

IMPERIAL COLLEGE LONDON
INSTITUTE OF REPRODUCTIVE AND DEVELOPMENTAL BIOLOGY

project completed at
THE DEPARTMENT OF LIFE SCIENCES,
NATURAL HISTORY MUSEUM, CROMWELL ROAD LONDON

Bioinformatic approaches to germline characterisation in the mouse
bile-duct tapeworm *Hymenolepis microstoma* (Platyhelminthes;
Cestoda)

By
CATHERINE ELLEN SEED, BSc

Submitted as partial fulfillment of the requirements for the MSc in
Reproductive and Developmental Biology, Imperial College London,
2012

Acknowledgements

Abbreviations

aby(*number*) - *Hymenolepis microstoma* transcript translation

EgG - *Echinococcus granulosus* gene model

EmW - *Echinococcus multilocularis* gene model

HmN - *Hymenolepis microstoma* gene model

LTR - long terminal repeat

Outgroup - a group of taxa assumed to lie outside the monophyly of the taxa used in the analysis. Used to provide direction for character state change

PBS - Phosphate buffered saline

miRNA -micro RNA

piRNA - repetitive element derived Piwi interacting RNAs

Piwi - P-element induced wimpy testes

siRNA - small interfering RNA

TsM - *Taenia solium* gene model

ds - double stranded

Abstract

Word Count: 9854

Vasa, Piwi and Nanos are proposed to have a conserved role in germ cell maintenance and specification throughout Metazoa. Recent evidence has supported this role within flatworms, where they have been shown to play an additional role within populations of multipotent stem cells and regeneration. However, evidence of their roles within parasitic species, known for their reduced regeneration capacity, is scarce.

This project utilised bioinformatic approaches to explore the representation of these proteins as well as commonly documented interacting proteins in the genome of the mouse bile duct tapeworm, *Hymenolepis microstoma*. We demonstrate the loss of Vasa-related helicases as well as changes in the expression profile of well documented targets of Vasa regulation. *Hymenolepis* also displays the loss of the Piwi clade of small RNA associated proteins, in addition to potential changes to the siRNA pathway; both of which are associated with protection of the germline from selfish DNA elements. Analyses confirmed conservation of only a single Nanos RNA binding domain which is expressed at an extremely low level, along with a less well conserved repression-associated motif. The Nanos interacting protein Pumilio was also identified, and found to display less conservation in its Nanos-interacting motif. We were able to isolate members of both the Argonaute and DEAD box helicase families and created DIG labelled probes for future expression studies. These data unveil dynamic changes in the germline specification machinery of *Hymenolepis microstoma* as well as parasitic flatworms in general.

IMPERIAL COLLEGE LONDON
DISSERTATION

CONTENTS

Abstract	
Introduction	
Methods and Materials	
Results	
Discussion	
References	
Appendixes	
List of Figures and Illustrations	
Figure 1 Phylogeny of the Metazoa	
Figure 2 The <i>Hymenolepis</i> life-cycle	
Figure 3 Diagram of larval development in <i>Hymenolepis microstoma</i>	
Figure 4 <i>Hymenolepis</i> larval images	
Figure 5 The <i>Hymenolepis</i> germline	
Figure 6 Domain architecture of Vasa	
Figure 7 The piRNA ping-pong cycle	
Figure 8 Diagrammatic representation of adult <i>Hymenolepis microstoma</i> used for RNA-seq collection	
Figure 9 Multiple alignment of the conserved helicase domain in Vasa	
Figure 10 Bayesian phylogenetic analysis of DEAD box helicases in <i>Hymenolepis</i>	
Figure 11 Alignment of Nanos zinc finger domains	
Figure 12 Bayesian phylogenetic tree analysis of zinc-finger motifs in representatives of the Nanos family, the closely related E2 ubiquitin ligase zinc finger domain and query sequences in <i>Hymenolepis</i>	

Figure 13 Alignment of repeats 4-8 in Pumilio	
Figure 14 Bayesian phylogenetic analysis of repeats 4-7 of Pumilio family members	
Figure 15 Bayesian phylogenetic analysis of Argonaute family members and query sequences in Hymenolepis	
Figure 16 Domain organisation in RNase III proteins	
Figure 17 Bayesian phylogenetic analysis of Tudor domains and Hymenolepis gene model results	
Figure 18 Bayesian phylogenetic analysis of the DAZ/Boule family members and query sequences in Hymenolepis	
Table 1 Categories of small RNAs found across taxa	
Table 2 Potential germline markers	
Table 3 Primers for amplifying DEAD box helicases and Argonaute family members	
Table 4 Identification of the eight conserved domains of Vasa in <i>Hymenolepis</i> transcript data.	
Table 5 RNA-Seq analysis from larval, whole adult and sectioned <i>Hymenolepis microstoma</i>	
Table 6 Non-redundant database results for Pumilio-like gene models	
Table 7 Domain organisation in RNase III proteins	

Introduction

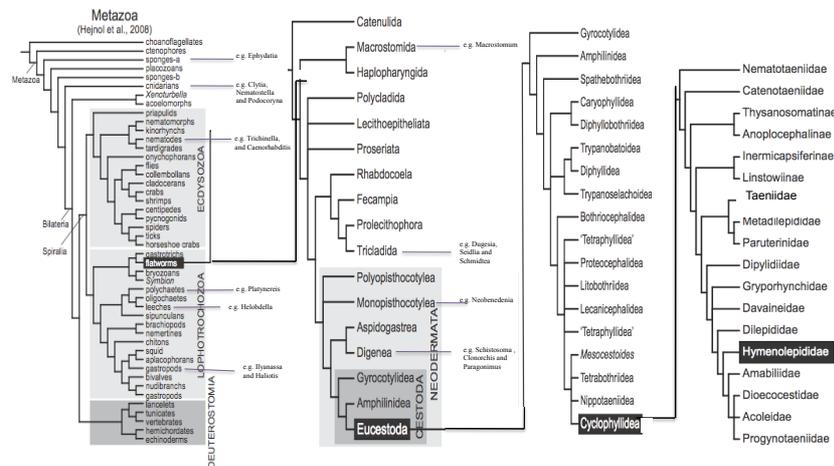
Parasites represent both a real and extensive economic and health burden in the world today, as well as a group of academic interest, typifying a rare evolutionary strategy with specific environmental and evolutionary implications. The World Health Organisation classifies parasites within the top six most harmful infective diseases worldwide (Northrop-Cleaves and Shaw, 2000), while intestinal helminthes are estimated to infect more than a quarter of the world's population (Miguel and Kremer, 2004). In previous decades, parasites were often overlooked in the laboratory, due to problems establishing laboratory populations, as well as the perception of parasitic species as being of little wider usefulness beyond that of a health burden. As a result of these factors, while there is a large collection of published data on parasitic life-cycles and infective strategies, many areas within Parasitology have remained relatively isolated from the emerging fields of molecular and developmental biology.

This dissertation aims to utilise data on pathways of importance in the establishment of the germline across taxa (i.e. *Vasa*, *Piwi* and *Nanos*) in order to a) molecularly characterise the germline of the mouse bile duct tapeworm *Hymenolepis* for future expression studies and the development of *in vitro* and transgenesis protocols. b) gain a deeper understanding of how this parasitic species has adapted/evolved on a genomic level. c) contextualise germ cell development of *Hymenolepis* within the parasitic order Neodermata, and the superphylum lophotrochozoa as a whole. This information will highlight the importance of developing *in vitro* protocols, and will reflect how the unique features of parasitic flatworms can expand the relevance of studying this fascinating species.

Introduction to parasitic flatworms

The lophotrochozoa comprise the third major branch in the grouping of bilaterians (i.e. bilaterally symmetrical organisms) along with deuterostomia and ectysozoa (Tessmar-Raible and Arendt, 2003). Numerically, the lophotrochozoa account for more animal phyla than any other clade of the metazoa, totaling around half of all animal phyla (Giribet, 2008). The lophotrochozoa contains a great amount of morphological diversity, including the well documented body plan diversity within the molluscs and annelids (Giribet, 2008).

Figure 1. Phylogeny of the Metazoa, with particular attention to the position of Hymenolepis. Key species used in Bayesian analysis are included as a guide to available taxon sampling (Olson et al., (2012);13; Hejnoj et al 2008; Littlewood, 2008: Olson et al., 2008; Hoberg et al 1999).



Within the Lophotrochozoa, is contained the phylum Platyhelminthes (Figure 1) encompassing all species of obligate parasitic flatworms in a group termed the Neodermata (Littlewood, 2006), assumed to represent a classic monophyletic group; obligate parasitism in the flatworms having arisen only once. The parasitic flatworms are

grouped into three major taxa; Monogenea, Trematoda, and Cestoda (Littlewood, 2006). The Cestodes are characterised by a lack of organs for a digestive system (Littlewood, 2006) and as a group exhibit a diverse range of morphological traits, hosts and host ranges, microhabitats, complexities of life-cycle as well as behaviour (Perkins et al., 2010). Within the Cestodes is grouped our genus of interest; *Hymenolepis*.

The genus *Hymenolepis* contains three species of scientific study; the rat tapeworm *H. diminuta*, the dwarf tapeworm *H. nana*, and the mouse bile duct tapeworm *H. microstoma*. All three have been maintained through their lifecycles in the laboratory, but *H. microstoma* is unique in that it locates specifically to the bile duct (Bogitsch, 1963, Bogitsch, 1966). *Hymenolepis microstoma* has proved very useful as a model tapeworm in that it is relatively easy to maintain within the laboratory (both hosts are themselves model organisms) and is refractory to human infection (Cunningham and Olson, 2010).

The life-cycle of *Hymenolepis microstoma* is complex. The species has two hosts; one arthropod, and the second mammalian (Figure 2.). Natural intermediate hosts include *Tribolium confusum* (confused flour beetle), *Tribolium castaneum* (red flour beetle) and *Orzaephilus surinamensis* (sawtoothed grain beetle) (www.olsonlab.com). Natural definitive hosts include voles, mice and gerbils, while laboratory life-cycles have been reported in rats, while infection led to pathology in the hamster *Mesocricetus auratus* (Cunningham and Olson, 2010; Bogitsch, 1963).

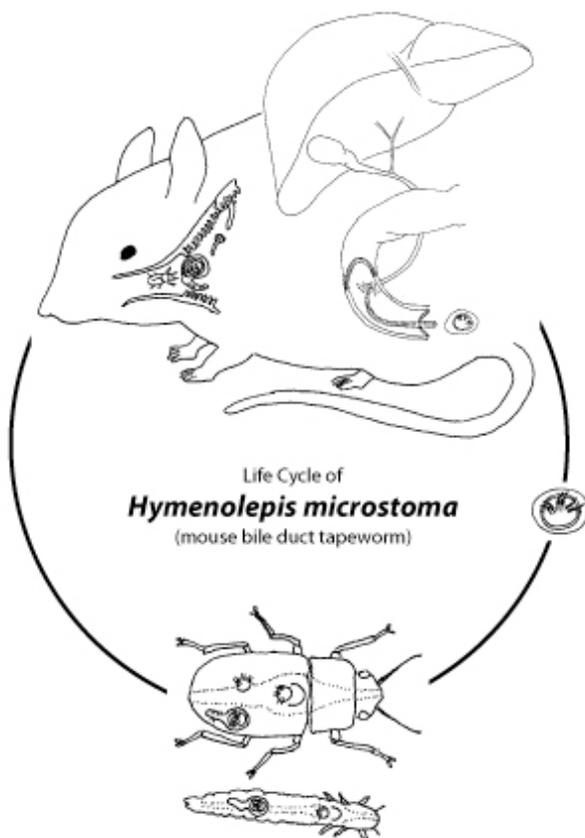


Figure 2. Hymenolepis life cycle The Nottingham stain of *Hymenolepis* is maintained in *Mus musculus* and *Tribolium castaneum* (From Olson and Cunningham (2011; 128)

The cycle is maintained thusly; the arthropod host ingests eggs excreted in the faeces of the definitive host. Once in the

beetle, oncospherical larvae are released from their egg shell and through physical (hooks) and chemical means (enzymes), penetrate the gut. Once in the haemocoel (Figure 2), they metamorphose (Figure 3) over 7-10 days. Upon reaching the infective cysticeroid stage, the cysticeroids are infective when the intermediate host is consumed by the definitive host (Cunningham and Olson, 2010). Hydrochloric acid and pepsin present in the stomach of the mammalian host leads to ruptures in the cysticeroid membranes, activating the scolex (the head) and causing excystation from the cyst (Goodchild and Stullken, 1970). Entering the bile duct of the host induces strobilation. Juvenile worms become established in the gut within 3 days, and within 7 days are sexually mature, releasing eggs (embryos) into the lumen of the intestine, to be released with the excrement of the host.

Figure 3. Diagram of larval development in *Hymenolepis microstoma*; from the egg stage, through development in the intermediate arthropod host e.g. *Tribolium confusum*, finishing with the infective cysticercoid stage ready for infection in the definitive host e.g. *Mus musculus*. 1) Egg (80-90 microns). Released into the faeces of the definitive host, for ingestion by the arthropod host. 2) Ingestion releases the egg shell (pictured) 3) Oncosphere post shell-shedding (30 microns) 4) Development in the embryo of the internal cavity and hooks (130 microns) 5) *Hymenolepis microstoma* forms a pear shaped structure (204 microns) 6) Body divisions are created through the processes of growth and elongation 7) Sucker (not pictured) and rostellum development occur (336 microns) 8) Anterior section withdraws 9) Anterior section withdraws into the tail bud (254 microns). 10) Tail extends, (Dvorak et al., 1961; Cunningham and Olson, 2010; Goodchild and Davis, 1972).

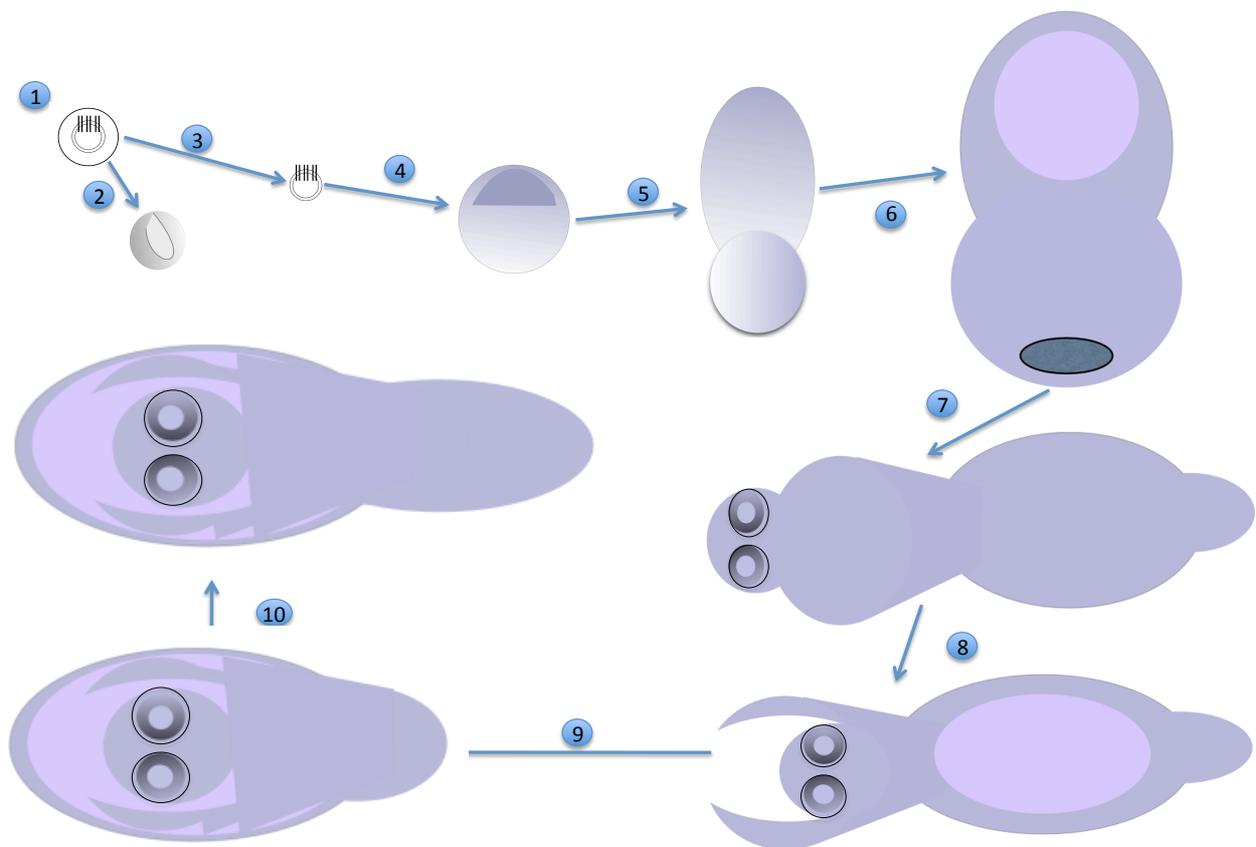
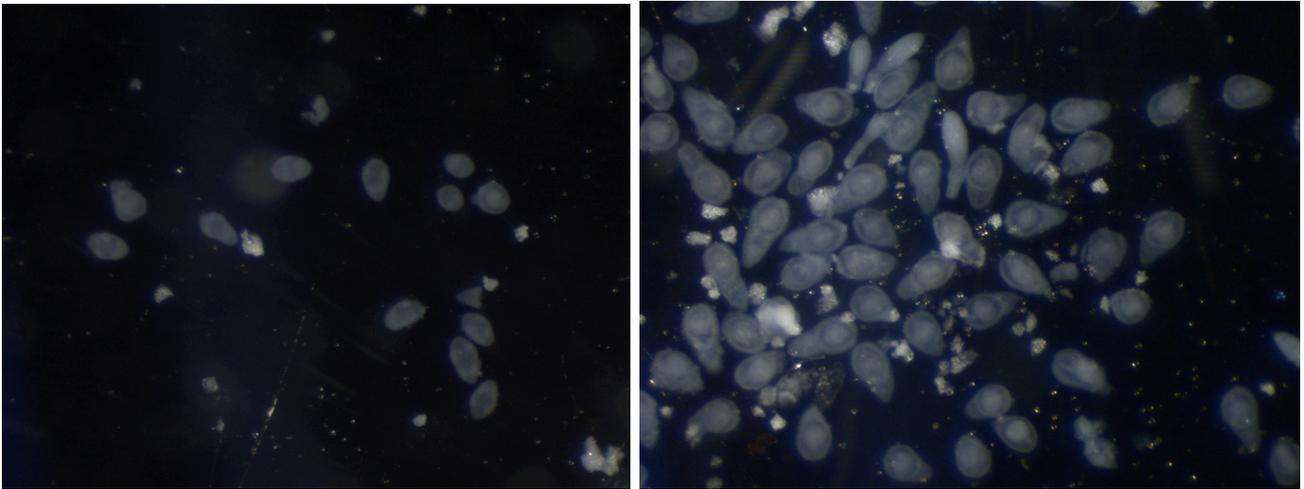


Figure 4. *Hymenolepis* larval images. Left: Photo of mid metamorphosis *Hymenolepis* larva.

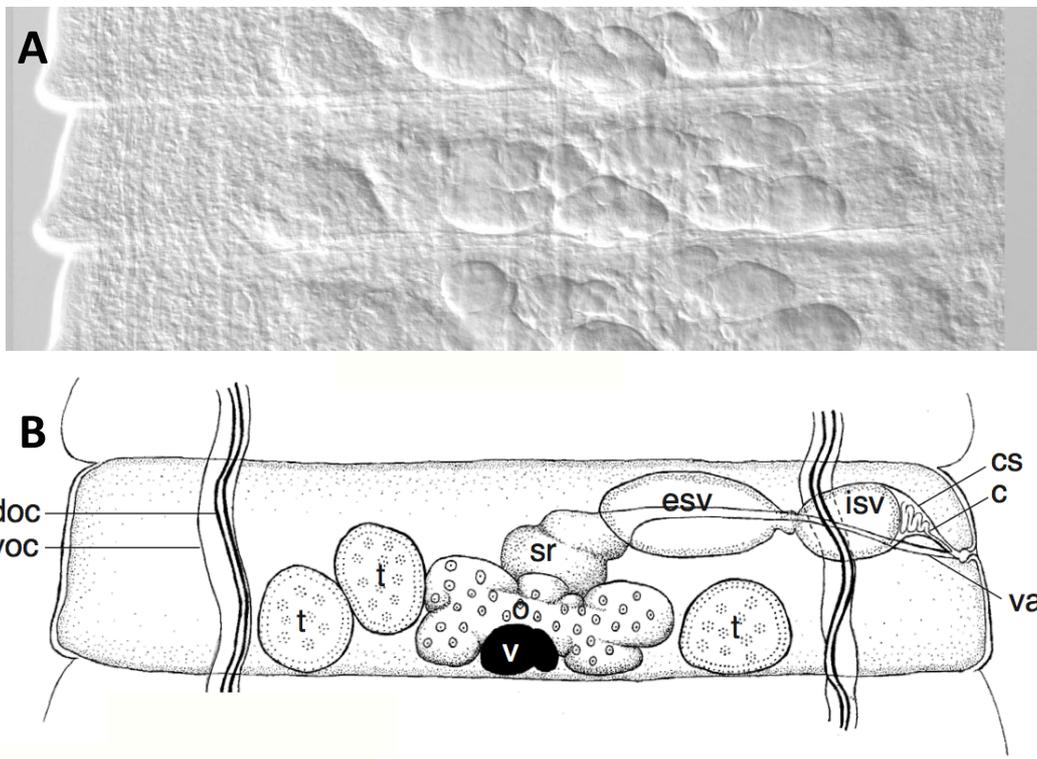
Right: Infective cysticercoids. Note the extension of the tail.



Germline development in *Hymenolepis*

Segmentation in *Hymenolepis* is induced anteriorly from the neck/scolex region. Gradual posteriorisation of segments is associated with germline development and reproductive output. *Hymenolepis* is a hermaphroditic tapeworm, with each of the segments containing male and female reproductive organs as confirmed in Tritiated thymidine studies in *H. dimimuta* (Nollen, 1975). However, development of the testes and ovaries are not concomitant, with male reproductive organs initiating their development prior to the female reproductive organs; a process which characterises the species as Protandrous. The posterior gravid proglottid segments are characterised by the senescence of organs, as the segments become loads with gametes for distribution in the feces of the host.

Diagram 5. The *Hymenolepis* germline. A) Light microscopy image of sexually mature adult segments in *Hymenolepis microstoma* B) Diagrammatic representation of an adult hymenolepis segment, Abbreviations: c, cirrus; cs, cirrus sac; doc, dorsal osmoregulatory canal; esv, external seminal vesicle; isv, internal seminal vesicle; o, ovary ;sr, seminal receptacle; t, testis; u, uterus; va, vagina; voc, ventral osmoregulatory canal. Scale bar 100 μ m. Adapted from Cunningham and Olson, (2010;125) Image A has been flipped horizontally to match the diagram.



Key players in Germline development

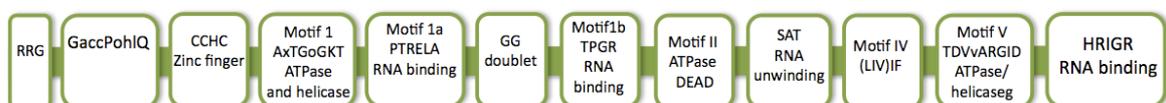
The extensive body of literature surrounding germ cell lineages across taxa highlight the usefulness of Vasa, Nanos and Piwi as defacto germline markers (Rebscher et al., 2012) (Weisblat, 2006). Below we explore the identification of these proteins and closely related family proteins, their known roles in the germline and stem cells, and evidence for their expression in sampled lophotrochozoan lineages.

Vasa

The DEAD-box helicase Vasa is the most widely used germ cell marker across Metazoa (Ewen-Campen et al., 2009; Pfister et al., 2008). A member of the largest of RNA helicase families (Skinner et al., 2012), the DEAD-box helicases are so named after the Asp(D)-Glu(E)-Ala(A)-Asp(D) motif present in their amino acid chain (Cordin et al., 2006). They function to catalyze the separation of double stranded RNA using ATP-derived energy (Cordin et al., 2006).

The DEAD box family are divided into the eIF4A, p68, PL10 and Vasa subfamilies, which have functions as wide ranging as RNA editing, pre-mRNA splicing, translation initiation and mRNA degradation, and in the Vasa subfamily, also cell cycle progression and piRNA biosynthesis (Mochizuki et al., 2001, Skinner et al., 2012). A number of conserved domains are often employed in the identification and inference of function of DEAD box family members (Diagram).

Figure 6 Domain architecture of Vasa showing the 8 conserved domains and other frequently noted motifs in DEAD box helicases; key to amino acids: a-aromatic amino acid c-charged amino acid o-alcohol amino acid h-hydrophobic amino acid l-aliphatic amino acid (Cordin et al., 2006)



The Vasa and PL10 subfamilies are considered to be closely related, and phylogenetic evidence suggests that Vasa genes emerged from an ancestral DEAD box helicase related to the extant PL10 family members (Skinner et al., 2012) after which they acquired the CCHC-zinc knuckle domains which are now considered characteristic (Gustafson and Wessel, 2010).

Vasa has been found to be associated with the ovaries and testes in species ranging from the Catfish *Clarias gariepinus* (Raghuveer and Senthikumar, 2010), the coral *Euphyllia ancora* (Shikina et al (2012) and the crab *Eriocheir sinensis* (Wang et al 2012). Evidence of a physical interaction between Vasa and a translation initiation factor supports a role for Vasa as a translational regulator (Ewen-Campen et al., 2009). Evidence of a possible association with pluripotency in the germline comes from evidence that Vasa expression correlates inversely with the temporal development of germ cells (Raghuveer and Senthikumar, 2010). Support for this pluripotency role as a being a wider role for Vasa in free-living flatworms, comes from evidence of Vasa expression in the stem cells, testes and ovaries of *Macrostomum lligano* (Pfister et al., 2008). Evidence for an additional developmentally early role for Vasa in Platyhelminthes comes from expression of a Vasa-like gene in the early cleavage embryo of *Schmidtea polychroa* (Solana and Romero, 2009).

Vasa is often associated with the nuage, a germline and pluripotency associated cytological structure, also known as the chromatoid body (Gustafson and Wessel, 2010), and is molecularly characterised to target other germline markers e.g. Oskar (homologs of which have not been found in non-dipteran species) (Solana et al 2009) and through Oskar the RNA binding protein Bruno and Germ-Cell-Less (Moore et al., 2009), and well as other components of chromatoid bodies/nuage, such as Gurken and Nanos (Tomancak et al 1998). Thus Vasa is considered an important regulator of the pluripotency associated functions of the chromatoid body.

PL10 members of the DEAD box helicase family have also been shown to show some roles in the germline. While PL10 proteins often have wider expression profiles than Vasa e.g. not exclusive to the germline and planarian stem cells, they have been shown

to play a role in the spermatogenesis of mice (Mochizuki et al., 2001) and gametogenesis in *Drosophila* (Johnstone et al., 2005). In spiralian lineages (Figure 1). PL10 has also been shown to be expressed in the germline of the annelid *Platynereis* and planarians (Shibata et al., 1999) Rebscher et al., 2007) suggesting that Vasa may not be the only member of the DEAD-box family to have germline roles in Platyhelminthes.

Nanos

The zinc finger containing protein Nanos (Kang et al., 2002) is known to be associated with the germline of all Metazoa, with samples including vertebrates, cnidaria, nematodes and leeches (Agee et al., 2006). Known roles of Nanos include translational regulation (Juliano et al., 2009), epigenetic regulation as shown in *C. elegans* and *Drosophila*, repression of apoptosis, and finally, repression of somatic lineage gene expression (Lai et al., 2011).

In contrast to an exclusive association with the germline in a number of different taxa, the role of Nanos in some platyhelminth lineages appears to be associated intimately with stem cell properties. In *Platynereis dumerilii*, Nanos expression correlates with both the multipotent cells contributing to both somatic and germline cells (Juliano et al., 2009), whereas in planarians, Nanos is induced in cells undergoing regeneration (Wang et al., 2007). However, in an asexual form of the planarian *Dugesia japonica* and *Dugesia ryukyuensis*, a Nanos homolog can be detected in presumptive gonadal tissue and gonads up to pre-meiotic spermatogonia during sexualisation; cells which do not participate in regeneration. This expression is interpreted to indicate that the roles that Nanos plays in the soma and germline stem cells are separated functionally (Sato et al., 2006; Nakagawa et al., 2012). Support for stem cell and germ line roles in lophotrochozoa come from the spatial expression in *Dugesia* and *Schmidtea mediterranea* (Hanberg-Thorsager and Saló, 2007).

In many species Nanos is known to complex with the protein Pumilio in regulating translation (Hanberg-Thorsager and Saló, 2007) where Pumilio aids in the targeting of

mRNAs, by binding the 3' UTR of mRNAs through its Puf domain (Ewen-Campen et al., 2009). This interaction was first described in *Drosophila*, but has also been described in *C. elegans* and humans, with Pumilio homologs being identified across taxa (Parisi and Lin, 2000; Ewen-Campen et al., 2009). In *Dugesia japonica*, a Pumilio homolog had also been identified in neoblasts (Salvetti et al., 2005), and Koziol et al., (2008) have demonstrated the presence of two Pumilio homologs in parasitic Platyhelminthes; one of which appears to be a gene duplication that occurred before the last common ancestor of Neodermatans.

Unique stem cells of the flatworms; Neoblasts

Neoblasts are the stem cells of the Platyhelminthes, which, based on their regenerative capacity, and structural characteristics, are considered unique to Platyhelminthes (Webb 2004). These cells have been shown to be the only mitotically active cells in planarians (i.e. transit amplifying properties are not seen in these species) (Sato et al., 2006), and are responsible for the remarkable regenerative capacity of planarians as shown in irradiation rescue experiments (Baguña et al., 2012). Neoblasts share characteristics of germline stem cells e.g. the presence of chromatoid bodies (Sato et al., 2006), and cytologically are characterised as small cells with a large nucleocytoplasmic ratio (Baguña, 2012).

Argonaute proteins

The past decade has seen an expansion in our understanding of the roles that small non-coding RNAs play in the cell and during development, as well as their capacity to regulate cellular events in a sequence specific manner (Batista and Marques, 2011). Most roles played by small RNAs are effected through their preparation from hairpin or dsRNA after which they are able to complex with a member of the Argonaute protein family, forming a ribonucleoprotein complex utilising the RNA-Sequence as a guide to regulate the path of complementary sequences (Batista and Marques, 2011; Kawaoka et al., 2011). Effects of this union can be epigenetic, such as DNA and histone methylation; or cellular, such as the regulation of translation or of specific endonuclease activities

(Houwing et al., 2007). However, the major role for these complexes is the targeted degradation of mRNA through the complementary of sense and antisense RNA, and the routine degradation of dsRNA in the cell (Seto et al., 2007; Cerutti et al., 2000, Palakodeti et al., 2007).

Members of the Argonaute family are characterized by the presence of two domains; a Piwi domain that creates a ribonuclease fold into which the small RNA can be held; and a PAZ domain that creates a hydrophobic pocket to bind the 3' end of the RNA (Batista and Marques, 2011) Farazi et al., 2008). Sequence similarity in the Argonaute family separates its members into three clades: AGO, WAGO (a *C. elegans* associated clade), and Piwi (Carmell et al., 2002; Garcia-Silva et al., 2010; Yogit et al., 2006). This sequence trend also follows a functional division as evidenced in cellular studies (Table 1). AGO members bind the double stranded siRNA (which targets foreign sequences) and miRNA (endogenous gene regulation), and these RNAs require processing into mature small RNAs by the double stranded ribonuclease III enzyme, Dicer, and may require initial processing by the related Drosha (Wei et al., 2012; Carthew and Sontheimer, 2009; Ishizuka et al 2003).

Table 1 : Categories of small RNAs found across taxa. Information from Seto et al., (2007) Chu and Rana, (2007); Schwalbe, (2011); Farazi et al., (2008); Schwab et al., (2009); Pane et al., (2007)

Small RNA	Mode of Action	Length and composition	Association with Dicer	Associated Argonaute protein
miRNA	Post-transcriptional silencing	Hairpin precursors PAZ domain 21-23nt	Yes	Argonaute
siRNA	Transcription and post-transcriptional silencing	PAZ domain dsRNA 21-23nt	Yes	Argonaute

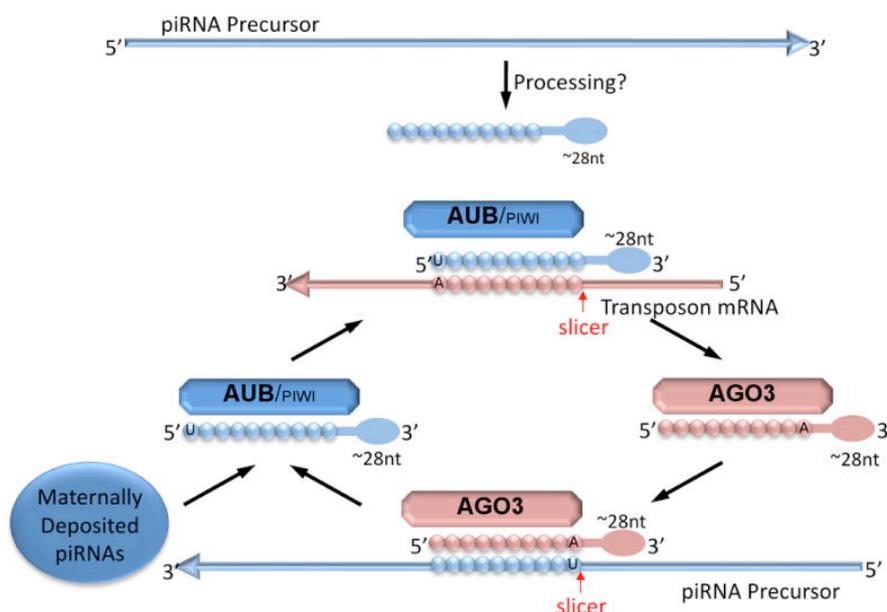
Small RNA	Mode of Action	Length and composition	Association with Dicer	Associated Argonaute protein
piRNA	Epigenetic post-translational	Single stranded.RNA 26-30nt 5' phosphate 2' O-methylation at 3' end	No	Piwi
tasiRNA	Function in trans- on genes different to the gene of origin	21-22nt dsRNA	Yes	Ago1
scnRNA	histone methylation and DNA elimination	28nt	?	?
rasiRNA	Regulation of chromatin structure	26-31nt ssRNA 2'-omethyl at 3'	No	Piwi

Piwi, in contrast only binds piRNAs and rasiRNAs whose biosynthesis is Dicer-independent (Table 1., Figure 7.) (Bak et al., 2011). One of the key characteristics of Piwi and piRNAs is that, throughout the animal kingdom they have been shown to be associated with the germline, almost exclusively. Piwi proteins have been shown to be key to spermatogenesis in mammals (Houwing et al., 2007; Deng and Lin, 2002) and oogenesis in *Drosophila* (Cox et al., 1998), whereas disruption of the pathway in *Caenorhabditis* leads to hermaphroditic sperm, and in *Xenopus* leads to germ cell apoptosis (Das et al., 2008; Houwing et al., 2007); a role which has been extended to stem cells in *Drosophila* and planarians (Cox et al., 1998; Das et al., 2008). In lophotrochozoans, SMED-WI-2, a Piwi like gene in *Schmidtea mediterranea*, when knocked down with RNAi, does not eliminate neoblasts or proliferation, but instead plays

a role in maintaining homeostasis in the neoblast population (Reddien et al., 2005). After irradiation in *Dugesia*, djPiwi has been similarly shown to be associated with dividing cells (Rossi et al., 2006).

Like other Argonaute family members, the role that Piwi plays in the cell can be multifaceted. Roles for Piwi include; epigenetic regulation, such as modulating Position-effect-variegation in *Drosophila* (Thomson and Lin, 2009). Piwi is also involved in the regulation or participation in double stranded- break repair during meiosis, as evidenced by defective repair in Piwi mutants (Thomson and Lin, 2009). However, their most documented role is as a line of defense against transposons in the germline.

Figure 7. The piRNA ping-pong cycle. From Thomson and Lin, 2009;376). piRNA biosynthesis occurs independently of the RNase III Dicer. A shortened version of transcribed piRNA precursors are produced by an undefined mechanism. Amplification of piRNAs occurs by cycling between AUB/PIWI and Ago3 (which groups phylogenetically with Piwi) Argonaute family members. AUB/PIWI bound piRNAs correspond to the antisense DNA strand to produce mature antisense piRNAs which bind AGO3. AGO3 associated piRNAs in contrast match the sense strand, guiding cleavage of antisense transcripts to create antisense piRNA, which can then bind AUB/PIWI act to bind sequences of active transposons (Bak et al 2011, Senti and Brennecke, 2010)). Thus they form a forward amplification loop. The cycle is biased, so that more AUB binds than AGO3 , and thus there are more antisense piRNAs in comparison to transposons (Zhang et al., 2011).



piRNAs have been shown to be either maternally derived (Brennecke et al., 2008) or derived bidirectionally from a number of specific genomic loci composed of clustered fragments of transposons which have lost their capacity to undergo transposition (Girard et al., 2006; Ariand and Kai, 2012). Pyrosequencing in mammals has shown that 17% of piRNAs map directly to transposable elements e.g. SINEs, LINEs, retrotransposons and LTRs (Ariand and Kai, 2012; Girard et al., 2006; Houwing et al., 2007), and that Piwi *C. elegans* mutants show 100x increase in transposon excision of Tc3 family transposons as indicated by the reversion of the onc022 transposon insertion allele (Das et al., 2008). A role for Piwi in transposon control in the germline is therefore strongly supported. A role in suppressing transposons may also be supported by Piwi's role in double stranded break repair, as unchecked double stranded breaks increase transposition rate of mobile genetic elements (Thomson and LIn, 2009).

Transposition of these 'genomic parasites' (Senti and Brennecke, 2010) is expected to be strongly selected against. Thus defense mechanisms such as Piwi have developed to protect the genome and germline in particular from gene disruption, transcriptional misregulation and loss of genome integrity (Ariang and Kai, 2012). Piwi is not however the only member of the Argonaute family that plays a role in the suppression of transposons, as siRNAs also protect the genome from foreign sequences. They often originate from transposons, are cleaved into 21-23nt sequences by Dicer, and bind Argonaute. (Senti and Breenecke, 2010; Miyoshi et al 2005). This is supported by laboratory use of transposons are partially down regulated through siRNA mediated RNAi (Rauschhuber and Ehrhardt, 2012)

Dicer/Drosha

RNase III proteins function to introduce staggered cuts in dsRNA (Lee et al., 2003). Due to maternal and genome specific origins of piRNAs, and a poorly characterised processing mechanism, piRNA/Piwi functions occur independently of any RNase III proteins. However, RNase III proteins are required for the production of both miRNAs (Lee et al., 2003) and siRNAs (Batista and Marques, 2011; Korhonen et al., 2011) before their association with Argonaute.

RNase III proteins group phylogenetically into two clades: Dicer and Drosha. Dicer is involved in the production of mature miRNAs, and Drosha may be involved in the production of their precursors (Lee et al., 2003). Evidence for the importance of these in cellular processes from the transcriptomic differences observed in Dicer and Drosha mutants (Wu et al., 2012).

Tudor domain containing proteins

Tudor domains are members of the 'Royal family' of protein interacting domains, along with MBT, PWWP, Chromo and Agenet subfamilies (Ying and Chen, 2012). Functionally, Tudor domain containing proteins are involved in RNA metabolism, germ cell development, DNA damage response and histone methylation (Chen et al., 2011). These domains bind symmetrical, dimethylated arginines such as those found on the N terminus of Piwi (Solana et al., 2009; Chen et al., 2011) and are believed to be important in the construction of macromolecular complexes (Pek et al., 2012)

A role for Tudor domain containing proteins in transposon silencing comes from mutant Tudor domain containing 9 (Tdrd9) mice, where antisense piRNA corresponding to LINE1 sequences decreases in the experimental condition (Bak et al., 2011). Indeed they have been shown to affect transposon activity independently of piRNA activity as well (Tanaka et al., 2011).

Tudor domain containing proteins have been shown to be expressed in the neoblasts, germline and central nervous system of *Schmidtea polychroa* (Solana et al.,

2009); are chromatoid body-associated proteins (Tanaka et al., 2011) and have been shown to be expressed in the neoblasts of *Schmidtea polychroa* (Solana et al., 2009).

Ying and Chen, (2012) have shown that proteins containing Tudor domains can be further divided based on their function into 4 groups. Thus it is the structural differences that characterise the role of different Tudor domain containing proteins in; chromatin regulation, snRNP synthesis, miRNA biosynthesis and piRNA biosynthesis (Ying and Chen, 2012). Only those designated 'group four' have been shown to play a role in the germline and piRNA biosynthesis. In particular, group 4 is characterised by multiple Tudor repeats, which may be related to their recruitment and structural roles in the formation of large protein complexes.

Germline specification programme

Given the widespread roles of Vasa, Nanos, PL10 and Piwi proteins in the germline of Metazoa, it is not surprising that evidence in a range of taxa shows interweaved relationships in their regulation. For example, in *Drosophila*, Vasa and Piwi have been shown to display a functional interaction (Alié et al., 2011); Vasa has been shown to regulate Nanos localisation (Skinner et al., 2012) and early piRNA synthesis (Yajima and Wessel, 2011) while Piwi has also been shown to regulate Nanos indirectly (Rossi et al., 2006). In addition, Nanos has been shown to also regulate Vasa abundance in mutants of the sea urchin (Juliano et al., 2009).

Often it appears it is the specific interactions between this conserved machinery that account for documented species to species variation in germline specification gene regulatory networks (Ewen-Campen et al., 2009). It is considered that the interaction between Vasa, Nanos and Piwi may therefore represent an evolutionarily 'stable' interaction (Ewen-Campen, 2009); network required for specification of a structure with strong selective constraints.

One structural component to the inter-regulation of components of germline specification may be the chromatoid body. Chromatoid bodies are electron dense cytological structures observed in diverse species, from *Drosophila* germ cells (termed nuage) to planarians neoblasts (Agata, 2003) which can be seen to decrease in size with differentiation (Shibata et al., 1999). The chromatoid body is often located close to nuclear pore complexes, suggesting a role in processing or screening of mRNAs (Yajima and Wessel, 2011). This is supported by evidence of RNase H sensitivity which suggests that RNA plays an important role in these structures (Sato et al., 2006).

The previously mentioned germline markers are often associated with these structures, coordinating the events required for germline specification. Piwi, Vasa, Bruno, Pumilio, Nanos and Tudor domain containing proteins are all associated with the chromatoid body (Houwing et al 2007; Agata (2008). Thus it has been hypothesised that Vasa and Piwi function to patrol dsRNA and protect the genome from transposon, while Nanos regulated translation of germline genes. This is facilitated by organisation in a complex, organised by recruitment characteristics of Tudor-domain containing proteins.

Other potential germline markers

Outside of the 'universal' three germ cell markers that appear in the literature, and their often associated proteins e.g. Tudor domain and Pumilio, there is also a large volume of data covering other key components of germline identity/specification. Transcriptome and microarray analyses have revealed a range of germline markers, many of which correlate with germ cell differentiation such as Boule, Elav, Granulin, Germinal histone and GLD-1 (Table2) and proteins which have been shown to be regulated by Vasa, such as Bruno, and Germ-Cell-Less or which regulate Nanos e.g. Smaug.

Boule is the ancestral member of Deleted in Azoospermia family, and binds the 3'UTR of mRNAs with short polyA tails (Kuales et al., 2011). Germline expression in flatworms has been suggested with ovary and testes expression of two out of three homologs in *Macrostomum lignano* (Kuales et al., 2011).

Smaug is a post-transcriptional regulator identified in yeast and humans, found to repress the translation of Nanos by binding Smaug recognition elements in the 3' of the transcript (Tadros et al (2007); Jeske et al 2011).

Bruno is an RNA recognition motif containing protein, associated with PL10 and Piwi, and is involved in planarian neoblasts and germ stem cell maintenance (Alié et al., 2011;Guo et al 2006). It is believed to act as a translational repressor (Moore et al 2009). Bruno contains three RRM's (Moore et al 2009), and regulates Oskar, Sexlethal, gurken and Germ-Cell-Less (Moore et al 2009). Germ-Cell-Less has been shown in *Drosophila*, to be negatively related by Bruno (Moore et al 2009), and is known to be expressed in the ovaries and testes of Zebrafish (Li et al., 2006) and the ovary in medaka (Sholz et al., 2004).

Table 2: Potential germline markers

Marker	Germline expression (Species)	Reference
Elav	Spermatid differentiation (Macrostomum)	Kuales et al., (2011)
Granulin	Sperm ducts and seminal vesicles (Schmidtea)	Chong et al., (2011)
Germinal histone	Spermatogonia (Planaria)	Collins et al., (2010)
GLD-1	Germline (Caenorhabditis)	Barbee and Evans (2006)

Materials and Methods

Bioinformatics

The amino acid sequences for Vasa, Piwi and Nanos from a range of deuterostome genera as well as the lophotrochozoans *Schistosoma japonicum*, *Schistosoma mansoni* and *Helobdella robusta* were used as query sequences to search the genomic and transcriptomic data from the Nottingham strain of *Hymenolepis microstoma* (Cunningham and Olson, 2010) in order to identify possible orthologs. Genome data were downloaded by FTP from Sanger Helminth Group www.sanger.ac.uk/resources/downloads/helminthes/hymenolepis-microstoma.html and visualised using Geneious v5.6 (Drummond et al., 2010) .

The *Hymenolepis* genome and gene model database were searched using query sequences for Pumilio (GAA27968, CAG25892, ABX58015, ABX58009 and CCD61586), Boule/Daz (AAH71023, AAK69026, AEK69207, XP_002920065 and XP_002575519), Dicer/Drosha (ADF47441, ADQ02405, CBL30926 AEF32762, AAF59169, AAZ80928, ADB65770 and CAP07635), and edited Tudor domain containing proteins (D5JG62, D5JG63, B3W673, C0L3L0, C0L3LI, C1M000).

Known Vasa Piwi, Nanos, Pumilio, Boule, DAZ and Dicer/Drosha sequences as well as *Hymenolepis* gene models were used as BLASTP queries to identify orthologs in other available parasitic flatworm gene model databases, including *Echinococcus*, *Schmidtea*, *Taenia* and *Macrostomum*.

Sequences similar to the genes of interest were obtained by BLAST searches through the genome databases. Known annotated genes, *Hymenolepis* gene models and other flatworm gene models were aligned with MUSCLE, whereas Nanos sequences were aligned with MAFFT with the L-ins-I algorithm and Blosum62 scoring matrix.

Vasa was aligned along the DEAD box helicase domain, while Nanos was aligned with the Nanos RNA binding domain. ‘Royal family’ domain analysis was first aligned in a subfamily based alignment (PWWP, Tudor, Chromo, MBT, Agenet). *Hymenolepis* gene models were aligned with Lophotrochozoan Tudor domain containing proteins and trimmed. Domains were trimmed, guided by prosite ClustalW alignments. These were then combined and aligned for Bayesian analysis.

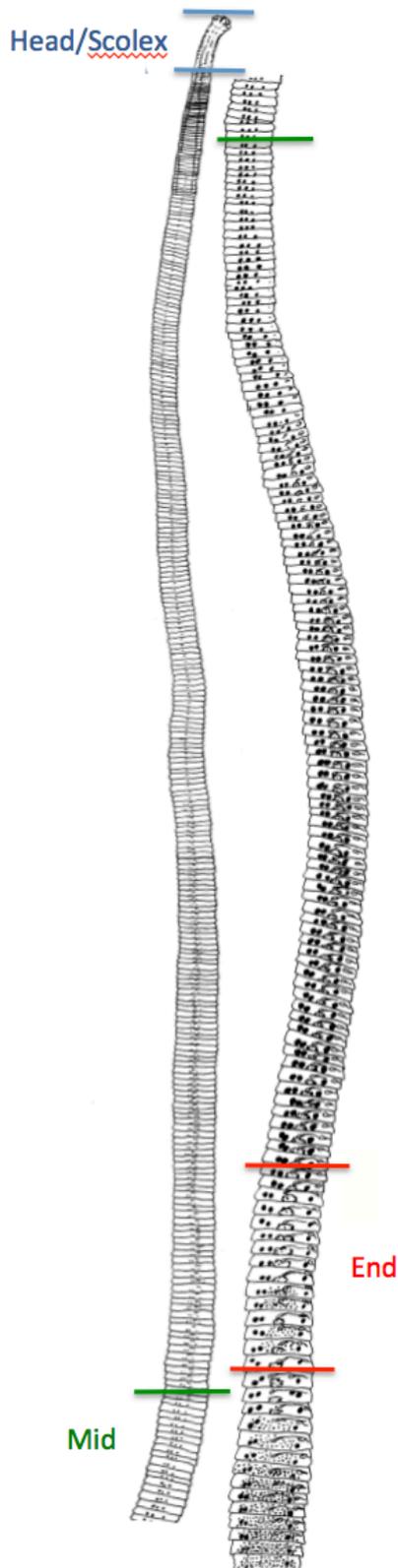
Alignment gaps were removed and the regions concatenated. Sequences with sequence divergence to the extent that homology could not be inferred were removed.

Bayesian inference (Ronquist and Huelsenbeck, 2003) analyses were performed using the MrBayes program, under the WAG model (Whelan and Goldman, 2001) of amino acid substitution including gamma among-site rate variation. Analyses were run for 1100000 generations with samples saved every 200 generations and the first 110000 generations eliminated as burn-in prior to forming a consensus tree (Riddiford and Olson, 2011). Expression was confirmed using unpublished RNA-Seq data to validate primer design. Artemis was used to view and quantify RNA-Seq data (Carver et al., 2008).

Domain analysis

Transcriptome sequences corresponding to the gene models identified in earlier analyses from the Bayesian analysis were screened against both the NCBI conserved domain database and Interproscan protein signature recognition software to search for functional domains. This was used for possible Vasa, Piwi, Nanos, Tudor, Boule/DAZ, Pumilio, Dicer and Drosha models/transcripts.

RNA-Seq



Comparisons were made of Illumina deep-sequencing RNA-seq data provided by Sanger, representing adult (four replicates) and larval tissue. Quantitative data had been processed into FKPM so that measurements mapped the number of reads mapped to the gene model per million reads sequenced. This corrected for sample variation in length of transcript and molar concentration.

Figure 8. Diagrammatic representation of adult *Hymenolepis microstoma* used for RNA-seq. Boxes indicate section removed for RNA-seq comparisons (see Table 5 for results), and to whole organism values. Image adapted from (Cunningham and Olson , 2010;125)

Genome size and transposon representation

Flatworm genome sizes were taken from (Olson et al., 2012) and composition analyses were provided by Tsai (pers. comm) where genomic data was analysed to identify repetitive elements using Repeat Modeler (<http://www.repearmasker.org/repeatmodeler.html>) and Transposon PSI (<http://transposonpsi.sourceforge.net>).

Specimens and sampling

A population of *Hymenolepis microstoma* larva were maintained in *Tribolium castaneum* (flour beetles) as a primary host at the Natural History Museum, London, and BALB/c mice as definitive hosts. Eggs were collected by maceration of adult worms freshly dissected from mice hosts, and seeded onto blotting paper in order to be fed to prestarved beetles. Mid-metamorphosis larva were obtained by dissecting *Tribolium* fed with *Hymenolepis* embryos five days after infection. Adult worms were obtained by dissecting the infected *Tribolium* and aliquoting the resulting *Hymenolepis* larva to be fed to mice by mouth-drip to mice. Mice were dissected at least 14 days post infection. Bile duct and intestine were dissected from the infected host, and fixed in situ in heated saline and moved to 4% PFA. Worms were then carefully removed using a dissecting microscope (Leica), and preserved in 100% ethanol.

RNA extraction

Total RNA was extracted from adult worms preserved in RNA later using RNeasy mini kit (Qiagen, Crawley, UK). Purity was assessed using the Nanodrop. The PCR SMART[™] RACE cDNA amplification kit (Clontech, Saint-Germain-en-Laye, France) was used to create total cDNA.

Primer design

Transcript sequences were used as templates for primer design. Primers were designed using Primer 3 (Rozen and Skaletsky, 2000) in order to amplify the largest transcript sequence possible.

PCR

PCR amplification was carried out under the following conditions, 95°C for 5min, 95 °C for 15 sec, 69°C for 1 min 30 sec, 72°C for 2 min for 35 cycles. This was finished with a 10 minute elongation step. Amplification of DBH2, Arg2 and PI102 was performed with a 64°C annealing temperature. PCR amplified cDNA fragments were gel purified from a 1% agarose gel using the QIA quick gel extraction kit (Qiagen), Products were extended with a 3' overhang protocol, purified with the Microcon (Millipore, Billerica, US) and cloned into a TOPO-TA cloning vector (Invitrogen) . Plasmids were extracted, checked with M13 primers using the PCR reaction of 94°C, and (94°C 55°C 68°C) for 34 cycles . They were then sequenced in order to determine insert size and directionality.

Table 3 Primers for amplifying DEAD box helicase and Argonuate family members

	Primer Sequence	Corresponding gene model/transcript
Forward	Hm_Arg-1_F1 AACGTCACTCGACGCTCGGC	HmN658200/ aby67247
Reverse	Hm_Arg-1_R1 AGGCCGGGGCAGGAATGGAT	
Forward	Hm_Arg-2_F TGCGTACGATAGCAGAAGTTGGG	HmN114300/ aby226620
Reverse	Hm_Arg-2_R CCGAACCAACAGCAGCCGCA	
Forward	Hm_HymDBH2_F TGTTTCTCTTGCGTCTCGTCGTG	HmN481800/ aby155633
Reverse	Hm_HymDBH2_R TGACGGGACAGGGGGCATCT	
Forward	Hm_HymDBH3_F TCGAAAAGGCGGGCAGTCGC	HmN15000/ aby3602
Reverse	Hm_HymDBH3_R TGTCCAGACCACGAGCGCCA	

	Primer Sequence	Corresponding gene model/transcript
Forward	Hm_PL10-1_F1 CGTGGAACGCCACCACGGTT	HmN1021800/ HmN61900/ aby156031
Forward	Hm_PL10-1_F2 GGTGCCCCACCTGAAAGCGG	
Reverse	Hm_PL10-1_R ACCACGAGCAGCAACAGCCG	

Whole mount In-situ hybridisation

Sense and antisense RNA probes were synthesized from the TOPO plasmids with T7 or Sp6 polymerase. DNA templates were removed by DNase treatment. DIG labeled RNA probes were synthesized from the plasmids **using a kit**. Insert size was determined by digestion with the appropriate restriction enzymes. Sense and antisense probes were synthesized.

In-situ hybridisation protocols were tested using hox gene probes on adult and mid-metamorphosis larval samples as described on www.olsonlab.com/data.data.html. Alterations were made to the permeabilisation step so that 14-22 mg/ml Proteinase K was used for a period of 10-15 min in adult samples and 2 min for larval samples.

Results

Vasa-like proteins are absent in *Hymenolepis*

Reciprocal blast searches of the *Hymenolepis* and flatworm genomes identified seven transcripts in the *Hymenolepis* transcriptome with similarity to Vasa. These transcripts are predicted to contain a range of the 8 domains characteristic of Vasa (Table 4, Figure 9.). Only one transcript, aby158900 contained all 8 motifs. No fasta files revealed CCHC repeats which are characteristic of Vasa.

Phylogenetic analysis indicate that both of these sequences group phylogenetically with PL10 family members (Figure 9.), while the other identified transcripts (tree is based on corresponding gene model) belong to either the closely related eIF4a family, or other dead box helicase families. This analysis revealed a flatworm specific group of PL10 proteins, as well as nested Digenea and Cestoda clades with the more basally joined Monogenea branch, a topology which matches commonly supported phylogenies.

Table 4 Identification of the Eight Conserved domains of Vasa in *Hymenolepis* transcript data.

	AQTGGSGKT /AXXXGKT Motif I	PTRELA Motif II	GG Motif III	TPGR Motif IV	DEAD Motif V	SAT Motif VI	ARGVD/ARGXD Motif VII	HRIGRTGR /HRIGR Motif VIII
Aby_155633	✓	✓	✓	✓	✓	✓		
Aby_156031			✓	✓	✓	✓	✓	
Aby_158900	✓	✓	✓	✓	✓	✓	✓	✓
Aby_257173	✓	✓	✓	✓	✓	✓		
Aby_3602	✓	✓	✓	✓	✓	✓	✓	
Aby_240046	✓	✓	✓	✓	✓	✓	✓	
Contig00660R			✓			✓	✓	✓

Figure 9. Multiple Alignment of the conserved helicase domain in Vasa; proteins from *Drosophila*, *Xenopus*, *Mus* and *Danio*, as well as the PL10 grouping DEAD box helicases in *Hymenolepis*. Well conserved motifs are highlighted. HmN=gene model aby=transcript. Annotations taken from Pfister et al (2008);148. Accession numbers available in Appendix

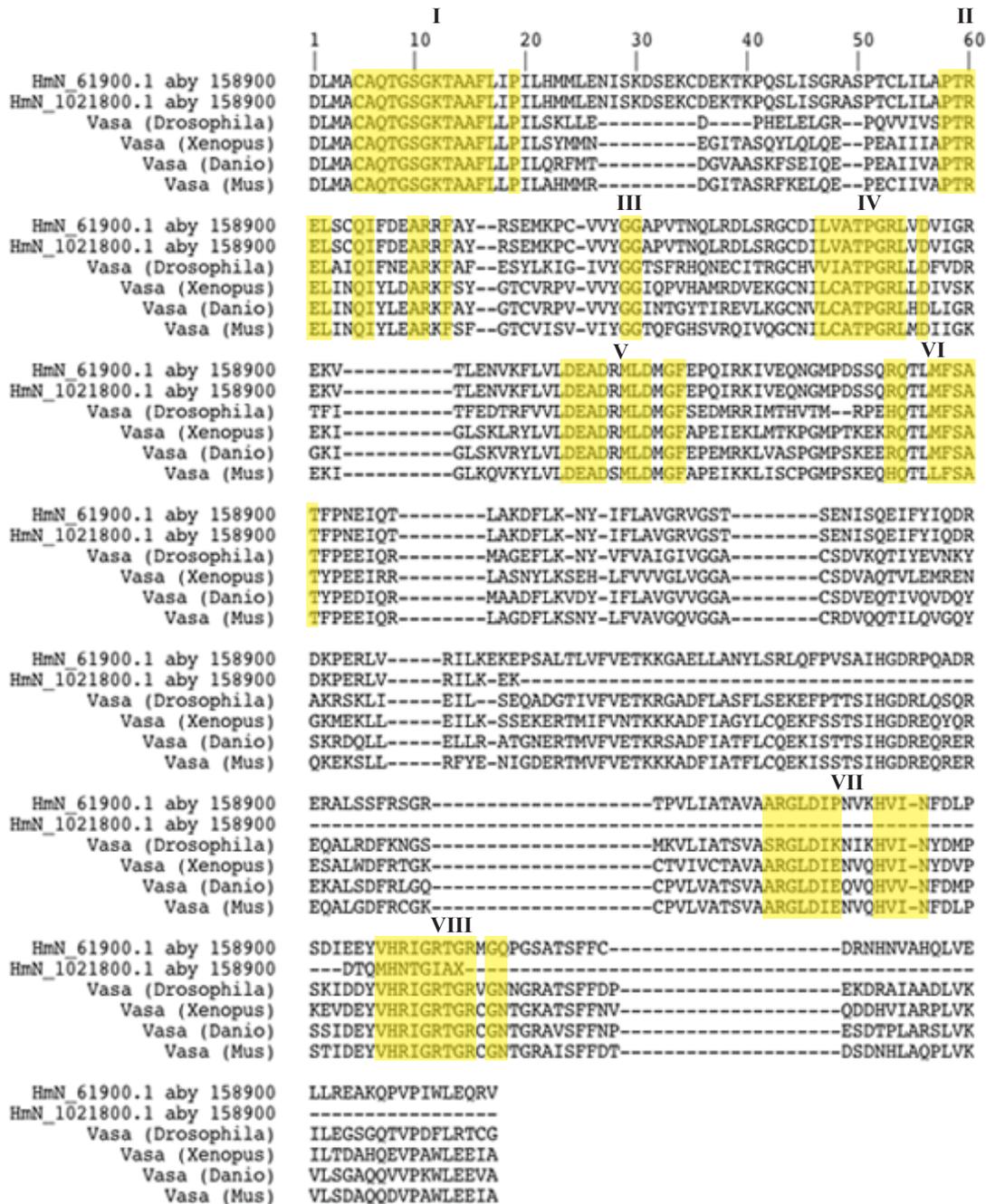
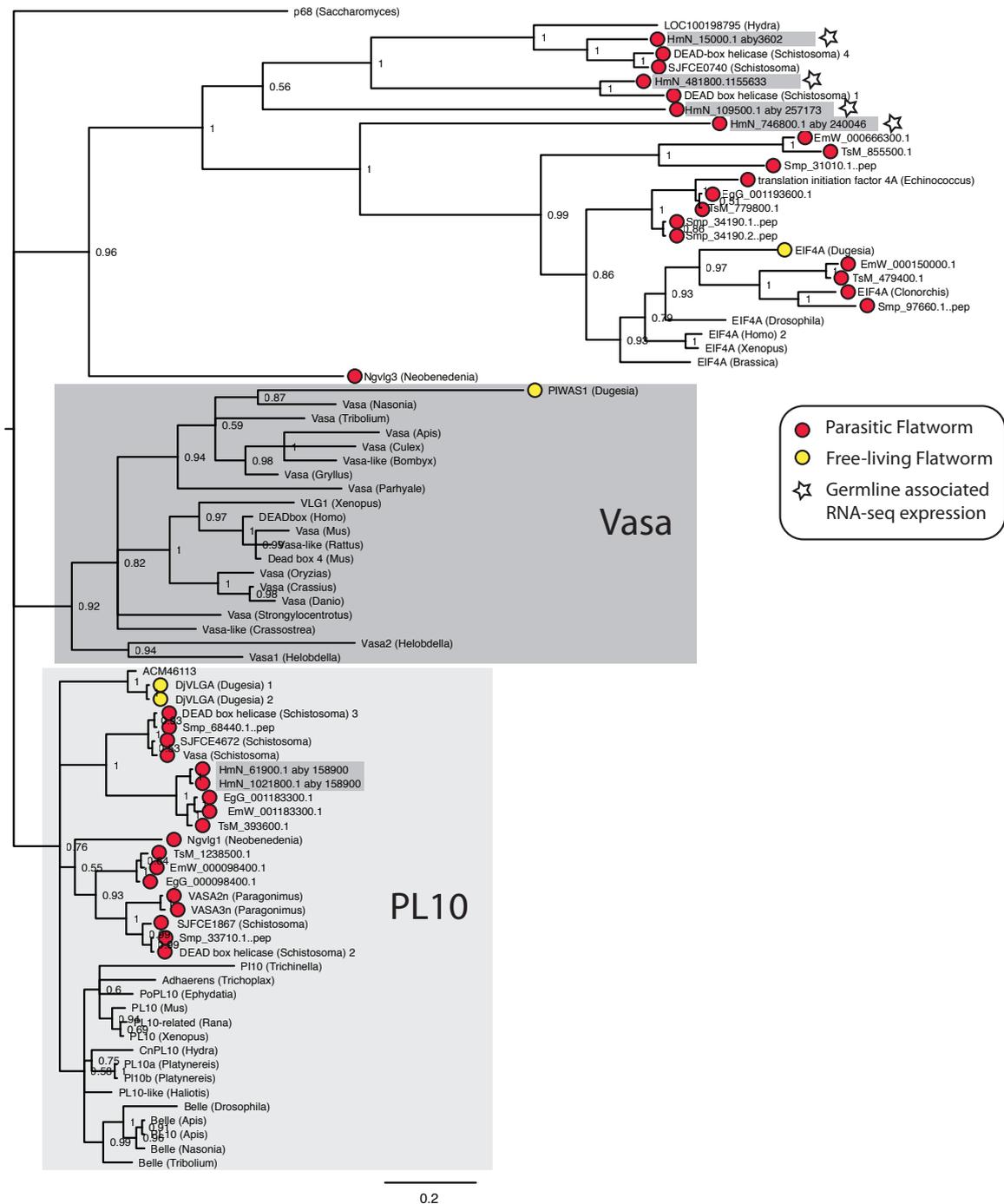


Figure. 10 Bayesian phylogenetic analysis of DEAD box helicases in and query sequences in *Hymenolepis*. p68 was used as an outgroup to root the tree. Tree is based on 179 amino acids. Accession numbers available in Appendix 1.



RNA-seq expression data confirmed that no members of the DEAD box helicases described were associated exclusively with the germline (Table 5), and that the flatworm specific clade grouping with EIF4a display both a phylogenetic and functional grouping; all are down-regulated in the end region compared to the head and scolex.

Table 5 RNA-Seq analysis of *Hymenolepis microstoma*. Data given in FKPM

ID	GENE/PROTEIN	GERMLINE?	LOCUS	LENGTH	larva_FPKM	whole_adult_scolex_mekc_mid_FPKM	end_FPKM	end_2PKM	whole_adult_vs_larva	mid_vs_scolex/mekc	end_vs_scolex/mekc	end_vs_mid
PIVA/ARGONAUTE												
HmM_000114200	eukaryotic translation initiation factor 2c	GERM	pathogen_HVM_scaffold_10:195110-201050	3655	0.2	83.8	1.1	169.1	13.4	UP	UP	DOWN
HmM_000685200	Eukaryotic translation initiation factor 2c 2	UBQ	pathogen_HVM_scaffold_122:31564-45298	3621	47.1	68.6	19.7	30.3	22.6	NO	NO	NO
VACA/PI3K												
HmM_000155000	DEAD box ATP dependent RNA helicase	UBQ	pathogen_HVM_scaffold_2:571-3109	1590	75.2	107.8	91.3	67.4	53.3	NO	DOWN	NO
HmM_000619300	ATP dependent RNA helicase DDX3X	UBQ	pathogen_HVM_scaffold_5:271685-28422	2736	94.0	46.0	22.5	30.9	22.2	NO	NO	NO
HmM_000079700	Hypothetical protein	UBQ	pathogen_HVM_scaffold_6:1282696-1330211	921	13.5	26.5	11.9	32.9	18.8	NO	NO	NO
HmM_000195900	ATP dependent RNA helicase DDX49	UBQ	pathogen_HVM_scaffold_9:911328-913920	1434	24.4	29.8	32.7	27.1	15.5	NO	DOWN	NO
HmM_000481800	ATP dependent RNA helicase DDX27	UBQ	pathogen_HVM_scaffold_65:436744-441492	1805	53.1	44.6	43.4	35.7	21.6	NO	DOWN	NO
HmM_000716800	ATP dependent RNA helicase mek31b	UBQ	pathogen_HVM_scaffold_157:184420-185836	1341	368.8	304.0	248.5	258.7	124.9	NO	DOWN	NO
HmM_001021800	DEAD box ATP dependent RNA helicase	UBQ	pathogen_HVM_scaffold_802:1135-37180	1356	41.0	23.6	16.2	15.4	13.3	NO	NO	NO
WAKES												
HmM_000753700	nanos protein	NONIE	pathogen_HVM_scaffold_160:258220-266421	630	1.7	1.6	1.7	0.6	0.0	NO	NO	NO
PUMILIO												
HmM_000758500	pumilio 2	NONIE	pathogen_HVM_scaffold_163:201100-207516	3129	174.5	166.2	52.0	78.7	93.2	NO	UP	NO
HmM_000516200	pumilio 2	GERM	pathogen_HVM_scaffold_78:234736-242242	3192	31.4	73.9	13.9	51.5	20.6	UP	DOWN	NO
HmM_1255500	pumilio domain containing		pathogen_HVM_scaffold_11:273275-276921	2337	60.1	80.1	66.8	72.0	31.5	NO	DOWN	NO
BOULE												
HmM_000762300	protein boule	germ	pathogen_HVM_scaffold_165:128225-131599	1590	1.9	97.4	0.6	118.0	9.4	UP	UP	DOWN
GERMINAL HISTONE												
HmM_000016000	Histone H4	UBQ	pathogen_HVM_scaffold_1:155763-156340	453	3.4	8.7	0.7	0.8	1.4	NO	NO	NO
HmM_000170200	Histone H4	NONIE	pathogen_HVM_scaffold_16:769294-769606	312	0.0	0.0	0.0	0.0	0.0	NO	NO	NO
HmM_000547400	Histone H4	UBQ	pathogen_HVM_scaffold_87:86105-86417	312	63.5	55.8	26.7	45.2	6.6	NO	NO	NO
HmM_000548200	Histone H4	UBQ	pathogen_HVM_scaffold_87:119007-119319	312	56.4	36.5	64.3	50.7	29.7	NO	NO	NO
GLD-1												
HmM_000886400	hypothetical protein	UBQ	pathogen_HVM_scaffold_249:57231-77030	1734	8.0	48.6	26.4	21.4	28.7	UP	NO	NO
HmM_000886900	hypothetical protein	UBQ	pathogen_HVM_scaffold_250:45637-61643	1344	6.8	37.3	26.3	19.4	26.4	UP	NO	NO
GERANULIN												
HmM_000317200	geranulins	UBQ	pathogen_HVM_scaffold_36:680469-684275	1287	122.7	309.3	174.2	218.3	217.1	NO	NO	NO
HmM_000317300	geranulin	UBQ	pathogen_HVM_scaffold_36:686690-690841	1056	162.8	309.9	190.8	173.5	187.6	NO	NO	NO
ELAV												
HmM_000517200	ELAV protein 2,3;4	GERM	pathogen_HVM_scaffold_78:418062-420279	2217	0.0	273.6	0.8	254.2	31.3	NO	UP	DOWN
HmM_000765000	ELAV protein 2		pathogen_HVM_scaffold_139:487710-210697	1734	22.5	31.1	29.9	22.4	15.7	NO	DOWN	NO
HmM_000788200	1 serine dehydratase1 threonine deaminase		pathogen_HVM_scaffold_179:40591-43255	2538	121.7	77.1	36.0	55.0	14.2	NO	DOWN	NO
HmM_000839300	ELAV		pathogen_HVM_scaffold_208:11526-46289	2052	34.2	43.9	16.4	21.7	47.1	NO	UP	NO
HmM_000599600	CUGBP Elav family		pathogen_HVM_scaffold_102:208214-219584	477	0.0	3.7	2.2	2.3	2.2	NO	NO	NO
BRUNO												
HmM_000599500	bruno 4	UBQ	pathogen_HVM_scaffold_102:196880-201224	1809	20.5	38.2	51.5	24.1	34.7	NO	NO	NO
TUDOR DOMAIN												
HmM_000122200	Kil domain containing protein	GERM	pathogen_HVM_scaffold_10:1218566-1221228	2214	77.5	137.2	71.0	217.8	41.5	NO	UP	DOWN
HmM_000184100	Survival of motor neuron signaling factor zinc finger CCHC type and RNA binding		pathogen_HVM_scaffold_17:995388-996194	630	58.3	128.2	164.3	114.0	94.3	NO	DOWN	NO
HmM_000272300	zinc finger CCHC type and RNA binding		pathogen_HVM_scaffold_31:7157-8915	678	8.1	21.2	11.5	15.3	8.9	NO	NO	NO
HmM_000687200	Stable domain containing protein		pathogen_HVM_scaffold_133:1021029-105549	3006	287.7	188.7	85.3	107.8	100.7	NO	NO	NO
DICER												
HmM_000353400	dicer 1		pathogen_HVM_scaffold_27:51078-53458	5694	28.1	23.1	10.4	14.4	5.3	NO	DOWN	DOWN
HmM_000260100	endonuclease Dicer	UBQ	pathogen_HVM_scaffold_20:89593-97513	3552	36.1	109.5	50.3	86.8	46.6	NO	NO	NO
HmM_000260200	endonuclease Droscha	GERM	pathogen_HVM_scaffold_20:58436-101652	1827.0	19.4	47.6	11.9	15.7	21.1	UP	NO	DOWN
HmM_000061600	ribonuclease Droscha	UBQ	pathogen_HVM_scaffold_5:225299-263232	3606.0	8.8	23.6	11.5	23.5	9.0	NO	NO	DOWN
SMAUS												
HmM_000064600	protein Smaug 1	UBQ	pathogen_HVM_scaffold_5:676370-684663	3330	37.7	34.2	12.2	20.3	16.5	NO	NO	NO
GERM CELL LESS												
HmM_000775500	germ cell less protein 1	UBQ	pathogen_HVM_scaffold_172:76569-81765	1587	17.1	23.4	18.7	16.9	14.8	NO	NO	NO

Bruno

Query searches identified one protein with similarity to Bruno. Conserved domain analysis confirmed the presence of two RNA recognition motifs. RNA-seq data showed little spatial variation in expression levels; it is ubiquitously expressed (Table 5).

Germ-Cell-Less

Query sequences identified one sequence with similarity to Germ cell Less, which RNA-seq data showed to be ubiquitously expressed throughout the worm (Table 5).

Conservation of the Nanos zinc finger domain

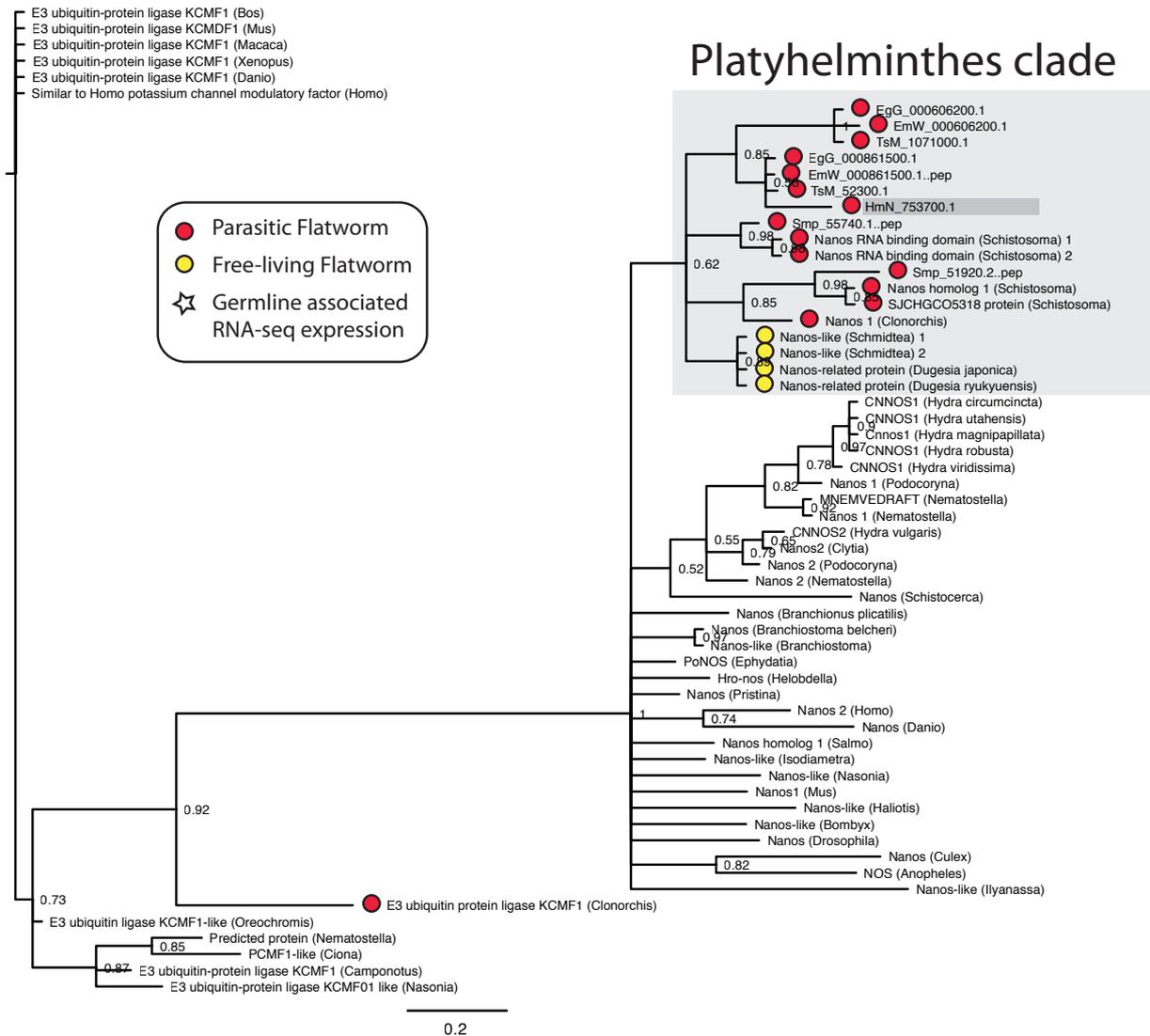
Due to high divergence around the Nanos domain across species (not shown) alignments of along the Nanos domain of Nanos proteins across Sponges, cnidarians and Lophotrochozoans were used to create query sequences to search the Hymenolepis data. One gene model was identified. This was realigned with the query sequences (Figure 11.). The alignment shows little variation in the Nanos domain in comparison to sampled vertebrate taxa. The aligned sequences were used to construct a phylogenetic tree by Bayesian analysis (Figure 12). The consensus tree, rooted by the closely related E3 ubiquitin ligase domain shows Nanos sitting within the Nanos clades, in a flatworm specific group. There is 100% bayesian posterior probability support for the Nanos node.

Figure 11. Alignment of the Nanos zinc finger domain



In other species Nanos usually has multiple zinc finger domains (Juliano et al., 2009) Sunanaga et al., 2008, Fujii et al., 2006, Mochizuki et al., 2000), whereas HmN753700 only has one zinc finger domain. Deletion analysis shows that a 14a.a. region in the N terminus is responsible for repressive activity (Lai et al., 2011). Alignments of this region (Appendix 3) highlight extensive variation in this region in *Schistosoma*, *Dugesia*, *Schmidtea* and *Hymenolepis* when compared to those of *Danio* and *Mus*. *Danio* and *Mus* showed 47.1% identity, while comparisons among mammals, *Schistosoma*, *Schmidtea*, *Dugesia* and *Hymenolepis* sequences varied between 5.9% and 23.5%; highlighting the extensive variation in these sequences.

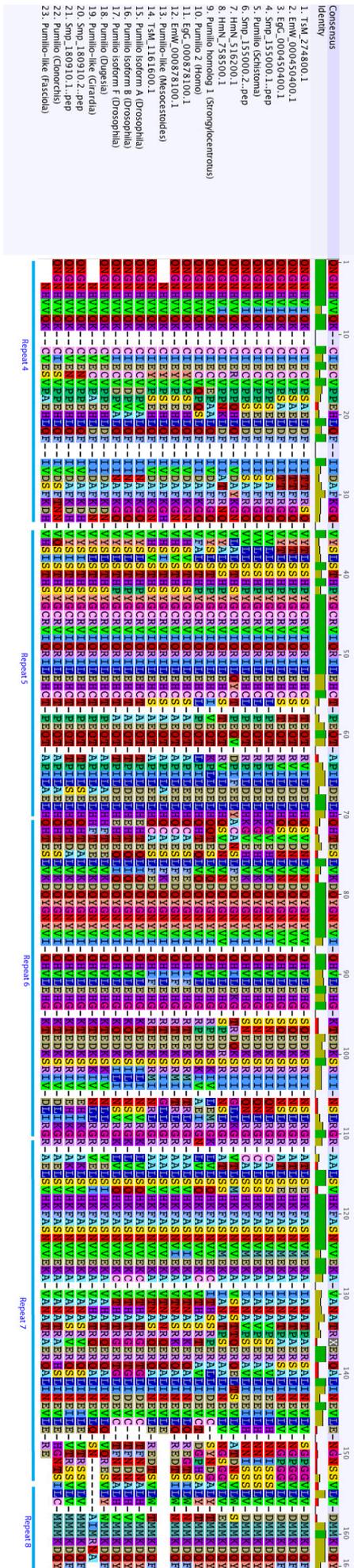
Figure 12. Bayesian phylogenetic tree analysis of zinc-finger motifs in representatives of the Nanos family, the closely related E2 ubiquitin ligase zinc finger domain and query sequences in *Hymenolepis*. Tree is based on alignment of 44 amino acids. Accession numbers available in Appendix.



RNA-Seq analysis showed extremely low levels of Nanos expression in *Hymenolepis*, and therefore no variation in expression through the life-cycle or organism (Table 5).

Pumilio

Figure 13. Alignment of repeats 4-8 in Pumilio.
(Annotations from Koziol et al., 2008)



Three gene models were identified in a bioinformatic screen for Pumilio in the the *Hymenolepis* genome. Blast searches against the non-redundant data based identified all of these as having similarity to Pumilio or Pumilio domain containing proteins (table 6).

Alignments show similar variation in the Nanos association domain (repeat 8) compared to other lophotrochozoan species (Figure 13 Appendix 2). Koziol et al (2008) noted that repeats 7-8 (which in *Drosophila* are associated with Nanos recruitment) show the greatest sequence divergence in lophotrochozoa genera such as *Echinococcus*, *Mesocostoides* and *Fasciola*. % identity comparisons with *Hymenolepis* Pumlilos show great domain conservation compared to *Echinococcus*, *Mesocostoides* and *Fasciola* (Appendix 4).

Figure 14. Bayesian phylogenetic analysis of Repeats 4-7 of Pumilio family members. Tree is rooted by... Tree is based on 112 amino acids. Accession numbers available in appendix.

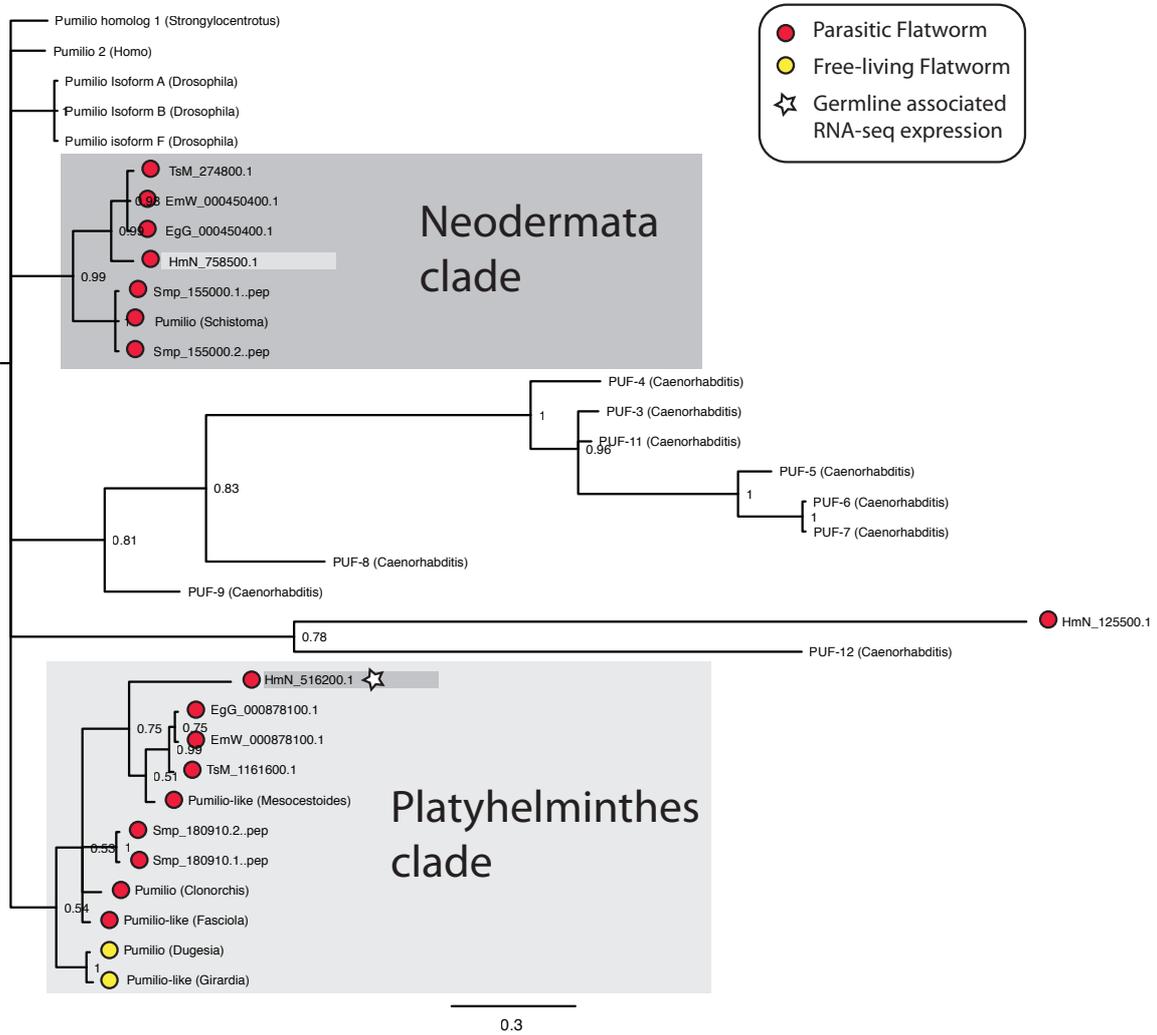


Table 6: Non redundant database results for Pumilio-like gene models.

Species	Gene Model	First hit on NCBI BlastX search	Accession number	E value	%identical sites
Hymenolepis	HmN516200.1	Pumilio homolog (Dugesia)	CAG25899	1.30E-128	50.2
Hymenolepis	HmN758500.1	Pumilio homolog (Schistosoma)	XP_002577387	3.06E-178	72.4
Hymenolepis	Hmn125500	Pumilio homolog domain containing protein family member 6 (Clonorchis)	GAA28293	6.25E-180	43.9

Alignments were selected based from analysis by Koziol (2008), using repeats 4-7 of the PUM homology domain involved in RNA binding. The aligned sequences were used to construct a phylogenetic tree by Bayesian analysis (Figure 14).

The tree shows a platyhelminth branch and parasitic branch, supporting Koziol (2008). RNA-seq data showed a spatial separation in the expression of the gene models, with member of the parasitic clade becoming expressed later in the development of reproductive organs compared to the flatworm clade member (Table 5). The gene model identified (Table 6) as being similar to Pumilio domain containing proteins, was highly expressed and is down regulated in the head compared to the other sections, suggesting a possible association with the germline (Table 5).

SMAUG

Query sequences identified one protein model with similarity to SMAUG. RNA-seq analysis showed this to be ubiquitously expressed (Table 5).

Argonaute family members in the *Hymenolepis* genome

Nine transcripts and their corresponding gene models were identified in a bioinformatic screen for Argonaute family proteins in the the *Hymenolepis* genome. Domain analysis showed that only aby226620 (HmN 114300), aby 67247 (HmN658200) contained both PAZ and Piwi domains which characterise the Argonaute family.

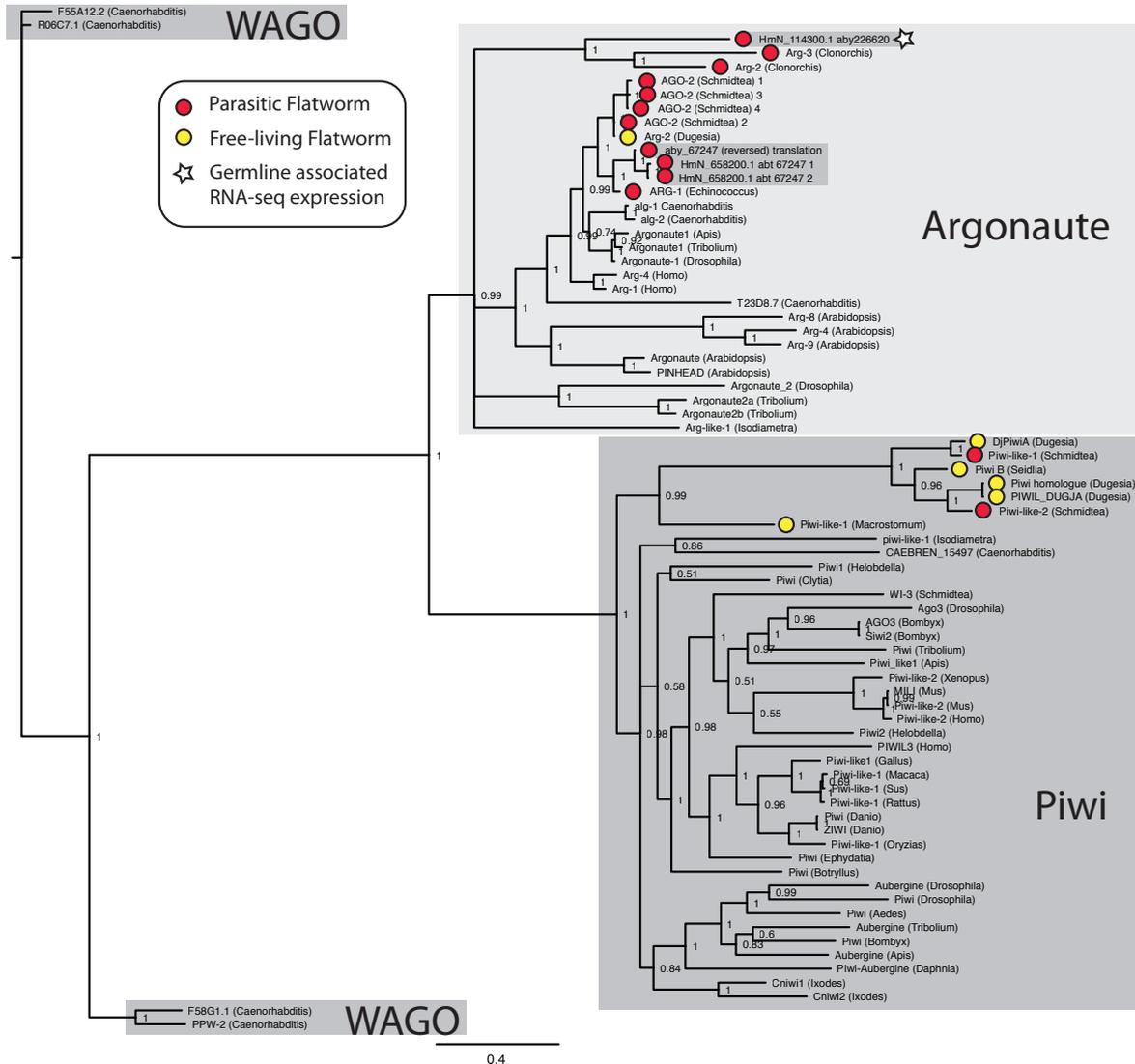
Piwi family members are not represented in the *Hymenolepis* genome, while Argonaute members are present and expressed.

The aligned sequences were used to construct a phylogenetic tree by Bayesian analysis (Figure 15). The consensus tree, rooted by the *C. elegans* gene F55A12.2 (a documented WAGO family member, Garcia-Silva et al., 2010) showed 3 major groups corresponding to Argonaute, Piwi, and a divided WAGO clade. There is 100% bayesian posterior probability support for the Piwi/Argonaute node.

The DDH motif of Piwi is used for slicer activity in piRNA synthesis (slicer is an RNA guides RNase activity) (Reuter et al., 2011). Sequence searches of the fasta files corresponding to the gene protein models failed to identify a DDH motif, suggesting lack of slicer activity in these models/transcripts.

RNA-seq analysis showed no signs of germline specific expression in the model HmN 658200 which grouped phylogenetically with closely with other Argonaute members (Table 5). However, differential expression in larval and adult, as well as higher expression in mid and end sections compared to the head, suggests that the gene model HmN114300 may be associated with the germline. In addition this model groups phylogenetically with another parasitic platyhelminth, the trematode *Clonorchis* (Figure 15).

Figure 15. Bayesian phylogenetic tree analysis of Argonaute family members and query sequences in *Hymenolepis*. From a MUSCLE alignment. WAGO proteins used as an outgroup. Tree is based on alignment of 441 amino acids. Accession numbers available in Appendix.



Genome size and transposon percentage

Comparisons of genome sizes among lophotrochozoan lineages highlights the small genome size in parasitic flatworm species (Olson et al., 2012; Tsai et al in prep). Thus, whereas the *Schistosoma* genome size is around 363Mb, the *Hymenolepis* genome is only 147Mb (Olson et al 2012).

Tsai (pers comm.) performed analysis to confirm that only 2% of the *Hymenolepis* genome exhibited characteristics similar to transposable elements, a characteristic that appears to be the largest contributor to the change in genome size in parasitic species compared to free-living species. The analysis also confirmed that Gypsy and Merlin are the most common transposon families of the LTR and retrotransposon classes represented in the *Hymenolepis* genome. .

Retainment of Dicer suggest that the miRNA and siRNA pathways may still be functional

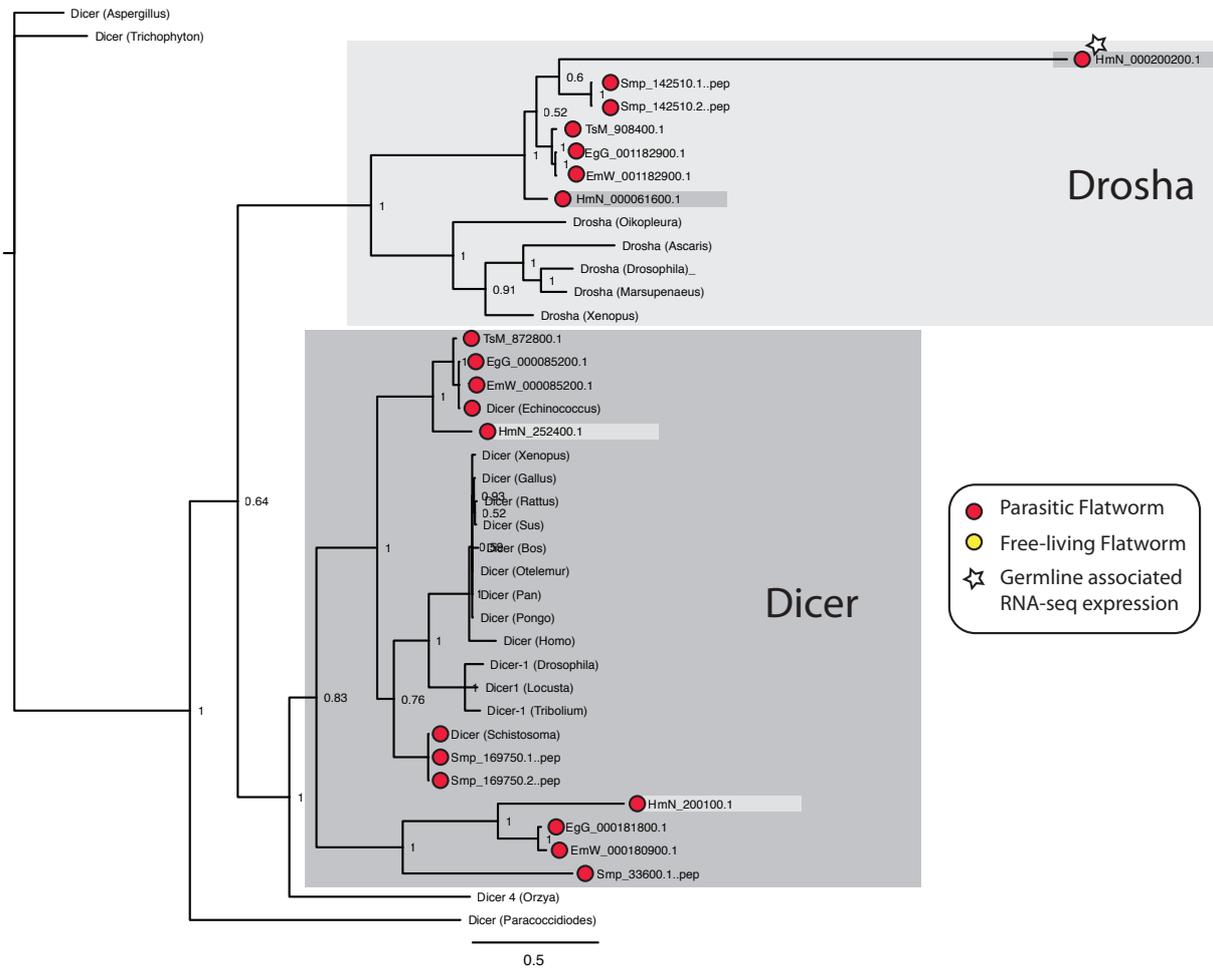
Four gene models were identified in a bioinformatic screen for Dicer/Drosha family proteins in the the *Hymenolepis* genome and other flatworm genomes.

Table 7 Domain organisation in RNase III proteins. Taken and adapted from Pfam results in Batista and Marques, 2011;8). and NCBI conserved domain database

Protein/protein model (Organism)	Helicase C domain	dsRNA-binding domain	PAZ domain	RNase III domain	dsRBM domain
Dicer (Homo)	1	1	1	2	0
Dicer 1 (Drosophila)	1	1	1	2	0
Dicer 2 (Drosophila)	1	1	1	2	1
Dicer (Schistosoma)	1	1	1	2	0
Dicer (Giardia)	0	0	1	2	0
Dicer-1 (Naumovia)	0	0	0	1	2
Smp 033600 (Schistosoma)	0	0	0	2	0
Drosha (Homo)	0	0	0	2	1
Drosha (Schistosoma)	0	0	0	2	1
Drosha (Drosophila)	0	0	0	2	1
HmN 252400 (Hymenolepis)	0	1	1	2	0
HmN 200100 (Hymenolepis)	0	0	0	2	0
HmN 61600.1 (Hymenolepis)	0	0	0	2	1
HmN 200200 (Hymenolepis)	1	1	0	0	0

Domain analysis supports HmN252400 and HmN200200 similarity in domain architecture to documented DICERS while HmN200100 and HmN61600.1 are more similar to Drosha (Table 7). However, phylogenetic analysis support HmN200200 and HmN61600 as Drosha homologs, with HmN200100 and HmN 252400 as Dicer representatives (Figure 16).

Figure 16. Bayesian phylogenetic analysis of Dicer and Drosha. Tree is based on alignment of 372 amino acids. Accession numbers available in Appendix 1.

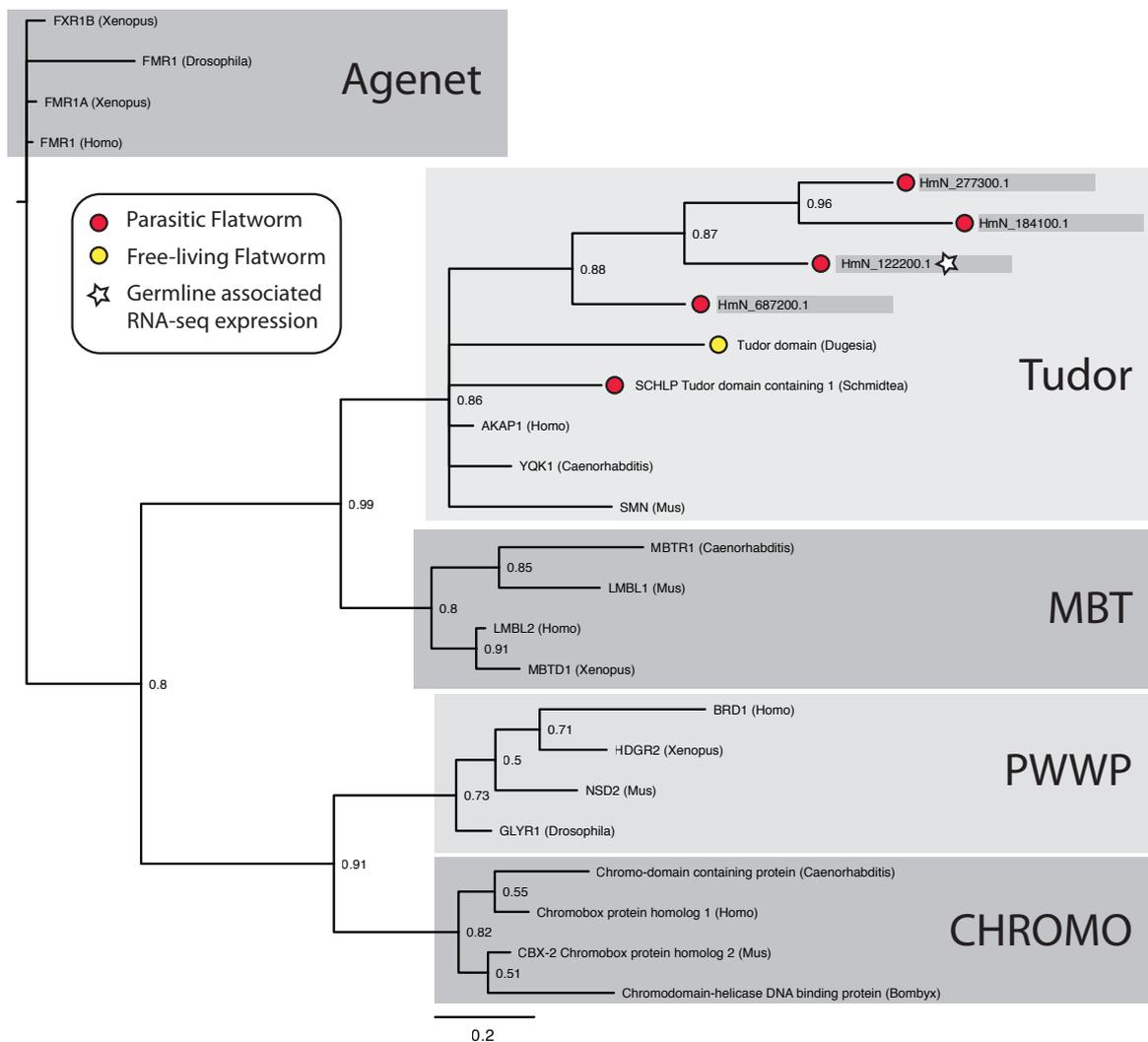


RNA-seq data showed all Dicer-like genes to be expressed. HmN200200 was shown to be expressed at low levels, and to have differential expression which correlates with germline expression (Table 5).

Tudor domain family is retained in Hymenolepis

Reciprocal blast searches of the Hymenolepis genome using the extracted Tudor domain from an alignment of lophotrochozoan Tudor domain containing proteins, identified four sequences with homology to the Tudor domain. Phylogenetic analysis places these sequences within the Tudor domain subfamily of the Royal family domains (Figure 17).

Figure 17; Bayesian analysis of Tudor domains and gene model results. Different family members are indicated in different colours. Tree is based on alignment of 69 amino acids. Accession numbers available in Appendix 1.



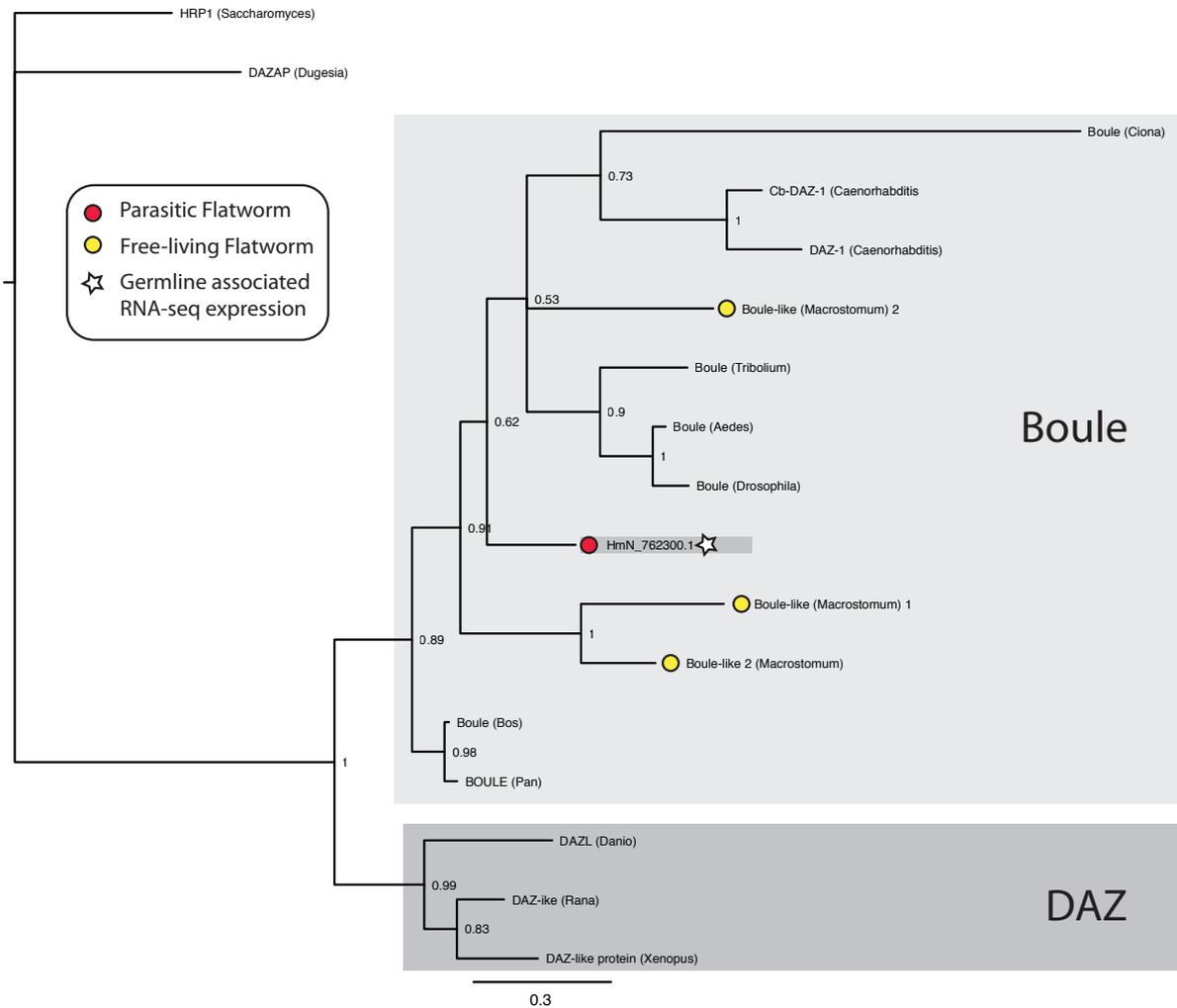
Correspondence with Jason Tsai (Sanger Centre) showed that these Tudor domain containing proteins of *Hymenolepis* and other parasitic flatworms do not belong to group 4 Tudor domain proteins of *Schmidtea polychroas* designated by (Ying and Chen, 2012). (Tsai et al., in prep). However, RNA-seq data highlighted the pattern of expression of the Tudor domain containing protein HmN122200, in a pattern suggestive of germline association (Table 5).

DAZL/BOULE

Macrostomum (AEK69207; ADL09421) and HmN762300.1 were used as query sequences for the flatworm genomes. The resulting sequences were blasted, and splicing factors and DAZ-associated proteins were removed before phylogenetic analysis. Bayesian analysis was performed (Figure 18)

RNA-seq data confirmed that Boule is not expressed in the larva of *Hymenolepis*, while it is highly expressed in the mid section where germ cells are forming and functional, no expression was found in the germ cell-free scolex, and low levels were shown in the senescing end section (Table 5).

Figure 18. Bayesian phylogenetic analysis of the DAZ/Boule family and query sequences in *Hymenolepis*. HRP1 was used as an outgroup. Tree is based on alignment of 96 amino acids.



Exploration of the genome in search of other well known germline markers supported exploration of the ELAV-like protein HmN517200 as a germline marker (Table 5), as well as the possible marker GLD-1 whose expression is higher in adults compared to larva (Table 5). In contrast Granulin was shown to be ubiquitously expressed, while no histone 4 variants showed possible exclusive germline expression (Table 5).

Discussion

This project set out to utilise widely conserved markers to characterise the germline of the mouse bile duct tapeworm, *Hymenolepis microstoma*, and through this, to contextualise *Hymenolepis* germline development within a phylogenetic framework. Results clearly indicate that the germline multipotency triad that has been catalogued across so many taxa, has been dramatically altered in the context of the mouse bile duct tapeworm, *Hymenolepis microstoma*. While both the *Hymenolepis* genome and transcriptome contain members of the Argonaute and DEAD-box families, the species has lost members of both the Vasa and Piwi clades; the latter of which appears to be more widely characteristic of parasitic flatworms. In addition, we have described an Argonaute clade specific to parasitic flatworms.

In order to investigate the fate of how components regulated by Vasa in other species in *Hymenolepis*, Nanos, Germ-Cell-Less and Bruno orthologs were searched for, revealing in RNA-Seq data that Germ cell less and Bruno are ubiquitously expressed throughout whole worm sections, and therefore are not spatially restricted.

While the *Hymenolepis* genome does retain a Nanos domain-containing protein as well as its interacting partner Pumilio, their coding characteristics, dynamics and roles in the specification of the germline require further consideration. Further work is required to explore the role that the 7-8th domains of Pumilio play in its interaction with Nanos and the effect this has on the germline. RNA-seq data has confirmed extremely low expression levels of Nanos. When combined with evidence of modifications in the repressive motif of Nanos, and the presence of only a single Nanos domain, the germline function of this protein in *Hymenolepis* must be questioned. In addition, Smaug, which is well characterised regulator of Nanos in other species is ubiquitously expressed in *Hymenolepis*. Whether this expression maintains its repressive effect on the low level of Nanos expression seen in the RNA-seq data is not known.

The presence and conservation of the Nanos binding domain further prompts us to ask whether this is the vital component of germline specification, or whether specific

characteristics of *Hymenolepis* have enabled conservation of its sequence, and perhaps role in the germline. The single copy of the Nanos domain in the identified gene and transcript may suggest that the roles of Nanos in *Hymenolepis* are also derived (for example changed regulation with regard to Vasa and Piwi), while the N terminus suggests that Nanos repressive function may be divergent.

Our results further support hypotheses proposed by Koziol et al., (2008) which support the presence of two Pumilio homologues in parasitic flatworms; one of which is exclusive to Neodermatans. Salvetti et al., (2005) noted the divergence in repeats 7-8 in Plathelminthes, and link this to the role that the region is known to play in recruiting Nanos in *Drosophila*. Sequence alignments suggest that repeat 8 in *Hymenolepis* is similar to other flatworms, and thus may diverge functionally from other characterised Pumilio homologs. We have also identified a Pumilio domain containing protein, which RNA-seq data suggests may be germline associated. However, the role that these proteins play in the germline is unclear, as their counterpart Nanos is expressed at extremely low levels, and is not spatially restricted.

These losses in the *Hymenolepis* genome highlight the potential difficulties in defining molecular markers based on key proteins other lineages. However, evidence of presence and differential expression of a Boule homology, the Arg-like HmN114300, as well as Pumilio, ELAV and GLD-1 highlight possible germline markers in *Hymenolepis* for use in comparative hybridisations and FACS based cell sorting. We have also been able to rule out the use of Germ Cell Less, Smaug, Bruno, Germinal Histone and Granulin as germline markers.

The loss of Piwi in *Hymenolepis* and parasitic lineages suggests that the dynamics of the Argonaute proteins in the organism may be different to their demonstrated roles in other species. This is supported by RNA-Seq data, indicating possible germline expression of the Arg-like HmN114300.

Absence of both Vasa and PIWI suggest that piRNAs may not have a large role in this species, Furthermore, the presence of two Dicer and Drosha genes, suggests that miRNA and siRNA pathways may still be intact. Therefore it is important to question

whether siRNAs, whose function includes transposon silencing (but to a lesser extent than piRNAs) maintain a transposon silencing operation to the same extent as in species with a functional piRNA pathway, or whether the changes in the size and composition of the *Hymenolepis* genome have relieved the burden of transposons on the genome, and germline. A clue to a more widespread modification of genome protection pathways comes from domain analysis of Argonaute clade associated proteins in *Hymenolepis*, both of which lack the DDH motif characteristic of the conserved active site in slicer activity. Slicer is the enzyme which cleaves the mRNA complementary to the small RNA in the groove which is required for both miRNA and siRNA guided cleavage (Song et al., 2004). This suggests that small RNA mediated transposon silencing mechanisms may be dramatically altered or completely lost in *Hymenolepis*.

The loss of Vasa and Piwi begs to question how vital they are to the maintenance of the germline, and whether it was specific environmental factors in the parasitic lifestyle, topology of the gene networks in the parasitic lineage, the presence of a neoblast phenotype or a combination of all of the above, which facilitated the loss of apparently vital germline markers. One key component to piecing together this quandary will come from furthering our understanding of the roles that closely related protein families can play in the maintenance of a developmental function, compensation for loss of a component, and more generally, in the control of a phenotype from a network perspective i.e. should we expect more ancient phenotypes or roles to be maintained by a larger or more complex network of closely related proteins? First and foremost, the related DEAD box helicases and Argonautes require in situ expression studies in *Hymenolepis* to confirm a possible germline function, and to corroborate RNA-Seq inferences.

Changes to Vasa, Piwi and Nanos-like genes also suggest possible changes have occurred to the makeup and role of chromatoid bodies in *Hymenolepis*. The absence of type 4 Tudor domains (Tsai, pers comm), and their roles in building macromolecular complexes also suggest the dynamics of the chromatoid bodies may be different in *Hymenolepis*. Indeed, cytological studies are required to confirm the presence of such large RNase sensitive structures in *Hymenolepis*. This change in the representation of Tudor

domains raises the question whether relaxed selection on Piwi led to relaxed selection on type 4 Tudor domains, or whether alternative pressures meant that type 4 domains were no longer selected for. RNA-seq evidence of a possible germline expression of one other Tudor domain containing protein, also suggests that we need to take a closer look at the determinants of 'germline' function in Tudor domain containing proteins.

Given the relationship that seems to exist between chromatoid bodies and regenerative capabilities in planarians, it is important to investigate the importance of chromatoid body components on regeneration, for example through transgenic planarians or the development of in vitro culture in *Hymenolepis*. Anecdotal evidence suggests that parasitic species have a reduced regenerative capability (Egger et al 2007). As regenerating flatworms owe this ability to the multipotent characteristics of neoblasts, as well as the extensive evidence described previously of the association of Vasa, Piwi and Nanos with neoblasts and regeneration, it is important to ask whether the loss of these gene families are functionally related to the reduction in regenerative capacity in parasitic flatworms.

Our attempt to catalogue Argonaute proteins in *Hymenolepis* and representative parasitic tapeworms, has been almost exclusively descriptive, raising a number of questions which need to be tackled, and corroborations and cellular experiments which must be completed to support our hypothesis. For example, the presence of piRNAs and miRNAs as well as their interaction with Argonautes must first be confirmed in *Hymenolepis* in order to test hypotheses on how germline protection has evolved in this species. This may be possible in yeast two hybrid assays, or through the use of antibodies to Argonaute in order to identify its targets (Beitzinger et al., 2007). In addition quantification of mRNA levels using quantitative PCR, and confirmation of mature proteins using western blot is an important step to moving the analysis from computational to cellular studies. As we have seen, the roles that Argonaute proteins play can be widespread, and while cell culture of *Hymenolepis* is not yet possible, understanding of its roles is still possible. For example, Piwi is known to play a role in fixing double strand breaks. Accumulation of a phosphorylated histone variant occurs in DNA damage, thus FISH detection of gammaH2Sv analysis is able to monitor the presence of dsDNA breaks in the genome (Thomson and Lin, 2009;Zhang et al., 2011). This

will enable us to investigate perhaps whether another pathway has taken over Piwi function, or if there has been a cost to the loss of Piwi.

Understanding the role of small RNAs in the germline and genome protection in *Hymenolepis* further requires the creation of small RNA libraries (Palekodedi et al., 2008), in order to compare the expression of small RNAs in the germline, somatic cells and neoblasts. Furthermore, deep sequencing of small RNA fractions and analysis with piRNA prediction methods using k-mer and genome aligning (Liu et al., 2012) would be of great interest. Tsai (pers comm) has been able to determine a loss of transposon sequences in the *Hymenolepis* genome. However, it is unclear whether and in what regard this change is related to changes in piRNA cluster composition and extent.

While it is currently not possible to ask whether the reduction in genome size is a cause or consequence of the loss of Piwi, increased sampling of lophotrochozoan lineages and in particular, parasitic species, as well as determination of genome size and Argonaute representation would be important. Combining this by investigating the relationship between the burden of transposable elements and the complexity of transposon surveillance mechanisms. may provide clues as to the relationships that exist between Piwi, transposons and genome size changes.

Characterisation of the expression of transposon families using whole tiling arrays may enable us to infer more about the factors surrounding the genomic changes in transposon distribution. For example Klattenhoff et al., 2009 demonstrated this capability which was followed up by Wei et al., (2012) who demonstrated the distribution of SINE-2, Gypsy, CRIII, CR1-13 and piggybac transposon families in the sea urchin.

Note

The results of this analysis also raise questions about the validity of some annotated protein models/transcripts submitted to the NCBI database. In particular, the Bayesian analysis of Vasa like proteins shows only two sequences grouping with the Piwi clade: *Helobdella* and *Dugesia dorotocephala*. The characteristic of interest in this topology is that the Lophotrochozoan lineages are more similar to more distantly related species than each

other. This may represent convergence, however as no other planarian sequences including the closely related *Dugesia japonica* have a Piwi like Argonaute member, we propose that the validity of the *Dugesia dorotocephala* 'Piwi' grouping sequence should be investigated.

In contrast, the DEAD-box analysis also raises questions about the validity of the labelling of *Neobenedenia* Argonaute members as Vasa-like based purely on neighbor joining phylogenetic analysis (Ohashi et al., 2007). Bayesian analysis showed Ngvlg2 to be an outlier to all clades in the analysis (not shown), while Ngvlg3 grouped with neither Vasa nor PL10, and was highly divergent from the closely related EIF4A family. We believe this error may be due to the adoption of neighbour joining methodology.

Evo-devo and the *Hymenolepis* germline

These insights in to parasitic flatworms and *Hymenolepis* have further relevance to questions within the discipline of Evo-devo. Thus the ability to explore a wider range of species through genome data and expression information can enable us to consider how the complex networks of temporal and spatial information involved in development are organised and how they have been able to evolve. Of particular interest is investigating how pathway components and the roles they play in the organism contribute to the flexibility of their relationships in an evolutionary context i.e. the molecular plasticity of the networks involved. Our investigations have highlighted the potential interest in looking at Nanos, Vasa and Piwi related protein families and networks, and their roles in the specification of the germline.

Non-coding RNAs also play an important evolutionary role by modulating the relationship between genotype and phenotype, and can regulate canalisation through working to regulate genetic and environmental variation and the effects of such variation on the cellular level (Juliano et al 2011). Thus from an evolutionary developmental perspective it is also important to ask whether the loss of Piwi, did as an event, alter the range of evolutionary possibilities i.e. the phenotypic space of germline development, as well as the range of possible environmental variations which can be experienced on the cellular level, and by cellular processes. Indeed, Littlewood (2006) raised the question of whether the

parasitic lifestyle has predisposed a different phenotypic space in the evolution of genetic structure, thus was it the parasitic lifestyle which facilitated the loss of seemingly vital germline markers? However, we should consider whether it was characteristic specific to the Argonaute family, and Piwi in particular that shaped the loss of Piwi in *Hymenolepis*, as Piwi is unique in the Argonautes, in that genomic changes are able to modulate changes (Piwi is dependent on specific genomic loci). Thus deeper consideration must be made of the contribution of network specific and environmental modifiable properties to evolutionary phenotypic space.

Technical issues

This project aimed to characterise Vasa and Piwi related proteins in *Hymenolepis* bioinformatically, and through the generation of expression data using RNA-seq and whole mount in situ hybridisation of mid-metamorphosis larvae and adult worms. Multiple problems were encountered with the cloning step of the protocol. Control experiments proved one batch of TOPO cells to be defective, causing rerunning of the protocol, while further issues prompted adoption of a 3' overhang protocol to ensure ligation capacity of the PCR products into the vector. Streamlining of the sequencing step was accomplished by adoption of a colony PCR to identify M13 inserts prior to commencement of sequencing reactions.

Issues collecting intact adult worms during dissection were overcome by developing an in situ heating and fixation protocol to enable careful extraction of samples from the mouse bile duct wall under a Leica dissecting microscope. This improved whole worm yields per infected mouse. Amplification of DEAD box and Argonaute-like transcripts required multiple PCR refinements, resulting in the generation of two separate PCR programs for sequences of interest.

Further future research

Bioinformatic approaches and expression studies are an important step towards piecing together the roles that such factors play in the genome, as well as conceptualising the evolutionary lability and tinkering within these networks across genera. However, the absence of cell culture and *in vitro* culture in *Hymenolepis* research limits our understanding of the spatial and temporal requirement for different developmental signaling systems. While *in vitro* culture protocols for *Hymenolepis microstoma* were first described in the 1960's (De Rycke and Bernstzen, 1967), attempts to recapitulate these methods produced limited growth (Pouchkina-Stantcheva et al., in prep), perhaps failing to recapitulate the *in vivo* conditions required for effective growth and reproductive capacity. This is indicative of the complex signaling and environmental requirements needed for the development of a parasitic species within its host(s).

Here we propose that integration of modern cell co-culture methodology with the chemical media employed in previous attempts, may help to establish growth conditions more representative of those experienced *in vivo*. Within the host, the parasitic environment is to a large extent biotic, and therefore is itself responsive to signals from the environment and the parasite. Thus co-culture methods employing mouse embryonic fibroblasts, immortal intestinal cell lines, or 3D culture, should be the next step in this attempt. Molecular characterisation of the functional interactions of network components may also benefit from the establishment of *Hymenolepis* cell cultures or immortal cell lines. Of particular interest would be exploration of the differing roles that small RNA pathways may play in the determination of male and female germ cells.

One of the main advancements in our understanding of gene function has been through the ability to modify gene expression, both transiently and in an inheritable fashion. Here we shall discuss areas of development for modification of gene function in parasitic flatworms. Forward genetic screens would enable identification of genes unlike those already known.

Transient change in expression can occur through dsRNA or antisense RNA which both lead to degradation of the transcript (Boyle and Yoshino, 2003). However the effectiveness of this strategy has been shown to vary within and between species, and even from a cell-to-cell manner. RNAi should be possible in *Hymenolepis*, as we have shown that components of the RNAi machinery are present. Transient expression has also been worked used plasmids, for example in luciferase expression in *Schistosoma mansoni* (Boyle and Yoshino, 2003). For example *Hydra vulgaris* embryos have allowed tracking of cells using B-actin driven EGFP (Wittlieb et al 2006). RNAi has also been demonstrated in cestodes, with down regulation of actin transcripts shown in *Moniezia expansa* (Pierson et al., 2010).

More permanent modifications in gene expression may come from the development of transgenic approaches. Generation of transgenic animals or cell cultures would facilitate exploring the effects of down regulation, tissue specific expression, conditional knockouts (e.g. Cre recombinase driven), and inducible knockouts (e.g. modified estrogen receptor based) on cell function (Qin et al 2008). However, it is important to note the requirement of effective *in vitro* protocols before application of many of these techniques. Evidence from *Schistosomes* highlights the structural sensitivity of cestodes, suggesting that methods of transfection may be more limited in this group. Indeed during attempts at generation of transgenic *Schistosomes*, the use of particle delivery system used in projectile bombardment played a particularly large role in effectiveness (Brindley et al 2007; Skinner et al 2012). This structural sensitivity of cestodes means that rather than bombardment, Lipofection or virus mediated (E.g. murine leukemia virus pseudotyped with vesicular stomatitis virus glycoprotein as evidenced in *S. mansoni*) transfection may be a more suitable approach.

Transgenic approaches are also complicated by creating effective integration of the sequence of interest into the genome, without creating multiple sites, interrupting a gene, or modifying the genome epigenetically or for example through affecting the production of small RNAs. Methods such as plasmid injection often lead to inefficient transfection and extrachromosomal DNA persisting through early development leading to mosaicism (Thermes et al 2002). Single site integration may be achieved for example by the

employment of meganucleases, whose recognition sites occur rarely in the genome due to their long length (Thermes et al 2002).

A further issue will be creating mutants, in order to characterise genes unlike those already characterised. One possible area of investigation for the production of insertional mutant flatworms is the use of transposon such as piggybac. While their effectiveness has not been shown in non-insect species, this enables an unbiased production of mutants (Boyle and Yoshino, 2003).

The above methods would also open up research into the role of differentiated cells and conserved signaling pathways in germline specification; broadening our understanding of the network relationships involved in germline specification, the cell interactions involved, and placing this expanded network into an evolutionary framework; selection acts on the phenotype, and thus interactions and cellular identities are all important components in conceptualising the changes that occurred during evolution.

References

Agata, K. (2008) 'Stem cells in Planarians'. IN Bosch, T.C.G. (ed) Stem cells; from Hydra to Man. Springer.

Agata, K. (2003) Regeneration and gene regulation in planarians. *Current Opinion in Genetics and Development*. 13: 493-496

Agee, S.J. Lyons, D.C. Weisblat, D.A. (2006) Maternal expression of a Nanos homolog is required for early development of the leech *Helobdella robusta*. *Developmental Biology*. 298: 1-11

Alié, A. Leclère, L. Jager, M. Dayraud, C. Chang, P. le Guyader, H. Quéinnec, E. Manuel, M. (2011) Somatic stem cells express Piwi and Vasa genes genes in an adult ctenophore: ancient association of 'germline genes' with stemness. *Developmental Biology*. 350: 183-197

Ariand, A. Kai, T. (2012) The Tudor domain protein Kumo is required to assemble the nuage and to generate germline piRNAs in *Drosophila*. *The EMBO Journal*. 31: 870-882

Baguña, J. (2012) The planarian neoblast: the rambling history of its origin and some current black boxes. *International Journal of Developmental Biology*. 56; 19-37

Bak, C.W. Yoon, T. Choi, Y. (2011) Functions of Piwi proteins in spermatogenesis. *Clinical and Experimental Reproductive Medicine*. 38 (2): 61-67

Batista, T.M. Marques, J.T. (2011) RNAi pathways in parasitic protists and worms. *Journal of Proteome Research*. doi: jprot.2011.02.032

Beitzinger, M. Peters, L. Zhu, J.Y. Kremmer, E. Meister, G. (2007) Identification of Human microRNA targets from isolated Argonaute Protein Complexes. *RNA Biology*. 4:2:

Bogisch, B.J. (1966) Histochemical Studies on *Hymenolepis microstoma* (Cestoda: Hymenolepididae) III Effects on the Bile Duct of the Mouse. *Transactions of the American Microscopical Society*. 85 (3): 372-377

Bogisch, B. J. (1963) Histochemical Studies in *Hymenolepis microstoma*. *The Journal for Parasitology*. 49 (6) 989-997

Boyle, J.P. Yoshino, T.P.(2003) Gene manipulation in parasitic helminths. *International Journal for Parasitology*. 33: 1259-1268

Brennecke, J. Malone, C.D. Aravin, A. Sadidanadam, R. Stark, A. Hannon, G.J. (2008) An epigenetic role for maternally inherited piRNAs in Transposon silencing. *Science*. 322:1387-1392

Brindley, P.J. Pearce, E.J. (2007) Genetic manipulation of schistosomes. *International Journal for Parasitology*. 37: 465-473

Carmell, M.A. Xuan, Z. Zhang, M.Q. Hannon, G.J. (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance and tumorigenesis. *Genes and Development*. 16: 2733-2742

Carthew, R.W. Sontheimer, E.J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*. 136. (4): 642-655

Carver, T. Berriman, M. Tivey, A. Patel, C. Bohme, U. Barrell, B.G. Parkhill, J. Rajandream, M.A. (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics*. 24L 2672-2676

Cerutti, L. Mian, L. Bateman A. (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends in Biochemical Sciences*. 481-482

Chen, C. Nott, T.J. Jin, J. Pawson, T. (2011) Deciphering arginine methylation: Tudor tells the tale. *Nature Reviews*. 12: 629-642

Chong, T. Stary, J.M. Wang, Y. Newmark, P.A. (2011) Molecular markers to characterize the hermaphroditic reproductive system of the planarian *Schmidtea mediterranea*. *BMC Developmental Biology*. 11: 69

Chu, C-Y. Rana, T.M. (2007) Small RNAs: Regulators and Guardians of the Genome. *Journal of Cellular Physiology*. 412-419

Collins, J.J. Hou, X. Romanova, E.V. Lambrus, B.G. Miller, C.M. Saberi, A. Sweedler, J.V. Newmark, P.S. (2010) Genome-wide analyses reveal a role for peptide hormones in planarian germline development. *PLoS Biology*. 8 (10): e1000509

Cordin, O. Branroques, J. Tanner, N.K. Linder, P. (2006) The DEAD-box protein family of RNA helicases. *Gene*. 367: 17-37

Cox, D.N. Chao, A. Baker, J. Chang, L. Quai, D. Lin, H. (1998) A novel class of evolutionarily conserved genes defined by Piwi are essential for stem cell self-renewal. *Genes and Development*. 12 3715-3727

Cunningham, L.J. Olson, P.D. (2010) Description of *Hymenolepis microstoma* (Nottingham strain): a classical tapeworm model for research in the genomic era. *Parasites and Vectors*. 2:123

Das, P.P. Bagjin, M.P. Goldstein, L.D. Woolford, J.R. Lehrbach, N.J. Sapetschnig, A. Buhecha, H.R. Gilchrist, M.J. Howe, K.L. Stark, R. Matthews, N. Berezikov, E. Ketting, R.F. Tawaré, S. Miska, E.A. (2008) Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon motility in the *Caenorhabditis elegans* germline. *Molecular Cell*. 31: 79-90

De Rycke, P.H. Berntsen, A.K. (1967) Maintenance and growth of *Hymenolepis microstoma* (Cestoda: Cyclophyllidea) *in vitro*. *The Journal for Parasitology*. 53 (2): 352-354

Deng, W. Lin, H. (2002) Miwi, a Murine Homolog of Piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Developmental Cell*. 2:819-830

Drummond, A.J. Ashton, B. Buxton, S. Cheung, M. Cooper, A. A. Duran, C. Heled, J. Kearse, M. Markowitz, S. Moir, R. Stones-Havas, S. Sturrock, S. Swidan, F. Thierer, T. Wilson, A. (2012) Geneious c5.6. Available from <http://www.geneious.com>

Dvorak, J.A. Jones, A.W. Kuhlman, H.H. (1961) Studies on the biology of *Hymenolepis microstoma*. *Parasites and Vectors*. 47 (5): 833-838

Egger, B. Gschwentner, R. Rieger, B. (2007) Free-living flatworms under the knife: past and present. *Development Genes and Evolution*. 217: 89-104

Evans, W.S. Movak, M. Basilevsky, A. (1985) Effects of Environmental Temperature, Sex and Infection with *Hymenolepis microstoma* on the Liver and Bile Duct Weights of Mice. *The Journal for Parasitology*. 71 (1): 106-109

Ewen-Campen, B. Schwager, E.E. Extravour, G.M. (2009) The molecular machinery of germ line specification. *Molecular Reproduction and Development*. 77: 3-18

Farazi, T.A. Juranek, S.A. Tuschl, T. (2008) The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development*. 135: 1201-1214

Fujii, T. Mitsunaga-Nakatsubo, K. Saito, I. Iida, H. Sakamoto, N. Akasaka, K. Yamamoto, T. (2006) Developmental expression of HpNanos, the *Hemicentrotus pulcherrimus* homologue of Nanos. *Gene expression Patterns*. 6: 572-577

Garcia-Silva, M.R. Tasar, J.P. Frugier, M. Pantano, S. Bonilla, B. Esteban, L. Serra, E. Roura, C. Robello, C. Cayota, C. (2010) Cloning, characterization and subcellular localization of a *Trypanosoma cruzi* Argonaute protein defining a new subfamily distinctive of trypanosomatids. *Gene*. 466: 26-35

Girard, A. Sachidanandam, R. Hannon, G.J. Carmell, M.A. (2006) A germline-specific class of small RNAs bind mammalian Piwi proteins. *Nature*. 442: 199-202

Giribet, G. (2008) Assembling the lophotrochozoan (=spiralian) tree of life. *Philosophical Transactions of the Royal Society B*. 363 (1496): 1513-1522

Goodchild, C.G. Davis, B.O. (1972) *Hymenolepis microstoma* Cysticercoid Activation and Excystation *in vitro* (Cestoda). *The Journal for Parasitology*. 58 (4): 735-741

Goodchild, C.G. Stullken, R.E. (1970) *Hymenolepis microstoma*: Cysticeroid Morphogenesis. *Transactions of the American Microscopical Society*. 89 (2): 224-229

Guo, T. Peters, A.H.F.M. Newmark, P.A. (2006) A bruno-like gene is required for stem cell maintenance in Planarians. *Developmental Cell*. 11: 159-169

Gustafson, E.A. Wessel, G.M. (2010) Vasa genes: Emerging roles in the germ line and in multipotent cells. *Bioessays*. 32: 626-637

Handberg-Thorsager, M. Salö, E. (2007) The planarian Nanos-like gene Smednos is expressed in germline and eye precursor cells during development and regeneration. *Development Genes and Evolution*. 217: 403-411

Hejnol A, Obst M, Stamatakis A, Ott, M. Rouse, G.W. Edgecombe, G.D. Martinez, P. Baguna, J. Bailly, X. Jondelius, U. Wiens, M. Muller, W.E.G. Seaver, E. Wheeler, W.C. Martindale, M.Q. Giribet, G. Dunn, C.W. (2009) Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings of the Royal Society B Biological Sciences*. **276**: 4261–4270.

Hoberg EP, Jones A & Bray RA. (1999) Phylogenetic analysis among the families of the Cyclophyllidea (Eucestoda) based on comparative morphology, with new hypotheses for co-evolution in vertebrates. *Systems Parasitology* **42**: 51–73.

Höck, J. Meister, G. (2008) The Argonaute protein family. *Genome Biology*. 9: 210

Houwing, S. Kamminga, L.M. Berezikov, E. Cronenbold, D. Girard, A. van den Eist, H. Filippov, D.V. Blaser, H. Raz, E. Moens, C.B. Plasterk, R.H.A. Hannon, G.J. Draper, B.W. Ketting, R.F. (2007) A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon silencing in Zebrafish. *Cell*. 129: 69-82

Ishizuka, H. Maezawa, T. Kawauchi, J. Nodono, H. Hirao, Y. Nishimura, O. Nakagawa, H. Sekii, K. Tasaka, K. Tanii, H. Agata, K. Hoshi, M. Kobayashi, K. Johnstone, O. Deuring, B. Bock, R. Linder, P. Fuller, M.T. Lasko, P. (2005) Belle is a *Drosophila* DEAD-box protein required for viability and in the germline. *Developmental Biology*. 277: 92-101

Jeske, M. Moritz, B. Anders, A. Whale, E. (2011) Smaug assembles an ATP-dependent stable complex repressing nanos mRNA translation at multiple levels. *EMBO Journal*. 30 (1): 90-103

Juliano, E. Wang, J. Lin, H. (2011) Uniting Germline and Stem cells: The Function of Piwi proteins and the piRNA pathway in Diverse Organisms. *Annual Review of Genetics*. 45: 447-469

Juliano, C.E. Yajima, M. Wessel, G.M. (2009) Nanos functions to maintain the fate of the small micromere lineage in the sea urchin embryo. *Developmental Biology*. 337 (2): 220-232

Kang, D. Pilon, M. Weisblat, D.A. (2002) Maternal and zygotic expression of a Nanos-class gene in the leech *Helobdella robusta*: Primordial germ cells arise from segmental mesoderm. *Developmental Biology*. 245: 28-41

Kawaoka, S. Izumi, N. Katsuma, S. Tomari, Y. (2011) 3' End Formation of Piwi-interacting RNAs *in Vitro*. *Molecular Cell*. 43: 1015-1022

Klattenhoff. C. Xi, H. Li, C. Lee, S. Xu, J. Khurana, J.S. Zhang, F. Schultz, N. Koppetsch, B.S. Nowosielska, A. Seitz, H. Zamore, P.D. Weng, Z. Theurkauf, W.E.

(2009) The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell*. 138 (6): 1137-1149

Knox, D.P. Geldof, P. Viser, A. Britton, C. (2007) RNA interference in parasitic nematodes of animals: a reality check? *TRENDS in Parasitology*. 23 (3): 105-108

Korhonen, H.M Meikar, O. Yadav, R.P. Papaioannou, M.D. Romero, Y. Ros, M.D. Herrera, P.L. Toppari, J. Nef, S. Kotaja, N. (2011) Dicer is required for Haploid male germ cell differentiation in mice. *PLoS ONE*. 6(9): e24821

Koziol, U. Marin, M. Castillo, E. (2008) Pumilio genes from the Platyhelminthes. *Development, Genes and Evolution*. 218;47-53

Kuales, G. De Mulder, K. Glashauser, J. Salvenmoser, W. Takashima, S. Hartenstein, V. Berezikov, E. Salzburger, W. Ladurner, P. (2011) Boule-like genes regulate male and female gametogenesis in the flatworm *Macrostomum lignano*. *Developmental Biology*. 357: 117-132

Lai, F. Zhou, Y. Luo, X. Fox, J. Kings, M.L. (2011) Nanos1 functions as a translational repressor in the *Xenopus* germline. *Mechanisms of Development*. 1-2: 153-163

Lee, Y. Ahn, C. Han, J. Choi, H. Kim, J. Yim, J. Lee, J. Provost, P. Radmark, O. Kim, S. Kim, N. (2003) The nuclear RNase III Drosha Initiates microRNA processing. *Nature*. 425 414-419

Lee, Y. Ahn, C. Han, J.J. Choi, H. Kim, J. Provost, P. Radmark, O. Kim, S. Kim, V.N. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*. 425 (6956): 415-419

Li, W. Deng, F. Wang, H. Zhen, Y. Xiang, F. Sui, Y. Li, J. (2006) Germ cell-less expression in Zebrafish embryos. *Development, Growth and Differentiation*. 48 (5): 333-338

Littlewood, D.T.L (2006) The Evolution of Parasitism in Flatworms. IN A.G.Maule, N.J. Marks (eds) *Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology and Physiology*. CAB pp1-37

Liu, G. Lei, B. Li, Y. Tong, K. Ding, Y. Luo, L. Xia, X. Jiang, S. Deng, C. Xiong, Y. Li, F. (2012) Discovery of potential piRNAs from next generation sequences of the sexually mature porcine testes. *PLoS ONE*. 7 (4): e34770

Miguel, E. Kremer, M. (2004) Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*. 72; 159-217

Miyoshi, K. Tsukumo, H. Nagami, T. Siomi, H. Siomi, M.V. (2005) Slicer function of *Drosophila* Argonautes and its involvements in RISC formation. *Genes and Development*. 19 (23): 2837-2848

Mochizuki, K. Nishimiya-Fujisawa, C. Fujisawa, T. (2001) Universal occurrence of the Vasa-related genes among metazoans and their germline expression in *Hydra* Development, *Genes and Evolution*. 211: 299-308

Mochizuki, K. Sano, H. Kobayashi, S. Nishimiya-Fujisawa, C. (2000) Expression and evolutionary conservation of Nanos-related genes in *Hydra*. *Development, Genes and Evolution*. 210: 591-602

Moore, J. Han, H. Lasko, P. (2009) Bruno negatively regulates germ-cell-less expression in a BRE-independent manner. *Mechanisms of Development*. 126: 503-516

Nakagawa, H. Ishizu, H. Chinone, A. Kobayashi, K. Matsumoto, M. (2012) The Dr-Nanos gene is essential for germ cell specification in the planarian *Dugesia ryukyuensis*. *International Journal of Developmental Biology*. 56: 165-171

Nollen, P.M. (1975) Studies on the reproductive system of *Hymenolepis diminuta* using autoradiography and transplantation. *The Journal for Parasitology*. 61 (1): 100-104

Northrop-Cleaves, A. Shaw, C. (2000) Parasites. *British Medical Bulletin*. 56 (1): 193-208

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 function by double-stranded RNA in the monogenean, *Neobenedenia girellae*. *International Journal for Parasitology*. 37: 515-523

Olovnikov, I. Aravin, A.A. Toth, K.F. (2012) Small RNAs in the nucleus: the RNA chromatin ping-pong. *Current Opinion in Genetics and Development*. 22 (2): 164-171

Olson PD, Poddubnaya LG, Littlewood DTJ & Scholz T. (2008) On the position of *Archigetes* and its bearing on the early evolution of the tapeworms. *Journal for Parasitology* ; **94**: 898–904.

Olson, P.D. Zarowiecki, M. Kiss, F. Brehm, K. (2012) Cestode genomics-progress and prospects for advancing basic and applied aspects of flatworm biology. *Parasite Immunology*. 34: 130-150

Olson, P.D. Zarowiecki, M. Kiss, D. Brehm, K. (2011) Cestode genomics-progress and prospects for advancing basic and applied aspects of flatworm biology. *Parasite Immunology*. 34 (2-3): 130-150

Palakodeti, D. Smielewska, M. Lu, Y.C. (2008) The Piwi proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA*. 14: 1174-1186

Pane, A. Wehr, K. Schüpbach (2007) zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Developmental Cell*. 12 (5): 851-862

Parisi, M. Lin, H. (2000) Translational repression: a duet of Nanos and Pumilio. *Current Biology*. 10: 81-83

Pek, J.W. Anand, A. Kai, T. (2012) Tudor domain proteins in development. *Development*. 139: 2255-2266

Perkins, E.M. Donnellan, S.C. Bertozzi, T. Whittington, I.D. (2010) Closing the mitochondrial circle on paraphyly of the Monogenea (Platyhelminthes) infers evolution in the diet of parasitic flatworms. *International Journal for Parasitology*. 40 (11); 1237-1245

Pfister, D. De Mulder, K. Hertenstein, V. Kuales, G. Borgonie, G. Marx, F. Morris, J. Laduma, P. (2008) Flatworm stem cells and the germline: developmental and evolutionary implications of macVasa expression in *Macrostomum lignano*. *Developmental Biology*. 319 (1): 146-159

Pierson, L. Mousley, A. Devine, L. Marks, N.J. Day, T.A. Maule, A.G. (2010) RNA interference in a cestode reveals specific silencing of selected highly expressed gene transcripts. *International Journal for Parasitology*. 40:605-615

Pouchkina-Stantcheva, N.N. Cuningham, L. Hrcakova, G. Olson, P.D. (in prep) RNA-mediated gene suppression and in vitro culture in *Hymenolepis microstoma*. *International Journal for Parasitology*.

Qin, J. Tsai, M.J. Tsai, S.Y. (2008) Essential Roles of COUP-TFII in Leydig cell differentiation and male fertility. *PLoS ONE*. 3 (9); 3285

Raghuveer, K. Senthilkumaran, B. (2010) Cloning and differential expression pattern of Vasa in the developing and recrudescing gonads of catfish, *Clarias gariepinus*. *Comparative biochemistry and Physiology- Part A: Molecular and Integrative Physiology*. 157 (1): 79-85

Rauschhuber, C. Ehrhardt, A. (2012) RNA interference is responsible for reduction of transgene expression after Sleeping Beauty Transposase mediated somatic integration. *PLoS ONE*. 7 (5): e35389

Rebscher, N. Lidke, A.K. Ackermann, C.F. (2012) Hidden in the crowd: primordial germ cells and somatic stem cells in the mesodermal posterior growth one of the polychaete *Platynereis dumerilii* are two distinct cell populations. *EvoDevo*. 3:9

Rebscher, N. Zelad-González, F. Banisch, T.U. Raible, F. 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*. *Developmental Biology*. 306: 599-611

Reddien, P.W. Oviedo, N.J. Jennings, J.R. Jenkin, J.C. Sánchez-Alvarado, A. (2005) SMEDWI-2 is a Piwi-like protein that regulates Planarian stem cells. *Science*. 310: 1327-1330

Reuter, M. Berninger, P. Churna, S. Shah, H. Hosokawa, M. Funaya, C. Antony, C. Sachidanandam, R. Pillai, R.S. (2011) Miwi catalysis is required for piRNA amplification independent LINE1 transposon silencing. *Nature*. 482: 264-269

Riddiford, N. Olson, P.D. (2011) Wnt gene loss in flatworms. *Development Genes and Evolution*. 221: 187-197

Ronquist, F. Huelsenbeck, J 2003 MRBAYES 3 Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574)

Rossi, L. Salvetti A. Lena, A. Batistoni, R. Deri, P. Pugliesi, C. Loreti, E. Gremigni, V. (2006) DjPiwi-1 a member of the PAZ-Piwi gene family defines a subpopulation of planarian stem cells. *Development Genes and Evolution*. 216: 335-346

Rozen, S. Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. IN Krawetz, S. Misener, S. (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ. pp365-386

Rutherford, K.M. Parkhill, J. Crook, J. Horsnell, T. Rice, P. Rajandream, M.A. Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics*. 16: 944-945

Saló, E. Abril, J.F. Adell, T. Cebria, D. Eckelt, K. Fernandez0Taboada, E. Handberg,- Thorsager, M. Iglesias, M. Molina, M.D. Rodriguez-Esteban, G. (2009) Planarian regeneration achievements and future directions after 2 years of research. *International Journal of Developmental Biology*. 53: 1317-1327

Salveti, A. Rossi, L. Lena, A. Batistoni, R. Deri, P. Rainaldi, G. Locci, M.T. Evangelista, M. Gremigni, V. (2005) DjPum, a homologue of Drosophila Pumilio, is essential to planarian stem cell maintenance. *Development*. 132: 1863-1874

Sato, K. Shibata, N. Orii, H. Amikura, R. Sakurai, T. Agata, K. Kobayashi, S. Watanabe, K. (2006) Identification and origin of the germline stem cells as revealed by the expression of Nanos-related gene in planarians. *Development Growth and Differentiation*. 48: 615-628

Scholz, S. Domaschke, H. Kanamori, A. Ostermann, K. Rodel, G. Gutzeit, H.O. (2004) Germ cell-less expression in medaka (*Oryzias latipes*). *Molecular Reproduction and Development*. 76 (1): 15-18

Schwab, R. Maizel, A. Ruiz-Ferrer, V. Garcia, D. Bayer, M. Crespi, M. Voinnet, O. Martienssen, R.A. (2009) Endogenous TasiRNAs mediate non-cell autonomous effects on gene regulation in *Arabidopsis thaliana*. *PLoS ONE*. 4 (6): e5980

Schwalbe, H. (2011) Double Take on Piwi protein/piRNA complex structure. *Structure*. 19: 141-142

Seidel, J.S. (1975) The Life Cycle *in vitro* of *Hymenolepis microstoma* (Cestoda). *The Journal for Parasitology*. 61 (4): 677-681

Senti, K.A. Brennecke, J. (2010) The piRNA pathway, a fly's perspective on the guardian of the genome. *Trends in Genetics*. 26 (12): 499-509

Seto, A.G. Kingston, R.E. Lau, N.C. (2007) The coming of age for Piwi proteins. *Molecular Cell*. 26: 603-609

Shibata, N. Umesono, Y. Orii, H. Sakurai, T. Watanabe, K. Agata, K. (1999) Expression of Vasa (vas)- related genes in germline cells and totipotent somatic stem cells of Planarians. *Developmental Biology*. 206: 73-87

Siomi, H. Siomi, M.C. (2008) Interactions between transposable elements and Argonautes have (probably) been shaping the *Drosophila* genome throughout evolution. *Current Opinion in Genetics and Development*. 18: 181-187

Skinner, D.E. Rinaldi, G. Suttiprapa, S. Mann, V.H. Smircich, P. Cogswell, A. A. Williams, D.L. Brindley, P.J. (2012) Vasa-like DEAD-Box RNA Helicases of *Schistosoma mansoni*. *PLoS Neglected Tropical Disease*. 6 (6): e1686

Solana, J. Lasko, P. Romero, R. (2009) Spoltud-1 is a chromatoid body component required for planarian long-term stem cell self-renewal. *Developmental Biology* .328 (2): 410-421

Solana, J. Romero, R. (2009) SpolvlgA is a DDX3/PL10-related DEAD-box RNA helicase expressed in blastomeres and embryonic cells in planarian embryonic development. *International Journal of Biological Sciences*. 5 (1): 64-73

Song, J.J. Smith, S.K. Hannon, G.J. Joshua-Tor, L. (2004) Crystal structure of Argonaute and its implications for RISC Slicer activity. *Science*. 305 (5689): 1434-1437

Sunanaga, T. Satoh, M. Kawamura, K. (2008) The role of Nanos homologue in gametogenesis and blastogenesis with special reference to male germ cell formation in the colonial ascidian, *Botryllus primigenus*. *Developmental Biology*. 324 (1): 31-40

Tadros, W. Goldman, A.L. Babak, T. Menzies, F. Vardy, L. Orr-Weaver, T. Hughes, T.R. Westwood, T. Smibert, C.A. Lipshitz, H.D. (2007) SMAUG is a major regulator of maternal mRNA destabilisation in *Drosophila* and its translation is activated by the PAN GU kinase. *Developmental Cell*. 12 (1): 143-155

Tanaka, T. Hosokawa, M. Vagin, V.V. Reuter, M. Hayashi, E. Mochizuki, A.L. Kitamura, K. Tamanaka, H. Kondoh, G. Okawa, K. Kuramochi-Miyagawa, S. Nakano, T. Sachidanandam, R. Hannon, G.J. Pillai, R.S. Nakatsuji, N. Chuma, S. (2011) Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodelling of chromatoid bodies during spermatogenesis. *PNAS*. 108 (26): 10579-10584

Tessmar-Raible, K. Arendt, D. (2003) Emerging systems: between vertebrates and arthropods: the Lophotrochozoa. *Current Opinion in Genetics and Development*. 13 (4): 331-340

Thermes, V. Grabher, C. Ristoratore, D. Bourrat, D. Choulika, A. Wittbrodt, J. Joly, J.S. (2002) *I-SceI* meganuclease mediates highly efficient transgenesis in Fish. *Mechanisms of Development*. 118 (1-2): 91-98

Thomson, T. Lin, H. (2009) The Biogenesis and Function of Piwi proteins and piRNAs: Progress and Prospect. *Annual Review in Cell and Developmental Biology*. 25: 355-376

Tomancak, P. Guichet, A. Zavorszky, P. Ephrussi, A. (1998) Oocyte polarity depends on regulation of Gurken by Vasa. *Development*. 125 (9): 1723-1732

Tsai et al (in prep)

Wang, Q. Fang, D.A. Sun, J.L. Wang, Y. Liu, L.H. (2012) Characterisation of the Vasa gene in the Chinese mitten crab *Eriocheir sinensis*: a germ line molecular marker. *Journal of Insect Physiology*. 58 (7): 960-965

Wang, Y. Zayas, R.M. Guo, T. Newmark, P.A. (2007) Nanos function is essential for development and regeneration of planarian germ cells. *PNAS*. 104 (14): 5901-5906

Webb, R.A. (2004) Studies on platyhelminthes: yesterday, today and tomorrow. *Canadian Journal of Zoology*. 82: 161-167

Wei, K. Wi, L. Chen, J. Chen, Y. Xie, D. (2012) Structural Evolution and Functional Diversification Analyses of Argonaute protein. *Journal of Cellular Biochemistry*. 113: 2576-2585

Wittleib, J. Khalturin, K. Lohmann, J.U. Anton-Erxleben, F. Bosch, T.C.G. (2006) Transgenic *Hydra* allow in vivo tracking of individual stem cells during morphogenesis. *Proceedings of the National Academy of Sciences U.S.A.* 103 (1): 6206-6211

Wu, Q. Song, R. Ortogero, N. Zheng, H. Evanoff, R. Small, CL. Griswold, M.D. Namekawa, S.H. Royo, H. Turner, J.M. Yan, W. (2012) The RNase III Enzyme Drosha is essential for microRNA production and spermatogenesis. *The Journal of Biological Chemistry*. 287 (30): 25173-25190

Wei, Z. Liu, X. Zhang, H. (2012) Identification and Characterisation of piRNA-like small RNAs in the gonad of Sea Urchin (*Strongylocentrotus nudus*). *Marine Biotechnology*. 14: 459-467

Weisblat, D.A. (2006) Germline Regeneration: The worm's turn. *Current Biology*. 16: 12. 454

Whelan S, Goldman , N. 2001 A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*. 18: 691-699

Yajima, M. Wessel, G. M. (2011)The multiple hats of Vasa: its functions in the germline and in cell cycle progression. *Molecular Reproduction and Development* 78: 861-867.

Ying, M. Chen, D. (2012) Tudor domain-containing proteins of *Drosophila melanogaster*. *Development Growth and Differentiation*. 54: 32-43

Yogit, E. Batista, P.J. Bel, Y. Pang, K.M. Chen, C.G. Tolia, N.H. Joshua-Tor, L. Mitani, S. Simard, M.J. Mello, C.C. (2006) Analysis of the *C.elegans* Argonaute family reveals that Distinct Argonautes act sequentially during RNAi. *Cell*. 127: 747-757

Zhang, D. Duarte-Guterman, P. Langlois, V.S. Trudeau, V.L. (2010) Temporal expression and steroidal regulation of piRNA pathway genes (Mael, Piwi, Vasa)

during *Silurana (Xenopus) tropicalis* embryogenesis and early larval development. *Comparative Biochemistry and Physiology Part C. Toxicology and Pharmacology*. 152 (2): 202-206

Zhang, Z. Koppetsch, B.S. Wang, J. Tipping, C. Ma, S. Weng, Z. Theurkauf, W.E. Zamore, P.D. (2011) Heterotypic piRNA Ping-pong requires Qin, a protein with both E3 Ligase and Tudor domains. *Molecular Cell*. 44 (4): 572-584

Appendix 1 Accession numbers

Accession numbers used in Phylogenetic analysis and Protein Alignments

Argonaute

F55A12.2 (Caenorhabditis) NP_491579
R06C7.1 (Caenorhabditis) NP_49245
F58G1.1(Caenorhabditis) CAB04519
PPW-2 (Caenorhabditis) CCD66211
Arg-like-1 (Isodiametra) CAQ05990
Arg-3 (Clonorchis) GAA55650
Arg-2 (Clonorchis) GAA47732.1
Argonaute_2 (Drosophila) NP_648775
Argonaute2a (Tribolium) NP_001107842.1
Argonaute2b (Tribolium) NP_001107828
Argonaute (Arabidopsis) AAC18440
PINHEAD (Arabidopsis) AED95011
Arg-8 (Arabidopsis) NP_180853
Arg-4 (Arabidopsis) NP_001189613
Arg-9 (Arabidopsis) CAD66636
T23D8.7 (Caenorhabditis) NP_492643
Arg-4 (Homo) NP_060099.2
Arg-1 (Homo) NP_036331
Alg-1 (Caenorhabditis) NP_510322
Alg-2 (Caenorhabditis) NP_493837
Argonaute-1 (Drosophila) NP_725341.1
Argonaute1 (Apis) XP_624444
Argonaute1 (Tribolium) XP_971295.2
ARG-1 (Echinococcus) CBL30927
AGO-2 (Schmidtea) JF263459
Piwi-like-1 (Macrostomum) CAQ03958.1
DjPiwiA (Dugesia) NP_001159378
Piwi-like-1 (Schmidtea) ABB77337
Piwi B (Seidlia) AEJ35111
Piwi-like-2 (Schmidtea) ABB77338
Piwi homologue (Dugesia) CAI26303
PIWIL-DUGJA (Dugesia) Q2PC95
Piwi-like-1 (Isodiametra) CAQ03959
CARBREN-15497 (Caenorhabditis) EGT57709
Cniwi1 (Ixodes) XP_002399390
Cniwi2 (Ixodes) XP_002411770.1
Piwi-Aubergine (Daphnia) EFX88764
Piwi (Aedes) EAT34539
Aubergine (Drosophila) NP_476734.1
Piwi (Drosophila) NP_476875.1
Aubergine (Apis) NP_001159378.1
Aubergine (Tribolium) XP_001811159.1
Piwi (Bombyx) BAF98574
Piwi1 (Helobdella)
Piwi (Clytia) ABY67112
Piwi (Botryllus) BAG69146
Piwi (Ephydatia) BAJ07610
PIWIL3 (Homo) BAC81343
Piwi-like-1 (Oryzias) ACB47463
Piwi (Danio) AAL57170
ZIWI (Danio) NP_899181.1
Piwi-like1 (Gallus) ABR09543
Piwi-like-1 (Rattus) NP_001102323

Piwi-like-1 (Macaca) NP_001182640
Piwi-like-1 (Sus) NP_001181902
WI-3 (Schmidtea) ACC97187
Piwi-like 1 (Apis) XP_001120996
Piwi (Tribolium) XP_968053.2
Ago3 (Drosophila) NP_001036627.2
AGO3 (Bombyx) NP_001108114.1
Siwi2 (Bombyx) NP_001098067.2
Piwi2 (Helobdella)
Piwi-like-2 (Xenopus) NP_001106470
Piwi-like-2 (Homo) NP_001129193
MILI (Mus) NP_067283
Piwi-like-2 (Mus) NP_067283

Royal family

FMR1 (Drosophila) NP-611645
FMR1 (Homo) AAH38998
FMR1a (Xenopus) NP_001079156
FXR1B (Xenopus) NP_001080422
Chromo domain (Caenorhabditis) NP_495652
CBX-2 (Mus) NP_031649
Chromodomain (Bombyx) NP_001106734
Chromobox (Homo) NP-006798
MBTD1 (Xenopus) NP_001006742
LBL1 (mus) NP_001074807
MBTR1 (Caenorhabditis) NP_001122542
LMBL2 (homo) NP_113676
BRD1 (homo) EAW73473
NSD2 (mus) NP_780440
GLYR1 (Drosophila) CG4747
HDGR2 (Xenopus) NP_001085132
SMN (Mus) CAX15798
Tudor domain (Dugesia) ADF47433
YQK1 (Caenorhabditis) NP_741190
AKAP1 (Homo) NP_003479
SCHLP (Schmidtea) ACN54319

Dicer/Drosha

Dicer (Paracoccidioides) EEH22894
Dicer (Trichophyton) EGD96414
Dicer (Aspergillus) EIT78563
Drosha (Oikopleura) CAP07635
Drosha (Xenopus) NP_001107152
Drosha (Ascaris) AEF32762
Drosha(Drosophila) AAF59169
Drosha (Marsupenaeus) ADB65770
Dicer 4 (Orzya) A7LFZ6
Dicer (Schistosoma) ABQ02405
Dicer (Echinococcus) CBL30926
Dicer1 (Drosophila) AAF56056
Dicer (Rattus) XP_001069041
Dicer (Sus) NP_001184123
Dicer (Otelemur) XP_003787066
Dicer (Xenopus) NP_001123390
Dicer (Gallus) NP_001035555
Dicer (Bos) NP_976235
Dicer (Homo) NP_001182502
Dicer (Pan) XP_003832879

Dicer (Pongo) XP_002825116

Vasa alignment

Drosophila (AAF53438) Danio (AAI29276) Mus (BAA03584) Xenopus (Q91372)

DEAD box helicases

Vasa2 (Helobdella)
Vasa1 (Helobdella)
Vasa (Strongylocentrotus) ACM80369
Vasa-like (Crassostrea) AAR37337
Vasa (Paryhale) ABX76969.1
Vasa (Tribolium) NP_001034520.2
PIWAS1 (Dugesia) BAB13313
Vasa (Nasonia) XP_001603956
Vasa (Gryllus) BAG65665.1
Vasa (Apis) NP_001035345
Vasa (Culex) EDS32555
Vasa-like (Bombyx) NP_001037347
Vasa (Oryzias) BAB1047
Vasa (Crassius) AAV70960
Vasa (Danio) AAI29276
VLG1 (Xenopus) AAI69679
DEADbox (Homo) AAF86585
Vasa (Mus) BAA03584.1
Vasa-like (Rattus) AAB33364.1
Dead box 4 (Mus) NP_001139357
Ngvlg3 (Neobenedenia) AB265787
DEAD box helicase (Schistosoma) 4
SJFCE0740 (Schistosoma) FN317190
Translation initiation factor 4A (Echinococcus) CAC18543
EIF4A (Brassica) AAF19805
EIF4A (Homo)2 NP_001036024
EIF4A (Xenopus) NP_001011139
EIF4A (Drosophila) AAF52317
EIF4A (Dugesia) BAF57644
EIF4A ((Clonorchis) GAA43007
DjVLGA (Dugesia) 1 AB017002
DEAD box helicase (Schistosoma) 3
SJFCE4672 (Schistosoma) FN326795
Vasa (Schistosoma) JQ353769
Ngvlg1 (Neobenedenia) AB265786
VASA2n (Paragonimus) ABM30180
VASA3n (Paragonimus) ABM30181
DEAD box helicase (Schistosoma) 2 CCD58866.1
PL10-like (Haliotis) GQ259891
CnPL10 (Hydra) AB047381
PL10a (Platynereis) AM048813
PL10b (Platynereis) AM048814
Belle (Tribolium) NP_001153721.1
Belle (Drosophila) NP_536783
Belle (Nasonia) CG9748-PA
Belle (Apis) CG9748-PA
PL10 (Apis) XP_391829
PL10 (Trichinella) XP_003378958
Adhaerens (Trichoplax) X M_002118017
PoPL10 (Ephydatia) AB047384
PL10 (Mus) NP_149068.1

PL10-related (Rana) CAH61467
PL10 (Xenopus) NP_001080283

Nanos alignment

Cnnos1 (Hydra magnipapillata) (XP_002161850)
Nanos 1 (Podocoryna) (AAU11513)
Nanos 2 (Clytia) (AFD28591)
Hro-nos (Helobdella) (AAB63111)
Nanos 2 (Homo) (NP_001025032)
Nanos1 (Mus) (AAH56473)
Nanos 1 (Clonorchis) GAA47235
Nanos-like (Schmidtea)1 (ABO52809)
Nanos-like (Schmidtea) 2 (ABO77964)
Nanos-related protein (Dugesia japonica) (BAD88623)
Nanos related protein (Dugesia ryukensis) (BAK88623)
Nanos RNA binding domain (Schistosoma) 1(CCD79863)
Nanos RNA binding domain (Schistosoma) 2 (Smp_055740)
Nanos homolog 1 (Schistosoma) (CAX69806)
SJCHGC053118 (Schistosoma) (AAW26680)
Nanos (Drosophila) (AAA28715)

Nanos

E3 ubiquitin ligase KCMF1 (Bos) DAA24608
E3 ubiquitin ligase KCMF1 (Mus) Q80UY2
E3 ubiquitin ligase KCMF1 (Macaca) AFE79271
E3 ubiquitin ligase KCMF1 (Xenopus) Q6GPB6
E3 ubiquitin ligase KCMF1 (Danio) 7T321
E3 ubiquitin ligase KCMF1 (Oreochromis) LOC100706944
E3 ubiquitin ligase KCMF1 (Camponotus) EFN73191
E3 ubiquitin ligase KCMF1 (Nasonia) LOC100122483
E3 ubiquitin ligase KCMF1 (Clonorchis) GAA54801
Predicted protein (Nematostella) EDO37139.1
PCMF1-like (Ciona) XP_002130812
Nanos (Branchionus plicatillis) ADD91653
PoNOS (Ephydatia) BAB19253
Hro-nos (Helobdella) AAB63111
Nanos (Pristina) ADE44350
Nanos homolog 1 (Salmo) ACI70036
Nanos-like (Isodiametra) CAX32468
Nanos-like (Nasonia) AAT94169
Nanos1 (Mus) AAH56473
Nanos-like (Haliotis) ACT35656
Nanos-like (Bombyx) ACI49631
Nanos (Drosophila) AAA28715
Nanos-like (Ilyanassa) ABV54788
Nanos (Branchiostoma belcheri) ADD25828
Nanos-like (Branchiostoma) ADM26639
Nanos 2 (Homo) AAH42883
Nanos (Danio) AAH97090
Nanos (Culex) ACB20970
NOS (Anopheles) AAS93543
Nanos (Schistocerca) AAO38523
Nanos 2 (Nematostella) AAY67908
Nanos 2 (Podocoryna) AAU11513
CNNOS2 (Hydra vulgaris) BAB01492
Nanos2 (Clytia) AFD28591
MNEMVEDRAFT (Nematostella) XM_001637125

Nanos1 (Nematostella) AAW29070
Nanos1 (Podocoryna) AAU11513
CNNOS (Hydra viridissima) BAJ23289
CNNOS (Hydra utahensis) BAJ23284
CNNOS (Hydra circumcincta) BAJ23286
Cnnos1 (Hydra magnipapillata) XP_002161850
CNNOS1 (Hydra robusta) BAJ23267
Nanos RNA binding domain (Schistosoma) 1 CCD79863
Nanos RNA binding domain (Schistosoma) 2 XM_002576802
Nanos 1 (Clonorchis) GAA47235
Nanos homolog 1 (Schistosoma) CAX69806
SJCHGCO5318 (Schistosoma) AAW26680
Nanos-like (Schmidtea) 1 ABO52809
Nanos-like (Schmidtea) 2 ABO77964
Nanos-related protein (Dugesia japonica) BAD88623
Nanos-related protein (Dugesia ryukyuensis) BAK57419

Pumilio alignment

Pumilio 2 (Homo) AF315592_1
Pumilio 5 (Arabidopsis) AEE76352
Pumilio6 (Arabidopsis) AEE85129
Pumilio 11 (Arabidopsis) AEE82681
Pumilio (Clonorchis) GAA27968
Pumilio (Dugesia) CAG25892
Pumilio (Schistosoma) CCD59841
Pumilio homolog 1 (Strongylocentrotus) XP_794621
Pumilio isoform A (Drosophila) NP_731314
Pumilio isoform B (Drosophila) NP_731316
Pumilio isoform F (Drosophila) NP_001247002
Pumilio-like (Fasciola) ABX58009
Pumilio-like (Girardia) ABX58014
Pumilio-like (Mesocestoides) ABX58015

Pumilio

PUF-3 (Caenorhabditis) CAB63369
PUF-4 (Caenorhabditis) CCD73134
PUF-5 (Caenorhabditis) CAA90254
PUF-6 (Caenorhabditis) NP_496773
PUF-7 (Caenorhabditis) CCD61586
PUF-8 (Caenorhabditis) CCD66262
PUF-9 (Caenorhabditis) NP_508980
PUF-11 (Caenorhabditis) CCD74185
PUF-12 (Caenorhabditis) NP_496178
Pumilio 2 (Homo) AF315592_1
Pumilio 5 (Arabidopsis) AEE76352
Pumilio6 (Arabidopsis) AEE85129
Pumilio 11 (Arabidopsis) AEE82681
Pumilio (Clonorchis) GAA27968
Pumilio (Dugesia) CAG25892
Pumilio (Schistosoma) CCD59841
Pumilio homolog 1 (Strongylocentrotus) XP_794621
Pumilio isoform A (Drosophila) NP_731314
Pumilio isoform B (Drosophila) NP_731316
Pumilio isoform F (Drosophila) NP_001247002
Pumilio-like (Fasciola) ABX58009
Pumilio-like (Girardia) ABX58014
Pumilio-like (Mesocestoides) ABX58015

DAZ/Boule

HRP1 (Saccharomyces) CAA91142
DAZAP (Dugesia) BAG15905
Boule (Ciona) XP_002124906
Cb-DAZ-1 (Caenorhabditis) BAE93142
DAZ-1 (Caenorhabditis) BAA88577
Boule-like (Macrostomum) 2 AEK69206
Boule (Tribolium) EFA05679
Boule (Aedes) XP_001659950
Boule (Drosophila) ACZ94665
Boule-like (Macrostomum) 1 ADL09421
Boule-like 2 (Macrostomum) AEK69206
Boule (Bos) ACO07307
Boule (Pan) CAG30556
DAZL (Danio) AAH76423
DAZ-like (Rana) AAV30542
DAZ-like protein (Xenopus) AAC41242

Appendix 2 Alignment of Pumilio domains 7-8 as determined by Koziol et al (2008)

Consensus Identity	1	10	17
	F S X W S B Y M X L A X X S G X		
1. Nanos (Danio)	F O P W K D Y M G L A D M I R G M		
2. Nanos1 (Mus)	F S S W N D Y L G L A T L I T R A		
3. Nanos homolog 1 (Schistosoma)	F O T L A R R G C E S P I T C G V		
4. HmN_753700.1	V S V S D S L N S Y A T V T P I S		
5. Nanos-like (Schmidtea)	P F I W S N N V S Y S S N K S N V		
6. Nanos-related protein (Dugesia japonica)	P S L W S N I M C L N S S K S G S		

Appendix 3 Alignment of Nanos N terminal domains corresponding to amino acids 7-34 in Lai et al., 2011)

Consensus Identity	1	10	20	30	39
	F A S N V V E K A V A N A T R X E R C A L I N E V I N G X S S V L V D M M K D				
1. TsM_274800.1	F A S N V M E K A I A N A T P A E R S A L I N E V I T S G P G G V L V D M M K D				
2. EmW_000450400.1	F A S N V M E K A I A N A T P A E R S A L I N E V I N G C P G G V L V D M M K D				
3. EgG_000450400.1	F A S N V M E K A I A N A T P A E R S A L I N E V I N G P G G V L V D M M K D				
4. Smp_155000.1..pep	F A S N V M E K A I A N A V P S E R A V L I E E I L N N I S S S L V D M M K D				
5. Pumilio (Schistoma)	F A S N V M E K A I A N A V P S E R A V L I E E I L N N I S S S L V D M M K D				
6. Smp_155000.2..pep	F A S N V M E K A I A N A V P S E R A V L I E E I L N N I S S S L V D M M K D				
7. HmN_516200.1	F A S N V V E K A V S N S T T Q E R C E L T S E V I L A T D N S V L W S M M K D				
8. HmN_758500.1	F A S N V M E K A I A N A T P S E R A A L I N E V I N G S G G V L V E M M K D				
9. Pumilio 2 (Homo)	F A S N V V E K C V T H A S R T E R A V L I D E V C D G P H S A L Y T M M K D				
10. EgG_000878100.1	F A S N V I E K A V T N A S R K E R C A L I N E V I R E G T S I L W N M M K D				
11. EmW_000878100.1	F A S N V I E K A V T N A S R K E R C A L I N E V I R E D T S I L W N M M K D				
12. Pumilio-like (Mesocostoides)	F A S N V V E K A V T N A N R Q E R C A L I N E V I L				
13. TsM_1161600.1	F A S N V V E K A V T N A S R Q E R C A L I N E V I R E D T S V I W T M M K D				
14. Pumilio Isoform A (Drosophila)	F A S N V V E K C V T H A T R G E R T G L I D E V C T F N D N A L H V M M K D				
15. Pumilio Isoform B (Drosophila)	F A S N V V E K C V T H A T R G E R T G L I D E V C T F N D N A L H V M M K D				
16. Pumilio isoform F (Drosophila)	F A S N V V E K C V T H A T R G E R T G L I D E V C T F N D N A L H V M M K D				
17. Pumilio (Dugesia)	F A S N V V E K A V A H A T R Q E R C A L I N E V I V Q R E S V L V W M M K D				
18. Pumilio-like (Girardia)	F A S N V V E K A V A H A T R Q E R C A L I N E V I T S N - - - - - A I M R N				
19. Smp_180910.2..pep	F A S N V V E K A V A N A S R V E R C S L I N E I I L V T R S S V L V M M M K D				
20. Smp_180910.1..pep	F A S N V V E K A V A N A S R V E R C S L I N E I I L V T R S S V L V M M M K D				
21. Pumilio (Clonorchis)	F A S N V V E K A V A N A T R A E R H S L I N E V I H G N T S I L C M M M K D				
22. Pumilio-like (Fasciola)	F A S N V V E K A V A N A T R A E R C A L I N E V I R E				

