAR COMPONENTS OF GERM LINE SPECIFICATION

Bmps

In contrast to organisms where germ cell determination relies on the inheritance of germ plasm (reviewed by Extavour and Akam, 2003), in the mouse this process requires inductive signals (Tam and Zhou, 1996) (Fig. 1). The first germ cell-inducing signal in mouse embryos was identified only a decade ago as Bmp4, a member of the *B*one *m*orphogenetic *p*rotein family. Prior to PGC induction in the proximal epiblast, Bmp4 is expressed in the tissue directly adjacent to the epiblast, the extraembryonic ectoderm (ExE). This expression is essential for PGC determination, as *Bmp4* mutant mice do not form PGCs (Lawson et al., 1999).

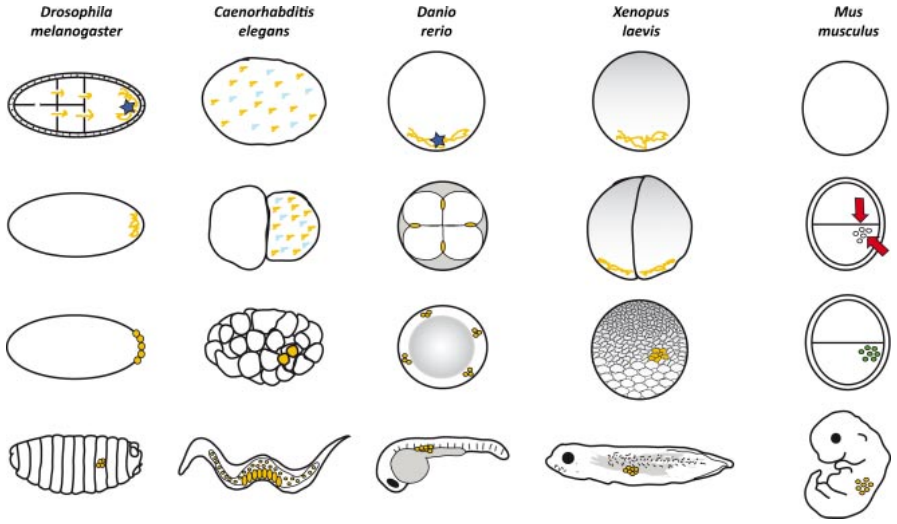


Figure 1. Localization of germ line specification molecules throughout animal development. Developmental progression in time goes from top to bottom. Germ cell specification and the localization of molecules discussed in this review are schematized for the five genetic laboratory organisms that have contributed the most to our current understanding of the molecular mechanisms underlying germ cell specification. In all organisms except for mice (*Mus musculus*), germ cell-specific gene products (yellow), including mRNAs and/or proteins of *vasa*, *nanos*, *pumilio*, *piwi*, and *tudor*, are localized to the cytoplasm of germ cells either late in oogenesis or early during embryogenesis. The fruit fly (*Drosophila melanogaster*) protein Oskar and the zebrafish (*Danio rerio*) protein Bucky ball (dark blue star) can autonomously assemble many of these germ plasm components in oocytes and early embryos. The nematode (*Caenorhabditis elegans*) protein PIE-1 (light blue) plays an important role in regulating germ line gene expression (yellow). In mice, somatic signals (red) trigger the expression of Blimp1 (green) in PGCs, followed by the expression of conserved germ line genes (yellow).

Similar to Bmp4, Bmp8b is also expressed in the ExE, and *Bmp8b* mutants lack or show very reduced numbers of PGCs. However, the effects of Bmp4 and Bmp8b are not additive (Ying et al., 2000). Very recently it was shown that the role of Bmp8 is very different from that of Bmp4. Rather than being directly required for induction of PGCs, Bmp8b signaling from the ExE restrains growth of the anterior visceral endoderm (AVE), which sends still-unidentified signals to the epiblast, thereby inhibiting Bmp4 (Ohinata et al., 2009).

Unlike Bmp4 and Bmp8b, Bmp2 is predominantly expressed in the visceral endoderm (VE). *Bmp2* mutants show a reduced number of PGCs, and Bmp2 and Bmp4 together have an additive effect on PGC development, while Bmp2 and Bmp8b together do not. Thus, Bmp signals from both the VE (Bmp2) and the ExE (Bmp4 and 8b) are required for PGC induction (Ying and Zhao, 2001).

Bmp ligands exert their function by binding and recruiting transmembrane type I and II Bmp serine/threonine kinase receptors on the cell surface. These receptors transmit the Bmp signal by phosphorylating Smad1, Smad5, or Smad8, which enter the nucleus as heterodimers with Smad4 and then serve as transcriptional regulators (Shi and Massagué, 2003). The precise downstream molecules that transduce Bmp signals and result in PGC formation are still largely unknown. There are only reports of three type I Bmp receptors possibly being involved in PGC determination, *Alk2*, *Alk3*, and *Alk6* (de Sousa Lopes et al., 2004; Ohinata et al., 2009). Several Smads have been shown to be involved in PGC formation. *Smad4*, *Smad1*, and *Smad5* mutants show reduced numbers or a complete lack of PGCs (Chang and Matzuk, 2001; Tremblay et al., 2001; Hayashi et al., 2002; Chu et al., 2004; Arnold et al., 2006). Furthermore, Smad1 and Smad5 are sufficient for PGC induction in combination with Bmp4 and Alk3 or Alk6 (Ohinata et al., 2009). However, a type II receptor involved in this process, as well as genes that are directly regulated by the Smads during PGC determination, remain to be identified.

Bmp signals have not yet been reported to be involved in germ line specification outside of mice, even though the epigenetic mode of germ cell formation has been hypothesized to be ancestral in metazoans (Extavour and Akam, 2003). More investigation will be necessary to provide functional evidence for or against this hypothesis, to determine whether the Bmp signal is conserved for germ cell induction, or whether different signaling pathways can be used by different animals to induce germ cells.

Blimp1

The earliest marker of mouse PGCs, *Blimp1* (*B l*ymphocyte *i*nduced *m*aturation *p*rotein *1*, also known as *Prdm1*) was discovered only recently (Ohinata et al., 2005). It was first described in the context of plasma cell differentiation (Turner et al., 1994). The histone methyl transferase Blimp1 contains a PR domain and a proline-rich region at the N-terminus, five C2H2 zinc fingers, and a C-terminal acidic domain (Turner et al., 1994; Tunyaplin et al., 2000). *Blimp1* orthologs are found in many bilaterian animals (see e.g. de Souza et al., 1999; Tunyaplin et al., 2000; Hinman and Davidson, 2003; Ng et al., 2006; Arenas-Mena, 2008), but expression data do not suggest a role in germ cell specification outside of mammals. Among a variety of lethal defects, *Blimp1* mutants exhibit only a very small number of PGC founder cells that fail to proliferate or migrate (Ohinata et al., 2005; Vincent et al., 2005; Robertson et al., 2007). These PGC-like cells do not show the wild-type pattern of *Hox* gene repression (Ohinata et al., 2005), and *Blimp1* mutant cells fail to repress other somatic genes (Kurimoto et al., 2008). Consistent with this observation, in other cellular differentiation processes, Blimp1 has been observed to act as a transcriptional repressor that recruits a complex of Groucho-family proteins (Ren et al., 1999) as well as a histone deacetylase (Yurke et al., 2000). The molecular interactions of Blimp1 during germ cell specification are still unknown. Interestingly, Blimp1 co-immunoprecipitates with the mammalian ortholog of *Drosophila* Capsuleen/dart5 (Prmt5, discussed above), which is required for Tudor localization. This complex of Blimp1 and Prmt5 might play a role in germ line maintenance during PGC migration, as both proteins co-localize to the nucleus during this process. After migration they co-localize in the cytoplasm, which coincides with H2A/H4R3me2s downregulation in PGCs (Ancelin et al., 2006). As mentioned above, fly Capsuleen also binds Tudor and is necessary for its localization to the germ plasm in *Drosophila*. However, whether mouse Tudor-related proteins bind the Blimp1–Prmt5 in a functional complex during PGC specification remains an open question.

Pie-1 and Polar Granule Component

In *C. elegans*, one of the best-understood germ line determinants is the PIE-1 protein. The *C. elegans pie-1* mutant (*p*haryngeal and *i*ntestinal *e*xcess) was discovered in a screen as being required for the PGC precursor (Fig. 1) to follow germ line rather than somatic fate (Mello et al., 1992). The PIE-1 protein contains two CCCH zinc fingers (ZF1 and ZF2), which are separated by arginine–serine dipeptide repeats. PIE-1 is expressed maternally, is then asymmetrically distributed to the germ line blastomeres, and continues to be expressed in the germ line throughout development (Mello et al., 1996).

The ZF2 domain of PIE-1 is required for the translation of NOS-2 protein from maternal *nos-2* mRNA in the germ line (Tenenhaus et al., 2001). However, PIE-1's principal role is to mediate transcriptional repression in the *C. elegans* germ line from P2 onwards (Seydoux et al., 1996). mRNA transcription requires phosphorylation of the C-terminal domain (CTD) of RNA polymerase II on Serine 5 (Ser5) for transcriptional initiation, and on Serine 2 (Ser2) for elongation (reviewed in Peterlin and Price, 2006; Saunders et al., 2006; Corden, 2007). The PIE-1 C-terminal region contains a motif that resembles the CTD, but has no phosphorylatable sites. This CTD-like motif, an HLX homology region, and additional C-terminal repeats, are sufficient for transcriptional repression (Batchelder et al., 1999). The PIE-1 CTD-like domain is thought to directly compete with the CTD for binding cyclin T (CycT), thereby inhibiting Ser2 phosphorylation of the CTD by the kinase CDK9 (*C*yclin-*d*ependent *k*inase 9). CycT and CDK9 together form the

*p*ositive *t*ranscription *e*longation *f*actor *b* (P-TEFb) (Zhang et al., 2003). The CTD-like motif of PIE-1 is essential for inhibiting Ser2 phosphorylation, but does not play a role in Ser5 phosphorylation; the latter activity is mediated by sequences around the CTD-like motif (Ghosh and Seydoux, 2008).

Pgc (*p*olar *g*ranule *c*omponent) accomplishes a similar transcriptional repression function by inhibition of RNA PolII via P-TEFb inhibition in *Drosophila* pole cells (Nakamura et al., 1996; Martinho et al., 2004). However, the 71-amino-acid Pgc protein does not bear any resemblance to *C. elegans* PIE-1, even though it interacts with P-TEFb and represses CTD Ser2 phosphorylation, thereby inhibiting its recruitment to transcription sites (Hanyu-Nakamura et al., 2008). Furthermore, *pgc* is essential only for germ cell migration and not for pole cell specification (Nakamura et al., 1996; Martinho et al., 2004). This is therefore an interesting case of independent evolution of two unrelated proteins that play an analogous role in the same molecular pathway. The PIE-1/Pgc relationship parallels that of two other proteins involved in germ line specification, Oskar and Bucky ball (discussed below).

oskar and *bucky ball*

All genes discussed thus far are either components of germ plasm (*vasa*, *nanos*, *pumilio*, *piwi*, *tudor*, *pie-1*, *Blimp1*), or molecules that induce the accumulation of germ plasm components (BMPs). In the case of the germ plasm components described above, these genes are necessary but not sufficient for germ cell specification and function. The BMPs and their downstream effectors, in contrast, are both necessary and sufficient for germ cell specification, but are not themselves germ plasm components. There do exist, however, two molecules that are not only germ plasm components, but are also both necessary and sufficient for germ plasm formation. These genes, *oskar* and *bucky ball*, are lineage-restricted genes with independent, recent evolutionary histories, whose shared molecular function makes them of special interest in the context of this review.

oskar (*osk*) mRNA accumulates in the posterior cytoplasm during oogenesis in *Drosophila* (Ephrussi et al., 1991; Kim-Ha et al., 1991), and its translation is likewise confined to the posterior germ plasm (Kim-Ha et al., 1995). Loss-of-function mutants do not form germ cells (Lehmann and Nüsslein-Volhard, 1986). The sufficiency of *osk* for germ plasm assembly and germ cell formation was demonstrated in elegant experiments that drove *osk* expression in ectopic embryonic locations (Ephrussi et al., 1991). This showed that *osk* gene products can autonomously recruit germ plasm components, resulting in ectopic germ cells that are capable of functional gametogenesis (Ephrussi et al., 1991). In *vas* or *tud* mutants, however, ectopic *osk* does not lead to ectopic PGCs (Ephrussi and Lehmann, 1992), consistent with the hypothesis that the role of Osk is to recruit germ plasm components rather than to induce PGC fate directly.

osk mRNA localizes to the posterior pole during stages 8–10 of oogenesis, via a mechanism that requires Staufen (Table 1), microtubules, and the plus end-directed motor protein kinesin (Lehmann and Nüsslein-Volhard, 1986; Brendza et al., 2000; Zimyanin et al., 2008). *osk* translation is confined to the posterior cytoplasm both by positive regulation of localized transcripts and by negative regulation of unlocalized transcripts (Kim-Ha et al., 1995; Wilson et al., 1996; Micklem et al., 2000; Chekulaeva et al., 2006; Klattenhoff et al., 2007; Klattenhoff and Theurkauf, 2008). When *osk* mRNA is translated, alternative start codons in the *osk* message result in two isoforms of Osk protein, Short Osk and Long Osk, which have separable roles in germ plasm assembly (Markussen et al., 1995; Rongo et al., 1995; Breitwieser et al., 1996; Vanzo and Ephrussi, 2002). Both Oskar isoforms are phosphorylated by the Par-1 kinase (Table 1) (Riechmann et al., 2002), which is enriched at the posterior in an actin-dependent but microtubule-independent step during oogenesis (Doerflinger et al., 2006). Par-1-dependent phosphorylation is thought to stabilize Osk protein in the pole plasm (Riechmann et al., 2002), where it recruits Par-1 and thereby participates in a positive feedback loop that reinforces its posterior localization (Shulman et al., 2000; Zimyanin et al., 2007).

oskar's highly upstream position in the *Drosophila* germ cell specification pathway stems from its ability to ectopically induce germ plasm assembly (Ephrussi et al., 1991; Ephrussi and Lehmann, 1992). Accordingly, Short Osk protein has been shown to directly interact with Staufen and Vasa proteins (Breitwieser et al., 1996), and recruits *nanos* mRNA (Ephrussi et al., 1991; Kim-Ha et al., 1995).

The recently described zebrafish gene *bucky ball* (*buc*) has biological properties that are remarkably similar to those of *oskar*. *buc* transcripts are localized to the vegetal pole during oogenesis, together with other germ cell-specific molecules, and Buc protein is subsequently localized to germ plasm in early cleavage stage embryos (Marlow and Mullins, 2008). Loss-of-function mutations in *buc* lead to defects in both anterior–posterior patterning (Marlow and Mullins, 2008) and germ cell formation, including a failure of *vas*, *dazl*, *nos*, and *buc* mRNAs, and other germ plasm organelles, to localize to germ cells (Bontems et al., 2009). Ectopic expression of *buc* in non-germ line cell lineages of early embryos results in supernumerary germ cells that are derived from the cells containing ectopic *buc* (Bontems et al., 2009). In summary, *buc*, like *osk*, appears to be both necessary and sufficient for germ plasm assembly and germ cell specification.

These two genes share another striking similarity: they are both very recently evolved and do not contain any recognizable functional domains. Both genes encode novel proteins, and while *osk* is restricted to the Diptera (two-winged flies), *buc* is restricted to the vertebrates (Bontems et al., 2009). Despite the presence of germ plasm and pole cells in several other insects, *osk* is not found in nonfly insect genomes (Extavour, unpublished observations). The fact that the Diptera are the insect order furthest removed from the last common insect ancestor suggests that *oskar* may be a recent evolutionary innovation associated with germ cell segregation only in this derived lineage. Similarly, early determination of germ cells is observed in some nonvertebrate deuterostome lineages, but *buc* is not found in their genomes. Despite their evolutionary unrelatedness and

lineage restriction, their biological function is highly similar. This appears to be because their molecular interactors, all of which are conserved across metazoans, are the same in both cases: for example, both Osk and Buc recruit gene products of the *vasa* and *nanos* loci to form germ plasm. More biochemical studies on the transcriptional and translational regulation of *buc*, and on its direct physical molecular interaction partners, will be necessary to determine the extent of the apparent similarity between the biological functions of *buc* and *osk*. Given that orthologs of *buc* exist across vertebrates, it will be fascinating to see whether this gene plays a role in germ cell specification of mice, whose germ cells are specified through inductive signalling rather than the cytoplasmic inheritance of germ plasm. It is conceivable, for example, that interactions between *buc* and germ line factors are conserved in the germ cells of mice, but that the expression of *buc* itself is induced through BMP signalling rather than through the localization of maternal *buc*.

SUMMARY

We have seen that many genes involved in germ cell specification are conserved across evolution, and expression studies have demonstrated some similarities in their modes of localization to germ cells (Fig. 1). Moreover, these molecules often also interact biochemically in similar ways in phylogenetically distant animals. The Nanos/Pumilio complex, the Tudor domain/PIWI family interaction, and the Tudor/spliceosome component association may therefore represent evolutionarily ancient interactions. In other cases, proteins that are not homologous serve analogous functions in germ cells: PIE-1 in nematodes and Polar granule component in flies both regulate transcriptional elongation by inhibiting RNA polymerase II phosphorylation. Moreover, some of these molecules, and in some cases their molecular roles, are conserved not just in germ cells but are also found in pluripotent cells of many types, and in RNA-processing bodies of somatic cells (see e.g. Lin et al., 2006; Megosh et al., 2006; Gallo et al., 2008; Jud et al., 2008; Lykke-Andersen et al., 2008).

There is still too little biochemical information for us to be able to know how extensive a putative metazoan germ line GRN could be. However, the conserved protein–protein and protein–RNA interactions described above could represent components of an ancestral pluripotency module, which would have likely contained Tudor domain protein, PIWI family members, and a DEAD box helicase. In early multicellular animals where a pluripotent stem cell population produced gametes, those stem cells that entered into gametogenesis would have tailored this module by the addition of unique germ cell genes, such as *nanos*, *vasa*, and *Aub*/*Ago3*. With the advent of dedicated germ cells in metazoans, this specialized germ line module would have come under the control of cytoplasmic inheritance or inductive mechanisms that operated exclusively in the germ line, preventing somatic cells (including somatic stem cells) from producing gametes.

One of the predicted consequences of modularity in development is that modules themselves can remain highly conserved throughout evolution, while their upstream effectors and downstream targets can evolve independently. The germ line specification pathways fulfill this prediction: the robustness of the molecular interactions between the conserved germ line gene products links them together into a module that can be either induced by BMP signals (mouse), assembled autonomously in oocytes (nematode, frog), or possibly even nucleated by a single molecule (fruit fly, zebrafish). In fact, the mechanisms that localize germ line determinants to germ cells appear to be relatively flexible not only on an evolutionary time scale, but also even within developing individuals. A fascinating study has recently demonstrated that several germ line genes become ectopically expressed in somatic tissues of long-lived *C. elegans* mutants that lack insulin signalling (Curran et al., 2009). Although the restriction of germ line factors such as PIE-1 to the germ cells of *C. elegans* is normally achieved through asymmetric cell divisions, in this case the ectopic expression of PIE-1 is effected at the transcriptional level. Moreover, these somatically-expressed germ line factors appear to serve a crucial function, as knocking down any of these genes drastically reduces the longevity of these mutants. Such apparent flexibility in the spatial and temporal deployment of multiple functional germ line factors further supports the notion that such factors may operate as an interacting module, capable of being induced by a variety of upstream signals. Similarly, the downstream targets of germ line factors have evolved in lineage-specific ways.

Although we have been able to identify a few conserved molecular interaction motifs among germ line specification gene products, we still have far less biochemical data than we do genetic data on these mechanisms. We have knowledge of local interactions between a few pairs of molecules, but still lack information on how the entire suite of genes is linked together biochemically. To improve our definitions of the extent and limits of this modular network, many more biochemical studies, whose results are placed into evolutionary context, are needed.

In addition to molecular studies on the traditional laboratory model organisms that have provided us with the most data thus far, work on the physical interactions of germ line genes in ''nonmodel'' organisms will also be extremely informative and should be pursued in future. A thorough understanding of the genetic control of germ line development in any organism clearly requires adequate functional tools (Sommer, 2009). However, understanding the extent of the evolutionary conservation of biochemical interactions between germ line molecules is not dependant on a complete knowledge of the developmental genetics of germ line specification. Such studies therefore need not be confined to organisms for which functional genetic analysis tools have been established.

Finally, construction of a GRN for the germ line will require a somewhat different approach to those that have been undertaken thus far. Many of the powerful GRNs that have been constructed to model aspects of somatic differentiation rely largely on transcriptional regulation (see, e.g., Davidson et al., 2002; Loose and Patient, 2004; Koide et al., 2005; Imai et al., 2006). However, most of the molecules for which functional biochemical data are available appear to be

involved in translational regulation and protein–protein interactions, suggesting that post-transcriptional gene regulation is particularly crucial in the germ line (reviewed in Cinalli et al., 2008; see e.g. Merritt et al., 2008). Moreover, it is becoming increasingly apparent that several other recently discovered mechanisms of gene regulation play a critical, albeit not yet well-defined, role in the germ line. These include piRNA-mediated transposon silencing, RNP formation to repress translation, and chromatin architecture-mediated gene regulation. We may therefore need new ways of building GRNs in order to create a framework for understanding the molecular network of the germ line.

ACKNOWLEDGMENTS

Our laboratory receives funding from the National Science Foundation (NSF; IOS-0817678). B.E.C. is supported by an NSF predoctoral award (1000081823). We apologize to colleagues whose work we have not been able to include due to space limitations.

REFERENCES

Amikura R, Kashikawa M, Nakamura A, Kobayashi S. 2001. Presence of mitochondria-type ribosomes outside mitochondria in germ plasm of *Drosophila* embryos. Proc Natl Acad Sci USA 98:9133–9138.

Amikura R, Sato K, Kobayashi S. 2005. Role of mitochondrial ribosome-dependent translation in germline formation in *Drosophila* embryos. Mech Dev 122:1087–1093.

Ancelin K, Lange U, Hajkova P, Schneider R, Bannister A, Kouzarides T, Surani MA. 2006. Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. Nat Cell Biol 8:623–630.

Anderson RA, Fulton N, Cowan G, Coutts S, Saunders PT. 2007. Conserved and divergent patterns of expression of DAZL, VASA and OCT4 in the germ cells of the human fetal ovary and testis. BMC Dev Biol 7:136.

Anne J, Mechler BM. 2005. Valois, a component of the nuage and pole plasm, is involved in assembly of these structures, and binds to Tudor and the methyltransferase Capsuléen. Development 132:2167–2177.

Arenas-Mena C. 2008. The transcription factors HeBlimp and HeTbrain of an indirectly developing polychaete suggest ancestral endodermal, gastrulation, and sensory cell-type specification roles. J Exp Zool B Mol Dev Evol 310:567–576.

Arkov A, Wang JYS, Ramos A, Lehmann R. 2006. The role of Tudor domains in germline development and polar granule architecture. Development 133:4053–4062.

Arnold SJ, Maretto S, Islam A, Bikoff EK, Robertson EJ. 2006. Dose-dependent Smad1, Smad5 and Smad8 signaling in the early mouse embryo. Dev Biol 296:104–118.

Artavanis-Tsakonas S, Rand MD, Lake RJ. 1999. Notch signaling: Cell fate control and signal integration in development. Science 284:770–776.

Asaoka-Taguchi M, Yamada M, Nakamura A, Hanyu K, Kobayashi S. 1999. Maternal Pumilio acts together with Nanos in germline development in *Drosophila* embryos. Nat Cell Biol 1:431–437.

Barbee SA, Evans TC. 2006. The Sm proteins regulate germ cell specification during early *C. elegans* embryogenesis. Dev Biol 291:132–143.

Barbee S, Lublin A, Evans T. 2002. A novel function for the Sm proteins in germ granule localization during *C. elegans* embryogenesis. Curr Biol 12:1502–1506.

Batchelder C, Dunn MA, Choy B, Suh Y, Cassie C, Shim EY, Shin TH, Mello C, Seydoux G, Blackwell TK. 1999. Transcriptional repression by the *Caenorhabditis elegans* germ-line protein PIE-1. Genes Dev 13:202–212.

Bontems F, Stein A, Marlow F, Lyautey J, Gupta T, Mullins MC, Dosch R. 2009. Bucky ball organizes germ plasm assembly in zebrafish. Curr Biol 19:414–422.

Boswell RE, Mahowald AP. 1985. *Tudor*, a gene required for assembly of the germ plasm in *Drosophila melanogaster*. Cell 43:97–104.

Braat AK, van de Water S, Goos H, Bogerd J, Zivkovic D. 2000. Vasa protein expression and localization in the zebrafish. Mech Dev 95:271–274.

Braat AK, van de Water S, Korving J, Zivkovic D. 2001. A zebrafish vasa morphant abolishes vasa protein but does not affect the establishment of the germline. Genesis 30:183–185.

Brahms H, Meheus L, de Brabandere V, Fischer U, Lührmann R. 2001. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. RNA 7:1531–1542.

Breitwieser W, Markussen F-H, Horstmann H, Ephrussi A. 1996. Oskar protein interaction with Vasa represents an essential step in polar granule assembly. Genes Dev 10: 2179–2188.

Brendza RP, Serbus LR, Duffy JB, Saxton WM. 2000. A function for kinesin I in the posterior transport of *oskar* mRNA and Staufen protein. Science 289:2120–2122.

Carmell M, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon G. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell 12:503–514.

Carrera P, Johnstone O, Nakamura A, Casanova J, Jackle H, Lasko P. 2000. VASA mediates translation through interaction with a *Drosophila* yIF2 homolog. Mol Cell 5:181–187.

Chang H, Matzuk MM. 2001. Smad5 is required for mouse primordial germ cell development. Mech Dev 104:61–67.

Cheeks RJ, Canman JC, Gabriel WN, Meyer N, Strome S, Goldstein B. 2004. *C. elegans* PAR proteins function by mobilizing and stabilizing asymmetrically localized protein complexes. Curr Biol 14:851–862.

Chekulaeva M, Hentze MW, Ephrussi A. 2006. Bruno acts as a dual repressor of *oskar* translation, promoting mRNA oligomerization and formation of silencing particles. Cell 124: 521–533.

Chu GC, Dunn NR, Anderson DC, Oxburgh L, Robertson EJ. 2004. Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. Development 131: 3501–3512.

Chuma S, Hosokawa M, Kitamura K, Kasai S, Fujioka M, Hiyoshi M, Takamune K, Noce T, Nakatsuji N. 2006. Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. Proc Natl Acad Sci USA 103:15894–15899.

Cinalli RM, Rangan P, Lehmann R. 2008. Germ cells are forever. Cell 132:559–562.

Corden JL. 2007. Transcription. Seven ups the code. Science 318:1735–1736.

Costa Y, Speed RM, Gautier P, Semple CA, Maratou K, Turner JM, Cooke HJ. 2006. Mouse MAELSTROM: The link between meiotic silencing of unsynapsed chromatin and microRNA pathway? Hum Mol Genet 15:2324–2334.

Curran SP, Wu X, Riedel CG, Ruvkun G. 2009. A soma-to-germline transformation in long-lived *Caenorhabditis elegans* mutants. Nature 459:1079–1084.

Curtis D, Treiber DK, Tao F, Zamore PD, Williamson JR, Lehmann R. 1997. A CCHC metal-binding domain in Nanos is essential for translational regulation. EMBO J 16:834–843.

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Schilstra MJ, Clarke PJ, Rust AG, Pan Z, Arnone MI, Rowen L, Cameron RA, McClay DR, Hood L, Bolouri H. 2002. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev Biol 246:162–190.

de la Luna S, Allen KE, Mason SL, La Thangue NB. 1999. Integration of a growth-suppressing BTB/POZ domain protein with the DP component of the E2F transcription factor. EMBO J 18: 212–228.

de Sousa Lopes SM, Roelen BA, Monteiro RM, Emmens R, Lin HY, Li E, Lawson KA, Mummery CL. 2004. BMP signaling mediated by ALK2 in the visceral endoderm is necessary for the generation of primordial germ cells in the mouse embryo. Genes Dev 18:1838–1849.

de Souza FS, Gawantka V, Gómez AP, Delius H, Ang SL, Niehrs C. 1999. The zinc finger gene *Xblimp1* controls anterior endomesodermal cell fate in Spemann's organizer. EMBO J 18: 6062–6072.

Deng W, Lin H. 2002. *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. Dev Cell 2:819–830.

Deshpande G, Calhoun G, Yanowitz JL, Schedl PD. 1999. Novel functions of *nanos* in downregulating mitosis and transcription during the development of the *Drosophila* germline. Cell 99: 271–281.

Doerflinger H, Benton R, Shulman JM, St Johnston D. 2003. The role of PAR-1 in regulating the polarised microtubule cytoskeleton in the *Drosophila* follicular epithelium. Development 130:3965–3975.

Doerflinger H, Benton R, Torres IL, Zwart MF, St Johnston D. 2006. *Drosophila* anterior-posterior polarity requires actin-dependent PAR-1 recruitment to the oocyte posterior. Curr Biol 16: 1090–1095.

Edwards T, Pyle S, Wharton R, Aggarwal A. 2001. Structure of Pumilio reveals similarity between RNA and peptide binding motifs. Cell 105:281–289.

Ephrussi A, Lehmann R. 1992. Induction of germ cell formation by *oskar*. Nature 358:387–392.

Ephrussi A, Dickinson LK, Lehmann R. 1991. Oskar organizes the germ plasm and directs localization of the posterior determinant *nanos*. Cell 66:37–50.

Eulalio A, Behm-Ansmant I, Izaurralde E. 2007. P bodies: At the crossroads of post-transcriptional pathways. Nat Rev Mol Cell Biol 8:9–22.

Extavour C. 2007. Evolution of the bilaterian germ line: Lineage origin and modulation of specification mechanisms. Integr Comp Biol 47:770–785.

Extavour C, Akam ME. 2003. Mechanisms of germ cell specification across the metazoans: Epigenesis and preformation. Development 130:5869–5884.

Findley SD, Tamanaha M, Clegg NJ, Ruohola-Baker H. 2003. *Maelstrom*, a *Drosophila* spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Development 130:859–871.

Forbes A, Lehmann R. 1998. *Nanos* and *Pumilio* have critical roles in the development and function of *Drosophila* germline stem cells. Development 125:679–690.

Friesen WJ, Dreyfuss G. 2000. Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN). J Biol Chem 275:26370–26375.

Fujiwara Y, Komiya T, Kawabata H, Sato M, Fujimoto H, Furusawa M, Noce T. 1994. Isolation of a DEAD-family protein gene that encodes a murine homolog of *Drosophila vasa* and its specific expression in germ cell lineage. Proc Natl Acad Sci USA 91: 12258–12262.

Gallo CM, Munro E, Rasoloson D, Merritt C, Seydoux G. 2008. Processing bodies and germ granules are distinct RNA granules that interact in *C. elegans* embryos. Dev Biol 323: 76–87.

Ghosh D, Seydoux G. 2008. Inhibition of transcription by the *Caenorhabditis elegans* germline protein PIE-1: Genetic evidence for distinct mechanisms targeting initiation and elongation. Genetics 178:235–243.

Goldstrohm AC, Hook BA, Seay DJ, Wickens M. 2006. PUF proteins bind Pop2p to regulate messenger RNAs. Nat Struct Mol Biol 13:533–539.

Gonsalvez GB, Rajendra TK, Tian L, Matera AG. 2006. The Sm-protein methyltransferase, dart5, is essential for germ-cell specification and maintenance. Curr Biol 16:1077–1089.

Gruidl ME, Smith PA, Kuznicki KA, McCrone JS, Kirchner J, Roussell DL, Strome S, Bennett KL. 1996. Multiple potential germ-line helicases are components of the germ-line-specific P granules of *Caenorhabditis elegans*. Proc Natl Acad Sci USA 93:13837–13842.

Guo S, Kemphues KJ. 1995. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell 81:611–620.

Gupta YK, Lee TH, Edwards TA, Escalante CR, Kadyrova LY, Wharton RP, Aggarwal AK. 2009. Co-occupancy of two Pumilio molecules on a single hunchback NRE. RNA 15: 1029–1035.

Hanyu-Nakamura K, Sonobe-Nojima H, Tanigawa A, Lasko P, Nakamura A. 2008. *Drosophila* Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. Nature 451: 730–733.

Hartig JV, Tomari Y, Förstemann K. 2007. piRNAs—The ancient hunters of genome invaders. Genes Dev 21:1707–1713.

Hashimoto Y, Maegawa S, Nagai T, Yamaha E, Suzuki H, Yasuda K, Inoue K. 2004. Localized maternal factors are required for zebrafish germ cell formation. Dev Biol 268:152–161.

Hay B, Ackerman L, Barbel S, Jan LY, Jan YN. 1988a. Identification of a component of *Drosophila* polar granules. Development 103:625–640.

Hay B, Jan LY, Jan YN. 1988b. A protein component of *Drosophila* polar granules is encoded by *vasa* and has extensive sequence similarity to ATP-dependent helicases. Cell 55:577–587.

Hayashi K, Kobayashi T, Umino T, Goitsuka R, Matsui Y, Kitamura D. 2002. SMAD1 signaling is critical for initial commitment of germ cell lineage from mouse epiblast. Mech Dev 118:99–109.

Hayashi Y, Hayashi M, Kobayashi S. 2004. Nanos suppresses somatic cell fate in *Drosophila* germ line. Proc Natl Acad Sci USA 101:10338–10342.

Hayashi K, de Sousa Lopes SM, Surani MA. 2007. Germ cell specification in mice. Science 316:394–396.

Hinman VF, Davidson EH. 2003. Expression of *AmKrox*, a starfish ortholog of a sea urchin transcription factor essential for endomesodermal specification. Gene Expr Patterns 3:423–426.

Hosokawa M, Shoji M, Kitamura K, Tanaka T, Noce T, Chuma S, Nakatsuji N. 2007. Tudor-related proteins TDRD1/MTR-1, TDRD6 and TDRD7/TRAP: Domain composition, intracellular localization, and function in male germ cells in mice. Dev Biol 301:38–52.

Houston DW, King ML. 2000a. Germ plasm and molecular determinants of germ cell fate. Curr Top Dev Biol 50:155–181.

Houston DW, King ML. 2000b. A critical role for *Xdazl*, a germ plasm-localized RNA, in the differentiation of primordial germ cells in *Xenopus*. Development 127:447–456.

Houston DW, Zhang J, Maines JZ, Wasserman SA, King ML. 1998. A *Xenopus* DAZ-like gene encodes an RNA component of germ plasm and is a functional homologue of *Drosophila boule*. Development 125:171–180.

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, Plasterk RH, Hannon G, Draper BW, Ketting RF. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 129:69–82.

Ikenishi K, Tanaka TS. 1997. Involvement of the protein of *Xenopus vasa* homolog (*Xenopus vasa-like gene 1*, XVLG1) in the differentiation of primordial germ cells. Dev Growth Diff 39:625–633.

Ikenishi K, Yamakita S. 2003. A trial for induction of supernumerary primordial germ cells in *Xenopus* tadpoles by injecting RNA of *Xenopus vasa* homologue into germline cells of 32-cell embryos. Dev Growth Diff 45:417–426.

Imai KS, Levine M, Satoh N, Satou Y. 2006. Regulatory blueprint for a chordate embryo. Science 312:1183–1187.

Irion U, Adams J, Chang CW, St Johnston D. 2006. Miranda couples *oskar* mRNA/Staufen complexes to the *bicoid* mRNA localization pathway. Dev Biol 297:522–533.

Irish V, Lehmann R, Akam M. 1989. The *Drosophila* posterior-group gene *nanos* functions by repressing hunchback activity. Nature 338:646–648.

Jaruzelska J, Kotecki M, Kusz K, Spik A, Firpo M, Reijo Pera RA. 2003. Conservation of a Pumilio-Nanos complex from *Drosophila* germ plasm to human germ cells. Dev Genes Evol 213:120–126.

Johnson AD, Bachvarova RF, Drum M, Masi T. 2001. Expression of axolotl DAZL RNA, a marker of germ plasm: Widespread maternal RNA and onset of expression in germ cells approaching the gonad. Dev Biol 234:402–415.

Johnstone O, Lasko P. 2004. Interaction with eIF5B is essential for Vasa function during development. Development 131: 4167–4178.

Jongens TA, Hay B, Jan LY, Jan YN. 1992. The *germ cell-less* gene product: A posteriorly localized component necessary for germ cell development in *Drosophila*. Cell 70:569–584.

Jud MC, Czerwinski MJ, Wood MP, Young RA, Gallo CM, Bickel JS, Petty EL, Mason JM, Little BA, Padilla PA, Schisa JA. 2008. Large P body-like RNPs form in *C. elegans* oocytes in response to arrested ovulation, heat shock, osmotic stress, and anoxia and are regulated by the major sperm protein pathway. Dev Biol 318:38–51.

Kadyrova LY, Habara Y, Lee TH, Wharton RP. 2007. Translational control of maternal *Cyclin B* mRNA by Nanos in the *Drosophila* germline. Development 134:1519–1527.

Karashima T, Sugimoto A, Yamamoto M. 2000. *Caenorhabditis elegans* homologue of the human azoospermia factor DAZ is required for oogenesis but not for spermatogenesis. Development 127:1069–1079.

Kashikawa M, Amikura R, Kobayashi S. 2001. Mitochondrial small ribosomal RNA is a component of germinal granules in *Xenopus* embryos. Mech Dev 101:71–77.

Kedde M, Strasser MJ, Boldajipour B, Oude Vrielink JA, Slanchev K, le Sage C, Nagel R, Voorhoeve PM, van Duijse J, Ørom UA, Lund AH, Perrakis A, Raz E, Agami R. 2007. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell 131: 1273–1286.

Kim-Ha J, Smith JL, Macdonald PM. 1991. *oskar* mRNA is localized to the posterior pole of the *Drosophila* oocyte. Cell 66:23–35.

Kim-Ha J, Kerr K, Macdonald PM. 1995. Translational regulation of *oskar* mRNA by bruno, an ovarian RNA-binding protein, is essential. Cell 81:403–412.

Kimura T, Yomogida K, Iwai N, Kato Y, Nakano T. 1999. Molecular cloning and genomic organization of mouse homologue of *Drosophila germ cell*-less and its expression in germ lineage cells. Biochem Biophys Res Commun 262:223–230.

Kimura T, Ito C, Watanabe S, Takahashi T, Ikawa M, Yomogida K, Fujita Y, Ikeuchi M, Asada N, Matsumiya K, Okuyama A, Okabe M, Toshimori K, Nakano T. 2003. Mouse *germ cell-less* as an essential component for nuclear integrity. Mol Cell Biol 23: 1304–1315.

Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, Rigoutsos I, Jongens TA, Mourelatos Z. 2009. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nat Cell Biol 11:652–658.

Klattenhoff C, Theurkauf W. 2008. Biogenesis and germline functions of piRNAs. Development 135:3–9.

Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE. 2007. *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. Dev Cell 12:45–55.

Kloc M, Bilinski S, Chan AP, Etkin LD. 2001. Mitochondrial ribosomal RNA in the germinal granules in *Xenopus* embryos revisited. Differentiation 67:80–83.

Kloc M, Dougherty MT, Bilinski S, Chan AP, Brey E, King ML, Patrick CW, Etkin LD. 2002. Three-dimensional ultrastructural analysis of RNA distribution within germinal granules of *Xenopus*. Dev Biol 241:79–93.

Knaut H, Pelegri F, Bohmann K, Schwarz H, Nusslein-Volhard C. 2000. Zebrafish *vasa* RNA but not its protein is a component of the germ plasm and segregates asymmetrically before germline specification. J Cell Biol 149:875–888.

Kobayashi S, Okada M. 1989. Restoration of pole-cell-forming ability to u.v.-irradiated *Drosophila* embryos by injection of mitochondrial lrRNA. Development 107:733–742.

Kobayashi S, Amikura R, Okada M. 1993. Presence of mitochondrial large ribosomal RNA outside mitochondria in germ plasm of *Drosophila melanogaster*. Science 260:1521–1524.

Kobayashi S, Yamada M, Asaoka M, Kitamura T. 1996. Essential role of the posterior morphogen *nanos* for germline development in *Drosophila*. Nature 380:708–711.

Koide T, Hayata T, Cho KW. 2005. *Xenopus* as a model system to study transcriptional regulatory networks. Proc Natl Acad Sci USA 102:4943–4948.

Komiya T, Itoh K, Ikenishi K, Furusawa M. 1994. Isolation and characterization of a novel gene of the DEAD box protein family which is specifically expressed in germ cells of *Xenopus laevis*. Dev Biol 162:354–363.

Kopan R, Ilagan MX. 2009. The canonical Notch signaling pathway: Unfolding the activation mechanism. Cell 137: 216–233.

Koprunner M, Thisse C, Thisse B, Raz E. 2001. A zebrafish *nanos*-related gene is essential for the development of primordial germ cells. Genes Dev 15:2877–2885.

Kosaka K, Kawakami K, Sakamoto H, Inoue K. 2007. Spatiotemporal localization of germ plasm RNAs during zebrafish oogenesis. Mech Dev 124:279–289.

Kraemer B, Crittenden S, Gallegos M, Moulder G, Barstead R, Kimble J, Wickens M. 1999. NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in *Caenorhabditis elegans*. Curr Biol 9:1009–1018.

Kuramochi-Miyagawa S, Kimura T, Ijiri TW, Isobe T, Asada N, Fujita Y, Ikawa M, Iwai N, Okabe M, Deng W, Lin H, Matsuda Y, Nakano T. 2004. *Mili*, a mammalian member of *piwi* family gene, is essential for spermatogenesis. Development 131:839–849.

Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M. 2008. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev 22:1617–1635.

Kuznicki KA, Smith PA, Leung-Chiu WM, Estevez AO, Scott HC, Bennett KL. 2000. Combinatorial RNA interference indicates GLH-4 can compensate for GLH-1; these two P granule components are critical for fertility in *C. elegans*. Development 127:2907–2916.

Lai EC. 2004. Notch signaling: Control of cell communication and cell fate. Development 131:965–973.

Lasko PF, Ashburner M. 1988. The product of the *Drosophila* gene *vasa* is very similar to eukaryotic initiation factor-4A. Nature 335: 611–617.

Lasko PF, Ashburner M. 1990. Posterior localization of Vasa protein correlates with, but is not sufficient for, pole cell development. Genes Dev 4:905–921.

Lavial F, Acloque H, Bachelard E, Nieto MA, Samarut J, Pain B. 2009. Ectopic expression of *Cvh* (*Chicken Vasa homologue*) mediates the reprogramming of chicken embryonic stem cells to a germ cell fate. Dev Biol 330:73–82.

Lawson KA, Dunn NR, Roelen BA, Zeinstra LM, Davis AM, Wright CV, Korving JP, Hogan BL. 1999. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev 13:424–436.

Leatherman JL, Kaestner KH, Jongens TA. 2000. Identification of a mouse *germ cell-less* homologue with conserved activity in *Drosophila*. Mech Dev 92:145–153.

Lehmann R, Nüsslein-Volhard C. 1986. Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of *oskar*, a maternal gene in *Drosophila*. Cell 47:141–152.

Lehmann R, Nüsslein-Volhard C. 1991. The maternal gene *nanos* has a central role in posterior pattern formation of the *Drosophila* embryo. Development 112:679–691.

Li W, Deng F, Wang H, Zhen Y, Xiang F, Sui Y, Li J. 2006. *Germ cell-less* expression in zebrafish embryos. Dev Growth Diff 48: 333–338.

Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, Kittler EL, Zapp ML, Klattenhoff C, Schulz N, Theurkauf WE, Weng Z, Zamore PD. 2009a. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell 137:509–521.

Li JZ, Zhou YP, Zhen Y, Xu Y, Cheng PX, Wang HN, Deng FJ. 2009b. Cloning and characterization of the SSB-1 and SSB-4 genes expressed in zebrafish gonads. Biochem Genet 47: 179–190.

Liang L, Diehl-Jones W, Lasko P. 1994. Localization of Vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. Development 120:1201–1211.

Lim AK, Kai T. 2007. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. Proc Natl Acad Sci USA 104:6714–6719.

Lin H, Spradling AC. 1997. A novel group of *pumilio* mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. Development 124:2463–2476.

Lin MD, Fan SJ, Hsu WS, Chou TB. 2006. *Drosophila* decapping protein 1, dDcp1, is a component of the *oskar* mRNP complex and directs its posterior localization in the oocyte. Dev Cell 10: 601–613.

Linder P, Lasko PF, Ashburner M, Leroy P, Nielsen PJ, Nishi K, Schnier J, Slonimski PP. 1989. Birth of the D-E-A-D box. Nature 337:121–122.

Liu N, Dansereau DA, Lasko P. 2003. Fat facets interacts with Vasa in the *Drosophila* pole plasm and protects it from degradation. Curr Biol 13:1905–1909.

Loose M, Patient R. 2004. A genetic regulatory network for *Xenopus* mesendoderm formation. Dev Biol 271:467–478.

Lykke-Andersen K, Gilchrist MJ, Grabarek JB, Das P, Miska E, Zernicka-Goetz M. 2008. Maternal Argonaute 2 is essential for early mouse development at the maternal-zygotic transition. Mol Biol Cell 19:4383–4392.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon G. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. Cell 137:522–535.

Markussen FH, Michon AM, Breitwieser W, Ephrussi A. 1995. Translational control of *oskar* generates short OSK, the isoform that induces pole plasm assembly. Development 121: 3723–3732.

Marlow FL, Mullins MC. 2008. *Bucky ball* functions in Balbiani body assembly and animal-vegetal polarity in the oocyte and follicle cell layer in zebrafish. Dev Biol 321:40–50.

Martinho RG, Kunwar PS, Casanova J, Lehmann R. 2004. A noncoding RNA is required for the repression of RNApolII-dependent transcription in primordial germ cells. Curr Biol 14: 159–165.

Megosh HB, Cox DN, Campbell C, Lin H. 2006. The role of PIWI and the miRNA machinery in *Drosophila* germline determination. Curr Biol 16:1884–1894.

Mello CC, Draper BW, Krause M, Weintraub H, Priess JR. 1992. The *pie-1* and *mex-1* genes and maternal control of blastomere identity in early *C. elegans* embryos. Cell 70:163–176.

Mello CC, Schubert C, Draper B, Zhang W, Lobel R, Priess JR. 1996. The PIE-1 protein and germline specification in *C. elegans* embryos. Nature 382:710–712.

Merritt C, Rasoloson D, Ko D, Seydoux G. 2008. 3′ UTRs are the primary regulators of gene expression in the *C. elegans* germline. Curr Biol 18:1476–1482.

Micklem DR, Adams J, Grünert S, St Johnston D. 2000. Distinct roles of two conserved Staufen domains in *oskar* mRNA localization and translation. EMBO J 19:1366–1377.

Miller MT, Higgin JJ, Hall TM. 2008. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. Nat Struct Mol Biol 15:397–402.

Mishima Y, Giraldez AJ, Takeda Y, Fujiwara T, Sakamoto H, Schier AF, Inoue K. 2006. Differential regulation of germline mRNAs in soma and germ cells by zebrafish *miR-430*. Curr Biol 16: 2135–2142.

Monteiro A, Podlaha O. 2009. Wings, horns, and butterfly eyespots: How do complex traits evolve? PLoS Biol 7:e37.

Moore FL, Jaruzelska J, Fox MS, Urano J, Firpo MT, Turek PJ, Dorfman DM, Pera RA. 2003. Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (deleted in AZoospermia) and DAZ-like proteins. Proc Natl Acad Sci USA 100:538–543.

Nakahata S, Katsu Y, Mita K, Inoue K, Nagahama Y, Yamashita M. 2001. Biochemical identification of *Xenopus* Pumilio as a sequence-specific cyclin B1 mRNA-binding protein that physically interacts with a Nanos homolog, Xcat-2, and a cytoplasmic polyadenylation element-binding protein. J Biol Chem 276: 20945–20953.

Nakamura A, Seydoux G. 2008. Less is more: Specification of the germline by transcriptional repression. Development 135: 3817–3827.

Nakamura A, Amikura R, Mukai M, Kobayashi S, Lasko PF. 1996. Requirement for a noncoding RNA in *Drosophila* polar granules for germ cell establishment. Science 274:2075–2079.

Ng T, Yu F, Roy S. 2006. A homologue of the vertebrate SET domain and zinc finger protein Blimp-1 regulates terminal differentiation of the tracheal system in the *Drosophila* embryo. Dev Genes Evol 216:243–252.

Nieuwkoop PD, Sutasurya LA. 1979. Primordial germ cells in the chordates. Cambridge:Cambridge University Press, p. 187.

Nieuwkoop PD, Sutasurya LA. 1981. Primordial germ cells in the invertebrates: From epigenesis to preformation. Cambridge:-Cambridge University Press, p. 258.

Nili E, Cojocaru GS, Kalma Y, Ginsberg D, Copeland NG, Gilbert DJ, Jenkins NA, Berger R, Shaklai S, Amariglio N, Brok-Simoni F, Simon AJ, Rechavi G. 2001. Nuclear membrane protein LAP2beta mediates transcriptional repression alone and together with its binding partner GCL (germ-cell-less). J Cell Sci 114: 3297–3307.

Noce T, Okamoto-Ito S, Tsunekawa N. 2001. *Vasa* homolog genes in mammalian germ cell development. Cell Struct Funct 26: 131–136.

Nüsslein-Volhard C, Frohnhöfer HG, Lehmann R. 1987. Determination of anteroposterior polarity in *Drosophila*. Science 238: 1675–1681.

Ohashi H, Umeda N, Hirazawa N, Ozaki Y, Miura C, Miura T. 2007. Expression of *vasa* (*vas*)-related genes in germ cells and specific interference with gene functions by double-stranded RNA in the monogenean, *Neobenedenia girellae*. Int J Parasitol 37: 515–523.

Ohinata Y, Payer B, O'Carroll D, Ancelin K, Ono Y, Sano M, Barton SC, Obukhanych T, Nussenzweig M, Tarakhovsky A, Saitou M, Surani MA. 2005. Blimp1 is a critical determinant of the germ cell lineage in mice. Nature 436:207–213.

Ohinata Y, Ohta H, Shigeta M, Yamanaka K, Wakayama T, Saitou M. 2009. A signaling principle for the specification of the germ cell lineage in mice. Cell 137:571–584.

Olivas W, Parker R. 2000. The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. EMBO J 19:6602–6611.

Özhan-Kizil G, Havemann J, Gerberding M. 2009. Germ cells in the crustacean *Parhyale hawaiensis* depend on Vasa protein for their maintenance but not for their formation. Dev Biol 327: 239–320.

Peterlin BM, Price DH. 2006. Controlling the elongation phase of transcription with P-TEFb. Mol Cell 23:297–305.

Ponting CP. 1997. Tudor domains in proteins that interact with RNA. Trends Biochem Sci 22:51–52.

Ramasamy S, Wang H, Quach HN, Sampath K. 2006. Zebrafish Staufen1 and Staufen2 are required for the survival and migration of primordial germ cells. Dev Biol 292:393–406.

Raz E. 2000. The function and regulation of *vasa*-like genes in germ -cell development. Genome Biol 1:1–6.

Raz E. 2003. Primordial germ-cell development: The zebrafish perspective. Nat Rev Genet 4:690–700.

Reese KJ, Dunn MA, Waddle JA, Seydoux G. 2000. Asymmetric segregation of PIE-1 in *C. elegans* is mediated by two complementary mechanisms that act through separate PIE-1 protein domains. Mol Cell 6:445–455.

Ren B, Chee KJ, Kim TH, Maniatis T. 1999. PRDI-BF1/Blimp-1 repression is mediated by corepressors of the Groucho family of proteins. Genes Dev 13:125–137.

Reynolds N, Collier B, Maratou K, Bingham V, Speed RM, Taggart M, Semple CA, Gray NK, Cooke HJ. 2005. Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of *Mvh* in germ cells. Hum Mol Genet 14:3899–3909.

Riechmann V, Gutierrez GJ, Filardo P, Nebreda AR, Ephrussi A. 2002. Par-1 regulates stability of the posterior determinant Oskar by phosphorylation. Nat Cell Biol 4:337–342.

Robertson EJ, Charatsi I, Joyner CJ, Koonce CH, Morgan M, Islam A, Paterson C, Lejsek E, Arnold SJ, Kallies A, Nutt SL, Bikoff EK. 2007. *Blimp1* regulates development of the posterior forelimb, caudal pharyngeal arches, heart and sensory vibrissae in mice. Development 134:4335–4345.

Rocak S, Linder P. 2004. DEAD-box proteins: The driving forces behind RNA metabolism. Nat Rev Mol Cell Biol 5:232–241.

Rongo C, Gavis ER, Lehmann R. 1995. Localization of *oskar* RNA regulates *oskar* translation and requires Oskar protein. Development 121:2737–2746.

Saffman EE, Lasko P. 1999. Germline development in vertebrates and invertebrates. Cell Mol Life Sci 55:1141–1163.

Saitou M. 2009. Specification of the germ cell lineage in mice. Front Biosci 14:1068–1087.

Sato K, Hayashi Y, Ninomiya Y, Shigenobu S, Arita K, Mukai M, Kobayashi S. 2007. Maternal Nanos represses *hid*/*skl*-depen-

dent apoptosis to maintain the germ line in *Drosophila* embryos. Proc Natl Acad Sci USA 104:7455–7460.

Saunders PT, Turner JM, Ruggiu M, Taggart M, Burgoyne PS, Elliott D, Cooke HJ. 2003. Absence of *mDazl* produces a final block on germ cell development at meiosis. Reproduction 126: 589–597.

Saunders A, Core LJ, Lis JT. 2006. Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol 7:557–567.

Schaner CE, Deshpande G, Schedl PD, Kelly WG. 2003. A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. Dev Cell 5:747–757.

Scholz S, Domaschke H, Kanamori A, Ostermann K, Rödel G, Gutzeit HO. 2004. Germ cell-less expression in medaka (*Oryzias latipes*). Mol Reprod Dev 67:15–18.

Schüpbach T, Wieschaus E. 1986. Maternal-effect mutations altering the anterior-posterior patterns of the *Drosophila* embryo. Roux's Arch Dev Biol 195:302–317.

Seydoux G, Braun RE. 2006. Pathway to totipotency: Lessons from germ cells. Cell 127:891–904.

Seydoux G, Mello CC, Pettitt J, Wood WB, Priess JR, Fire A. 1996. Repression of gene expression in the embryonic germ lineage of *C. elegans*. Nature 382:713–716.

Shi Y, Massagué J. 2003. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell 113:685–700.

Shulman JM, Benton R, St Johnston D. 2000. The *Drosophila* homolog of *C. elegans* PAR-1 organizes the oocyte cytoskeleton and directs *oskar* mRNA localization to the posterior pole. Cell 101:377–388.

Sommer RJ. 2009. The future of evo-devo: Model systems and evolutionary theory. Nat Rev Genet 10:416–422.

Sonoda J, Wharton RP. 1999. Recruitment of Nanos to *hunchback* mRNA by Pumilio. Genes Dev 13:2704–2712.

Sonoda J, Wharton RP. 2001. *Drosophila* brain tumor is a translational repressor. Genes Dev 15:762–773.

Soper SF, van der Heijden GW, Hardiman TC, Goodheart M, Martin SL, de Boer P, Bortvin A. 2008. Mouse *maelstrom*, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. Dev Cell 15:285–297.

Spike C, Meyer N, Racen E, Orsborn A, Kirchner J, Kuznicki K, Yee C, Bennett K, Strome S. 2008. Genetic analysis of the *Caenorhabditis elegans* GLH family of P-granule proteins. Genetics 178:1973–1987.

St Johnston D, Beuchle D., Nüsslein-Volhard C. 1991. Staufen, a gene required to localize maternal RNAs in the Drosophila egg. Cell 66:51–63.

St Johnston D, Brown NH, Gall JG, Jantsch M. 1992. A conserved doublestranded RNA-binding domain. Proc Natl Acad Sci USA 89:10979–10983.

Strasser MJ, Mackenzie NC, Dumstrei K, Nakkrasae L, Stebler J, Raz E. 2008. Control over the morphology and segregation of Zebrafish germ cell granules during embryonic development. BMC Dev Biol 8:58.

Strome S. 2005. Specification of the germ line. WormBook, pp. 1–10.

Strome S, Lehmann R. 2007. Germ versus soma decisions: Lessons from flies and worms. Science 316:392–393.

Styhler S, Nakamura A, Swan A, Suter B. 1998. *vasa* is required for GURKEN accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. Development 125:1569–1578.

Styhler S, Nakamura A, Lasko P. 2002. VASA localization requires the SPRY-domain and SOCS-box containing protein, GUSTAVUS. Dev Cell 3:865–876.

Subramaniam K, Seydoux G. 1999. *nos-1* and *nos-2*, two genes related to *Drosophila nanos*, regulate primordial germ cell development and survival in *Caenorgabditis elegans*. Development 126:4861–4871.

Sunanaga T, Watanabe A, Kawamura K. 2007. Involvement of *vasa* homolog in germline recruitment from coelomic stem cells in budding tunicates. Dev Genes Evol 217:1–11.

Tam PP, Zhou SX. 1996. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. Dev Biol 178: 124–132.

Tan CH, Lee TC, Weeraratne SD, Korzh V, Lim TM, Gong Z. 2002. *Ziwi*, the zebrafish homologue of the *Drosophila piwi*: Co-localization with *vasa* at the embryonic genital ridge and gonad-specific expression in the adults. Gene Expr Patterns 2:257–260.

Tanaka SS, Toyooka Y, Akasu R, Katoh-Fukui Y, Nakahara Y, Suzuki R, Yokoyama M, Noce T. 2000. The mouse homolog of *Drosophila vasa* is required for the development of male germ cells. Genes Dev 14:841–853.

Tautz D, Pfeifle C. 1989. A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. Chromosoma 98:81–85.

Tenenhaus C, Subramaniam K, Dunn MA, Seydoux G. 2001. PIE-1 is a bifunctional protein that regulates maternal and zygotic gene expression in the embryonic germ line of *Caenorhabditis elegans*. Genes Dev 15:1031–1040.

Thomas MG, Tosar LJM, Desbats MA, Leishman CC, Boccaccio GL. 2009. Mammalian Staufen 1 is recruited to stress granules and impairs their assembly. J Cell Sci 122:563–573.

Thomson T, Lasko P. 2004. *Drosophila tudor* is essential for polar granule assembly and pole cell specification, but not for posterior patterning. Genesis 40:164–170.

Tian AG, Deng WM. 2009. Par-1 and Tau regulate the anterior-posterior gradient of microtubules in *Drosophila* oocytes. Dev Biol 327:458–464.

Toyooka Y, Tsunekawa N, Takahashi Y, Matsui Y, Satoh M, Noce T. 2000. Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. Mech Dev 93:139–149.

Tremblay KD, Dunn NR, Robertson EJ. 2001. Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. Development 128:3609–3621.

Tsuda M, Sasaoka Y, Kiso M, Abe K, Haraguchi S, Kobayashi S, Saga Y. 2003. Conserved role of nanos proteins in germ cell development. Science 301:1239–1241.

Tung JY, Luetjens CM, Wistuba J, Xu EY, Reijo Pera RA, Gromoll J. 2006. Evolutionary comparison of the reproductive genes, DAZL and BOULE, in primates with and without DAZ. Dev Genes Evol 216:158–168.

Tunyaplin C, Shapiro MA, Calame KL. 2000. Characterization of the B lymphocyte-induced maturation protein-1 (*Blimp-1*) gene, mRNA isoforms and basal promoter. Nucleic Acids Res 28: 4846–4855.

Turner CA, Mack DH, Davis MM. 1994. Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. Cell 77: 297–306.

Vaccari T, Ephrussi A. 2002. The fusome and microtubules enrich Par-1 in the oocyte, where it effects polarization in conjunction with Par-3, BicD, Egl, and dynein. Curr Biol 12:1524–1528.

Vagin VV, Wohlschlegel J, Qu J, Jonsson Z, Huang X, Chuma S, Girard A, Sachidanandam R, Hannon GJ, Aravin AA. 2009. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. Genes Dev 23:1749–1762.

Vanzo NF, Ephrussi A. 2002. Oskar anchoring restricts pole plasm formation to the posterior of the *Drosophila* oocyte. Development 129:3705–3714.

Venables JP, Ruggiu M, Cooke HJ. 2001. The RNA-binding specificity of the mouse Dazl protein. Nucleic Acids Res 29: 2479–2483.

Vincent SD, Dunn NR, Sciammas R, Shapiro-Shalef M, Davis MM, Calame K, Bikoff EK, Robertson EJ. 2005. The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. Development 132:1315–1325.

Wagner G, Pavlicev M, Cheverud J. 2007. The road to modularity. Nat Rev Genet 8:921–931.

Wang C, Lehmann R. 1991. Nanos is the localized posterior determinant in *Drosophila*. Cell 66:637–647.

Wang X, McLachlan J, Zamore PD, Hall TM. 2002. Modular recognition of RNA by a human pumilio-homology domain. Cell 110:501–512.

Wang J, Saxe JP, Tanaka T, Chuma S, Lin H. 2009. Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. Curr Biol 19:640–644.

Wilson JE, Connell JE, Macdonald PM. 1996. *aubergine* enhances *oskar* translation in the *Drosophila* ovary. Development 122: 1631–1639.

Wreden C, Verrotti AC, Schisa JA, Lieberfarb ME, Strickland S. 1997. *Nanos* and *pumilio* establish embryonic polarity in *Drosophila* by promoting posterior deadenylation of *hunchback* mRNA. Development 124:3015–3023.

Xu H, Li M, Gui J, Hong Y. 2007. Cloning and expression of medaka *dazl* during embryogenesis and gametogenesis. Gene Expr Patterns 7:332–338.

Ying Y, Zhao GQ. 2001. Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. Dev Biol 232:484–492.

Ying Y, Liu XM, Marble A, Lawson KA, Zhao GQ. 2000. Requirement of *Bmp8b* for the generation of primordial germ cells in the mouse. Mol Endocrinol 14:1053–1063.

Yoon C, Kawakami K, Hopkins N. 1997. Zebrafish *vasa* homologue RNA is localized to the cleavage planes of 2- and 4 -cell-stage embryos and is expressed in the primordial germ cells. Development 124:3157–3166.

Yoon YJ, Mowry KL. 2004. *Xenopus* Staufen is a component of a ribonucleoprotein complex containing Vg1 RNA and kinesin. Development 131:3035–3045.

Yurke B, Tuberfield AJ, Mills APJ, Simmel FC, Neumann JF. 2000. A DNA-fueled molecular machine made of DNA. Nature 406: 605–608.

Zamore PD, Williamson JR, Lehmann R. 1997. The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. RNA 3:1421–1433.

Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, Wickens MP. 1997. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. Nature 390:477–484.

Zhang F, Barboric M, Blackwell TK, Peterlin BM. 2003. A model of repression: CTD analogs and PIE-1 inhibit transcriptional elongation by P-TEFb. Genes Dev 17:748–758.

Zhang D, Xiong H, Shan J, Xia X, Trudeau VL. 2008. Functional insight into Maelstrom in the germline piRNA pathway: A unique domain homologous to the DnaQ-H 3′-5′ exonuclease, its lineage-specific expansion/loss and evolutionarily active site switch. Biol Direct 3:48.

Zimyanin V, Lowe N, St Johnston D. 2007. An *oskar*-dependent positive feedback loop maintains the polarity of the *Drosophila* oocyte. Curr Biol 17:353–359.

Zimyanin VL, Belaya K, Pecreaux J, Gilchrist MJ, Clark A, Davis I, St Johnston D. 2008. In vivo imaging of *oskar* mRNA transport reveals the mechanism of posterior localization. Cell 134:843–853.

MRD

Appendix B:

Preliminary experiments regarding a possible role for *oskar* in the *Drosophila* nervous system

INTRODUCTION

In Chapter 2 and Chapter 5 of this thesis, I present evidence that *oskar* has a function in the nervous system of the cricket *Gryllus bimaculatus*. Specifically, I show that *Gb-oskar* functions in embryonic neuroblasts (Chapter 2), and also in neuroblasts present in the adult brain (Chapter 5). These data, together with reports of neural function of *oskar* in *Drosophila* (Dubnau et al., 2003; Xu et al., 2013) suggest that a role for *oskar* in the nervous system, but not in the germ line, is conserved between these two species, and may therefore be the ancestral function of *oskar*.

However, the precise roles of *oskar* in the *Drosophila* nervous system are not understood, making it unclear to what extent the specific neural roles are in fact conserved between *Gryllus* and *Drosophila*. Two studies have reported a phenotype for *oskar* in the *Drosophila* nervous system. Xu *et al* (2013) demonstrate that *osk* RNA particles, like *nanos* RNA particles, are motile in larval dendritic arborization (da) neurons, and further show that *osk* mutant larvae display defective motility of *nanos* particles. Furthermore, da-specific *osk* RNAi leads to defects in response to a mechanical stimulation (Xu et al., 2013). However, a direct comparison of this function between *Drosophila* and *Gryllus* is difficult to conceptualize, as hemimetabolous insects such as *Gryllus* do not pass through a larval stage at all.

The other report of a role for *oskar* comes from a forward genetic mutagenesis screen for defects in olfactory learning (Dubnau et al., 2003). However, these *oskar* data are ambiguous, as the putative *oskar* allele (dubbed *norka*), obtained via GAL4 insertional mutagenesis, does not map within the known *osk* locus (Fig 1). In fact, the *norka* insertion is markedly closer to *polychaetoid* (*pyd*), which encodes a cell adhesion molecule with known neural function in

development and in learning (Chen et al., 1996; Eddison et al., 2012; Seppa et al., 2008). In addition, the *norka* allele does not drive expression in the germ line (**Figure A.1**), where *oskar* has a well-characterized role (Ephrussi and Lehmann, 1992; Jenny et al., 2006). Despite these facts, Dubnau *et al* (2003) present northern blot data suggesting that the *osk* transcript is unusually long in *norka* mutants. Clarifying these issues will require additional studies with *bona fide* alleles of *osk*.

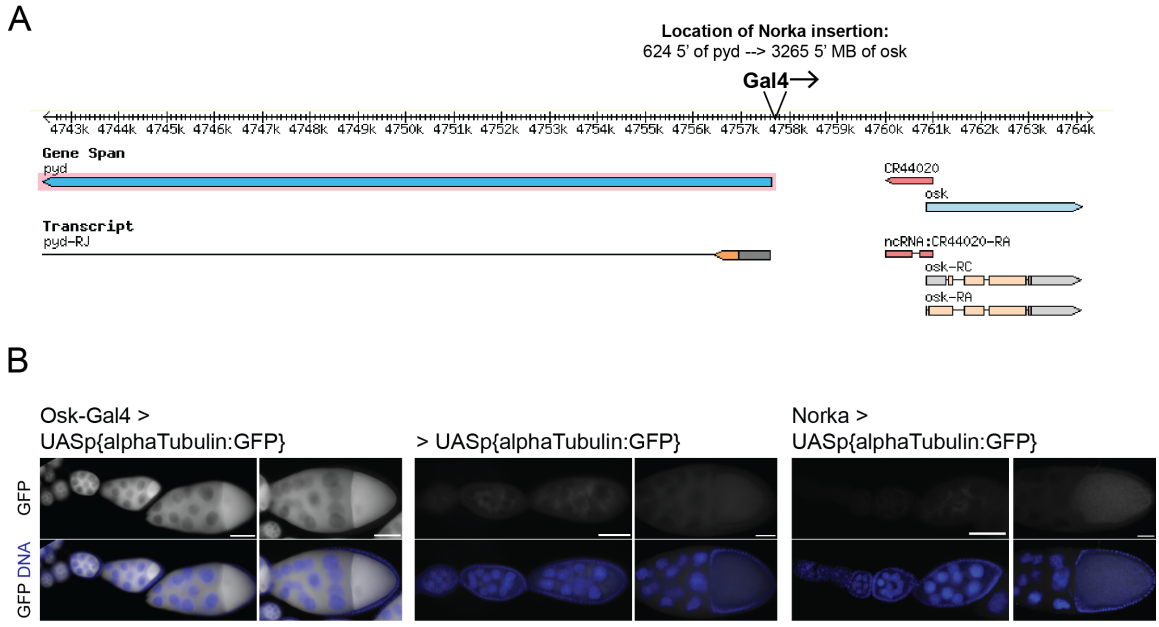


Figure A.1 The *norka* mutant may not be a true *osk* allele. (A) The genomic location of the Gal4 insertion in the *norka* allele. Data from (Dubnau et al., 2003). Note that the insertion is far upstream of *osk*, outside of any known transcription unit, but is quite close to *pyd*, a gene with several known neural functions (see text). (B) The *norka* GAL4 insertion does not drive GFP expression in the germ line, where Oskar is expressed. An Osk-Gal4 line (containing 1,743 bp of genomic sequence upstream of *oskar* fused to Gal4) drives expression throughout oogenesis (left panels), whereas *norka* (right panels) does not drive expression above background levels (compare to the no control, middle panel). Anterior is to the right. Scale bars = 50 μm.

In a preliminary attempt to clarify some of these issues, and to test whether there may be a more direct conservation of *oskar* function between *Drosophila* and *Gryllus*, I have undertaken preliminary studies of *oskar* in the *Drosophila* nervous system. In this Appendix, I present data suggesting that Oskar protein is undetectable by antibody staining in the adult brain. I then show

that an Oskar-Gal4 line does not drive expression in neurons of the adult brain. Finally, I present preliminary data suggesting that neither *oskar* RNAi nor *oskar* over-expression in the embryonic neuroblasts disrupt CNS development, in contrast to *Gryllus*. In sum, these initial data do not indicate a function for *oskar* in the embryonic nervous system or in the adult brain, but these studies are far from complete. I suggest several additional experiments that could help extend these data.

METHODS

Drosophila strains

The following stocks were used:

osk^{norka} (Dubnau et al., 2003) - gift of B. di Bivort, Harvard University.

oskar-Gal4 – a 1,743-bp genomic region upstream of *oskar* fused to Gal4 (Telley et al., 2012) - (Bloomington 44241, 44242)

oskar^{54}*/oskar*^{84} – strong *osk* hypomorphic transheterozygote (Lehmann and Nüsslein-Volhard, 1986; Vanzo and Ephrussi, 2002) (gift of R. Lehmann, NYU)

oskar^{A87}*/* Df(3R)PXT103 – *osk* null, over a deficiency (Jenny et al., 2006) (gift of E. Gavis, Princeton)

pJFRC7 = p{*20XUAS-IVS-:mCD8::GFP*}attP2 – Membrane-GFP reporter for somatic expression (Pfeiffer et al., 2010) (gift of J. Tuthill, R. Wilson Lab, Harvard Medical School)

UASp:GFP::alphaTub (*w ;* p{w[+mC]=UASp-GFPS65C-alphaTub84B) – Cytoplasmic GFP reporter for germ line (Bloomington 7373)

UAS-*osk* (w ; p[w+, UAS:osk]) – Line for *osk* over-expression (Zimyanin et al., 2007) (gift of D. St. Johnston, Gurdon Institute)

osk-RNAi – RNAi line for *osk* knockdown (VDRC 107546)

asense-Gal4 / CyO (Zhu et al., 2006) – Gal4 driver for neuroblasts (gift of T. Lee, Janelia Farms)

OK107 (*w ; P{w[+mW.hs]=GawB}ey[OK107]* – Gal4 driver for the mushroom body. (Gift of W. Tobin, R. Wilson Lab, Harvard Medical School)

Antibody staining

Adult brains were dissected and de-sheathed using fine forceps in ice-cold PBS. Brains were fixed for 60-90 minutes in 4% paraformaldehyde, and were permeabilized with 1% Triton-X in PBS prior to beginning antibody staining following standard protocols. Primary incubation was overnight at room temperature, and secondary incubation was overnight at 4ºC. Primary antibodies were *nc82* (1:50; labels neuropil, Iowa Developmental Hybridoma Bank), anti-Eve 28B (1:30, Iowa Developmental Hybridoma Bank), anti-HRP Alexa647 congujate (1:50, gift of S. Kunes), anti-Oskar (1:1500, preabsorbed against *Drosophila* embryos, gift of A. Ephrussi), and anti-GFP Alexa488 conjugate (1:250; Molecular Probes).

Embryonic CNS Scoring

Embryos were double-stained for Eve and HRP, then individually mounted on slides, oriented ventral side towards the coverslip. Confocal z-stacks covering the entirety of the embryonic CNS (all visible Eve+ and HRP staining) were captured, and were manually scored for any discernable defects in any of the EL, U/QC, RP2, aCC/pCC, and axonal scaffold. Between five and seven segments were scored per embryo, and the percentage of affected segments for each neuron cluster was used as the single data point for that embryo. Maximum intensity projections were made of representative embryos, but scoring was done on the full stacks to ensure that all Eve+ cells could be identified.

RESULTS

Oskar protein cannot be detected in the adult brain by antibody staining

Dubnau *et al* (2003) report that the *norka* mutation represents an allele of *oskar*. Because the allele was generated via GAL4 insertional mutagenesis, they were able to use a UAS:GFP reporter to reveal expression driven by the *norka* insertion. They show that *norka* drives expression in a variety of neurons in the adult brain, including the mushroom body, the

anatomical substrate associated with olfactory learning and memory (see Figure 3C in Dubnau et al., 2003).

Given the ambiguity of whether the *norka* allele is a true *osk* allele, I wished to test whether Oskar protein could be detected in the mushroom body or other regions of the adult brain via antibody staining. Using an antibody that is specific to Oskar protein in ovaries (**Figure A.2A**), I was unable to detect specific signal in adult brains above background levels observed in two *osk* mutants (**Figure A.2B**). These data are not consistent with the conclusion that the expression driven by the *norka* GAL4 insertion captures wildtype Oskar expression, although it remains a formal possibility that levels of Oskar below the detection limit of antibody staining may function in the adult brain.

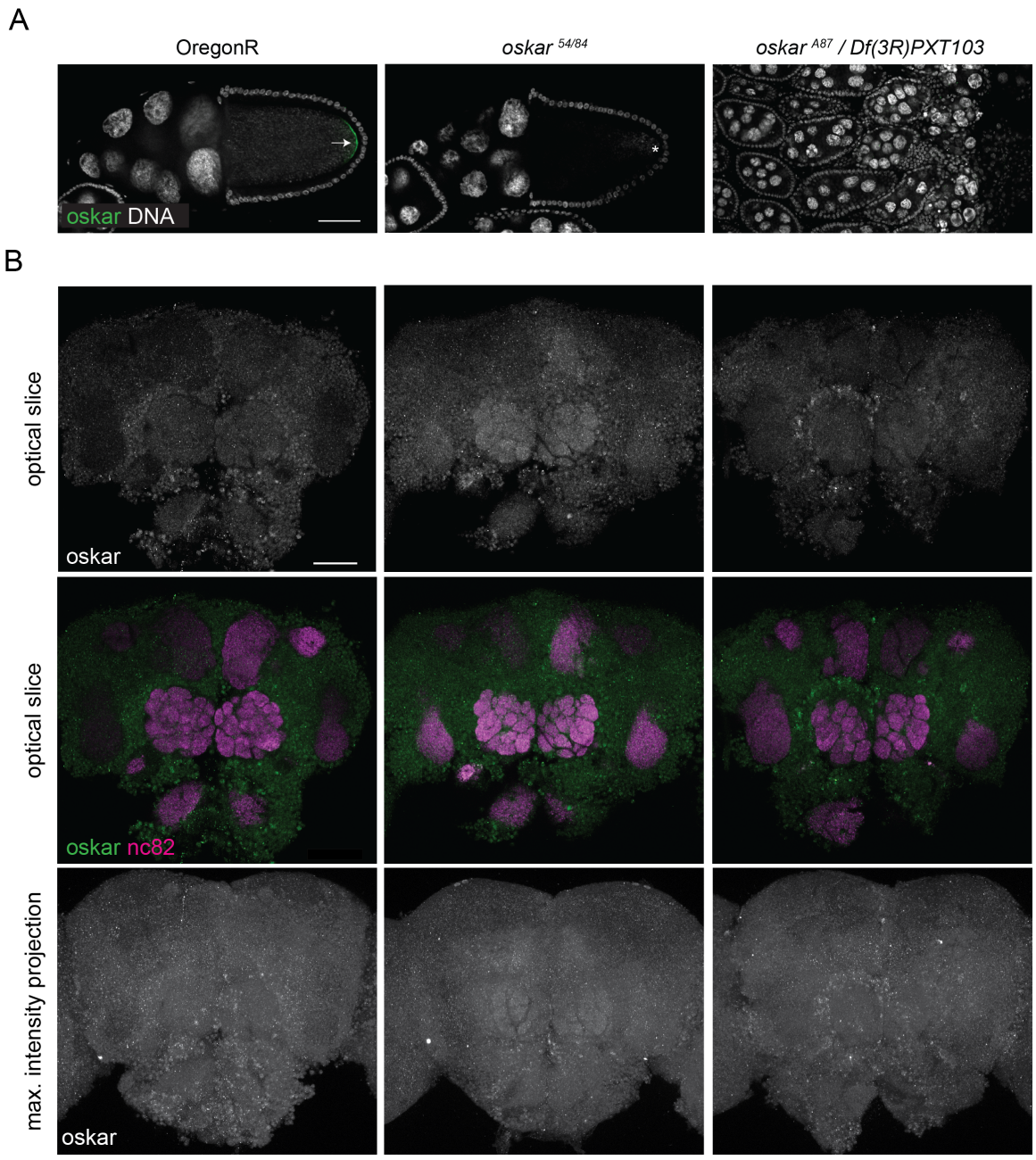


Figure A.2. Oskar protein cannot be detected above background levels in the brain using antibody staining. (A) anti-Oskar antibody is highly specific to Oskar protein in oocytes (left panel), and gives no signal in two strong *oskar* mutant ovaries (middle and right panels). Note that oogenesis arrests at stage 7 in *oskar*[A87] ovaries, as this is an RNA null (Jenny et al., 2006). Anterior is to the left. (B) In brains, Oskar protein cannot be detected above the background levels present also in *oskar* mutants. Signal is shown in single optical slices (top row), overlaid with the neuropil maker *nc82* (middle row) to see brain structure, and in maximum intensity projection (bottom row). Scale bars = 50 μm.

Osk-Gal4 does not drive expression in neurons of the adult brain

As an additional test for Oskar expression in the *Drosophila* brain, I crossed a published Osk-Gal4 line (made by fusing 1,743 bp of genomic DNA upstream of *oskar* to Gal4)(Telley et al., 2012) to a membrane-bound GFP reporter. A mushroom body driver, *OK107*, was used as a positive control, and a reporter-only fly was used as a negative control **(Figure A.3A**).

Osk-Gal4 did not drive GFP in any neurons of the adult brain (**Figure A.3B**). Surprisingly, GFP was instead observed in the non-neuronal sheath that surrounds the brain, which is typically ripped during dissection to aid in antibody penetration. The biological importance of this expression is unknown. Regardless, these experiments are consistent with the antibody staining (**Figure A.2**) and suggest that Oskar expression is either absent from the neurons of the adult brain, or, if present at low level, is driven by an enhancer outside of the 1,743 bp upstream of the *oskar* locus.

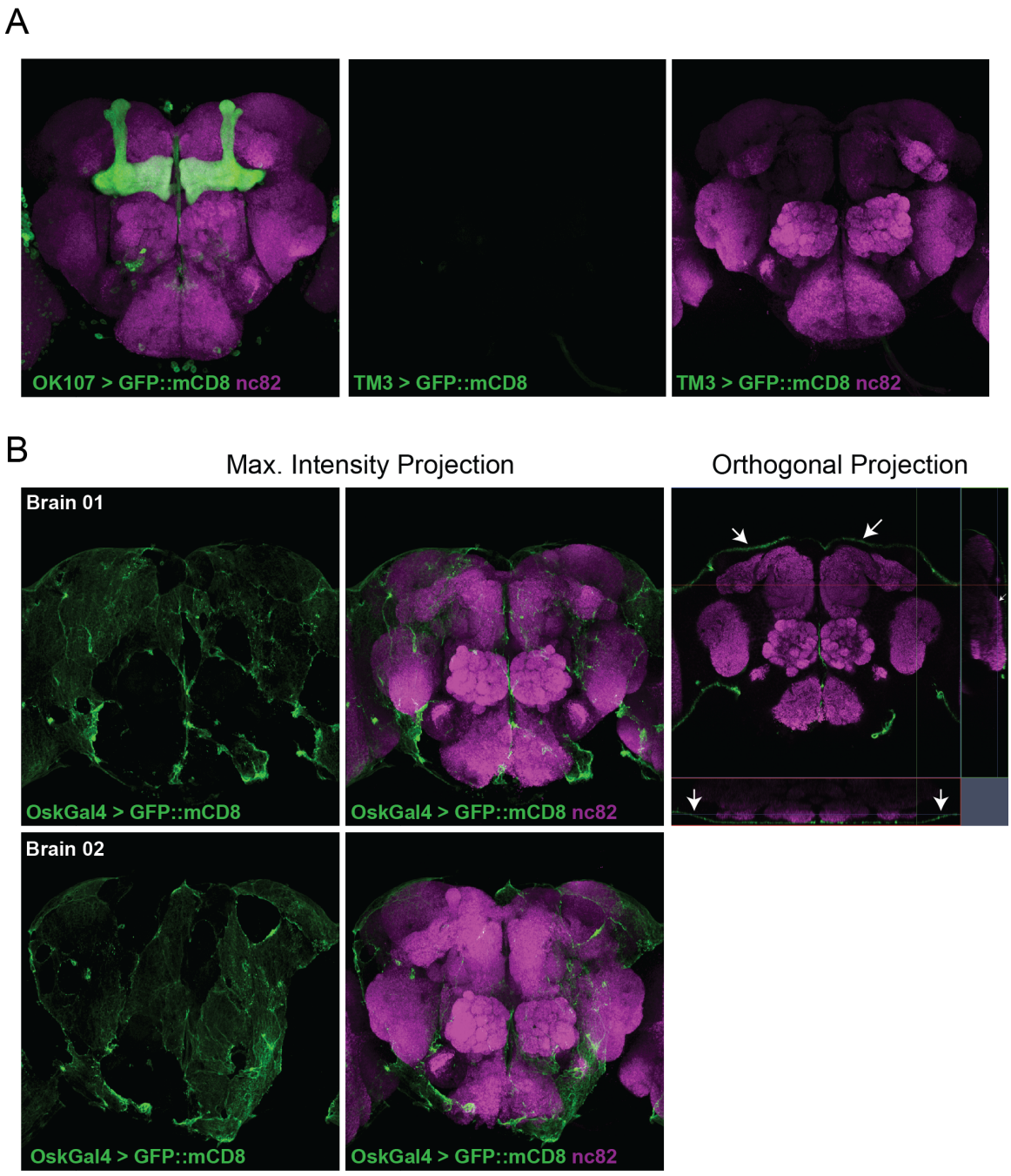


Figure A.3. Osk-Gal4 does not drive expression in neurons of the adult brain. (A) Positive and negative controls for GFP reporter construct. A mushroom body driver, OK107, drives strong, membrane-bound GFP expression in the mushroom body (left panel), and the reporter construct has no leaky expression in the absence of a driver (middle and right panel). (B) Osk-Gal4 drives expression in the sheath surrounding the brain (white arrows). Neurons themselves are not stained, as revealed in an orthogonal projection (top row, right panel). Two separate brains are shown to demonstrate that the staining pattern in the sheath results from mechanical ripping done during dissection to allow antibody penetration into the brain.

Preliminary evidence suggests that oskar RNAi and over-expression in the embryonic neuroblasts does not perturb nervous system development

We have previously shown that *Gb-oskar* is expressed in embryonic neuroblasts of crickets, and that *Gb-oskar* RNAi interferes with the development of the central nervous system (CNS) (Chapter 2; Ewen-Campen et al., 2012). I therefore wished to test for a comparable function of *oskar* in the embryonic neuroblasts of *Drosophila*, as a direct comparison to the cricket.

I drove expression of *osk*-RNAi and UAS-*osk* (an *osk* over-expression construct) using an Asense-GAL4 line (Zhu et al., 2006) which is expressed in neuroblasts and their progeny. I assayed CNS development using two metrics: (1) *even-skipped* antibody staining, which labels a small number of well-characterized neurons present in each segment, and which are often disrupted in mutants that affect neuroblast division (e.g. Dorfman et al., 1991; Ikeshima-Kataoka et al., 1997) and (2) the axonal scaffolding, visualized via anti-HRP staining, which can also reveal neuroblast defects (e.g. Ikeshima-Kataoka et al., 1997).

Neither *osk*-RNAi nor UAS-*osk* led to a detectable phenotype in *even-skipped*-positive cells nor in the morphology of the axonal tracts compared to controls (**Figure A.4**). In contrast to mutants defective for neuroblast division, which disrupt or abolish Eve+ neurons, all Eve+ cells could be readily identified in both *osk* manipulations, and were indistinguishable from wild type in location and appearance (**Figure A.4**). In addition, breaks or defects in the axonal tract, which can result from defects in neuroblast division cells (e.g. Dorfman et al., 1991; Ikeshima-Kataoka et al., 1997), were not observed (**Figure A.4**). Thus, these data did not reveal a function for *oskar* in the embryonic neuroblasts.

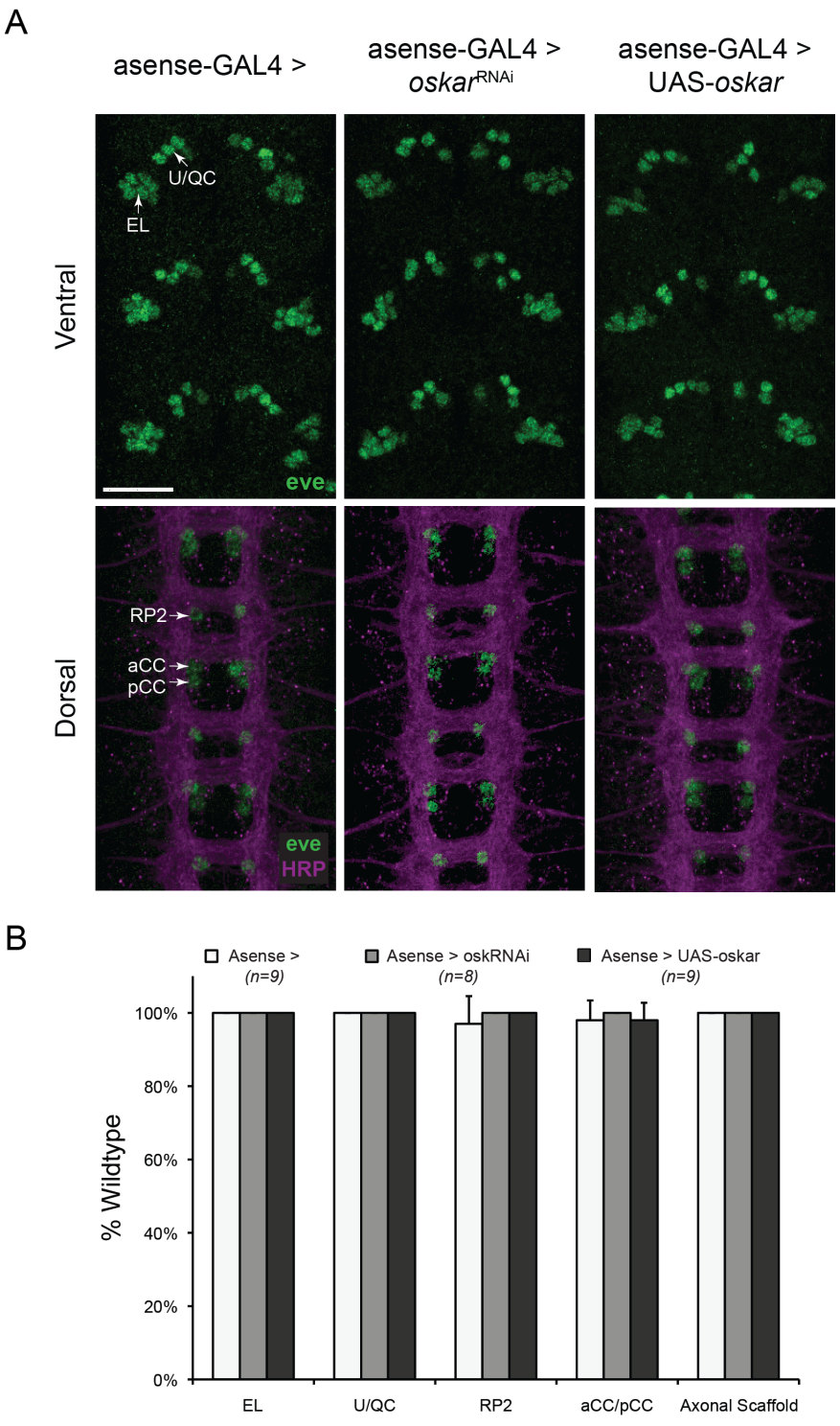


Figure A.4. CNS development is unperturbed in asense:GAL4 > *osk*-RNAi and asense:GAL4 > UAS:*osk* (A) Example images of *eve*-positive cells in the indicated genotypes. Two focal planes are shown, revealing ventral clusters (EL and U/QC) and dorsal motor neurons (RP2, aCC/pCC), as well as the axonal tracts visualized with HRP. (B) Quantification of scoring for each neuron cluster/axonal scaffold in each of the three genomes. Scale bar = 20 μm.

DISCUSSION

I have presented several initial investigations into possible roles for *oskar* in the *Drosophila* nervous system, in order to extend a recent report of *oskar* function in the larval da neurons (Xu et al., 2013). I was unable to detect Oskar protein in the adult brain via antibody staining (**Figure A.2**), and I also showed that an Osk-Gal4 reporter construct does not drive expression in the neurons of the adult brain (**Figure A.3**). These results conflict with the interpretation of the data obtained from the *norka* mutant (**Figure A.1**) (Dubnau et al., 2003), which drives expression in neurons including those of the mushroom body, and are consistent with the hypothesis that the *norka* mutant may not in fact be a *bona fide* allele of *oskar*. Furthermore, these expression data do not immediately suggest that *oskar* plays a role in the adult brain, although functional studies of additional *oskar* alleles are required to directly test whether or not this is the case. Specifically, it would be interesting to assay for olfactory learning and memory phenotypes in known *oskar* mutants (which are viable, as the gene is a maternal effect gene), and in flies containing UAS:*osk*-RNAi under the control of an adult brain-specific GAL4 driver.

In addition, I have presented preliminary data showing that the embryonic CNS develops normally when *oskar* levels are perturbed in neuroblasts using an asense-Gal4 driver. With the important caveat that the sample sizes were <10 per genotype for these experiments, these data did not reveal a phenotype for *osk*-RNAi or *UAS:oskar*. It will be important to repeat these experiments using additional GAL4 drivers and *oskar*-RNAi constructs, in addition to testing whether additional Dicer can increase RNAi efficacy. The *elav*-GAL4 driver, which drives expression in all embryonic neurons and a small number of glial cells (Berger et al., 2007), may be a good candidate, as are several others including *deadpan*-GAL4 (Lin et al., 2010), *prospero*-GAL4 (Atwood et al., 2007), and *worniu*-GAL4 (Atwood et al., 2007). If, following such

additional experiments, no embryonic CNS phenotype is observed in *Drosophila*, it would suggest that function of *Gb-oskar* in the *Gryllus* embryonic CNS is not conserved.

REFERENCES

Atwood, S.X., Chabu, C., Penkert, R.R., Doe, C.Q., Prehoda, K.E., 2007. Cdc42 acts downstream of Bazooka to regulate neuroblast polarity through Par-6 aPKC. J Cell Sci 120, 3200–3206.

Berger, C., Renner, S., Lüer, K., Technau, G.M., 2007. The commonly used marker ELAV is transiently expressed in neuroblasts and glial cells in the Drosophila embryonic CNS. Dev. Dyn. 236, 3562–3568.

Chen, C.M., Freedman, J.A., Bettler, D.R., Manning, S.D., Giep, S.N., Steiner, J., Ellis, H.M., 1996. Polychaetoid is required to restrict segregation of sensory organ precursors from proneural clusters in Drosophila. Mech Dev 57, 215–227.

Dorfman, R., Doe, C.Q., Shilo, B.Z., Chu-LaGraff, Q., Wright, D.M., Scott, M.P., 1991. The prospero gene specifies cell fates in the Drosophila central nervous system. Cell 65, 451–464.

Dubnau, J., Chiang, A.-S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U., Broger, C., Tully, T., 2003. The staufen/pumilio pathway is involved in Drosophila long-term memory. Curr Biol 13, 286–296.

Eddison, M., Belay, A.T., Sokolowski, M.B., Heberlein, U., 2012. A Genetic Screen for Olfactory Habituation Mutations in Drosophila: Analysis of Novel Foraging Alleles and an Underlying Neural Circuit. PLoS ONE 7, e51684.

Ephrussi, A., Lehmann, R., 1992. Induction of germ cell formation by oskar. Nature 358, 387–392.

Ewen-Campen, B., Srouji, J.R., Schwager, E.E., Extavour, C.G., 2012. oskar Predates the Evolution of Germ Plasm in Insects. Curr Biol 22, 2278–2283.

Ikeshima-Kataoka, H., Skeath, J.B., Nabeshima, Y., Doe, C.Q., Matsuzaki, F., 1997. Miranda directs Prospero to a daughter cell during Drosophila asymmetric divisions. Nature 390, 625–629.

Jenny, A., Hachet, O., Závorszky, P., Cyrklaff, A., Weston, M.D.J., Johnston, D.S., Erdélyi, M., Ephrussi, A., 2006. A translation-independent role of oskar RNA in early Drosophila oogenesis. Development 133, 2827–2833.

Lehmann, R., Nüsslein-Volhard, C., 1986. Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila. Cell 47, 141–152.

Lin, S., Lai, S.-L., Yu, H.-H., Chihara, T., Luo, L., Lee, T., 2010. Lineage-specific effects of

Notch/Numb signaling in post-embryonic development of the Drosophila brain. Development 137, 43–51.

Pfeiffer, B.D., Ngo, T.-T.B., Hibbard, K.L., Murphy, C., Jenett, A., Truman, J.W., Rubin, G.M., 2010. Refinement of tools for targeted gene expression in Drosophila. Genetics 186, 735–755.

Seppa, M.J., Johnson, R.I., Bao, S., Cagan, R.L., 2008. Polychaetoid controls patterning by modulating adhesion in the Drosophila pupal retina. Dev Biol 318, 1–16.

Telley, I.A., Gáspár, I., Ephrussi, A., Surrey, T., 2012. Aster migration determines the length scale of nuclear separation in the Drosophila syncytial embryo. The Journal of Cell Biology 197, 887–895.

Vanzo, N.F., Ephrussi, A., 2002. Oskar anchoring restricts pole plasm formation to the posterior of the Drosophila oocyte. Development 129, 3705–3714.

Xu, X., Brechbiel, J.L., Gavis, E.R., 2013. Dynein-dependent transport of nanos RNA in Drosophila sensory neurons requires Rumpelstiltskin and the germ plasm organizer Oskar. J. Neurosci. 33, 14791–14800.

Zhu, S., Lin, S., Kao, C.-F., Awasaki, T., Chiang, A.-S., Lee, T., 2006. Gradients of the Drosophila Chinmo BTB-Zinc Finger Protein Govern Neuronal Temporal Identity. Cell 127, 409–422.

Zimyanin, V., Lowe, N., St Johnston, D., 2007. An oskar-dependent positive feedback loop maintains the polarity of the Drosophila oocyte. Curr Biol 17, 353–359.