Annelid functional genomics reveal the origins of bilaterian life cycles

Francisco M. Martín-Zamora[1,§], Yan Liang[1,§], Kero Guynes[1], Allan M. Carrillo-Baltodano[1], Billie E. Davies[1], Rory D. Donnellan[1], Yongkai Tan[2], Giacomo Moggioli[1], Océane Seudre[1], Martin Tran[1,3], Kate Mortimer[4], Nicholas M. Luscombe[2], Andreas Hejnol[5,6], Ferdinand Marlétaz[7,8,*], José M. Martín-Durán[1,*]

§ These authors contributed equally.

[1]School of Biological and Behavioural Sciences, Queen Mary University of London, London, United Kingdom
[2]Genomics and Regulatory Systems Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan
[3]Department of Infectious Disease, Imperial College London, London, United Kingdom
[4]Department of Natural Sciences, Amgueddfa Cymru – National Museum Wales, Cardiff, United Kingdom
[5]Department of Biological Sciences, University of Bergen, Bergen, Norway
[6]Institute of Zoology and Evolutionary Research, Friedrich-Schiller-University Jena, Jena, Germany
[7]Molecular Genetics Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan
[8]Department of Genetics, Evolution and Environment, University College London, London, United Kingdom

* Correspondence: José M. Martín-Durán (chema.martin@qmul.ac.uk), Ferdinand Marlétaz (f.marletaz@ucl.ac.uk)

Abstract

Indirect development with an intermediate larva exists in all major animal lineages[1], making larvae central to most scenarios of animal evolution[2-12]. Yet how larvae evolved remains disputed. Here we show that temporal shifts (i.e., heterochronies) in trunk formation underpin the diversification of larvae and bilaterian life cycles. Combining chromosome-scale genome sequencing in the slow-evolving annelid *Owenia fusiformis*[13] with transcriptomic and epigenomic profiling during the life cycles of this and two other annelids, we found that trunk development is deferred to pre-metamorphic stages in the feeding larva of *O. fusiformis*, but starts after gastrulation in the non-feeding larva with gradual metamorphosis of *Capitella teleta* and the direct developing embryo of *Dimorphilus gyrociliatus*. Accordingly, the embryos of *O. fusiformis* develop first into an enlarged anterior domain that forms larval tissues and the adult head. Notably, this also occurs in the so-called "head larvae" of other bilaterians[14,15], with whom *O. fusiformis* larva shows extensive transcriptomic similarities. Together, our findings suggest that the temporal decoupling of head and trunk formation, as maximally observed in "head larvae", allowed larval evolution in Bilateria, thus diverging from prevailing scenarios that propose either co-option[10,11] or innovation[12] of gene regulatory programmes to explain larva and adult origins.

Main text

Many animal embryos develop into an intermediate, often free-living stage termed larva, which later metamorphoses into the sexually competent adult[1,2]. Larvae are morphologically diverse and display an array of lifestyles, from active predation and drastic metamorphoses (e.g., planktotrophic larvae) to maternal nourishment and a more gradual transition into the adult phase (e.g., lecithotrophic larvae)[1,2] (Fig. 1a). Given their broad phylogenetic distribution[2], larvae are central to major scenarios of animal evolution[2-12]. Yet, these dissent on whether larvae are ancestral[2-7] or secondarily evolved[10,11] life stages (Fig. 1b). The "intercalation" hypothesis[10,11] suggests that larval stages were added to animal life cycles multiple times independently, by co-opting genes and genetic programmes originally expressed in the adult. Conversely, the "terminal addition" scenario[2,3,12] considers that the ancestral bilaterians resembled existing larvae, and thereby adults convergently evolved through the parallel evolution of adult-specific genetic programmes[12]. How these proposed mechanisms of co-option and innovation might translate to our current understanding of the genetic control of animal embryogenesis remains, however, unclear, and thus larval origins—and their importance to explain animal evolution—are still contentious.

The trochophore[16] is a widespread larval type generally characterised by an apical sensory organ and a pre-oral locomotive ciliary band[17] that is typically assigned to Annelida and Mollusca but also potentially to related clades within Lophotrochozoa[18] (Fig. 1c). Classically exemplified by those of annelid worms, trochophore larvae are pivotal to the "terminal addition" scenario[2,19], which regards this larval type a vestige of the last common adult ancestor to Protostomia[2,3], or even Bilateria[20] (Fig. 1c). Annelids, however, exhibit diverse life cycles and larval morphologies, including species with direct and indirect development and either planktotrophic or lecithotrophic larvae[21]. Notably, the groups Oweniidae and Magelonidae—which form Oweniida, the sister group to all other annelids[13]—exhibit distinctive planktotrophic larvae (Fig. 1c, d). In particular, the oweniid larva, commonly referred to as "mitraria"[22], has an enlarged pre-oral region and a single bundle of posterior chaetae, as well as a pair of nephridia and a long monociliated ciliary band alike those of phylogenetically distant larvae of echinoderms (e.g., sea urchins) and hemichordates (e.g., acorn worms)[23-25]. Yet, oweniids exhibit many developmental characters considered ancestral to Annelida, and even Lophotrochozoa as a whole[26,27], as well as similarities in larval molecular patterns with other annelid trochophore and bilaterian larvae[25,26,28,29]. Therefore, Annelida, with its diversity of life cycles and larval forms, but generally conserved early embryogenesis and adult body plans, emerges as an excellent model to investigate how larval traits evolve, and thereby formulate and assess hypotheses on the origin of animal life cycles.

Here, we characterise the reference chromosome-scale genome assembly of the oweniid *Owenia fusiformis* Delle Chiaje, 1844 (Fig. 1e) and perform a comprehensive comparative study of the developmental transcriptomes and regulatory genomes underpinning the formation of its planktotrophic larva, as well as the lecithotrophic larva and direct developing embryo of the annelids *Capitella teleta*[30] and *Dimorphilus gyrociliatus*[31], respectively. Diverging from traditional scenarios, our extensive dataset, in comparison with those from

other lophotrochozoan and bilaterian taxa, provides compelling evidence that heterochronic shifts in trunk development contribute to the evolution of life cycles in Bilateria.

O. fusiformis ***has a conservatively evolving genome***

To characterise the transcriptomic and genomic regulatory basis for larval development in Annelida, we first generated a chromosome-scale reference assembly for *O. fusiformis* combining PacBio long-reads, 10x Genomics linked-reads, optical mapping, and Hi-C scaffolding (Supplementary Fig. 1a). Consistent with flow cytometry and *k*-mer estimations, the haploid assembly spans 505.8 Mb (Supplementary Fig. 1b–e), exhibiting 12 chromosome-scale scaffolds that encompass 89.2% of the assembly (Supplementary Fig. 1f, g). Almost half of the assembly consists of repeats (43.02%; largely DNA transposons) acquired steadily during evolution (Extended Data Fig. 1a–c; Supplementary Table 1). Using transcriptomic data from 14 developmental stages and 9 adult tissues (Supplementary Fig. 1a), we annotated 26,966 protein-coding genes and 31,903 transcripts, representing a nearly complete (97.5%) set of metazoan BUSCO genes (Supplementary Fig. 1g). Gene family reconstruction and principal component analysis on gene content across 22 animal genomes nested *O. fusiformis* within other non-annelid lophotrochozoan species (Fig. 1f), supporting that *O. fusiformis* has fewer gene family gains and losses, and retains more ancestral metazoan orthogroups than other annelid taxa (Fig. 1g; Extended Data Fig. 1d–g; Supplementary Tables 2–7). Indeed, we identified an ortholog of *chordin*, a bone morphogenetic protein inhibitor involved in dorsoventral patterning across Bilateria[32] and thought to be lost in annelids[33] (Extended Data Fig. 2a–f), which is asymmetrically expressed around the blastopore lip of the gastrula and antero-ventral oral ectoderm of the larva in *O. fusiformis* (Extended Data Fig. 2g, h). At a high-order genomic organisation, *O. fusiformis* has globally retained the ancestral bilaterian linkage, exhibiting chromosomal fusions that are present in molluscs and even nemerteans (Fig. 1h; Extended Data Fig. 1h, i), and less lineage-specific chromosomal rearrangements than other annelids (Extended Data Fig. 1h). Therefore, *O. fusiformis* genome contains a more complete gene repertoire and ancestral syntenic chromosomal organisation than those reported for other annelid lineages, which together with its phylogenetic position and conserved early embryogenesis[26,27] makes it a key lineage to reconstruct the evolution of Annelida and Lophotrochozoa generally.

Heterochronic shifts in gene regulatory programmes underpin annelid life cycles

To identify transcriptomic changes underpinning distinct life cycles in Annelida, we compared temporal series of embryonic, larval and competent/juvenile transcriptomes of *O. fusiformis* and *C. teleta*, two indirect developers with planktotrophic and lecithotrophic[30] larvae, respectively (Fig 1d), and *D. gyrociliatus*, a direct developer[31,34] (Fig. 2a; Supplementary Fig. 2; Supplementary Tables 13–21). In these three species, gene expression increases with time (Fig. 2b; Supplementary Fig. 3) and transcriptional dynamics are overall similar, especially during early embryogenesis (Fig. 2c, d; Supplementary Table 22). Yet, while *C. teleta* and *D. gyrociliatus* show increasing transcriptomic divergence with each other as they develop into adult stages, the maximal transcriptomic divergence between these annelids and *O. fusiformis* occurs with the planktotrophic mitraria (Fig. 2c, d). Soft clustering of all expressed transcripts during development generated an optimal number of 12 distinct

clusters of temporally co-regulated genes in *O. fusiformis* and *C. teleta*, and 9 clusters in *D. gyrociliatus* (Extended Data Fig. 3a–c; Supplementary Tables 23–26). In all three species, only one cluster (1,457 transcripts, 4.6% in *O. fusiformis*; 2,206 transcripts, 5.54% in *C. teleta*; and 1,847 transcripts, 10.75% in *D. gyrociliatus*) showed a bimodal activation involving early embryonic stages (blastulae in *O. fusiformis* and *C. teleta*; early embryogenesis in *D. gyrociliatus*) and competent/juvenile/adult stages (Extended Data Fig. 3a–c). These clusters are enriched in gene ontology (GO) terms associated with core cellular processes, such as catabolism and cellular transport (in *O. fusiformis*), oxidative phosphorylation (OXPHOS) metabolism (in *C. teleta*) and RNA processing, metabolism, and catabolism (in *D. gyrociliatus*) (Extended Data Fig. 3d, e, Supplementary Fig. 4–11, Supplementary Tables 27–31). Indeed, translation and metabolism predominate in clusters of early developmental phases in the three annelids, while cell communication and signalling, morpho- and organogenesis are enriched in later stages of development (Extended Data Fig. 3d). Thus, regardless of their life cycle, annelids share global transcriptional dynamics during development, yet adults and the planktotrophic larva are the most transcriptionally distinct stages.

Divergence in larval developmental programmes (null expectation; Fig. 2e) or alternatively, changes in expression levels of otherwise conserved larval transcriptomes and temporal shifts (i.e., heterochronies), could explain transcriptomic differences between annelid life cycles (Fig. 2e). To test these, we performed pairwise inter-species comparisons of gene (Extended Data Fig. 4a, b) and transcription factor (Fig. 2f; Supplementary Fig. 3, Supplementary Tables 32, 33) composition among clusters of temporally co-regulated genes. In agreement with transcriptional similarity comparisons (Fig. 2c, d), early, followed by late clusters—pre- and post-larval in indirect developers, respectively (Extended Data Fig. 4c)—are the most conserved phases in the three comparisons when all genes are considered (Extended Data Fig. 4a–d). However, transcription factors that are deployed post-larva in indirect development are consistently shifted to early embryogenesis in direct development (Fig. 2f, dotted rectangles; Extended Data Fig. 4e), including 28 transcription factors common to *O. fusiformis* and *C. teleta* and that are involved in a variety of processes (Fig. 2g; Supplementary Tables 34, 35), from nervous system (e.g., *pax6*[35]) and mesoderm (e.g., *foxF*[29]) development to axial patterning (e.g., *Hox1* and *Hox4*[36]). Notably, the overall expression of these 28 genes is also temporally shifted between indirect developing annelids, with the maximum level of expression occurring earlier in *C. teleta* than in *O. fusiformis* (Fig. 2g). Consistently, 2,583 genes also exhibit temporal shifts between the larvae of *O. fusiformis* and *C. teleta* (Extended Data Fig. 4f). These include 105 transcription factors, but most are enzymes and structural genes (Extended Data Fig. 4f–g, Supplementary Fig. 12, 13, Supplementary Tables 36–42) that likely reflect the different biology of these two larvae. For example, genes involved in chitin biosynthesis are among those expressed early in *O. fusiformis* but late in *C. teleta* (Supplementary Fig. 14, Supplementary Tables 43, 44), in agreement with the conspicuous chaetae of the mitraria larva that are absent in the capitellid larva (Fig. 1d). Conversely, the autophagy pathway, which is involved in yolk catabolism[37], is deployed earlier in the lecithotrophic larva of *C. teleta* (Extended Data Fig. 4i, Supplementary Tables 45, 46).

Therefore, although it is likely a combination of factors that explain transcriptional differences, heterochronic changes—sometimes involving just a few key genes related to a range of genetic programmes—correlate, and might account for, life cycle and larval differences in Annelida.

***Trunk development is delayed to pre-metamorphosis in* O. fusiformis**

Homeodomain transcription factors control many, often conserved developmental processes[38] and are the largest class among the 28 transcription factors exhibiting temporal shifts between annelids with different life cycles (Fig. 2g). Indeed, homeodomain genes are enriched in the competent larva in *O. fusiformis* but are prevalent among genes active from stage 5 larva onwards in *C. teleta* (Extended Data Fig. 4j). Accordingly, *Hox* genes—a conserved family involved in anterior-posterior trunk regionalisation in Bilateria[39]—are among the most upregulated genes in the competent mitraria larva (Extended Data Fig. 5b). Similar to *C. teleta*[36] and *D. gyrociliatus*[34] (although the latter lacks *Lox2* and *Post1*), *O. fusiformis* has a conserved complement of 11 *Hox* orthologues arranged as a compact, ordered cluster in chromosome 1, except for *Post1*, which is located downstream on that same chromosome (Extended Data Fig. 5a, c, d; Supplementary Table 47). However, while *C. teleta* and *D. gyrociliatus* deploy *Hox* genes during or soon after gastrulation[34,36] (Extended Data Fig. 5e), *O. fusiformis* does not express *Hox* genes during embryogenesis to pattern the larval body (Fig. 3a; Extended Data Fig. 5e, f). Instead, *Hox* genes are expressed in the trunk rudiment during larval growth, already in an anterior–posterior staggered pattern that is retained in the juvenile after metamorphosis (Fig. 3a–c; Extended Data Fig. 5f). This late activation of *Hox* genes is not unique to *O. fusiformis*, but also occurs in the planktotrophic trochophore of the echiuran annelid *Urechis unicinctus* (Extended Data Fig. 5e; Supplementary Table 48). Therefore, the spatially collinear *Hox* code along the trunk is established at different developmental stages depending on the life cycle mode in Annelida.

To determine whether the different timings of trunk patterning is limited to *Hox* genes or involve broader genetic programmes associated with trunk development in annelids, we used tissue-specific adult transcriptomes to define a set of 1,655 anterior and 407 posterior/trunk genes in *O. fusiformis* (Extended Data Fig. 6a–c; Supplementary Tables 49–53). Anterior genes are significantly more expressed during embryogenesis (Fig. 3d; Extended Data Fig. 6d, e). On the contrary, posterior/trunk genes are upregulated at the mitraria stage and significantly outweigh the expression dynamics of anterior genes from that stage on (Fig. 3d; Extended Data Fig. 6d, f). Indeed, anterior, trunk, and posterior genes for which there is spatially resolved expression data follow different temporal dynamics in *O. fusiformis*, *C. teleta*, and *D. gyrociliatus* (Extended Data Fig. 6g–l; Supplementary Tables 54–56). In *O. fusiformis*, trunk (e.g., *nkx6*, *msx* paralogs[28]) and posterior genes (e.g., *evx*[29], *wnt1*[27]) concentrate in a small ventral area and around the anal opening of the larva[27-29] and increase in spatial range and expression level as the trunk forms. On the other hand, anterior genes (e.g., *foxQ2* genes[40], *gsc*[29], *otx*[29]) pattern most of the mitraria and their expression remain fairly stable during development (Fig. 3d; Extended Data Fig. 6g, h). In contrast, posterior and anterior genes follow similar dynamics in *C. teleta* (Extended Data Fig. 6i, j) and trunk

genes upregulate already post-gastrula in both *C. teleta* and *D. gyrociliatus* (Fig. 3e; Extended Data Fig. 6i–l). Therefore, trunk development, which initially occurs from lateral growth of the trunk rudiment[22,30,41], is deferred to pre-metamorphic stages in planktotrophic annelid trochophores[14,42] compared to annelids with lecithotrophic larvae[36,43] and direct developers[34].

Chromatin dynamics support heterochronic shifts in Hox *regulation between annelid larvae*

To investigate the genomic regulatory basis for the temporal shift in trunk development between annelid larvae, we profiled open chromatin regions with ATAC-seq at five equivalent developmental stages based on cross-species minimal transcriptomic distance in *O. fusiformis* and *C. teleta* (Fig. 4a; Supplementary Fig. 15; Supplementary Tables 57–60). In total, we identified 63,726 and 44,368 consensus regulatory regions in *O. fusiformis* and *C. teleta*, respectively, mostly abundant within gene bodies (69.69% and 51.53%, respectively; from transcription start to end sites) rather than in promoters (12.92% and 20.53%) and distal intergenic regions (17.38% and 27.93%) (Extended Data Fig. 7a, b; Supplementary Fig. 16). The largest changes in peak accessibility occur in the mitraria (in *O. fusiformis*) and stage 5 larva (in *C. teleta*) (Supplementary Fig. 16; Supplementary Tables 61–66), with a general increase in promoter peaks (in *O. fusiformis*) and distant intergenic regulatory elements (in both species) during development (Fig. 4b). Soft clustering revealed, however, that most regulatory regions act before the start of trunk formation in *O. fusiformis* (clusters 1 to 9; 47,652 peaks; 76. 72%), while the number of accessible regions with a maximum of accessibility before (clusters 1 to 7; 23,678 peaks; 51.45%) and after (clusters 8 to 12; 22,341; 48.56%) the onset of trunk development (stage 4tt larva) are comparable in *C. teleta* (Fig. 4c; Supplementary Tables 67–69). Moreover, regulation of genes involved in morpho- and organogenesis, as well as neurogenesis, concentrates in late clusters in *O. fusiformis*, but unfolds more continuously in *C. teleta* (Supplementary Fig. 17–21). Therefore, different dynamics of temporally co-regulated accessible chromatin regions occur during development and larva formation in *O. fusiformis* and *C. teleta*.

Divergence in the regulatory programmes controlling larva development (null expectation), or alternatively, temporal changes in the deployment of common trunk regulatory programmes might explain the heterochronic shifts in trunk development in annelid larvae (Fig. 4d). To assess these scenarios, we predicted transcription factor-binding motifs on ATAC-seq peaks in *O. fusiformis* and *C. teleta*, which we clustered into 95 and 91 consensus motifs, respectively (Supplementary Fig. 22; Supplementary Tables 72–74). Among these, 51 are common to both species, including 33 that are robustly assigned to a known transcription factor class based on sequence similarity to human motif archetypes (Supplementary Tables 74–76). Motif accessibility and binding—as proxies of regulatory activity[44]—are well self- and cross-species correlated in both species (Extended Data Fig. 7a–c; Supplementary Tables 77–80) and reveal clusters of co-regulatory activity during *O. fusiformis* and *C. teleta* development (Supplementary Fig. 23, 24). The comparison of accessibility and binding dynamics of the 33 common annotated motifs between *O. fusiformis* and *C. teleta* show a temporal shift of regulatory motifs acting at the mitraria and competent larva to the stage 4tt

larva of *C. teleta* (Fig. 4e, dotted rectangle; Supplementary Fig. 25–27). Seven motifs (21%) followed this pattern (Extended Data Fig. 7f), including one with high similarity to the human HOX/CDX/EVX motif archetype (Fig. 4f; Supplementary Fig. 27) and that is among the most upregulated and overrepresented motifs at the competent stage in *O. fusiformis* (Fig. 4g; Extended Data Fig. 7g; Supplementary Fig. 27). Indeed, motif binding in *Hox* regulatory elements support a change of global regulation of the *Hox* cluster at the competent and stage 4tt in *O. fusiformis* and *C. teleta*, respectively (Fig. 4h; Supplementary Fig. 29), largely mirroring the transcriptional onset of these genes and the start of trunk development in the two species[36] (Extended Data Fig. 5e). Motifs assigned to NKX and GATA factors, as well as two annelid-specific motifs of unclear assignment, are the most abundant bound motifs in the *Hox* cluster in both species (Fig. 4i; Supplementary Tables 81, 82). Notably, *nkx* and *gata* genes are expressed in the juvenile rudiment[28] and along the annelid trunk[28,45] in *O. fusiformis* and *C. teleta*, and the putative GATA motif follows similar binding dynamics in these species than the HOX/CDX/EVX motif archetype (Extended Data Fig. 7c; Supplementary Fig. 27). However, only 39 one-to-one orthologs with bound HOX/CDX/EVX motifs are common to *O. fusiformis* and *C. teleta* at the time of maximum motif binding, albeit those include *Hox* themselves and a reduced set of transcription factors—some involved in trunk development, such as *msx*[28]—and axial polarity genes (Fig. 4j; Supplementary Tables 83, 84). Together, our epigenomic data reinforces that different regulatory dynamics of the *Hox* cluster—possibly by a reduced common set of upstream regulators—result in temporal shifts of the binding activity of *Hox* genes and expression of trunk patterning genes, likely accounting for the developmental and morphological differences between planktotrophic and lecithotrophic annelid larvae.

The larva of **Owenia** *does not rely on novel genes*

Novel genes account for a significant proportion of some trochophore and other larval transcriptomes[7,46], as they might be associated with the development of larval-specific characters (e.g., ciliary bands)[7,46,47]. Therefore, recruitment of novel genes at larval stages could also explain transcriptomic—and morphological—differences at this stage (Fig. 2c, d; Extended Data Fig. 4a, b). To define the contribution of novel genes at each developmental stage in annelids with different life cycles, we classified all predicted transcripts in seven different phylostrata according to their time of origin (Extended Data Fig. 8a; Supplementary Tables 85–87). Genes of metazoan and pre-metazoan origin (phylostratum 1) tend to peak, dominate and be enriched at earlier developmental stages in the three species (Fig. 5a; Extended Data 8h–j), whereas younger genes are more highly expressed in competent and juvenile stages (Fig. 5a, Extended Data Fig. 8b–g). Notably, species-specific genes are more expressed and enriched in the juvenile stages of *O. fusiformis* and *D. gyrociliatus*, but in the blastula and gastrula of *C. teleta* (and to some extent also at the blastula stage in *O. fusiformis*; Fig. 5a). Indeed, species-specific genes follow lineage-specific dynamics (Supplementary Fig. 30). At larval stages, however, species-specific genes are only enriched in *C. teleta*, while the mitraria shows enrichment of genes of metazoan and annelid origin (Extended Data Fig. 8h, i). Therefore, genes of different evolutionary origins contribute to the development of annelid larvae, suggesting that the increased use of novel genes in some

lophotrochozoan larvae[7,46] might be due to lineage-specific traits, such as the shell primordium of molluscan trochophores[7] and perhaps even ciliary bands with multiciliated cells[48], which are absent in oweniid larvae.

Bilaterian planktotrophic larvae share maximal transcriptional similarity

To assess whether similar transcriptional dynamics to those found in annelids are observed across different life cycles in Metazoa, we extended our comparative transcriptomic approach to nine other animal lineages using single copy one-to-one orthologs and reduced sets of either conserved cross-species single copy orthologs or one-to-one transcription factors (Extended Data Fig. 9a, b; Extended Data Fig. 10a; Supplementary Tables 88–90). In relative terms, global transcriptional dynamics between *O. fusiformis* and other major animal groups tend to be more dissimilar at early development, increasing in similarity as embryos proceed towards juvenile and adult stages (Fig. 5b; Extended Data Fig. 9a, b). The exception is the direct developer *Danio rerio*, for which the mitraria larva is the most dissimilar stage (Fig. 5b), as also occurs when comparing *O. fusiformis* with the direct developing annelid *D. gyrociliatus* (Fig. 2d). However, *O. fusiformis* shares maximal transcriptomic similarities during larval phases with bilaterian species with planktotrophic ciliated larvae and even the planula stages of the cnidarians *Nematostella vectensis* and *Clytia hemisphaerica* (Fig. 5b; Extended Data Fig. 9a–e; Extended Data Fig. 10a). The expression of genes involved in core cellular processes, such as transcription, translation, OXPHOS metabolism, and the assembly of cellular components and complexes, directly contribute to these similarities, likely reflecting common structural and ecological needs of metazoan larvae (Extended Data Fig. 9f, g; Supplementary Tables 91, 92). Yet, regulatory programmes as inferred from similarities in transcription factor expression levels are also maximally similar between those species at larval phases (Extended Data Fig. 9a, b, e). Together, these findings suggest that adult development is generally more similar[10] than early embryogenesis across major animal lineages, maybe because adult bodies operate on many similar cell biological functions. Additionally, these results also reveal unexpected genome-wide transcriptional—and potentially regulative—similarities between phylogenetically distant animal larvae.

Discussion

In this study, we report the chromosome-scale genome assembly of the oweniid *O. fusiformis* and extensive gene expression and genomic regulatory profiling during embryonic and larval development in three annelids with distinct life cycles. Together, these datasets provide an unprecedented perspective on life cycle evolution in Bilateria. We show that the planktotrophic larva of *O. fusiformis* defers trunk differentiation to late pre-metamorphic stages and largely develops from anterior ectodermal domains (Fig. 5c). This occurs in other feeding annelid larvae[42] (Extended Data Fig. 5f), and likely in Chaetopteriformia[49-52] too, and thus the late differentiation of the adult trunk might be an ancestral trait to Annelida (Fig. 5e). Notably, delaying trunk development to post-larval stages also occurs in phylogenetically distant clades within Lophotrochozoa (e.g., nemerteans[53] and phoronids[54]), Ecdysozoa (pancrustaceans[15] and pycnogonids[55]), and Deuterostomia (e.g., echinoderms[56,57] and hemichordates[58]), whose larvae are generally referred to as “head larvae”[14,15]. By contrast,

non-feeding larvae[36,59] and direct developers[34] in both Annelida and other bilaterian taxa[60,61] start to pattern their trunks with or straight after the onset of anterior/head patterning (Fig. 5d, e; Extended Data Fig. 10b), which always takes place before gastrulation in bilaterians[62,63]. Therefore, heterochronic shifts in trunk development correlate with, and possibly account for, the evolution of different life cycles in animals. This mechanism thus differs from predictions that propose either co-option of adult genes into larval-specific gene regulatory programmes[10,11] (in line with the "intercalation" scenario) or the independent evolution of adult gene regulatory modules[2,3,19] (as it would be expected in the "terminal addition" scenario) as major drivers of larva and adult evolution, respectively.

The post-embryonic activation of genetic programmes involved in trunk differentiation observed in bilaterian "head larvae" could be lineage-specific innovations associated with the evolution of maximal indirect development[14,15,24,54]. Consequently, "head larvae" evolved convergently by repeatedly delaying trunk differentiation and *Hox* patterning (Fig. 5f). The observed similarities in larval molecular patterns[6,54,58,64] thus reflect ancient gene regulatory modules that were independently co-opted to develop analogous cell types and larval organs (e.g., apical organs and ciliary bands). In contrast, an ancestral post-embryonic onset of trunk differentiation and *Hox* expression might be the most parsimonious state for Bilateria, based on our current understanding of the timing of trunk-associated expression of *Hox* genes in major bilaterian clades and Cnidaria (Extended Data Figure 10c; Supplementary Fig. 31; Supplementary Tables 93, 94). This ancestral temporal decoupling between head and trunk genetic programmes could thus have facilitated the evolution of larvae, which would then originally share the use of anterior genetic modules for their development (Fig. 5f). Regardless of the scenario, however, and despite the limitations that whole-embryo transcriptomic and epigenomic datasets without spatial resolution have, our study highlights the importance of heterochronic changes in the deployment of ancient genetic programmes for the diversification of bilaterian life cycles, uncovering a reduced set of candidate genes (Fig. 2g) and regulatory motifs (Extended Data Fig. 7f) that might influence life cycle differences in Annelida, and perhaps even Bilateria. In the future, comparative functional studies of these and other genes are needed to thoroughly dissect the regulatory principles underlying head and trunk development and decipher how temporal changes in gene expression and regulation have shaped the evolution of larval and adult forms in Bilateria.

Main References

1 Hall, B. K. & Wake, M. H. in *The Origin and Evolution of Larval Forms* (eds Brian K. Hall & Marvalee H. Wake) Ch. 1, 1-19 (Academic Press, 1999).

2 Nielsen, C. Origin and evolution of animal life cycles. *Biol Rev* **73**, 125-155 (1998).

3 Nielsen, C. Animal phylogeny in the light of the trochaea theory. *Biological Journal of the Linnean Society* **25**, 243-299 (2008).

4 Garstang, W. The origin and evolution of larval forms. 77-98 (British Association for the Advancement of Science, London, 1928).

5 Jägersten, G. *Evolution of the Metazoan Life Cycle*. (Academic Press, 1972).

6 Marlow, H. *et al.* Larval body patterning and apical organs are conserved in animal evolution. *BMC Biol* **12**, 7 (2014).

7 Wang, J. *et al.* Evolutionary transcriptomics of metazoan biphasic life cycle supports a single intercalation origin of metazoan larvae. *Nat Ecol Evol* **4**, 725-736 (2020).
8 Salvini-Plawen, L. v. On the origin and evolution of the lower Metazoa. *Z Zool Syst Evol Forsch* **16**, 40-87 (1978).
9 Hazsprunar, G., Salvini-Plawen, L. v. & Rieger, R. M. Larval Planktotrophy — A Primitive Trait in the Bilateria? *Acta Zoologica* **76**, 141-154 (1995).
10 Raff, R. A. Origins of the other metazoan body plans: the evolution of larval forms. *Philos Trans R Soc Lond B Biol Sci* **363**, 1473-1479 (2008).
11 Sly, B. J., Snoke, M. S. & Raff, R. A. Who came first - larvae or adults? Origins of bilaterian metazoan larvae. *Int J Dev Biol* **47**, 623-632 (2003).
12 Davidson, E. H., Peterson, K. J. & Cameron, R. A. Origin of bilaterian body plans: evolution of developmental regulatory mechanisms. *Science* **270**, 1319-1325 (1995).
13 Rouse, G. W., Pleijel, F. & Tilic, E. *Annelida*. (Oxford University Press, 2022).
14 Lacalli, T. C. Protochordate body plan and the evolutionary role of larvae: old controversies resolved? *Can. J. Zool.* **83**, 216-224 (2005).
15 Strathman, R. Multiple origins of feeding head larvae by the Early Cambrian. *Can. J. Zool.* **98**, 761-776 (2020).
16 Hatschek, B. Studien über Entwicklungsgeschichte der Anneliden. Ein Beitrag zur Morphologie der Bilaterien. *Arbeiten aus dem Zoologischen Institute der Universität Wien und der Zoologischen Station in Triest* **1**, 277-404 (1878).
17 Rouse, G. W. Trochophore concepts: ciliary bands and the evolution of larvae in spiralian Metazoa. *Biological Journal of the Linnean Society* **66**, 411-464 (2008).
18 Marlétaz, F., Peijnenburg, K., Goto, T., Satoh, N. & Rokhsar, D. S. A New Spiralian Phylogeny Places the Enigmatic Arrow Worms among Gnathiferans. *Curr Biol* **29**, 312-318 e313 (2019).
19 Nielsen, C. *Animal Evolution. Interrelationships of the Living Phyla*. 3rd edn, (Oxford University Press, 2012).
20 Nielsen, C. Evolution of deuterostomy - and origin of the chordates. *Biol Rev Camb Philos Soc* **92**, 316-325 (2017).
21 Rouse, G. W. Polychaetes have evolved feeding larvae numerous times. *Bulletin of Marine Science* **67**, 391-409 (2000).
22 Wilson, D. P. IV. On the Mitraria Larva of *Owenia fusiformis* Delle Chiaje. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* **221**, 231-334 (1932).
23 Nielsen, C. Trochophora larvae: cell-lineages, ciliary bands and body regions. 2. Other groups and general discussion. *J Exp Zool B Mol Dev Evol* **304**, 401-447 (2005).
24 Smart, T. I. & Von Dassow, G. Unusual development of the mitraria larva in the polychaete *Owenia collaris*. *Biol Bull* **217**, 253-268 (2009).
25 Gasiorowski, L. *et al.* Molecular evidence for a single origin of ultrafiltration-based excretory organs. *Curr Biol* **31**, 3629-3638 e3622 (2021).
26 Carrillo-Baltodano, A. M., Seudre, O., Guynes, K. & Martin-Duran, J. M. Early embryogenesis and organogenesis in the annelid *Owenia fusiformis*. *Evodevo* **12**, 5 (2021).

27 Seudre, O., Carrillo-Baltodano, A. M., Liang, Y. & Martín-Durán, J. M. ERK1/2 is an ancestral organising signal in spiral cleavage. *Nat Commun* **13**, 2286 (2022).
28 Martin-Duran, J. M. *et al.* Convergent evolution of bilaterian nerve cords. *Nature* **553**, 45-50 (2018).
29 Martin-Duran, J. M., Passamaneck, Y. J., Martindale, M. Q. & Hejnol, A. The developmental basis for the recurrent evolution of deuterostomy and protostomy. *Nat Ecol Evol* **1**, 5 (2016).
30 Seaver, E. C., Thamm, K. & Hill, S. D. Growth patterns during segmentation in the two polychaete annelids, *Capitella* sp. I and *Hydroides elegans*: comparisons at distinct life history stages. *Evol Dev* **7**, 312-326 (2005).
31 Kerbl, A., Fofanova, E. G., Mayorova, T. D., Voronezhskaya, E. E. & Worsaae, K. Comparison of neuromuscular development in two dinophilid species (Annelida) suggests progenetic origin of *Dinophilus gyrociliatus*. *Front Zool* **13**, 49 (2016).
32 Niehrs, C. On growth and form: a Cartesian coordinate system of Wnt and BMP signaling specifies bilaterian body axes. *Development* **137**, 845-857 (2010).
33 Tan, S., Huan, P. & Liu, B. Molluscan dorsal-ventral patterning relying on BMP2/4 and Chordin provides insights into spiralian development and evolution. *Mol Biol Evol* (2021).
34 Martin-Duran, J. M. *et al.* Conservative route to genome compaction in a miniature annelid. *Nat Ecol Evol* **5**, 231-242 (2021).
35 Klann, M. & Seaver, E. C. Functional role of *pax6* during eye and nervous system development in the annelid *Capitella teleta*. *Dev Biol* **456**, 86-103 (2019).
36 Fröbius, A. C., Matus, D. Q. & Seaver, E. C. Genomic organization and expression demonstrate spatial and temporal Hox gene colinearity in the lophotrochozoan *Capitella* sp. I. *PLoS One* **3**, e4004 (2008).
37 Kuhn, H., Sopko, R., Coughlin, M., Perrimon, N. & Mitchison, T. The Atg1-Tor pathway regulates yolk catabolism in *Drosophila* embryos. *Development* **142**, 3869-3878 (2015).
38 Burglin, T. R. & Affolter, M. Homeodomain proteins: an update. *Chromosoma* **125**, 497-521 (2016).
39 Pearson, J. C., Lemons, D. & McGinnis, W. Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* **6**, 893-904, doi:10.1038/nrg1726 (2005).
40 Seudre, O. *et al.* The *Fox* gene repertoire in the annelid *Owenia fusiformis* reveals multiple expansions of the *foxQ2* class in Spiralia. *bioRxiv*, 2022.2003.2002.482670 (2022).
41 Prud'homme, B. *et al.* Arthropod-like expression patterns of *engrailed* and *wingless* in the annelid *Platynereis dumerilii* suggest a role in segment formation. *Curr Biol* **13**, 1876-1881 (2003).
42 Woltereck, R. Trochophora-Studien I. Histogie der larve und die Entstehung des Annelids bei den *Polygordius*-Arten der Nordsee. *Zoologica* **13** (1902).
43 Hui, J. H. *et al.* Extensive chordate and annelid macrosynteny reveals ancestral homeobox gene organization. *Mol Biol Evol* **29**, 157-165 (2012).
44 Bentsen, M. *et al.* ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* **11**, 4267 (2020).

45 Boyle, M. J. & Seaver, E. C. Developmental expression of *foxA* and *gata* genes during gut formation in the polychaete annelid, *Capitella* sp. I. *Evol Dev* **10**, 89-105 (2008).
46 Wu, L., Ferger, K. E. & Lambert, J. D. Gene Expression Does Not Support the Developmental Hourglass Model in Three Animals with Spiralian Development. *Mol Biol Evol* **36**, 1373-1383 (2019).
47 Wu, L. *et al*. Genes with spiralian-specific protein motifs are expressed in spiralian ciliary bands. *Nat Commun* **11**, 4171 (2020).
48 Marinković, M., Berger, J. & Jékely, G. Neuronal coordination of motile cilia in locomotion and feeding. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190165 (2020).
49 Irvine, S. Q. & Martindale, M. Q. Expression patterns of anterior Hox genes in the polychaete *Chaetopterus*: correlation with morphological boundaries. *Dev Biol* **217**, 333-351 (2000).
50 Peterson, K. J., Irvine, S. Q., Cameron, R. A. & Davidson, E. H. Quantitative assessment of *Hox* complex expression in the indirect development of the polychaete annelid *Chaetopterus* sp. *Proc Natl Acad Sci U S A* **97**, 4487-4492 (2000).
51 Helm, C., Schwarze, G. & Beckers, P. Loss of complexity from larval towards adult nervous systems in Chaetopteridae (Chaetopteriformia, Annelida) unveils evolutionary patterns in Annelida. *Organisms Diversity & Evolution* (2022).
52 Irvine, S. Q., Chaga, O. & Martindale, M. Q. Larval ontogenetic stages of *Chaetopterus*: developmental heterochrony in the evolution of chaetopterid polychaetes. *Biol Bull* **197**, 319-331 (1999).
53 Hiebert, L. S. & Maslakova, S. A. *Hox* genes pattern the anterior-posterior axis of the juvenile but not the larva in a maximally indirect developing invertebrate, *Micrura alaskensis* (Nemertea). *BMC Biol* **13**, 23 (2015).
54 Gasiorowski, L. & Hejnol, A. Hox gene expression during development of the phoronid *Phoronopsis harmeri*. *Evodevo* **11**, 2 (2020).
55 Brenneis, G., Bogomolova, E. V., Arango, C. P. & Krapp, F. From egg to "no-body": an overview and revision of developmental pathways in the ancient arthropod lineage Pycnogonida. *Front Zool* **14**, 6 (2017).
56 Arenas-Mena, C., Cameron, A. R. & Davidson, E. H. Spatial expression of *Hox* cluster genes in the ontogeny of a sea urchin. *Development* **127**, 4631-4643 (2000).
57 Arenas-Mena, C., Martinez, P., Cameron, R. A. & Davidson, E. H. Expression of the *Hox* gene complex in the indirect development of a sea urchin. *Proc Natl Acad Sci U S A* **95**, 13062-13067 (1998).
58 Gonzalez, P., Uhlinger, K. R. & Lowe, C. J. The Adult Body Plan of Indirect Developing Hemichordates Develops by Adding a Hox-Patterned Trunk to an Anterior Larval Territory. *Curr Biol* **27**, 87-95 (2017).
59 Kulakova, M. *et al.* Hox gene expression in larval development of the polychaetes *Nereis virens* and *Platynereis dumerilii* (Annelida, Lophotrochozoa). *Dev Genes Evol* **217**, 39-54 (2007).
60 Aronowicz, J. & Lowe, C. J. *Hox* gene expression in the hemichordate *Saccoglossus kowalevskii* and the evolution of deuterostome nervous systems. *Integr Comp Biol* **46**, 890-901 (2006).

61 Pascual-Anaya, J. *et al*. Broken colinearity of the amphioxus *Hox* cluster. *Evodevo* **3**, 28 (2012).

62 Carrillo-Baltodano, A. M. & Meyer, N. P. Decoupling brain from nerve cord development in the annelid *Capitella teleta*: Insights into the evolution of nervous systems. *Dev Biol* **431**, 134-144 (2017).

63 Sur, A., Magie, C. R., Seaver, E. C. & Meyer, N. P. Spatiotemporal regulation of nervous system development in the annelid *Capitella teleta*. *Evodevo* **8**, 13 (2017).

64 Hiebert, L. S. & Maslakova, S. A. Expression of *Hox*, *Cdx*, and *Six3/6* genes in the hoplonemertean *Pantinonemertes californiensis* offers insight into the evolution of maximally indirect development in the phylum Nemertea. *Evodevo* **6**, 26 (2015).

Figure Legends

Figure 1 | *Owenia fusiformis* has a unique larva and a conservatively evolving genome. **a**, Bilaterian life cycles and larval forms. Within indirect developers, lecithotrophic larvae rely on yolk nutrients during dispersal, while planktotrophic larvae feed on plankton. **b**, Major scenarios for the evolution of bilaterian life cycles and larval forms. The "intercalation" scenario deems bilaterian adults ancestral and larvae secondary specialisations that evolved independently in certain lineages. Opposed to this, the "terminal addition" scenario considers that bilaterian larvae are ancestral and that adult forms evolved secondarily. **c**, A trochophore larval type has been proposed to be ancestral to Spiralia or even Protostomia (Ecdysozoa + Spiralia) and give rise to the diversity of larval forms found in lophotrochozoan taxa. **d,** The larval forms of oweniids and magelonids are unlike other annelid larvae. Differential interface contrast (DIC) images and z-stack confocal laser scanning views of a stage 5 metatrochophore of *C. teleta*, a *Magelona* spp. Larva, and a *O. fusiformis* mitraria stained for DAPI and acetylated α-tubulin. **e**, Image of an adult of *O. fusiformis*. **f,** Principal component analysis of metazoan gene complements demonstrates that *Owenia* clusters with other slow-evolving lineages. See Extended Data Fig. 1g for a fully labelled graph. **g,** Percentage of retained pre-metazoan and metazoan orthogroups per species. Dotted vertical line represents the value for *O. fusiformis*. **h,** Karyotypic correspondence between *O. fusiformis* and *Pecten maximus*, which exemplifies the ancestral spiralian chromosome complement. Each colour represents an ancestral bilaterian linkage group. Schematic drawings are not to scale. At: apical tuft; an: anus; ch: chaetae; he: head; mo: mouth; pt: prototroch; tt: telotroch.

Figure 2 | Heterochronic shifts in gene regulatory programmes underpin annelid life cycle diversification. a, Experimental design of the comparative developmental RNA-seq time courses. Orange circles highlight stages of *O. fusiformis*, *C. teleta*, and *D. gyrociliatus* development sampled for bulk RNA-seq. **b,** Number of expressed genes during annelid development. See Supplementary Fig. 3c for fully labelled plots. **c, d**, Heatmaps of pairwise transcriptomic Jensen–Shannon Divergence (JSD) between *O. fusiformis* and *C. teleta* (**c**), and between *D. gyrociliatus* and either *O. fusiformis* (**d**, left) or *C. teleta* (**d**, right). Relative JSD of the *C. teleta* or *O. fusiformis* stages of minimal divergence to each corresponding stage is shown on top. **e,** Schematics of the orthogroup overlap strategy. Chronologically ordered clusters of co-expressed genes from two distinct species with similar developmental gene expression dynamics would render a similarity diagonal (null expectation). Alternative scenarios include either lack of overlap in divergence or innovation situations, or timing differences when co-opted or temporally shifted genetic programmes are present. **f,** Similarity heatmaps showcasing the orthogroup overlap between the transcription factors contained in clusters of co-regulated genes obtained by soft *k*-means clustering, between all three studied annelid taxa. Time points associated to key clusters are shown for all three species. Dotted black lines encompass the sharp timing expression differences of a significant number of transcription factors shifted from post-larval expression in indirect developers to early embryogenesis in *D. gyrociliatus* can be observed. **g,** Heatmaps of gene-wise expression (bottom) and average expression dynamics (top) during *O. fusiformis*, *C. teleta*, and *D. gyrociliatus* development of the set of 28 single copy ortholog transcription factors shifted

from late expression in both *O. fusiformis* and *C. teleta* to early expression in *D. gyrociliatus*. Up to 43 % (12) of these belong to the homeodomain class of transcription factors.

Figure 3 | Trunk development is delayed to pre-metamorphosis in *O. fusiformis*. **a,** Whole mount *in situ* hybridisation of *Hox* genes at the mitraria, pre-competent, and juvenile stages of *O. fusiformis*. Only *Hox3* appears to be expressed at the mitraria stage (white arrow). *Hox* genes show spatial collinearity along the anterior-posterior axis at the developing trunk (white arrows in the pre-competent larva) and juvenile. Dotted lines in the competent larva panels indicate background from the gut content. Black arrowheads in the juvenile panels indicate head to trunk boundary. cs: chaetal sack; mg: mid gut; mo: mouth. **b, c** Schematic representations of the expression of *Hox* genes in the trunk rudiment of the competent larva (**b**) and juvenile trunk (**c**). A: anterior; P: posterior. Drawings are not to scale, and schematic expression domains are approximate. **d,** Expression dynamics of anterior and posterior and trunk genes expressed in corresponding adult tissues during *O. fusiformis* development. *P*-values were derived from two-tailed Student's *t*-tests and adjusted with the Bonferroni method for multiple testing correction. n.s.: not significant. **e,** Expression dynamics of *in situ* hybridisation-validated trunk markers throughout *O. fusiformis* (top) and *C. teleta* (bottom) development. In contrast with the gradual onset from the stage 4tt larva stage in *C. teleta*, trunk markers deploy markedly but only after the mitraria larva stage in *O. fusiformis*.

Figure 4 | Chromatin dynamics support heterochronic shifts in *Hox* regulation between annelid larvae. **a,** Experimental design of the comparative developmental ATAC-seq time courses, highlighting with orange circles the stages of *O. fusiformis* and *C. teleta* development sampled for bulk ATAC-seq. *Hox* expression onset, trunk development, and larval trunk domains are indicated in blue. **b,** Stacked bar plots showing the proportion of called peaks per developmental stage classified by genomic feature for *O. fusiformis* (left) and *C. teleta* (right). **c,** Heatmap of normalised peak accessibility of the soft clustered consensus ATAC-seq peak sets of *O. fusiformis* (top) and *C. teleta* (bottom). **d,** Schematics of the cross-species ATAC-seq comparison strategy. Our null expectation entails lack of similarity between regulatory dynamics at larval stages. Alternatively, heterochronic shifts of regulatory programmes could support the heterochronic shifts in gene expression observed between *O. fusiformis* and *C. teleta*. **e,** Correlation matrices of motif archetype accessibility and transcription factor binding score (TFBS). Dotted black lines highlight the high TFBS correlation and subsequent heterochronic shift between the mitraria and competent larva of *O. fusiformis* and the stage 4tt larva of *C. teleta*. **f,** TFBS dynamics for the HOX/CDX/EVX motif archetype during *O. fusiformis* (purple) and *C. teleta* (blue) development. Sequence logo of the annelid archetype (top) shows the substantial similarity to the human homolog (bottom). **g,** Volcano plot of differential transcription factor binding between the mitraria larva and the competent larva stages of *O. fusiformis*. **h,** Average TFBS dynamics of all motifs in the peaks of the *Hox* genes clusters. **i,** Most abundant bound motifs in peaks of the *Hox* genes clusters. **j,** Downstream regulated genes by transcription factors bound to the HOX/CDX/EVX motif archetype.

Figure 5 | Gene novelties and the evolution of life cycles in Annelida and Bilateria. **a**, Expression dynamics of each phylostratum by developmental stage in all three annelids, calculated from the 75% percentile of a quantile-normalised matrix of gene expression levels. **b,** Heatmaps of pairwise normalised Jensen-Shannon divergence (JSD) between *O. fusiformis* and *C. gigas*, *S. purpuratus*, *D. rerio* and *N. vectensis.* Asterisks indicate the stages of minimal JSD of each species to the larval phase of *O. fusiformis*. Larval phases are highlighted in green. On top, relative JSD from stages of minimal divergence to each *O. fusiformis* developmental stage. Confidence intervals represent the standard deviation from 250 bootstrap resamplings of the single copy ortholog set. **c**, Schematic drawing of the life cycle and patterning events in a bilaterian with indirect development with a feeding larva like the annelid *O. fusiformis*. Embryogenesis results in a larva with mostly anterior ectodermal fates from which larval organs and the adult head forms. The onset of trunk differentiation and *Hox* gene expression occurs later, with larval growth pre-metamorphosis. **d**, Schematic drawings of the three main types of life cycles and the timing of *Hox* gene expression in bilaterians. Compared to indirect development with feeding larvae, lineages with non-feeding larvae and direct development pre-displace (i.e., initiate earlier) trunk differentiation and *Hox* gene expression. Larval organs are reduced in non-feeding larvae and absent in direct development. **e**, Proposed evolutionary scenario for larval and life cycle evolution in Annelida. Post-embryonic trunk patterning is likely an ancestral condition (1) with the convergent pre-displacement of trunk differentiation to embryogenesis (2) concurring with the evolution of indirect development with feeding larva and direct development. Drawings are not to scale. **f,** Alternative scenarios for the evolution of "head larvae" in Bilateria.

Methods

Adult culture, spawning and in vitro fertilisation

Sexually mature *O. fusiformis* adults were collected from subtidal waters near the Station Biologique de Roscoff and cultured in the lab as described before[26]. *In vitro* fertilisations and collections of embryonic and larval stages were performed as previously described[26]. *Capitella teleta* Blake, Grassle & Eckelbarger, 2009 was cultured, grown, and sifted, and its embryos and larvae were collected following established protocols[30]. *Magelona* spp. were collected in muddy sand from the intertidal of Berwick-upon-Tweed, Northumberland, NE England (~55.766781, -1.984587) and kept initially in aquaria at the National Museum Cardiff before their transfer to Queen Mary University of London, where they were kept in aquaria with artificial sea water.

Genome size measurements

To estimate the haploid DNA nuclear content of *O. fusiformis*, we used a flow cytometer Partex CyFlow Space fitted with a Cobalt Samba green laser (532 nm, 100 mW) as described for the annelid *Dimorphilus gyrociliatus*[26] and adult individuals of *Drosophila melanogaster* as reference. Additionally, we used Jellyfish v.2.3[65] to count and generate a 31-mer histogram from adaptor-cleaned, short-read Illumina reads (see section below), and GenomeScope 2.0[66] to obtain an in-silico estimation of the genome size and heterozygosity of *O. fusiformis*.

Genome sequencing, assembly, and quality check

Ultra-high molecular weight (UHMW) genomic DNA (gDNA) was extracted following the Bionano genomics IrysPrep agar-based, animal tissue protocol using sperm from a single *O. fusiformis* male. UHMW gDNA was cleaned up using a salt:chloroform wash following PacBio's recommendations before long-read sequencing using PacBio v3.0 chemistry at the University of California Berkeley. A total of 16 SMRT cells of PacBio Sequel were used for sequencing with 600 min movie time, producing a total of 170.07 Gb of data (10.72 million reads, N50 read length between 25.75 kb and 30.75 kb). In addition, we used UHMW gDNA of that same individual to generate a 10x Genomics linked reads library, which we sequenced in an Illumina HiSeq4000 at Okinawa Institute of Science and Technology (OIST) to produce 28.62 Gb of data (141.66 million read pairs). PacBio reads were assembled with CANU v.8.3rc2[67] assuming 'batOptions="-dg 3 -db 3 -dr 1 -ca 500 -cp 50' and 'correctedErrorRate=0.065'. Pacbio reads were remapped using pbalign v.0.3.2 and the assembly polished once using Arrow (genomicconsensus, v2.3.2). Then Illumina paired end reads generated with the 10x Genomics linked reads were extracted, remapped using bwa mem v.0.7.17[68] and used for polishing with Racon v.1.16[69]. Bionano Genomics optical mapping data was used to scaffold the PacBio-based assembly, which was de-haploidised with purge_haplotigs v.1.0.4[70] setting cut-offs at 35, 85 and 70x coverages to reconstruct a high-quality haploid reference assembly. HiC-based chromosome scaffolding was performed as described below. Merqury v.1.1[71] and BUSCO v.5[72] were used to assess genome completeness and evaluate the quality of the assembly.

Transcriptome sequencing

Fourteen samples spanning key developmental time points of *O. fusiformis* life cycle, including active oocyte, zygote, 2-cell, 4-cell, and 8-cell stages, 3 hours post-fertilisation (hpf), 4 hpf, coeloblastula (5 hpf), gastrula (9 hpf), axial elongation (13 hpf), early larva (18 hpf), mitraria larva (27 hpf), pre-metamorphic competent larva (3 weeks post-fertilisation, wpf) and post-metamorphic juvenile were collected in duplicates (except for the latter), flash frozen in liquid nitrogen and stored at –80 °C for total RNA extraction. Samples within replicates were paired, with each one containing ~300 embryos or ~150 larvae coming from the same *in vitro* fertilisation. Nine further samples from adult tissues and body regions (blood vessel, body wall, midgut, prostomium, head, ovary, retractor muscle, tail, and testes) were also collected as described above. Likewise, further five samples spanning post-cleavage time points of *C. teleta*, including 64 cells and gastrula stages, and stage 4tt, stage 5, and stage 7 larval stages, were also collected in duplicates. Total RNA was isolated with the Monarch Total RNA Miniprep Kit (New England Biolabs, NEB) following supplier's recommendations. Total RNA samples from developmental stages from both *O. fusiformis* and *C. teleta* were used to prep strand-specific mRNA Illumina libraries that were sequenced at the Oxford Genomics Centre (University of Oxford, UK) over three lanes of an Illumina NovaSeq6000 system in 2 × 150 bases mode. Adult tissue samples were sequenced at BGI on a BGISeq-500 platform in 2 × 100 bases mode. All samples were sequenced to a depth of ~50 M reads (Supplementary Tables 13, 16).

Annotation of repeats and transposable elements (TEs)

RepeatModeler v.2.0.1[73] and RepBase were used to construct a *de novo* repeat library for *O. fusiformis*, which was then filtered for *bona fide* genes using the predicted proteome of *C. teleta*. Briefly, we used DIAMOND v.0.9.22[74] with an *e*-value cut-off of 1e-10 to identify sequences in the *de novo* repeat library with significant similarity to protein coding genes in *C. teleta* that are not transposable elements. Sequences with a significant hit were manually inspected to verify they were not transposable elements and if so, they were manually removed from the *de novo* repeat library. The filtered consensus repeat predictions were then used to annotate the genome assembly of *O. fusiformis* with RepeatMasker "open-4.0". We next used LTR_finder v.1.07[75], a structural search algorithm, to identify and annotate Long Tandem Repeats (LTR). Finally, we generated a consensus set of repeats by merging RepeatMasker and LTR_finder predictions with RepeatCraft[76], using default parameters but a maximum LTR size of 25 kb (as derived from the LTR_finder annotation). The general feature format (gff) and fasta files with the annotation of TEs and repeats are available in the GitHub repository (see Data Availability section).

Gene prediction and functional annotation

We used SAMtools v.1.9[77] and the annotation of repeats to soft-mask *O. fusiformis* genome assembly before gene prediction. We then mapped all embryonic and adult transcriptomes and a publicly available dataset[78] (SRR1222288) with STAR v. 2.5.3a[79] after removing low-quality read pairs and read pairs containing Illumina sequencing adapters with trimmomatic v.0.39[80]. StringTie v.1.3.6[81] was used to convert STAR alignments into gene transfer format (GTF) files and Portcullis v.1.1.2[82] to generate a curated set of splice junctions. Additionally, we generated *de novo* transcriptome assemblies for all samples with Trinity v.2.5.1[83] with

default parameters, which were thereafter mapped to the soft-masked assembly with GMAP v.2020-04-08[84]. We then ran the default Mikado v.2.1 pipeline[85] to merge all transcriptomic evidence and reliable splice junctions into a single set of best-supported transcripts and gene models. From this merged dataset, we filtered full-length, non-redundant transcripts with a BLAST hit on at least 50 % of their length and at least two exons to obtain a gene set that we used to train Augustus v.3.2.3[86]. Simultaneously, we used the Mikado gene annotation and Portcullis splice junctions to generate confident sets of exon and intron hints, respectively. We also ran Exonerate v.2.4.0[87] to generate spliced alignments of the proteome of *C. teleta* proteome on *O. fusiformis* soft-masked genome assembly to obtain further gene hints. We then merged all exon and intron hints into a single dataset which we passed to Augustus v.3.2.3[86] for *ab initio* gene prediction. Finally, PASA v.2.3.3[88] was used to combine RNA-seq and *ab initio* gene models into a final gene set, from which spurious predictions with in-frame STOP codons (228 gene models), predictions that overlapped with repeats (5,779 gene models) and that had high similarity to transposable elements in the RepeatPeps.lib database (2,450 models) were removed. This filtered gene set includes 26,966 genes, encompassing 31,903 different transcripts. To assess the completeness of this annotation, we ran BUSCO v.5[72] in proteome mode, resulting in 97.7 % of the core genes present. Moreover, 31,678 out of the 31,903 (99.29%) of the filtered transcripts are supported by RNA-seq data and 80.69% of the transcripts have a significant BLAST match (*e*-value cut-off < 0.001) to a previously annotated annelid gene (database containing non-redundant proteomes of the high-quality annelid genomes of *C. teleta*, *D. gyrociliatus*, *E. andreii*, *L. luymesi*, *P. echinospica*, *R. pachyptila* and *S. benedictii*). A similar functional annotation approach was followed to re-annotate the genome of *C. teleta* with the new RNA-seq data, using as starting assembly the soft masked version available at Ensembl Metazoa. This resulted in 41,221 transcripts, 39,814 of which have RNA-seq support (96.59%). Additionally, 80.47% of the transcripts have a significant BLAST match (*e*-value cut-off < 0.001) to other well-annotated annelid genomes (see above).

Protein homologies for the filtered transcripts of *O. fusiformis* and *C. teleta* were annotated with BLAST v.2.2.31+[89] on the UniProt/SwissProt database provided with Trinotate v.3.0[90]. We used HMMER v.2.3.2[91] to identify protein domains using Trinotate's PFAM-A database and signalP v.4.1[92] to predict signal peptides. These functional annotations were integrated into a Trinotate database, which retrieved Gene Onthology (GO), eggNOG and KEGG terms for each transcript. In addition, we ran PANTHER HMM scoring tool to assign a PantherDB[93] orthology ID to each transcript. In total, we retrieved a functional annotation for 22,516 transcripts (63.86 %). Functional annotation reports are provided in the GitHub repository (see Data Availability section).

Chromosome-scale scaffolding

Sperm from a single *O. fusiformis* worm and an entire sexually mature male were used as input material to construct two Omni-C Dovetail libraries following manufacturer's recommendations for marine invertebrates. These libraries were sequenced in an Illumina NovaSeq6000 at the Okinawa Institute of Science and Technology (Okinawa, Japan) to a depth of 229 and 247 million reads. HiC reads were processed using the Juicer pipeline

r.e0d1bb7[94] to generate a list of curated contracts ('merged no dups') that was subsequently employed to scaffold the assembly using 3d-dna v.180419[95]. The resulting assembly and contact map were visually inspected and curated using Juicebox v.1.11.08[94] and adjustments submitted for a subsequent run of optimisation using 3d-dna. Finally, repeats and TEs were re-annotated in this chromosome scale assembly as described above, and the annotation obtained for the PacBio-based assembly was lifted over with Liftoff v.1.6.1[96]. All gene models but two were successfully re-annotated in the chromosome-scale assembly.

Gene family evolution analyses

We used the AGAT suite of scripts to generate non-redundant proteomes with only the longest isoform for a set of 21 metazoan proteomes (Supplementary Table 2). To reconstruct gene families, we used OrthoFinder v.2.2.7[97] using MMSeqs2[98] to calculate sequence similarity scores and an inflation value of 2. OrthoFinder gene families were parsed and mapped onto a reference species phylogeny to infer gene family gains and losses at different nodes and tips using the ETE 3 library[99], as well as to estimate the node of origin for each gene family. Gene expansions were computed for each species using a hypergeometric test against the median gene number per species for a given family employing previously published code[34] (Supplementary Tables 3–7). Principal component analysis was performed on the orthogroups matrix by metazoan lineage, given that orthogroups were present in at least three of the 22 analysed species, to eliminate taxonomically restricted genes. All single copy ortholog files derived from this analysis employed throughout the study are available in the GitHub repository (see Data Availability section).

Macrosynteny analyses

Single copy orthologues obtained using the mutual best hit (MBH) approach generated using MMseqs2[98] using the annotations of *Branchiostoma floridae*[100], *Pecten maximus*[101], *Streblospio benedictii*[102], and *Lineus longissimus*[103,104] were used to generate Oxford synteny plots comparing sequentially indexed orthologue positions. Plotting order was determined by hierarchical clustering of the shared orthologue content using the complete linkage method as originally proposed. Comparison of the karyotype of all four species was performed using the Rideogram package by colouring pairwise orthologues according to the ALG assignment in comparisons with *P. maximus* and *B. floridae*.

Evolutionary analysis of chordin *in annelids*

The identification of *chordin* (*chrd*) and *chordin-like* (*chrdl*) genes in *O. fusiformis* was based on the genome functional annotation (see above). To mine *chrd* orthologues, 81 annelid transcriptomic datasets were downloaded from SRA (Supplementary Table 8) and assembled with Trinity v.2.5.1[83] to create BLAST local nucleotide databases. We also created a nucleotide database for *C. teleta* using its annotated genome[105] (ENA accession number GCA_000328365.1). Human and *O. fusiformis* CHRD proteins were used as queries to find *chrd* orthologues following the MBH approach (*e*-value $\leq 10^{-3}$), obtaining 103 unique candidate *chrd* transcripts that were then translated (Supplementary Table 9). A single candidate CHRD protein for *Themiste lageniformis* (unpublished data, provided by Michael J Boyle) was included *ad hoc* at this step. In addition, 15 curated CHRD and CHRDL protein

sequences (and an outgroup) were fetched from various sources (Supplementary Table 10) and aligned together with *O. fusiformis* CHRD and CHRDL sequences in MAFFT v.7[106] with the G-INS-I iterative refinement method and default scoring parameters. From this mother alignment further daughter alignments were obtained using "mafft --addfragments"[107], the accurate "--multipair" method, and default scoring parameters. For orthology assignment, two phylogenetic analyses were performed on selected candidate sequences, which included the longest isoform for each species-gene combination, given that it included a 10-residue or longer properly aligned fragment in either the CHRD domains or the von Willebrand factor type C (VWFC) domains. vWFC and CHRD domains were trimmed and concatenated using domain boundaries defined by ProSITE domain annotation for the human chordin precursor protein (UniProt: Q9H2X0). Either all domains or the VWFC domains only were used for phylogenetic inference (Extended Data Figure 2c, d, Supplementary Tables 11, 12) with a WAG amino acid replacement matrix[108] to account for transition rates, the FreeRate heterogeneity model (R4)[109] to describe sites evolution rates, and an optimization of amino acid frequencies using maximum likelihood (ML) using IQ-TREE v.2.0.3[110]. 1,000 ultrafast bootstraps (BS)[111] were used to extract branch support values. Bayesian reconstruction in MrBayes v.3.2.7a[112] were also performed using the same WAG matrix but substituting the R4 model for the discrete gamma model[113], with 4 rate categories (G4). All trees were composed in FigTree v.1.4.4. Alignment files are available in the GitHub repository (see Data Availability section).

Gene expression profiling

We profiled gene expression dynamics from blastula to juvenile stages for *O. fusiformis*, from 64-cell to competent larva stages for *C. teleta*, from early development to female adult stages for *D. gyrociliatus*, and across the 9 adult tissues samples of *O. fusiformis*. Sequencing adaptors were removed from raw reads using trimmomatic v.0.39[80]. Cleaned reads were pseudo-aligned to the filtered gene models using kallisto v.0.46.2[114] and genes with an expression level above an empirically defined threshold of 2 transcripts per million (TPM) were deemed expressed. For each species, the DESeq2 v.1.30.1 package[115] was used to normalise read counts across developmental stages (Supplementary Tables 13–21) and adult tissues (Supplementary Tables 49–51) and to perform pair-wise differential gene expression analyses between consecutive developmental stages. *P*-values were adjusted using the Benjamini-Hochberg method for multiple testing correction. We defined a gene as significantly upregulated for a $\log_2$(fold-change) (LFC) > 1 or downregulated for a LFC < 1, given that adjusted *p*-value < 0.05. Principal component analyses were performed on the variance stabilising-transformed matrices of the normalised DESeq2 matrices. For the *O. fusiformis* adult tissues samples, genes specifically expressed (TPM > 2) in both the head and head plus two anteriormost segments samples only were classified as adult anterior genes, and those expressed in both the tail and the body wall only were classified as adult trunk and posterior genes (Supplementary Tables 52, 53). For all 3 annelid taxa, anterior, trunk, and posterior markers were defined as genes whose spatial expression pattern has been validated through *in situ* hybridisation in the literature (Supplementary Tables 54–56). TPM and DESeq2 gene expression matrices of developmental and adult tissue samples are also available in the GitHub repository (see Data Availability section).

Gene clustering and co-expression network analyses

Transcripts were clustered according to their normalised DESeq2 expression dynamics through soft *k*-means clustering (or soft clustering) using the mfuzz v.2.52 package[116] (Supplementary Tables 23–26). Out of the total number of transcripts, we discarded those which were not expressed at any developmental stage (225 out of 31,903 for *O. fusiformis*, 1,407 out of 41,221 for *C. teleta*, and 200 out of 17,388 for *D. gyrociliatus*). We then determined an optimal number of 12 clusters (*O. fusiformis* and *C. teleta*) and 9 clusters (*D. gyrociliatus*) for our datasets by applying the elbow method to the minimum centroid distance as a function of the number of clusters. For the construction of the gene co-expression networks for *O. fusiformis* and *C. teleta*, we used the WGCNA package v.1.70–3[117]. All transcripts expressed at any developmental stage were used to build a signed network with a minimum module size of 300 genes and an optimised soft-thresholding power of 16 and 8, for *O. fusiformis* and *C. teleta*, respectively. Block-wise network construction returned 15 gene modules for *O. fusiformis*, from which one module was dropped due to poor intramodular connectivity, and 19 gene modules for *C. teleta* (Supplementary Tables 23, 24). The remaining 14 gene modules of *O. fusiformis* (A–N) and 19 gene modules of *C. teleta* (A–O, W–Z) were labelled with distinct colours with unassigned genes labelled in grey. Random subsets consisting of the nodes and edges of 30 % of the transcripts were fed to Cytoscape v.3.8.2[118] for network visualisation. Module eigengenes were chosen to summarise the gene expression profiles of gene modules. Gene ontology (GO) enrichment analysis of each gene cluster and gene module was performed using the topGO v.2.44 package. We performed a Fisher's exact test and listed the top 30 (soft *k*-means clusters) or top 15 (WGCNA modules) significantly enriched GO terms of the class biological process (Supplementary Tables 27–31). To ease visualisation, all 486 non-redundant enriched GO terms from the 33 soft *k*-means clusters from all 3 species were clustered through *k*-means clustering by semantic similarity using the simplifyEnrichment v.1.2.0 package[119] (Supplementary Figures 7, 8). Full network nodes and edges files and the random 30 % network subset files are available in the GitHub repository (see Data availability section).

Transcription factor repertoire analysis

We selected a custom set of 36 transcription factor classes from all 9 transcription factor superclasses from the TFClass database[120]. Transcripts in *O. fusiformis*, *C. teleta*, and *D. gyrociliatus* were deemed transcription factors and classified into one or more of the 36 classes if they were a match for any of the corresponding PANTHER identifiers (Supplementary Tables 32–33, Supplementary Figure 3). Over- and underrepresentation of the different transcription factor classes in the gene expression clusters was tested through pair-wise two-sided Fisher's exact tests, for which we then adjusted the *p*-values using the Benjamini-Hochberg correction for multiple testing.

Orthogroup overlap analysis

We performed pair-wise comparisons between each possible combination of soft *k*-means clusters of all 3 annelid taxa. The numbers of overlapped orthogroups between either the full clusters or the transcription factors belonging to each cluster only were subjected to upper-tail

hypergeometric tests. *P*-values were then adjusted using the Benjamini-Hochberg method for multiple testing correction. For the simplified analyses by quadrants, clusters were classed as early/pre-larval (*O. fusiformis*: 1–6; *C. teleta*: 1–5; *D. gyrociliatus*: 1–3) or late/pre-larval (*O. fusiformis*: 8–12; *C. teleta*: 7–12; *D. gyrociliatus*: 5–7), thus rendering 4 different quadrants for each species pair-wise comparison: $\text{early}_{\text{species A}}$–$\text{early}_{\text{species B}}$, $\text{early}_{\text{species A}}$–$\text{late}_{\text{species B}}$, $\text{late}_{\text{species A}}$–$\text{early}_{\text{species B}}$, and $\text{late}_{\text{species A}}$–$\text{late}_{\text{species B}}$. Clusters corresponding to female adult expression in *D. gyrociliatus* (8 and 9) were discarded for comparison purposes. Relative similarity (RS) for each of the four quadrants was computed as the following ratio:

$$RS = \frac{\text{mean}(-\log_{10}(adj\ p\text{-}value)_{\text{quadrant}})}{\text{mean}(-\log_{10}(adj\ p\text{-}value)_{\text{total}})}$$

Values above 1 indicate a higher orthogroup overlap than average, whereas values below 1 represent a lower overlap than average. For genes under heterochronic shifts – i.e., with distinct temporal expression dynamics – between indirect and direct development, a gene set was constructed with the genes with a single copy ortholog in both *O. fusiformis* and *C. teleta* whose expression was shifted from post-larval clusters (*O. fusiformis*: 7–12; *C. teleta*: 8–12) to early clusters 2 and 3 in *D. gyrociliatus* (see Fig. 2f). For the characterisation of genes under heterochronic shifts between planktotrophic and lecithotrophic larvae, two gene sets were generated with the genes with $\text{early}_{\textit{O. fusiformis}}$–$\text{late}_{\textit{C. teleta}}$ and $\text{late}_{\textit{O. fusiformis}}$–$\text{early}_{\textit{C. teleta}}$ dynamics, as described above. Gene ontology (GO) enrichment analysis of both gene sets was performed using the topGO v.2.44 package. We performed a Fisher's exact test and listed the top 15 significantly enriched GO terms of the class biological process (Supplementary Table 40). BlastKOALA[121] server was used to assign a KEGG orthology number to one-to-one orthologs showing heterochronic sifts and KEGG mapper[122] to analyse the annotations (Supplementary Tables 41, 42).

Pathway analyses

Human genes involved in the animal autophagy pathway (map04140) were obtained from the KEGG pathway database[123]. *D. melanogaster* and *Saccharomyces cerevisiae* genes involved in the chitin synthesis pathway were fetched from FlyBase[124] and SGD[125] based on the enzyme nomenclature (EC) numbers of the pathway enzymatic activities[126]. Orthology in *O. fusiformis* and *C. teleta* for the autophagy pathway genes was determined from the single copy ortholog sets to the human genes, where one for both species existed (Supplementary Tables 43, 44). For the chitin synthesis pathway, and due to the high number of paralogs and expansions/losses of enzymatic activities of the chitin synthesis pathway, orthology was inferred from PANTHER family/subfamily identifiers to the corresponding enzymatic activities (Supplementary Tables 45, 46). We then used this orthology to reconstruct the chitin synthesis pathway in annelids. Timing across both species and the presence or lack thereof of heterochronic shifts between *O. fusiformis* and *C. teleta* was determined as described above.

Hox *genes orthology assignment*

129 curated Hox sequences were retrieved from various databases (Supplementary Table 47) and aligned with *O. fusiformis* Hox proteins with MAFFT v.7 in automatic mode. Poorly

aligned regions were removed with gBlocks v.0.91b[127] yielding the final alignments. Maximum likelihood trees were constructed using RAxML v.8.2.11.9[128] with an LG substitution matrix[129] and 1,000 ultrafast BS. All trees were composed in FigTree v.1.4.4. Alignment files are available in the GitHub repository (see Data Availability section).

Whole mount in situ *hybridisation and immunohistochemistry*

Fragments of *chordin* and *Hox* genes were isolated as previously described[27] using gene-specific oligonucleotides and a T7 adaptor. Riboprobes were synthesise with the T7 MEGAscript kit (ThermoFisher, AM1334) and stored at a concentration of 50 ng/µl in hybridisation buffer at –20 ºC. Whole mount *in situ* hybridisation in embryonic, larval, and juvenile stages were conducted as described elsewhere[27,29]. Antibody staining in larval stages of *O. fusiformis*, *Magelona* spp. and *C. teleta* was carried out as previously described[26,130]. DIC images of the colorimetric *in situs* were obtained with a Leica 560 DMRA2 upright microscope equipped with an Infinity5 camera (Lumenera). Fluorescently stained samples were scanned with a Nikon CSU-W1 Spinning Disk Confocal.

Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)

We performed two replicates of ATAC-seq from samples containing ~50,000 cells at the blastula (~900 embryos), gastrula (~500), elongation (~300), mitraria larva (~150 larvae) and competent larva (~40) stages for *O. fusiformis*, and the 64-cells stage (~500 embryos), gastrula (~200), stage 4tt larva (~120 larvae), stage 5 larva (~90) and stage 8 larva (~50) for *C. teleta* following the omniATAC protocol[131], but gently homogenising the samples with a pestle in lysis buffer and incubating them on ice for 3 min. Tagmentation was performed for 30 min at 37°C with an in-house purified Tn5 enzyme[132]. After DNA clean-up, ATAC-seq libraries were amplified as previously described. Primers used for both PCR and qPCR are listed in Supplementary Tables 57 and 59. Amplified libraries were purified using ClentMag PCR Clean Up Beads as indicated by the supplier and quantified and quality checked on a Qubit 4 Fluorometer (Thermo-Fisher) and an Agilent 2200 TapeStation system before pooling at equal molecular weight. Sequencing was performed on an Illumina HiSeq4000 platform in 2 × 75 bases mode at the Oxford Genomics Centre (University of Oxford, United Kingdom) (blastula, elongation and mitraria larva stages, and one replicate of the gastrula sample of *O. fusiformis*, as well as the 64 cells, gastrula, and stage 4tt larva stages of *C. teleta*) and on an Illumina NovoSeq6000 in 2 × 150 bases mode at Novogene (Cambridge, United Kingdom) (one replicate of gastrula and the two replicates of competent larva stages of *O. fusiformis* and the two replicates of stage 5 and stage 8 larva of *C. teleta*).

Chromatin accessibility profiling

We used cutadapt v.2.5[133] to remove sequencing adaptors and trim reads from libraries sequenced in 2 × 150 bases mode to 75 bases reads. Quality filtered reads were mapped using NextGenMap v.0.5.5[134] in paired-end mode, duplicates were removed using samtools v.1.9[135] and mapped reads were shifted using deepTools v.3.4.3[136] (Supplementary Tables 58, 60). Fragment size distribution was estimated from resulting BAM files and transcription start site (TSS) enrichment analysis was computed using computeMatrix and plotHeatmap commands in deepTools v.3.4.3. Peak calling was done with MACS2 v.2.2.7.1[137,138] (-f BAMPE --min-

length 100 --max-gap 75 and -q 0.01). Reproducible peaks were identified by irreproducible discovery rates (IDR) (IDR < 0.05) v.2.0.4. at each developmental stage. Peaks from repetitive regions were filtered with BEDtools v.2.28.0[139] at each developmental stage. Next, we used DiffBind v.3.0.14[140] to generate a final consensus peak set of 63,732 peaks in *O. fusiformis* and 46,409 peaks in *C. teleta*, which were normalised using DESeq2 method. Peak clustering according to accessibility dynamics was performed as described above for RNA-seq, using the same number of 12 clusters to make both profiling techniques comparable. Principal component analysis and differential accessibility analyses between consecutive developmental stages were also performed as described above. An LFC > 0 and a LFC < 0 indicates whether a peak opens or closes, respectively, given that the adjusted *p*-value < 0.05. Stage-specific and constitutive peaks were determined using UpSetR v.1.4.0[141] and both the consensus peak set and the stage-specific peak sets were classified by genomic region using HOMER v.4.11[142] and further curated. Visualisation of peak tracks and gene structures was conducted with pyGenomeTracks v.2.1[143] and deepTools v.3.4.3[136]. To correlate chromatin accessibility and gene expression, this genomic region annotation was used to assign peaks to their closest gene (63,726 peaks were assigned to 23,025 genes in *O. fusiformis* and 44,368 peaks were assigned to 23,382 genes in *C. teleta*). Pearson correlation coefficient between chromatin accessibility and gene expression was computed individually by peak, and gene ontology (GO) enrichment analyses of the gene sets regulated by peak clusters was performed using the topGO v.2.44 package. We performed a Fisher's exact test and listed the top 30 significantly enriched GO terms of the class biological process. To ease visualisation, all 242 non-redundant enriched GO terms were clustered through *k*-means clustering by semantic similarity using the simplifyEnrichment v.1.2.0 package[119] (Supplementary Tables 61–71). Coverage files and peak set files are available in the GitHub repository (see Data Availability section).

Motif identification, clustering, matching and curation

To identify transcription factor-binding motifs in chromatin accessible regions in the two species, we first used HOMER[142] (v.4.1) to identify known and *de novo* motifs in the consensus peak sets, which yielded 456 motifs for *O. fusiformis* and 364 motifs for *C. teleta* (Supplementary Tables 72, 73). We then used GimmeMotifs v.0.16.1[144] , with a 90% similarity cut-off to cluster the motifs predicted in *O. fusiformis* and *C. teleta* into 141 consensus motifs, which we matched against four motif databases to assign their putative identity (Gimme vertebrate 5.0[144], HOMER[142], CIS-BP[145] and a custom JASPAR2022[146] core motifs without plant and fungi motifs, Supplementary Fig. 22). We then used the human non-redundant TF motif database (https://resources.altius.org/~jvierstra/projects/motif-clustering-v2.0beta/) to manually curate the annotation. After removing motifs that likely represented sequence biases, we finally obtained 95 motif archetypes for *O. fusiformis* and 91 for *C. teleta* (Supplementary Table 74), which we then used to perform motif counts in peaks (Supplementary Tables 75, 76) and motif accessibility estimation (Supplementary Tables 77, 78) with GimmeMotifs v.0.16.1[144]. Data clustering was performed with mfuzz v.2.52[116]. Over- and underrepresentation of counts of the common curated motif archetypes in the peak accessibility soft clusters (see above) was tested through pair-wise two-sided

Fisher's exact tests, for which we then adjusted the *p*-values using the Bonferroni correction for multiple testing.

Transcription factor footprinting and Hox *gene regulatory network exploration*
To predict transcription factor binding, as a proxy of activity, we conducted footprinting analysis with TOBIAS[44] v.0.12.0 during development in the 95 and 91 motif archetypes for *O. fusiformis* and *C. teleta*, respectively (Supplementary Tables 79, 80). Transcription factor binding scores (TFBS) were clustered with mfuzz v.2.52[116]. Pearson correlation coefficients of motif accessibility and TFBS were calculated by stage and by motif separately based on the 33 common, curated motif archetypes. To reconstruct potential upstream regulators and downstream effectors of the *Hox* genes, we first subset ATAC-seq peaks annotated to the *Hox* genes in the *Hox* cluster (i.e., all but *Post1*) in *O. fusiformis* and *C. teleta* and extracted the bound motifs on those peaks (Supplementary Tables 81, 82). TFBS were sum up for each motifs to obtain global dynamics, and their temporal dynamics were then clustered with mfuzz v.2.52[116]. For the downstream genes regulated by *Hox*, we obtained genes annotated to ATAC-seq peaks with a bound HOX/EVX/CDX motif at the competent stage in *O. fusiformis* and stage 4tt larva in *C. teleta* (Supplementary Tables 83, 84). One-to-one orthologs were used to identified shared targets and PANTHER IDs to obtain their functional annotation.

Phylostratigraphy
To evaluate gene expression dynamics by phylostratum and developmental stage in all 3 annelid lineages, we used the OrthoFinder gene families and their inferred origins. We deemed all genes originating before and with the Cnidarian-Bilaterian ancestor of pre-metazoan and metazoan origin (Supplementary Tables 85–87). We then applied a quantile normalisation onto the DESeq2 normalised matrices of gene expression. The 75 % percentile of the quantile-normalised gene expression levels was used as the summarising measure of the gene expression distribution by developmental stage. Over- and underrepresentation of the different phylostrata in the gene expression clusters was tested through pair-wise two-sided Fisher's exact tests, for which we then adjusted the *p*-values using the Bonferroni correction for multiple testing. Gene expression dynamics of novel genes and genes of pre-metazoan and metazoan origin across selected metazoan lineages (see Comparative transcriptomics section below) were also evaluated as described above.

Comparative transcriptomics
Publicly available RNA-seq developmental time courses for the development of *Amphimedon queenslandica*, *Clytia hemisphaerica*, *Nematostella vectensis*, *Strongylocentrotus purpuratus*, *Branchiostoma lanceolatum*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Crassostrea gigas*, *Dimorphilus gyrociliatus*, and two stages of *Capitella teleta* were downloaded from the SRA (Supplementary Table 88), cleaned for adaptors and low-quality reads with trimmomatic v.0.39[80] and pseudo-aligned to their respective non-redundant genome-based gene repertoires – i.e., with a single transcript isoform, the longest, per gene model – using kallisto v.0.46.2[114]. We then performed a quantile transformation of TPM values using scikit-learn v.1.0.2[147] and calculated the Jensen-Shannon divergence (JSD) from (i) all single copy orthologs, (ii) the set single copy transcription factor orthologs, and (iii) the

set of common single copy orthologs across all lineages, either between all possible one-to-one species comparisons (i) or between all species and *O. fusiformis* (ii, iii), using the philentropy v.0.5.0 package[148]:

$$JSD_{\text{raw}}(P \parallel Q) = \frac{1}{2}\sum_{i=0}^{n} p_i \times \log_2\left(\frac{p_i}{\frac{1}{2}(p_i + q_i)}\right) + \frac{1}{2}\sum_{i=0}^{n} q_i \times \log_2\left(\frac{q_i}{\frac{1}{2}(p_i + q_i)}\right)$$

Transcriptomic divergences were calculated based on 250 bootstrap replicates, from which statistically robust mean values and standard deviations were obtained. Raw mean JSD values (JSD_{raw}) were adjusted (JSD_{adj}) by dividing by the number of single copy orthologs (i), single copy transcription factor orthologs (ii), or common single copy orthologs (iii) of each comparison (Supplementary Tables 89, 90), and normalised using the minimum and maximum adjusted JSD values from all one-to-one species comparisons as follows:

$$JSD_{\text{norm}}(P \parallel Q) = \frac{JSD_{\text{adj}}(P \parallel Q) - \min JSD_{\text{adj}}}{\max JSD_{\text{adj}} - \min JSD_{\text{adj}}};\ JSD_{\text{norm}} \in [0, 1]$$

Relative JSD values were obtained equally, using minimum and maximum adjusted JSD values from each one-to-one species comparison instead. Gene-wise JSD ($gwJSD$) between five key one-to-one larval stages comparisons was computed as follows:

$$gwJSD(P \parallel Q) = \frac{1}{2} \times p_i \times \log_2\left(\frac{p_i}{\frac{1}{2}(p_i + q_i)}\right) + \frac{1}{2} \times q_i \times \log_2\left(\frac{q_i}{\frac{1}{2}(p_i + q_i)}\right)$$

Similarity-driving genes – i.e., those with very low gwJSD – were subset as those below the threshold defined as 25 % of the point of highest probability density of the gwJSD distributions. Gene ontology (GO) enrichment analysis of the similarity-driving gene sets was performed using the topGO v.2.44 package. We performed a Fisher's exact test and listed the top 30 significantly enriched GO terms of the class biological process (Supplementary Table 91). To ease visualisation, all 51 non-redundant enriched GO terms from the 5 gene sets were clustered through *k*-means clustering by semantic similarity using the simplifyEnrichment v.1.2.0 package[119]. For comparative *Hox* gene expression dynamics profiling in metazoan lineages, the same non-redundant gene expression matrices were normalised using the DESeq2 v.1.30.1 package[115], unless *Hox* gene models where missing, in which case they were manually added *ad hoc* to the non-redundant genome-based gene repertoires (Supplementary Table 94). *Hox* gene expression profiling in *Urechis unicinctus* was performed as described for the rest of taxa but using the available reference transcriptome[149] instead (Supplementary Table 48). All gene expression matrices are available in the GitHub repository (see Data Availability section).

Data availability

All sequence data associated with this project are available at the European Nucleotide Archive (project PRJEB38497) and Gene Expression Omnibus (accession numbers GSE184126, GSE202283, GSE192478, GSE210813 and GSE210814). Genome assemblies, transposable element annotations, genome annotation files used for RNA-seq and ATAC-seq analyses, WGCNA nodes and edges files, alignment files used in orthology assignment, and other additional files are publicly available in https://github.com/ChemaMD/OweniaGenome.

Code availability

All code used in this study is available in https://github.com/ChemaMD/OweniaGenome.

Method References

65 Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764-770 (2011).

66 Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1432 (2020).

67 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).

68 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **1303.3997** (2013).

69 Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**, 737-746 (2017).

70 Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).

71 Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245 (2020).

72 Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38**, 4647-4654 (2021).

73 Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* **117**, 9451-9457 (2020).

74 Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366-368 (2021).

75 Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265-268 (2007).

76 Wong, W. Y. & Simakov, O. RepeatCraft: a meta-pipeline for repetitive element de-fragmentation and annotation. *Bioinformatics* **35**, 1051-1052 (2019).

77 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).

78 Weigert, A. *et al.* Illuminating the base of the annelid tree using transcriptomics. *Mol Biol Evol* **31**, 1391-1401 (2014).

79 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

80 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

81 Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295 (2015).

82 Mapleson, D., Venturini, L., Kaithakottil, G. & Swarbreck, D. Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *Gigascience* **7** (2018).
83 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652 (2011).
84 Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875 (2005).
85 Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L. & Swarbreck, D. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *Gigascience* **7** (2018).
86 Stanke, M., Schoffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
87 Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
88 Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654-5666 (2003).
89 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
90 Bryant, D. M. *et al.* A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep* **18**, 762-776 (2017).
91 Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195 (2011).
92 Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785-786 (2011).
93 Mi, H. *et al.* PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res* **49**, D394-D403 (2021).
94 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).
95 Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95 (2017).
96 Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639-1643 (2021).
97 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).
98 Steinegger, M. & Soding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).
99 Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol* **33**, 1635-1638 (2016).
100 Marletaz, F. *et al.* Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64-70 (2018).
101 Zeng, Q. *et al.* High-quality reannotation of the king scallop genome reveals no 'gene-rich' feature and evolution of toxin resistance. *Comput Struct Biotechnol J* **19**, 4954-4960 (2021).

102 Zakas, C., Harry, N. D., Scholl, E. H. & Rockman, M. V. The genome of the poecilogonous annelid *Streblospio benedicti*. *Genome Biology and Evolution* (2022).
103 Kwiatkowski, D. *et al*. The genome sequence of the bootlace worm, *Lineus longissimus* (Gunnerus, 1770). *Wellcome Open Res* **6**, 272 (2021).
104 Cannon, J. T. *et al.* Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530**, 89-93 (2016).
105 Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526-531 (2013).
106 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
107 Katoh, K. & Frith, M. C. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**, 3144-3146 (2012).
108 Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**, 691-699 (2001).
109 Soubrier, J. *et al.* The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol Biol Evol* **29**, 3345-3358 (2012).
110 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274 (2015).
111 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**, 518-522 (2018).
112 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).
113 Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* **39**, 306-314 (1994).
114 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).
115 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
116 Futschik, M. E. & Carlisle, B. Noise-robust soft clustering of gene expression time-course data. *J Bioinform Comput Biol* **3**, 965-988 (2005).
117 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
118 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
119 Gu, Z. & Hübschmann, D. simplifyEnrichment: an R/Bioconductor package for Clustering and Visualizing Functional Enrichment Results. *bioRxiv*, 2020.2010.2027.312116 (2021).
120 Wingender, E., Schoeps, T., Haubrock, M., Krull, M. & Donitz, J. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res* **46**, D343-D347 (2018).

121 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726-731 (2016).
122 Kanehisa, M., Sato, Y. & Kawashima, M. KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci* **31**, 47-53 (2022).
123 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
124 Gramates, L. S. *et al.* FlyBase: a guided tour of highlighted features. *Genetics* **220** (2022).
125 Cherry, J. M. *et al.* Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700-705 (2012).
126 Merzendorfer, H. The cellular basis of chitin synthesis in fungi and insects: common principles and differences. *Eur J Cell Biol* **90**, 759-769 (2011).
127 Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564-577 (2007).
128 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
129 Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol Biol Evol* **25**, 1307-1320 (2008).
130 Meyer, N. P., Carrillo-Baltodano, A., Moore, R. E. & Seaver, E. C. Nervous system development in lecithotrophic larval and juvenile stages of the annelid *Capitella teleta*. *Front Zool* **12**, 15 (2015).
131 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962 (2017).
132 Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**, 2033-2040 (2014).
133 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
134 Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791 (2013).
135 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
136 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).
137 Gaspar, J. M. Improved peak-calling with MACS2. *bioRxiv*, 496521 (2018).
138 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
139 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
140 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393 (2012).
141 Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940 (2017).

142 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
143 Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets *Bioinformatics* **37**, 422-423 (2020).
144 van Heeringen, S. J. & Veenstra, G. J. GimmeMotifs: a *de novo* motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics* **27**, 270-271 (2011).
145 Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431-1443 (2014).
146 Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, D165-D173 (2022).
147 Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825−2830 (2011).
148 Drost, H. G. Philentropy: Information Theory and Distance Quantification with R. *Journal of Open Source Software* **3**, 765 (2018).
149 Park, C. *et al.* The developmental transcriptome atlas of the spoon worm *Urechis unicinctus* (Echiurida: Annelida). *Gigascience* **7**, 1-7 (2018).

Acknowledgments
We thank Alex de Mendoza and Daria Gavriouchkina for their support and valuable comments on the manuscript, as well as the staff at Station Biologique de Roscoff for their help with collections and animal supplies and the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of RNA-seq and ATAC-seq sequencing data. We also thank Michael J Boyle for providing the *chordin* sequence for *Themiste lageniformis*, Joseph Deane for initial help with *Hox* gene characterisation in *O. fusiformis*, and the core technical staff at the Department of Biology at Queen Mary University of London for their constant support. This work used computing resources from Queen Mary University of London's Apocrita HPC facilities. This work was funded by the Horizon 2020 Framework Programme to JMM-D (European Research Council Starting Grant action number 801669) and AH (European Research Council Consolidator Grant action number 648861) and a Royal Society University Research Fellowship (URF\R1\191161) and a Japan Society for the Promotion of Science Kakenhi grant (JP 19K06620) to FM.

Author Contributions
JMM-D, FM, YL and FMM-Z conceived and designed the study; YL collected RNA-seq samples for *O. fusiformis* and *C. teleta*, performed ATAC-seq experiments and contributed to all data analyses; FMM-Z performed *chordin* orthology studies and contributed to all data analyses; KG conducted *in situ* hybridisation analyses of *Hox* genes; AC-B collected RNA-seq samples for *C. teleta*, performed immunostainings on larvae and gene expression analyses of *chordin*; BED and RDD contributed to computational analyses; YT performed OMNI-C libraries; GM performed repeat annotations and analyses; OS identified and performed *in*

silico analyses of *Hox* genes; MT performed genomic extractions and optical mapping; KM collected *Magelona* spp.; AH and NML contributed to sequencing efforts; FM and JMM-D assembled and annotated the genome and contributed to data analyses; YL, FMM-Z and JMM-D drafted the manuscript and all authors critically read and commented on the manuscript.

Competing Interests

The authors declare no competing interests.

Extended Data Figure Legends

Extended Data Figure 1 | The genome of *Owenia fusiformis* is conservatively evolving. a, b, Pie charts of the transposable element content and Kimura substitution plots of transposable element divergence for *O. fusiformis* and other selected annelid species belonging to different annelid clades as depicted in **c**. Unlike *H. robusta* and *L. luymesi*, which show bursts of transposable elements, *O. fusiformis* shows more steady rates of expansion. **d**, Gene family evolution analysis across 22 metazoan lineages under a consensus tree topology. Gains are shown in green, losses in violet. Gene family losses in *O. fusiformis* are like those of slow-evolving lineages. **e, f,** *O. fusiformis* has the lowest number of gene losses of all sampled annelids (**e**), and the least gene expansions (**f**) after the extremely compact genome of *D. gyrociliatus*. **g**, Principal component analysis from Fig. 1f, showing the full set of species. **h**, Macrosynteny analysis between *O. fusiformis*, and from top to bottom, the cephalochordate *Branchiostoma floridae*, the bivalve *Pecten maximus*, and the annelid *Streblospio benedicti*. *Owenia fusiformis* retains ancestral linkage groups but also exhibits annelid- and species-specific chromosomal arrangements. However, the karyotype of *O. fusiformis* is more conserved than that of the annelid *S. benedicti*. **i**, Macrosynteny analysis between the bivalve *P. maximus* and the nemertean worm *L. longissimus*. *Lineus longissimus* exhibits conserved ancestral bilaterian linkage groups, including three potential lophotrochozoan-specific chromosomal rearrangements (H+Q, J2+L and K+O2), plus a nemertean-specific fusion (G+C1).

Extended Data Figure 2 | *chordin* was lost multiple times in annelids. a, Domain organisation of Chordin (CHRD) and Chordin-like (CHRDL1/2) proteins, as inferred from human orthologs. **b,** AlphaFold protein structure prediction for human Chordin (UniProt: Q9H2X0) and Chordin-like 1 (UniProt: Q9BU40) revealed a previously unknown and uncharacterised domain in CHRDL1 (also depicted in **a**). **c**, **d**, Orthology assignment of *chordin* annelid candidates. From the multiple sequence alignment, candidate annelid sequences with a 10-residue or longer fragment aligned against either the CHRD (**c**; i.e., bona fide *chordin* genes) or the vWFC domains (**d**; i.e., putative *chordin* genes) were kept for further analysis. CHRDL cluster is shaded in blue; CHRD cluster, in red. Bootstrap support values (top) and posterior probabilities (bottom) are shown at key nodes. Sequences in red and blue are curated CHRD and CHRDL sequences, respectively. **e, f**, Summary phylogenetic trees of presence/absence of *chordin* (red) or putative chordin (light brown) across Annelida. **g,** Expression levels of *chordin*, which peaks at the blastula and gastrula stages, after the specification and inductive activity of the embryonic organiser. **h,** Whole mount *in situ* hybridisation of *chordin* at the blastula (5 hours post fertilisation, hpf), gastrula (9 hpf), and mitraria larva (27 hpf) stages. Asterisks mark the animal/anterior pole. gp: gastral plate; bp: blastopore, mo: mouth.

Extended Data Figure 3 | Gene expression dynamics during annelid life cycles. a–c, Soft *k*-means clustered heatmap of all transcripts whose expression was not null in at least one developmental stage into an optimal number of 12 clusters (*O. fusiformis*, **a**; and *C. teleta*, **b**) and 9 clusters (*D. gyrociliatus*, **c**). Soft clustering increased temporal resolution for the RNA-

seq time course of *D. gyrociliatus*. On the right of each heatmap, gene-wise expression dynamics (grey lines) and locally estimated scatterplot smoothing (coloured lines) for each cluster. Coloured shaded areas represent standard error of the mean. **d,** Enrichment analysis of biological process gene ontology (GO) terms for RNA-seq clusters. Each line represents a single GO term, for which the $-\log_{10}(p\text{–value})$ for each RNA-seq cluster is shown in a colour coded scale. GO terms were clustered into 15 distinct clusters based on semantic similarity (see Supplementary Fig. 7, 8). Clusters are shown on the bottom of the heatmaps. **e–g,** Bar plots depicting *p*–values of the top 15 enriched GO terms in clusters with bimodal dynamics (**e**), clusters spanning three or more developmental stages (**f**) and larval clusters (**g**) for *O. fusiformis* (left) and *C. teleta* (right). For the full list of GO terms and clusters, see Supplementary Fig. 4–6.

Extended Data Figure 4 | Heterochronic shifts in gene regulatory programmes between annelid life cycles. a, b, Similarity heatmaps showcasing the orthogroup overlap between the clusters of co-regulated genes (see Extended Data Fig. 3a–c), between the three annelids. **c,** Explanation of the orthogroup overlap analysis by quadrants. Clusters were classed as "early" (before dotted lines) or "late" (after dotted lines). Clusters of the female adult of *D. gyrociliatus* were disregarded. **d, e,** Heatmaps of relative similarity by quadrants of the orthogroup overlap analyses of the whole genomes (**d**) and transcription factors only (**e**). **f,** Single copy ortholog and transcription factor ortholog number by quadrants in the *O. fusiformis* and *C. teleta* comparison. Colour scale in **d–f** is the same as in **c**. **g,** KEGGbrite characterisation of the gene sets under heterochronic shifts (surrounded by dotted lines in **f**) between *O. fusiformis* and *C. teleta*. **h,** Bar plots depicting *p*-values of top biological process GO terms of genes shifted from late expression in *O. fusiformis* to early expression in *C. teleta*. Full list is available in Supplementary Fig. 12. **i,** Heterochronic shifts of genes involved in the autophagy pathway between *O. fusiformis* and *C. teleta*. 80 % of the genes under heterochronic shifts (19 of 24) are displaced from post-larval expression in *O. fusiformis* to pre-larval expression in *C. teleta*. **j,** Enrichment analysis of the number of transcription factors per class in clusters of co-transcribed genes of *O. fusiformis* (top), *C. teleta* (centre) and *D. gyrociliatus* (bottom). For each cluster and class combination, the Fisher's exact test adjusted *p*–value is shown. Cells in red represent overrepresented classes (odds ratio, OR > 1; adjusted *p*–value < 0.05); cells in blue, underrepresented classes (OR < 1, adjusted *p*–value < 0.05). Dotted lines highlight clusters of maximal enrichment of the homeodomain class. n.s.: not significant.

Extended Data Figure 5 | The *Hox* gene complement and expression in *O. fusiformis*. a, Orthology assignment of *O. fusiformis Hox* genes through maximum likelihood phylogenetic inference. Bootstrap support values are shown for major gene groups. Of: *O. fusiformis*. **b,** Volcano plot of the mitraria to competent larva transition, highlighting the marked upregulation of *Hox* genes. LFC: $\log_2$(fold-change). **c**, Chromosomal location of the *Hox* cluster and *Post1* gene in *O. fusiformis* (top) and schematic comparison of *Hox* cluster organisation in annelids and a mollusc (bottom). **d,** Schematic representation to scale of the genomic loci and intron–exon composition of *Hox* genes in *O. fusiformis*. **e,** Heatmaps of *Hox* gene expression during *C. teleta*, *O. fusiformis* and the echiuran annelid *Urechis*

unicinctus development. In the two annelid species with planktotrophic larvae, *Hox* genes only become expressed at the larval stage (dotted vertical line) and not during embryogenesis, as observed in *C. teleta*. **f,** Whole mount *in situ* hybridisation of *Hox* genes in the gastrula (lateral views) and in the mitraria, pre-competent, and juvenile stages (ventral views). The area encircled by a dotted white line at the pre-competent stage highlights a region of probe trapping from ingested food content. bp: blastopore; mo: mouth.

Extended Data Figure 6 | Transcriptomic profiling of adult tissues and anterior-posterior expression dynamics. a, Drawing of an *O. fusiformis* adult. Samples of highlighted body parts were used to define adult anterior and posterior/trunk genes. **b,** Correlation matrix of RNA-seq experiments from all nine adult tissues, calculated from a variance stabilising-transformed matrix of the normalised DESeq2 matrix. Tail and body wall samples, and head and head plus anteriormost segments samples are the most correlated samples to each other. **c,** Venn diagram showing the number of tissue-specific and shared expressed genes (TPM > 2). Gene sets highlighted with red text were defined as adult anterior, and adult trunk and posterior genes. **d,** Phylostratigraphic classification of adult anterior, and adult trunk and posterior genes, compared to the whole genome and a random subset of 1,000 genes. See Extended Data Figure 8a for phylostrata explanation. Adult genes are enriched in *O. fusiformis*-specific gene innovations. **e, f,** Expression dynamics of each phylostratum by developmental stage in the adult anterior (**e**) and adult trunk and posterior gene sets (**f**), calculated from the 75 % percentile of a quantile-normalised matrix of gene expression levels. Adult anterior genes of most phylostrata peak at the blastula, with the maximum expression of adult trunk and posterior genes of most phylostrata at post-larval stages. **g–l,** Expression dynamics of *in situ* hybridisation-validated anterior, trunk, and posterior markers throughout *O. fusiformis* (**g, h**), *C. teleta* (**i, j**), and *D. gyrociliatus* (**k, l**) development. Curves in **h**, **j,** and **l** are locally estimated scatterplot smoothings. Coloured shaded areas represent standard error of the mean. Key stages where expression of trunk markers is incipient are shown for both *O. fusiformis* and *C. teleta*.

Extended Data Figure 7 | Chromatin dynamics during the development of *O. fusiformis* and *C. teleta*. a, b, Genomic feature annotation of the consensus ATAC-seq peak sets of *O. fusiformis* (**a**) and *C. teleta* (**b**). **c, d,** Self-correlation matrices of normalised motif accessibility (**c**) and transcription factor binding score (**d**) during *O. fusiformis* (left) and *C. teleta* (right) development. Both matrices demonstrate distinct chromatin regulatory dynamics at each stage of development for both species. **e,** Correlation matrices of normalised motif accessibility to transcription factor binding score during *O. fusiformis* (top) and *C. teleta* (bottom) development. Matrices demonstrate a similarity diagonal between both variables for both species. **c, d,** and **e** further validate the non-triviality of the results obtained in Fig. 4e. **f,** Heatmap of normalised motif accessibility and transcription factor binding dynamics for each of the common annotated annelid motif archetypes during *O. fusiformis* and *C. teleta* development. Colour scale denotes transcription factor binding score dynamics, bubble size represents motif accessibility dynamics, both in a z-score scale. Motif archetypes highlighted in red are representative examples of the heterochronic shifts shown in bulk in Fig. 4e. **g, h,** Enrichment analysis of the number of occurrences of the common annotated

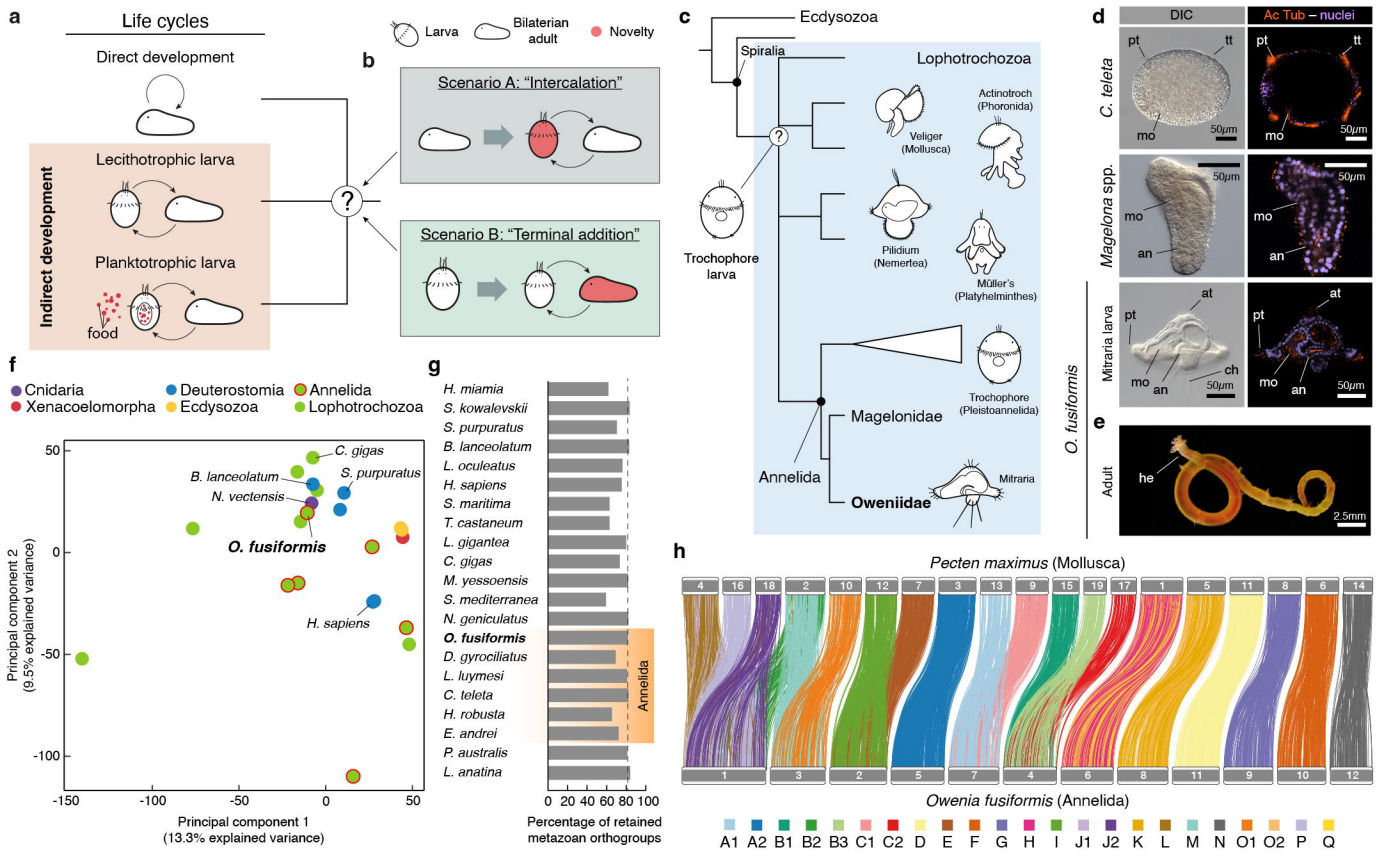
annelid motif archetypes in the peak clusters inferred through soft *k*-means clustering and shown in Fig. 4c, for *O. fusiformis* (**g**) and *C. teleta* (**h**). For each cluster and motif combination, the Bonferroni-adjusted *p*–value of the Fisher's exact test is shown. Red cells represent significantly overrepresented lineages (odds ratio, OR > 1, adjusted *p*–value < 0.05). Blue cells denote significantly underrepresented lineages (OR < 1, adjusted *p*–value < 0.05).

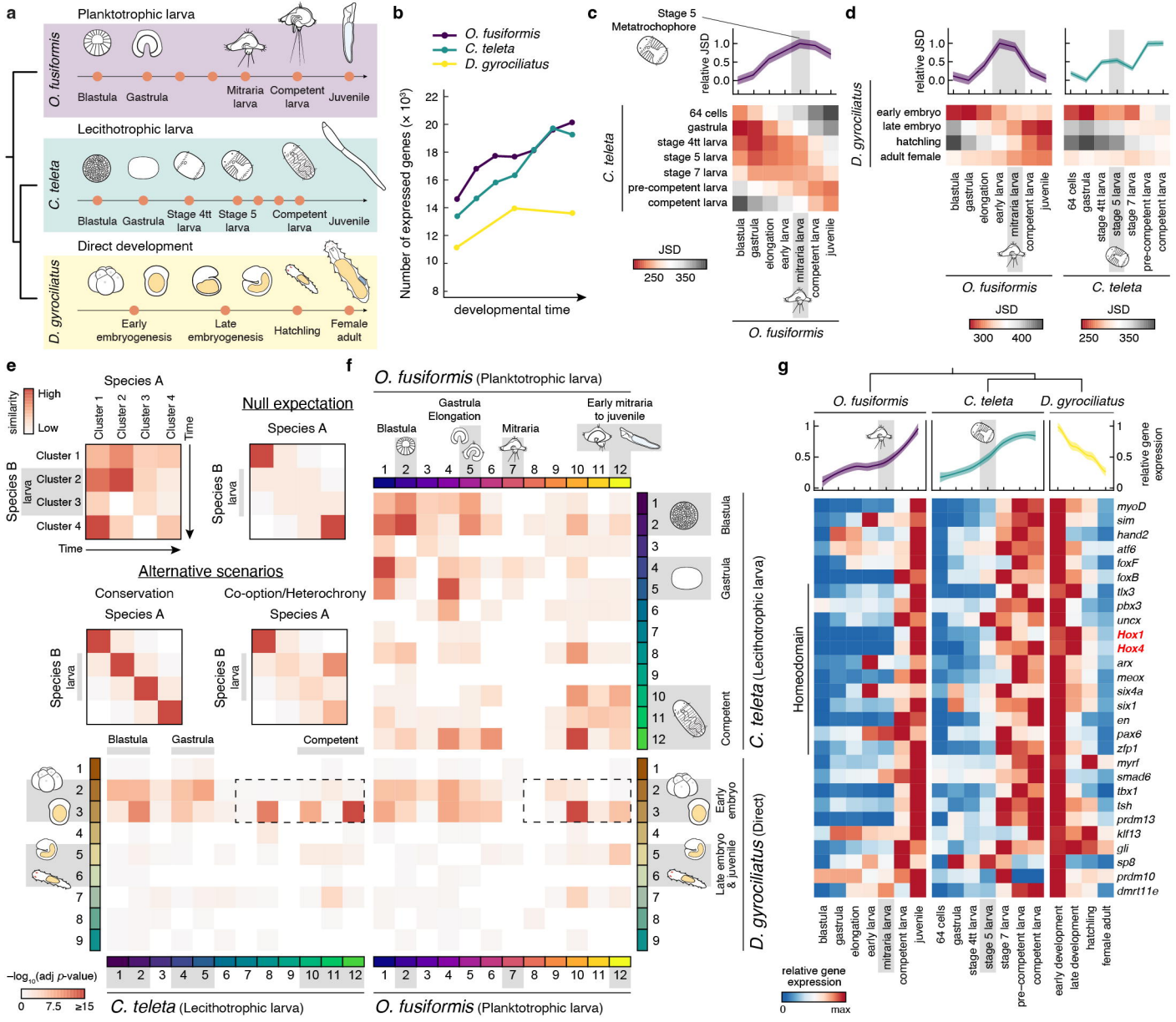
Extended Data Figure 8 | Phylostratography analyses in annelid life cycles. a, Cladogram of species used for comparative transcriptomics analyses, indicating on top the phylogenetic age of each phylostratum considered for phylostratigraphy analysis. For each species, the type of life cycle (direct/indirect), larval type (ciliated/juvenile-like) and larval nutritional mode (planktotrophic/lecithotrophy) are shown on the right. **b–d**, Expression contribution of each phylostratum by developmental stage in *O. fusiformis* (**b**), *C. teleta* (**c**), and *D. gyrociliatus* (**d**), calculated from 75% percentiles of quantile-normalised matrices of gene expression levels. Older genes are expressed at the highest levels across annelid development. **e–g,** Boxplots of quantile-normalised expression levels of genes classified by phylostratum across *O. fusiformis* (**e**), *C. teleta* (**f**), and *D. gyrociliatus* (**g**) development. A random subset of 2,000 genes is shown as a negative control. *N* denotes number of genes per phylostratum. **h–j,** Enrichment analysis of the number of genes per phylostratum in clusters of co-transcribed genes as inferred through soft *k*-means clustering and shown in Extended Data Fig. 3a–c, for *O. fusiformis* (**h**), *C. teleta* (**i**), and *D. gyrociliatus* (**j**). For each cluster and phylostratum combination, the Bonferroni-adjusted *p*–value of the Fisher's exact test is shown. Upper tables include significantly overrepresented lineages (odds ratio, OR > 1, adjusted *p*–value < 0.05). Lower tables include significantly underrepresented lineages (OR < 1, adjusted *p*–value < 0.05). Shaded grey areas indicate clusters of genes with peak expression at the mitraria larva (*O. fusiformis*) and stage 4–7 larval stages (*C. teleta*).

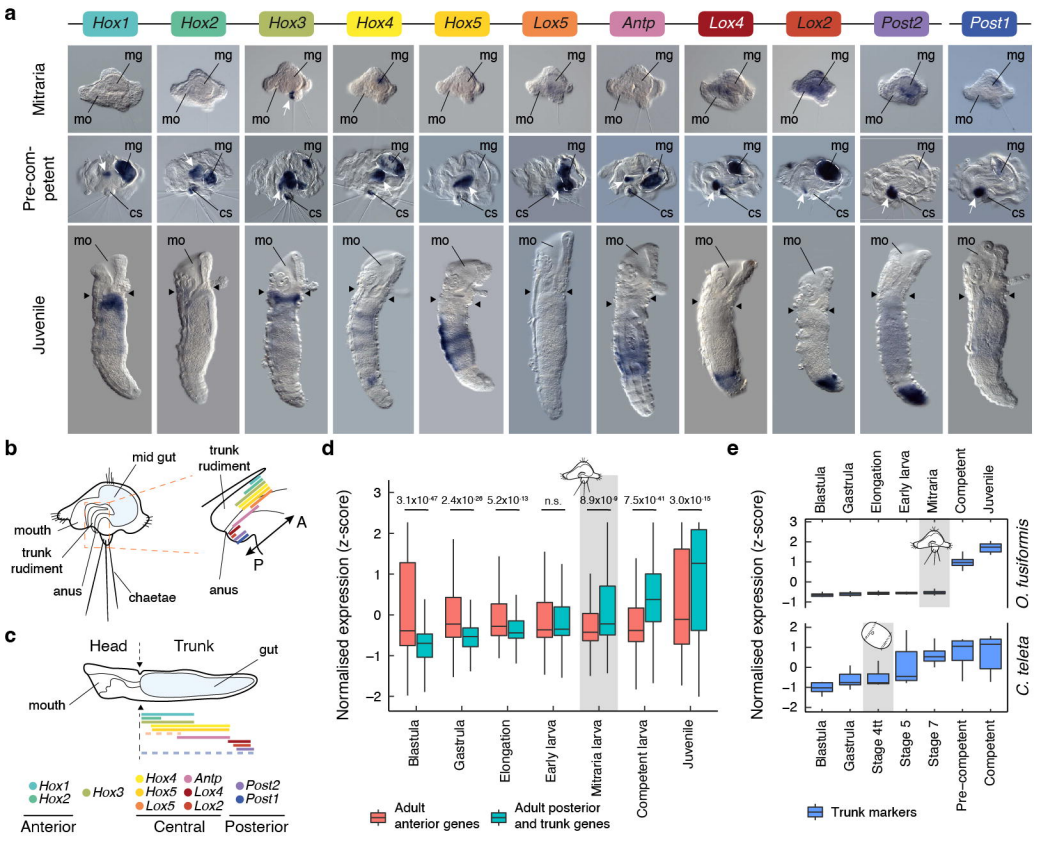
Extended Data Figure 9 | Bilaterian planktotrophic larvae and cnidarian larvae share maximal transcriptional similarity. a, Heatmaps of normalised transcriptomic Jensen–Shannon divergence (JSD) from pair-wise comparisons of all single copy one-to-one orthologs (left), the set of common orthologs to all species (centre), and all single copy one-to-one transcription factor orthologs (right), between *O. fusiformis* and ten other metazoan lineages with different life cycles. From top to bottom: the annelid *C. teleta*, the bivalve *C. gigas*, the nematode *C. elegans*, the insect *D. melanogaster*, the vertebrate *D. rerio*, the cephalochordate *B. lanceolatum*, the sea urchin *S. purpuratus*, the cnidarians *N. vectensis* and *C. hemisphaerica*, and the poriferan *A. queenslandica*. Larval stages are highlighted in green. **b,** Relative JSD for the datasets shown in **a**, from stages of minimal JSD to each *O. fusiformis* stage. Confidence intervals represent the standard deviation from 250 bootstrap resamplings of the ortholog sets. **c–e,** Stages of minimal JSD to each *O. fusiformis* stage, calculated from the one-to-one ortholog set (**c**), the common ortholog set (**d**), and the one-to-one transcription factor ortholog set (**e**). Larval stages are highlighted in green. **f**, Violin plots of the gene-wise Jensen Shannon divergence (gwJSD) distributions for the pair-wise comparisons of the one-to-one ortholog sets between the mitraria larva of *O. fusiformis* and the stages of minimal

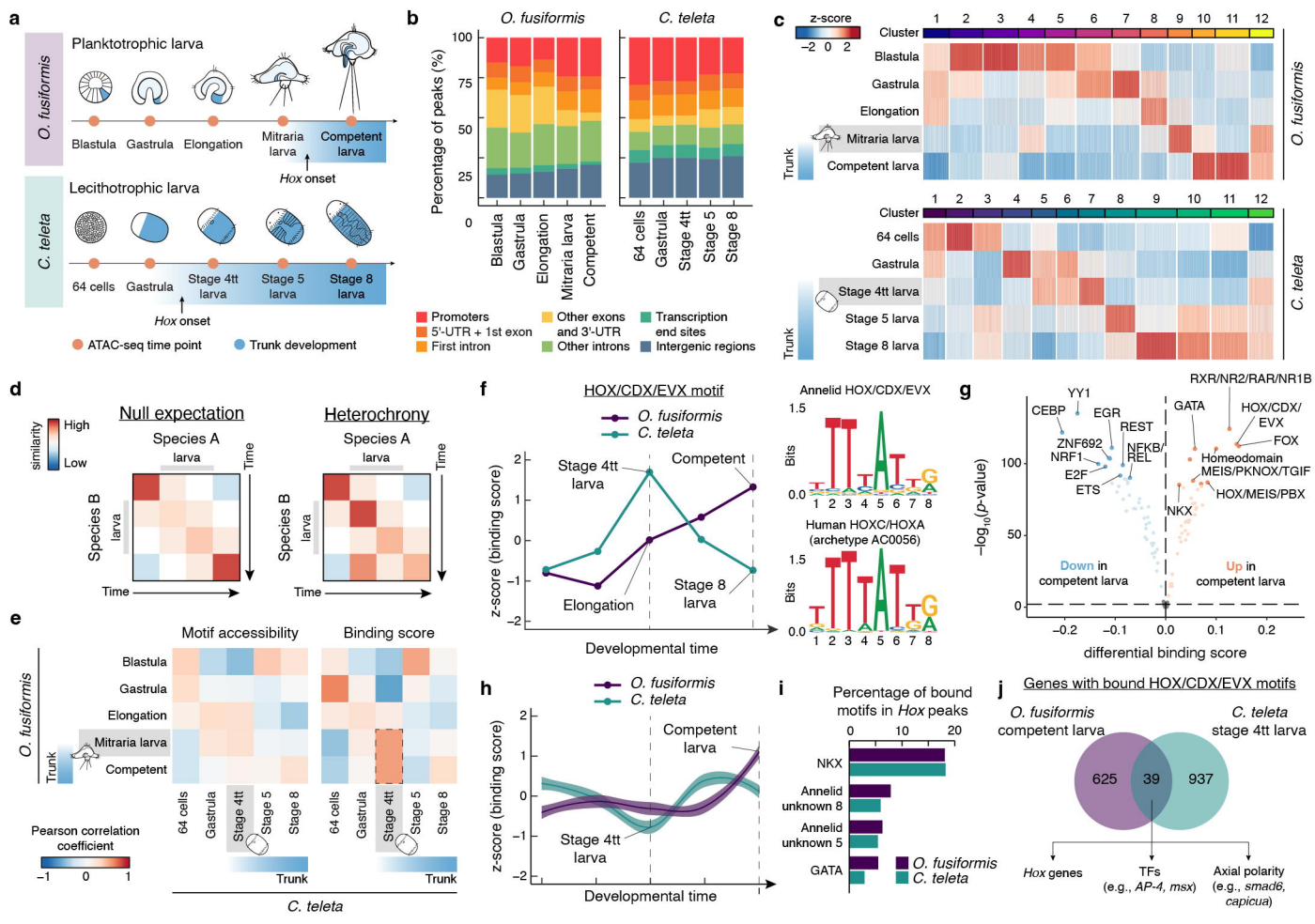
transcriptomic divergence as in **c**. for *C. gigas*, *C. hemisphaerica*, *C. teleta*, *N. vectensis*, and *S. purpuratus*. Boxes represent mean estimate and standard deviation. Dotted lines mark the point of highest probability density. Genes below ¼ of this point were subset as similarity-driving genes. **g,** Biological process GO terms enrichment of the five similarity-driving gene sets. GO terms were clustered by semantic similarity into 4 clusters. Each row represents a single GO term, for which the $-\log_{10}(p\text{–value})$ for each gene set is shown in a colour coded scale.

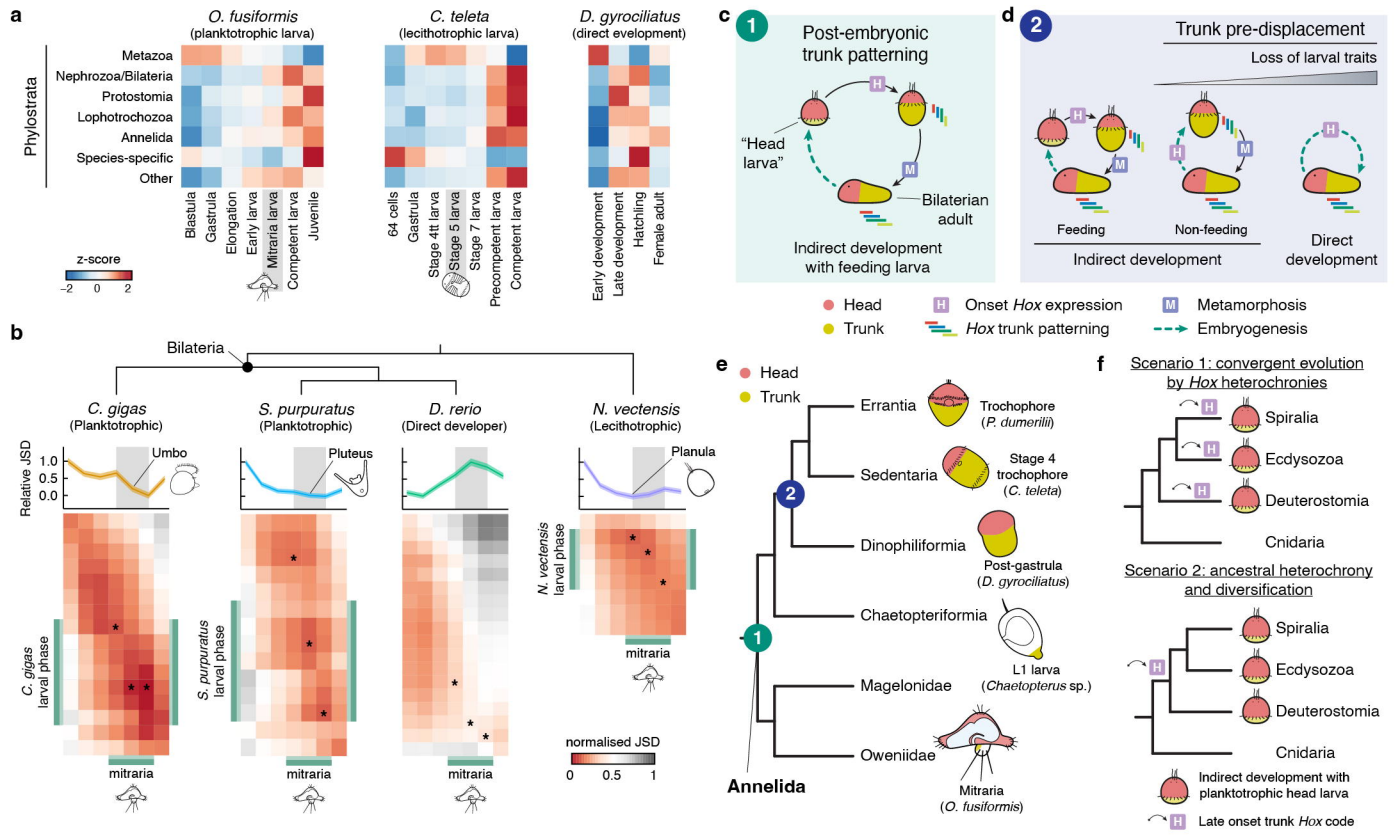
Extended Data Figure 10 | Comparative transcriptomic analysis of metazoan life cycles. **a,** Matrix of heatmaps of normalised transcriptomic Jensen–Shannon divergence (JSD) from pair-wise comparisons of all single copy one-to-one orthologs between all eleven metazoan lineages. From top to bottom and left to right: the annelids *O. fusiformis* and *C. teleta*, the bivalve *C. gigas*, the nematode *C. elegans*, the insect *D. melanogaster*, the vertebrate *D. rerio*, the cephalochordate *B. lanceolatum*, the sea urchin *S. purpuratus*, the cnidarians *N. vectensis* and *C. hemisphaerica*, and the poriferan *A. queenslandica*. **b,** Expression dynamics of *Hox* genes across the developmental RNA-seq time courses of all eleven species from **a** and the echiuran annelid *U. unicinctus*. Heatmaps were vertically aligned at the blastula, gastrula, and juvenile stages for all species. Lophotrochozoan lineages with trochophore larvae were also vertically aligned at the trochophore stage. Dotted lines encompass the larval stages of species with ciliated larvae. See Extended Data Figure 5d Supplementary Figure 32 for the fully labelled and non-deformed heatmaps. **c**, Alternative evolutionary scenarios for the deployment of *Hox* genes (as proxy for trunk patterning and assuming the staggered expression along the directive axis of cnidarians and anteroposterior axis of bilaterians is homologous, which does not necessarily imply homology of the two axes). Given our current understanding of *Hox* gene deployment in cnidarian and bilaterian taxa, a late post-embryonic *Hox* patterning ancestral to Bilateria and Cnidaria, as seen in extant lineages with maximal indirect development, is a more parsimonious scenario (on the right).

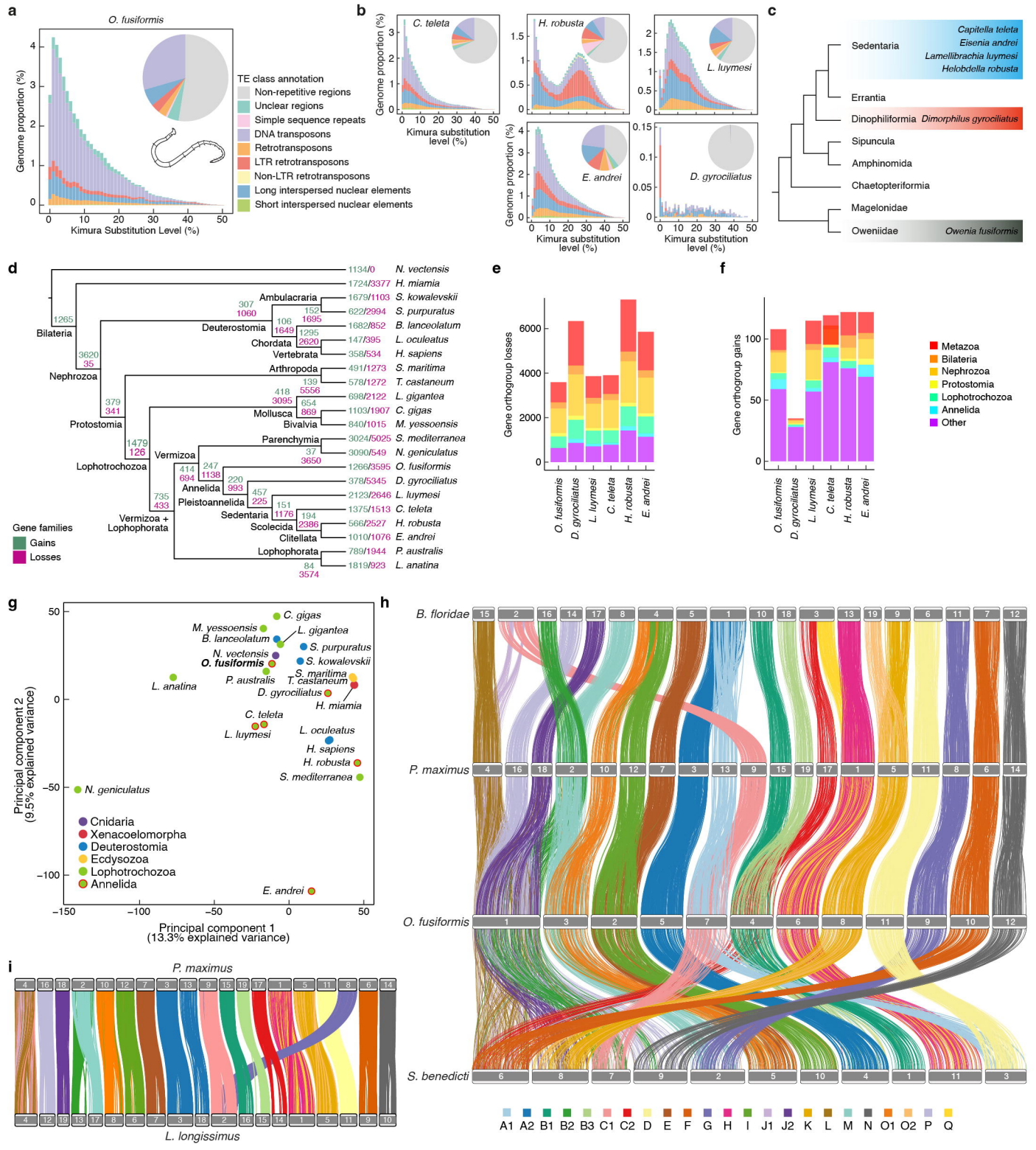

a
Life cycles
Larva
Bilaterian adult
Novelty
Direct development
Indirect development
Lecithotrophic larva
Planktotrophic larva
food
?
b
Scenario A: "Intercalation"
Scenario B: "Terminal addition"
c
Ecdysozoa
Spiralia
Lophotrochozoa
Actinotroch (Phoronida)
Veliger (Mollusca)
Pilidium (Nemertea)
Müller's (Platyhelminthes)
Trochophore larva
Trochophore (Pleistoannelida)
Magelonidae
Annelida
Oweniidae
Mitraria
d
DIC
Ac Tub – nuclei
C. teleta
Magelona spp.
O. fusiformis
Mitraria larva
pt
tt
mo
an
at
ch
50μm
e
Adult
he
2.5mm
f
Cnidaria
Deuterostomia
Annelida
Xenacoelomorpha
Ecdysozoa
Lophotrochozoa
C. gigas
B. lanceolatum
S. purpuratus
N. vectensis
O. fusiformis
H. sapiens
Principal component 2 (9.5% explained variance)
Principal component 1 (13.3% explained variance)
g
H. miamia
S. kowalevskii
S. purpuratus
B. lanceolatum
L. oculeatus
H. sapiens
S. maritima
T. castaneum
L. gigantea
C. gigas
M. yessoensis
S. mediterranea
N. geniculatus
O. fusiformis
D. gyrociliatus
L. luymesi
C. teleta
H. robusta
E. andrei
P. australis
L. anatina
Annelida
Percentage of retained metazoan orthogroups
h
Pecten maximus (Mollusca)
Owenia fusiformis (Annelida)
A1 A2 B1 B2 B3 C1 C2 D E F G H I J1 J2 K L M N O1 O2 P Q

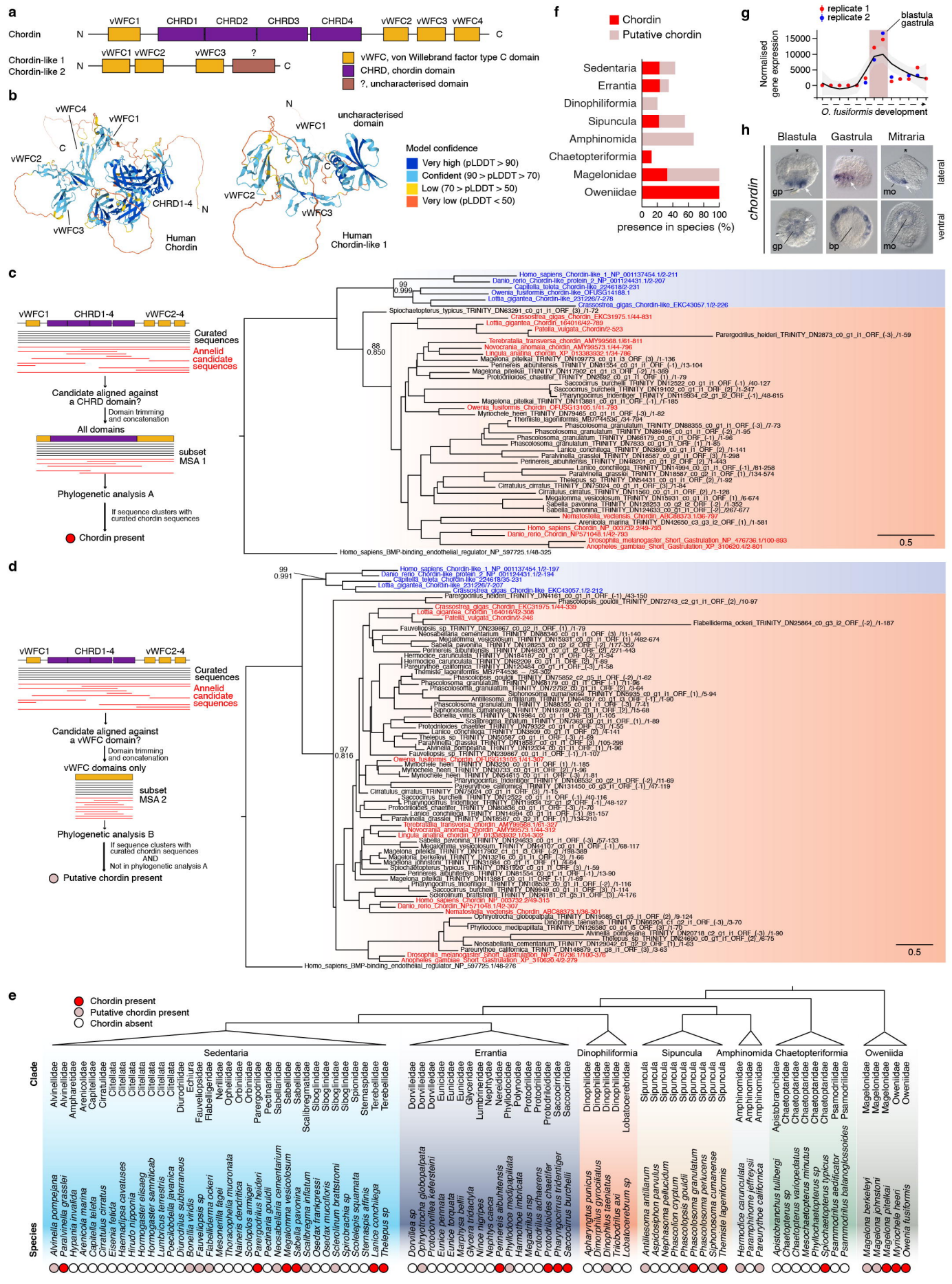

a
Planktotrophic larva
O. fusiformis
Blastula
Gastrula
Mitraria larva
Competent larva
Juvenile
Lecithotrophic larva
C. teleta
Blastula
Gastrula
Stage 4tt larva
Stage 5 larva
Competent larva
Juvenile
Direct development
D. gyrociliatus
Early embryogenesis
Late embryogenesis
Hatchling
Female adult
b
O. fusiformis
C. teleta
D. gyrociliatus
Number of expressed genes (× 10³)
developmental time
c
Stage 5
Metatrochophore
relative JSD
C. teleta
64 cells
gastrula
stage 4tt larva
stage 5 larva
stage 7 larva
pre-competent larva
competent larva
blastula
gastrula
elongation
early larva
mitraria larva
competent larva
juvenile
JSD
250 350
O. fusiformis
d
relative JSD
D. gyrociliatus
early embryo
late embryo
hatchling
adult female
blastula
gastrula
elongation
early larva
mitraria larva
competent larva
juvenile
64 cells
gastrula
stage 4tt larva
stage 5 larva
stage 7 larva
pre-competent larva
competent larva
O. fusiformis
C. teleta
JSD
300 400
JSD
250 350
e
Species A
similarity
High
Low
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Time
Species B
larva
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Time
Null expectation
Species A
Species B
larva
Alternative scenarios
Conservation
Co-option/Heterochrony
Species A
Species B
larva
Blastula
Gastrula
Competent
−log10(adj p-value)
0 7.5 ≥15
C. teleta (Lecithotrophic larva)
f
O. fusiformis (Planktotrophic larva)
Blastula
Gastrula Elongation
Mitraria
Early mitraria to juvenile
Blastula
Gastrula
Competent
C. teleta (Lecithotrophic larva)
Early embryo
Late embryo & juvenile
D. gyrociliatus (Direct)
O. fusiformis (Planktotrophic larva)
g
O. fusiformis
C. teleta
D. gyrociliatus
relative gene expression
Homeodomain
myoD
sim
hand2
atf6
foxF
foxB
tlx3
pbx3
uncx
Hox1
Hox4
arx
meox
six4a
six1
en
pax6
zfp1
myrf
smad6
tbx1
tsh
prdm13
klf13
gli
sp8
prdm10
dmrt11e
blastula
gastrula
elongation
early larva
mitraria larva
competent larva
juvenile
64 cells
gastrula
stage 4tt larva
stage 5 larva
stage 7 larva
pre-competent larva
competent larva
early development
late development
hatching
female adult
relative gene expression
0 max

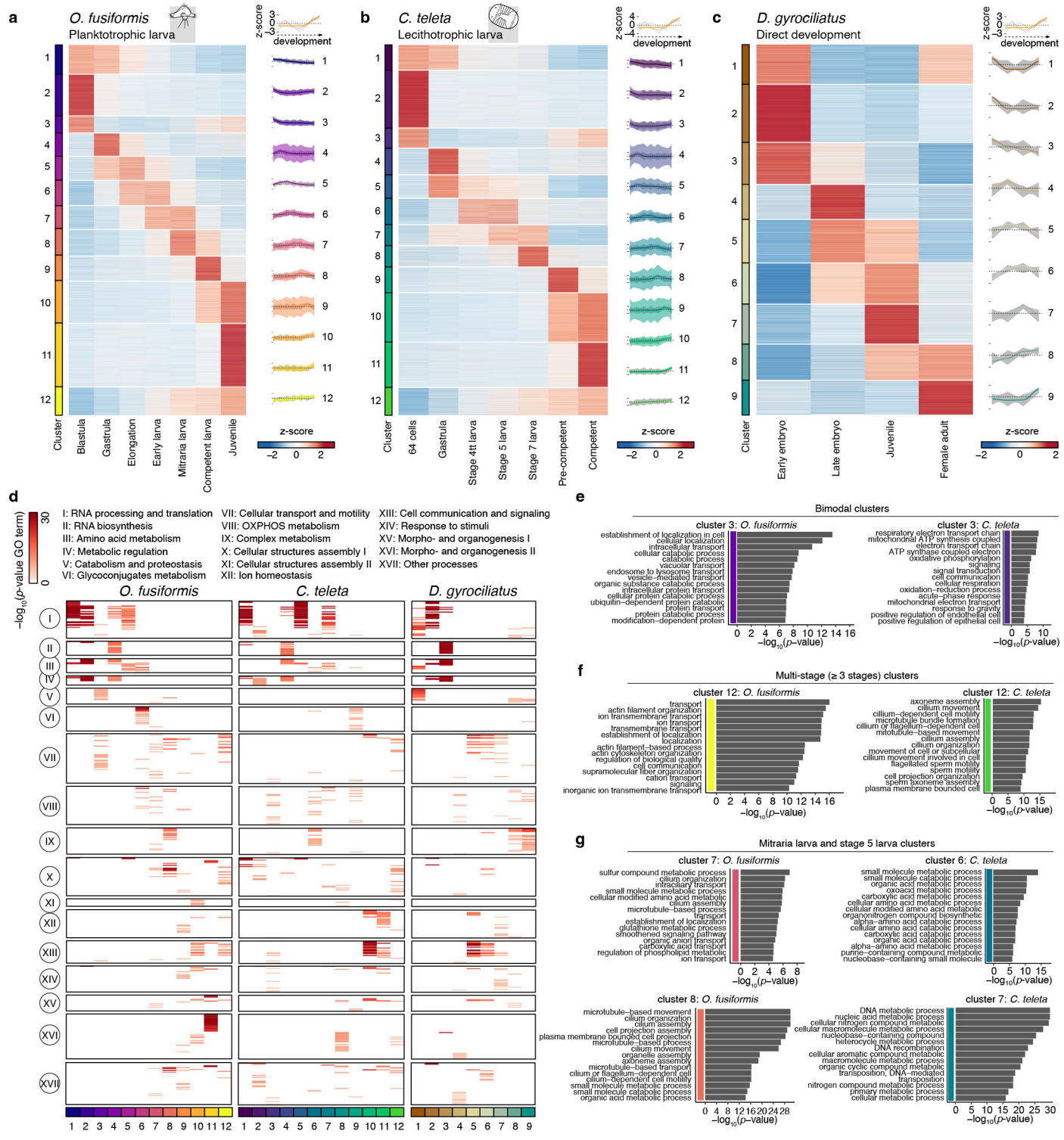

a
Hox1
Hox2
Hox3
Hox4
Hox5
Lox5
Antp
Lox4
Lox2
Post2
Post1
Mitraria
Pre-com-petent
Juvenile
mg
mo
cs
b
mid gut
trunk rudiment
mouth
trunk rudiment
anus
chaetae
anus
A
P
c
Head
Trunk
gut
mouth
Hox1
Hox2
Hox3
Hox4
Hox5
Lox5
Antp
Lox4
Lox2
Post2
Post1
Anterior
Central
Posterior
d
Normalised expression (z-score)
3.1x10-47
2.4x10-26
5.2x10-13
n.s.
8.9x10-9
7.5x10-41
3.0x10-15
Blastula
Gastrula
Elongation
Early larva
Mitraria larva
Competent larva
Juvenile
Adult anterior genes
Adult posterior and trunk genes
e
Blastula
Gastrula
Elongation
Early larva
Mitraria
Competent
Juvenile
O. fusiformis
C. teleta
Stage 4tt
Stage 5
Stage 7
Pre-competent
Competent
Trunk markers

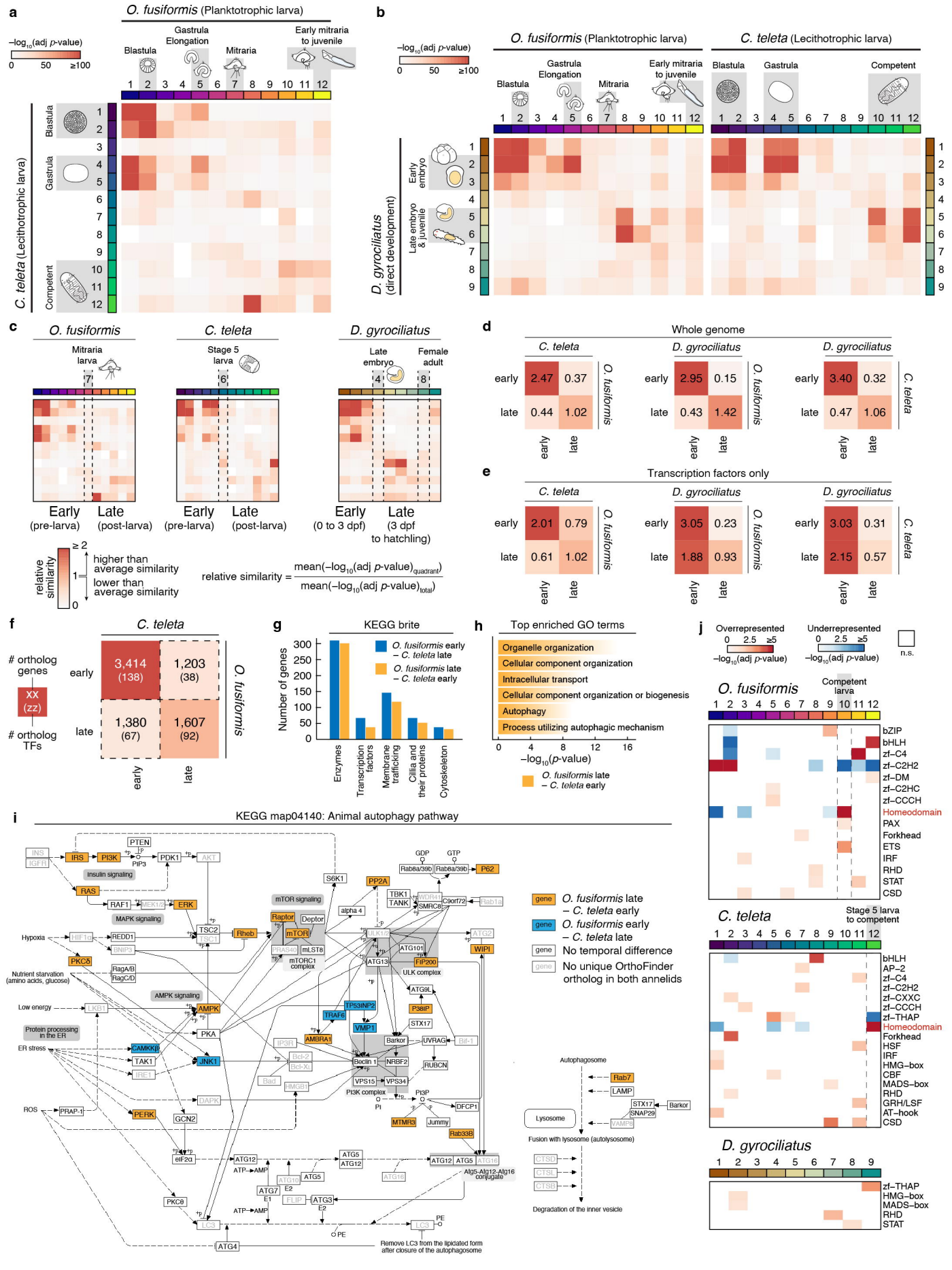

a
O. fusiformis
Planktotrophic larva
Blastula
Gastrula
Elongation
Mitraria larva
Competent larva
Hox onset
C. teleta
Lecithotrophic larva
64 cells
Gastrula
Stage 4tt larva
Stage 5 larva
Stage 8 larva
Hox onset
ATAC-seq time point
Trunk development
b
O. fusiformis
C. teleta
Percentage of peaks (%)
Blastula
Gastrula
Elongation
Mitraria larva
Competent
64 cells
Gastrula
Stage 4tt
Stage 5
Stage 8
Promoters
5'-UTR + 1st exon
First intron
Other exons and 3'-UTR
Other introns
Transcription end sites
Intergenic regions
c
z-score
Cluster
Blastula
Gastrula
Elongation
Mitraria larva
Competent larva
Trunk
O. fusiformis
64 cells
Gastrula
Stage 4tt larva
Stage 5 larva
Stage 8 larva
C. teleta
d
similarity
High
Low
Null expectation
Heterochrony
Species A
Species B
larva
Time
e
Motif accessibility
Binding score
Blastula
Gastrula
Elongation
Mitraria larva
Competent
O. fusiformis
Trunk
64 cells
Gastrula
Stage 4tt
Stage 5
Stage 8
Pearson correlation coefficient
C. teleta
f
HOX/CDX/EVX motif
O. fusiformis
C. teleta
Stage 4tt larva
Competent
Elongation
Stage 8 larva
z-score (binding score)
Developmental time
Annelid HOX/CDX/EVX
Human HOXC/HOXA (archetype AC0056)
Bits
g
YY1
CEBP
EGR
REST
ZNF692
NRF1
NFKB/REL
E2F
ETS
GATA
RXR/NR2/RAR/NR1B
HOX/CDX/EVX
FOX
Homeodomain
MEIS/PKNOX/TGIF
HOX/MEIS/PBX
NKX
Down in competent larva
Up in competent larva
−log10(p-value)
differential binding score
h
O. fusiformis
C. teleta
Competent larva
Stage 4tt larva
z-score (binding score)
Developmental time
i
Percentage of bound motifs in Hox peaks
NKX
Annelid unknown 8
Annelid unknown 5
GATA
O. fusiformis
C. teleta
j
Genes with bound HOX/CDX/EVX motifs
O. fusiformis competent larva
C. teleta stage 4tt larva
625
39
937
Hox genes
TFs (e.g., AP-4, msx)
Axial polarity (e.g., smad6, capicua)

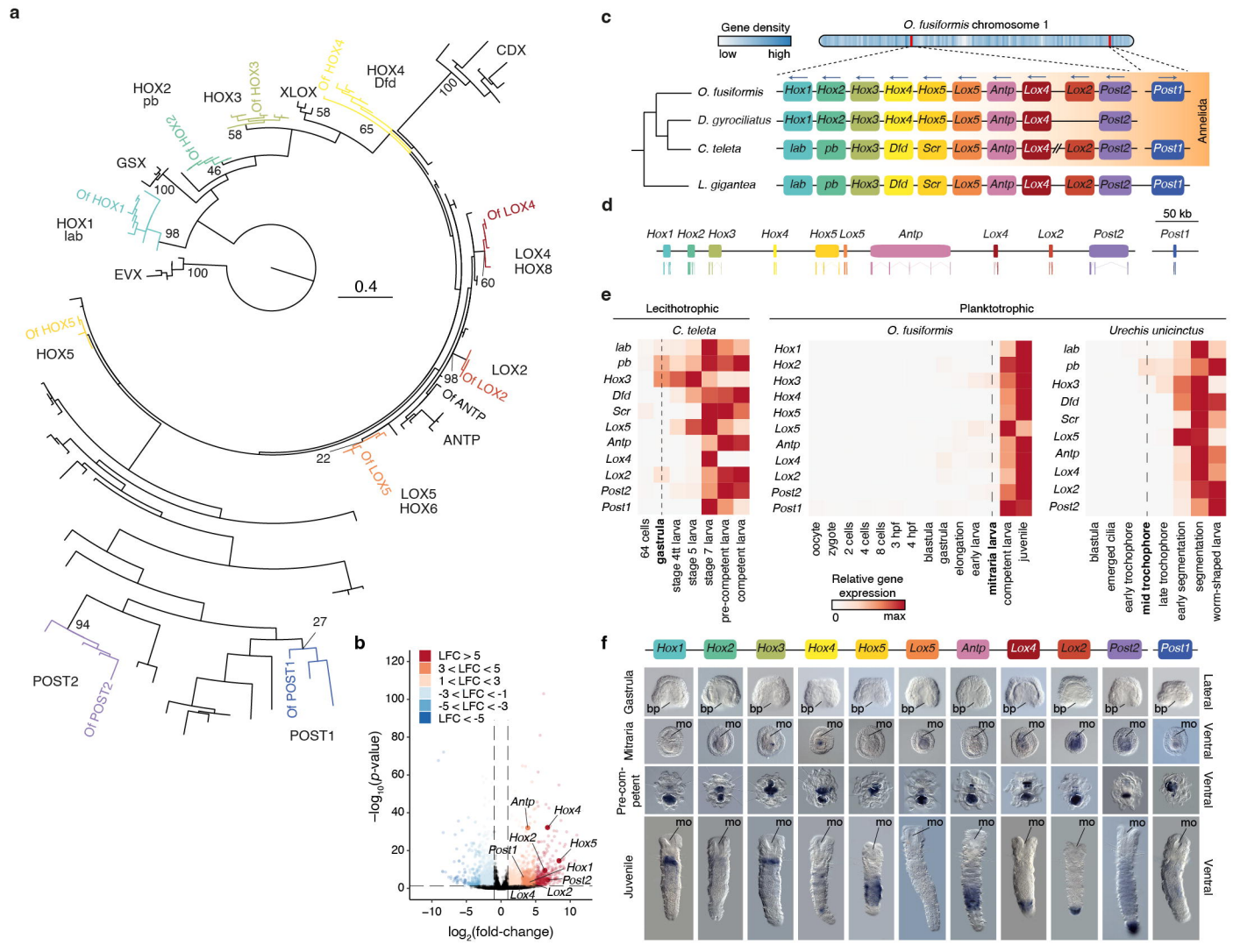

a
O. fusiformis
(planktotrophic larva)
C. teleta
(lecithotrophic larva)
D. gyrociliatus
(direct development)
Phylostrata
Metazoa
Nephrozoa/Bilateria
Protostomia
Lophotrochozoa
Annelida
Species-specific
Other
Blastula
Gastrula
Elongation
Early larva
Mitraria larva
Competent larva
Juvenile
64 cells
Gastrula
Stage 4tt larva
Stage 5 larva
Stage 7 larva
Precompetent larva
Competent larva
Early development
Late development
Hatchling
Female adult
z-score
−2
0
2
b
Bilateria
C. gigas
(Planktotrophic)
S. purpuratus
(Planktotrophic)
D. rerio
(Direct developer)
N. vectensis
(Lecithotrophic)
Relative JSD
1.0
0.5
0.0
Umbo
Pluteus
Planula
C. gigas
larval phase
S. purpuratus
larval phase
N. vectensis
larval phase
mitraria
normalised JSD
0
0.5
1
c
1
Post-embryonic
trunk patterning
"Head
larva"
H
M
Bilaterian
adult
Indirect development
with feeding larva
d
2
Trunk pre-displacement
Loss of larval traits
Feeding
Non-feeding
Indirect development
Direct
development
Head
Trunk
Onset Hox expression
Hox trunk patterning
Metamorphosis
Embryogenesis
e
Head
Trunk
Errantia
Sedentaria
Dinophiliformia
Chaetopteriformia
Magelonidae
Oweniidae
Annelida
Trochophore
(P. dumerilii)
Stage 4
trochophore
(C. teleta)
Post-gastrula
(D. gyrociliatus)
L1 larva
(Chaetopterus sp.)
Mitraria
(O. fusiformis)
f
Scenario 1: convergent evolution
by Hox heterochronies
Spiralia
Ecdysozoa
Deuterostomia
Cnidaria
Scenario 2: ancestral heterochrony
and diversification
Spiralia
Ecdysozoa
Deuterostomia
Cnidaria
Indirect development with
planktotrophic head larva
Late onset trunk Hox code

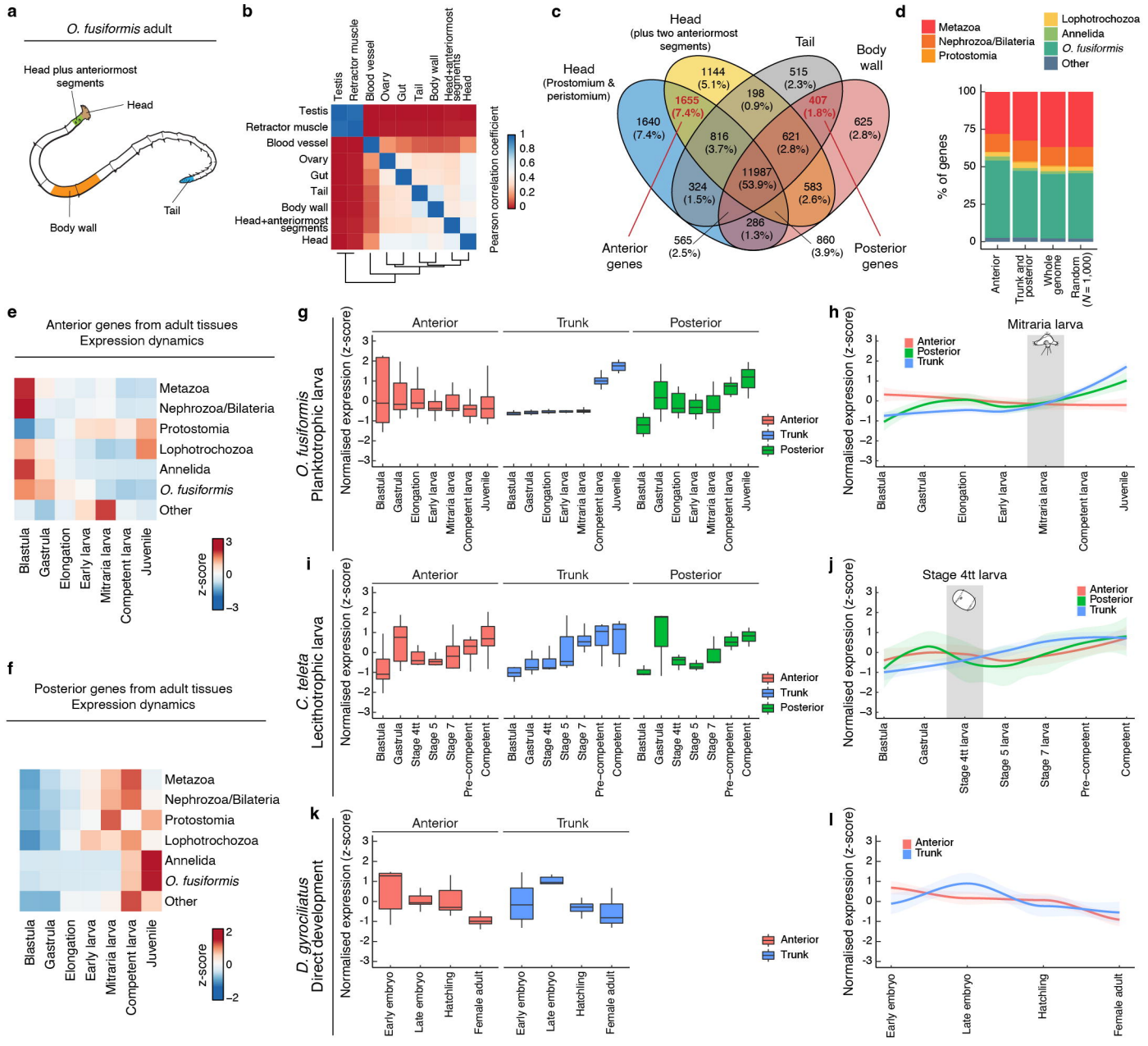

a
O. fusiformis
Genome proportion (%)
Kimura Substitution Level (%)
TE class annotation
Non-repetitive regions
Unclear regions
Simple sequence repeats
DNA transposons
Retrotransposons
LTR retrotransposons
Non-LTR retrotransposons
Long interspersed nuclear elements
Short interspersed nuclear elements
b
C. teleta
H. robusta
L. luymesi
E. andrei
D. gyrociliatus
Kimura substitution level (%)
c
Sedentaria
Capitella teleta
Eisenia andrei
Lamellibrachia luymesi
Helobdella robusta
Errantia
Dinophiliformia
Dimorphilus gyrociliatus
Sipuncula
Amphinomida
Chaetopteriformia
Magelonidae
Oweniidae
Owenia fusiformis
d
Bilateria
Nephrozoa
Protostomia
Lophotrochozoa
Vermizoa + Lophophorata
Deuterostomia
Chordata
Vertebrata
Arthropoda
Mollusca
Bivalvia
Parenchymia
Vermizoa
Annelida
Pleistoannelida
Sedentaria
Scolecida
Clitellata
Lophophorata
N. vectensis
H. miamia
S. kowalevskii
S. purpuratus
B. lanceolatum
L. oculeatus
H. sapiens
S. maritima
T. castaneum
L. gigantea
C. gigas
M. yessoensis
S. mediterranea
N. geniculatus
O. fusiformis
D. gyrociliatus
L. luymesi
C. teleta
H. robusta
E. andrei
P. australis
L. anatina
Gene families
Gains
Losses
e
Gene orthogroup losses
O. fusiformis
D. gyrociliatus
L. luymesi
C. teleta
H. robusta
E. andrei
f
Gene orthogroup gains
Metazoa
Bilateria
Nephrozoa
Protostomia
Lophotrochozoa
Annelida
Other
g
Principal component 2 (9.5% explained variance)
Principal component 1 (13.3% explained variance)
Cnidaria
Xenacoelomorpha
Deuterostomia
Ecdysozoa
Lophotrochozoa
Annelida
h
B. floridae
P. maximus
O. fusiformis
S. benedicti
A1 A2 B1 B2 B3 C1 C2 D E F G H I J1 J2 K L M N O1 O2 P Q
i
P. maximus
L. longissimus

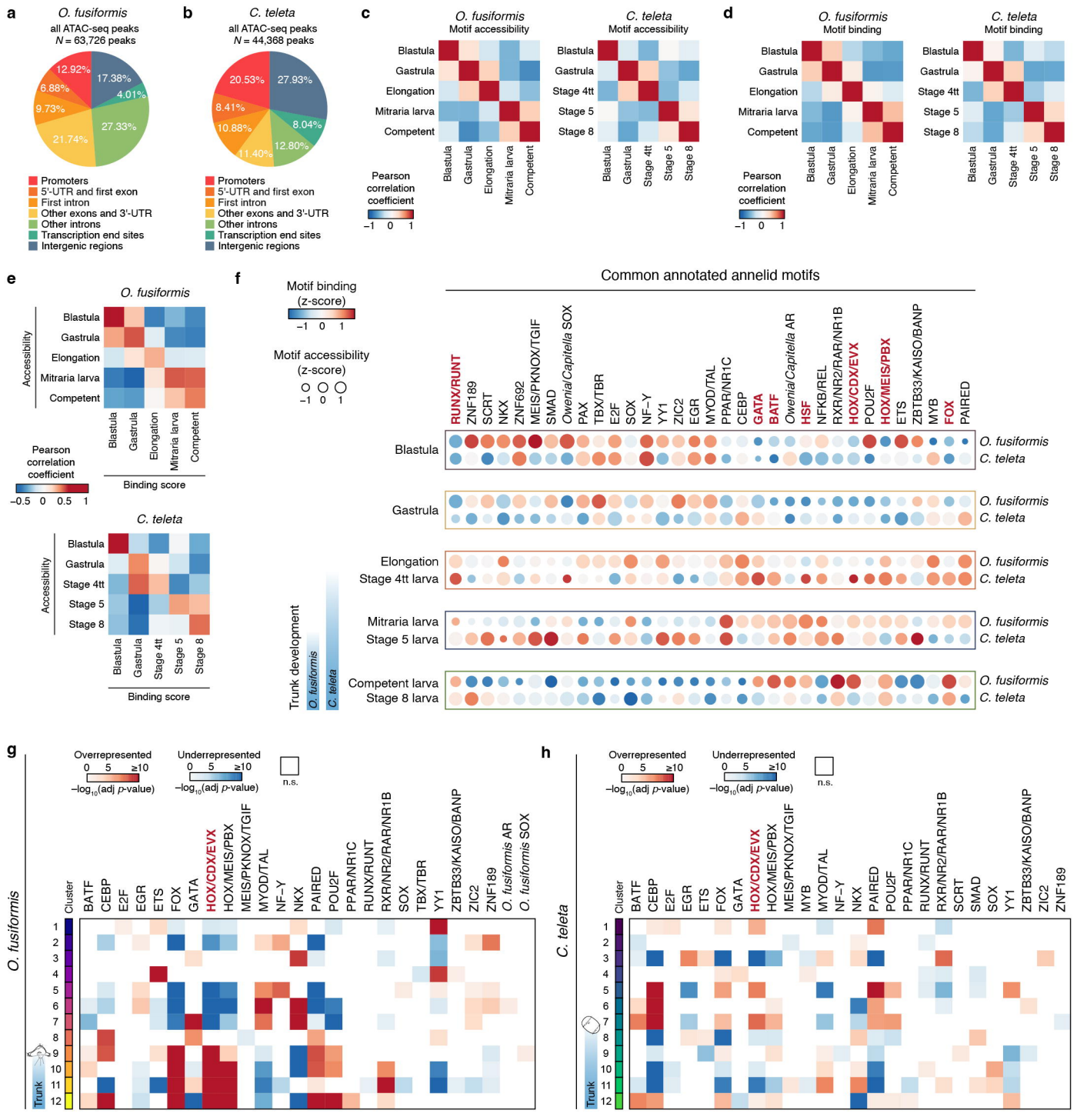

a
Chordin
N
vWFC1
CHRD1
CHRD2
CHRD3
CHRD4
vWFC2
vWFC3
vWFC4
C
Chordin-like 1
Chordin-like 2
vWFC1
vWFC2
vWFC3
?
vWFC, von Willebrand factor type C domain
CHRD, chordin domain
?, uncharacterised domain
b
vWFC4
vWFC1
vWFC2
CHRD1-4
vWFC3
Human Chordin
uncharacterised domain
Human Chordin-like 1
Model confidence
Very high (pLDDT > 90)
Confident (90 > pLDDT > 70)
Low (70 > pLDDT > 50)
Very low (pLDDT < 50)
f
Chordin
Putative chordin
Sedentaria
Errantia
Dinophiliformia
Siputncula
Amphinomida
Chaetopteriformia
Magelonidae
Oweniidae
presence in species (%)
g
replicate 1
replicate 2
blastula
gastrula
Normalised gene expression
O. fusiformis development
h
Blastula
Gastrula
Mitraria
chordin
lateral
ventral
gp
mo
bp
c
vWFC1
CHRD1-4
vWFC2-4
Curated sequences
Annelid candidate sequences
Candidate aligned against a CHRD domain?
Domain trimming and concatenation
All domains
subset MSA 1
Phylogenetic analysis A
If sequence clusters with curated chordin sequences
Chordin present
Homo_sapiens_Chordin-like_1_NP_001137454.1/2-211
Danio_rerio_Chordin-like_protein_2_NP_001124431.1/2-207
Capitella_teleta_Chordin-like_224618/2-231
Owenia_fusiformis_chordin-like_OFUSG141188.1
Lottia_gigantea_Chordin-like_2312267/2-278
Crassostrea_gigas_Chordin-like_EKC43057.1/2-226
Spiochaetopterus_typicus_TRINITY_DN63291_c0_g1_i1_ORF_(3)_/1-72
Crassostrea_gigas_Chordin_EKC31975.1/44-831
Lottia_gigantea_Chordin_164016/42-789
Patella_vulgata_Chordin/2-523
Parergodrilus_heideri_TRINITY_DN2873_c0_g1_i1_ORF_(-3)_/1-59
Terebratalia_transversa_chordin_AMY99568.1/61-811
Novocrania_anomala_chordin_AMY99573.1/44-796
Lingula_anatina_chordin_XP_013383932.1/34-786
Magelona_pitelkai_TRINITY_DN109773_c0_g1_i3_ORF_(3)_/1-136
Perinereis_albuhitensis_TRINITY_DN81554_c0_g1_i1_ORF_(-1)_/13-104
Magelona_pitelkai_TRINITY_DN117902_c1_g1_i3_ORF_(-2)_/1-389
Protodriloides_chaetifer_TRINITY_DN2692_c0_g1_i1_ORF_(1)_/1-79
Saccocirrus_burchelli_TRINITY_DN12522_c0_g1_i1_ORF_(-1)_/40-127
Saccocirrus_burchelli_TRINITY_DN19102_c0_g1_i1_ORF_(2)_/1-247
Pharyngocirrus_tridentiger_TRINITY_DN119934_c2_g1_i2_ORF_(-1)_/48-615
Magelona_pitelkai_TRINITY_DN113881_c0_g1_i1_ORF_(-1)_/1-185
Owenia_fusiformis_Chordin_OFUSG13105.1/41-793
Myriochele_heeri_TRINITY_DN79465_c0_g1_i1_ORF_(-3)_/1-82
Themiste_lageniformis_MB7P44536_/34-794
Phascolosoma_granulatum_TRINITY_DN88355_c0_g1_i1_ORF_(-3)_/7-73
Phascolosoma_granulatum_TRINITY_DN89496_c0_g1_i1_ORF_(-2)_/1-95
Phascolosoma_granulatum_TRINITY_DN68179_c0_g1_i1_ORF_(-1)_/1-96
Phascolosoma_granulatum_TRINITY_DN7833_c0_g1_i1_ORF_(1)_/1-85
Lanice_conchilega_TRINITY_DN53809_c0_g1_i1_ORF_(2)_/1-141
Paralvinella_grasslei_TRINITY_DN18587_c0_g1_i1_ORF_(3)_/1-298
Perinereis_albuhitensis_TRINITY_DN84201_c0_g1_i2_ORF_(2)_/1-443
Lanice_conchilega_TRINITY_DN114994_c0_g1_i1_ORF_(-1)_/81-258
Paralvinella_grasslei_TRINITY_DN18587_c0_g2_i1_ORF_(1)_/134-574
Thelepus_sp_TRINITY_DN54431_c0_g1_i1_ORF_(2)_/1-92
Cirratulus_cirratus_TRINITY_DN75024_c0_g1_i1_ORF_(3)_/1-84
Cirratulus_cirratus_TRINITY_DN11560_c0_g1_i1_ORF_(2)_/1-128
Megalomma_vesiculosum_TRINITY_DN15931_c0_g1_i1_ORF_(1)_/6-674
Sabella_pavonina_TRINITY_DN126253_c0_g2_i2_ORF_(-2)_/1-352
Sabella_pavonina_TRINITY_DN124633_c0_g1_i1_ORF_(-2)_/267-677
Nematostella_vectensis_Chordin_ABC88373.1/36-797
Arenicola_marina_TRINITY_DN42650_c3_g3_i2_ORF_(1)_/1-581
Homo_sapiens_Chordin_NP_003732.2/49-793
Danio_rerio_Chordin_NP571048.1/42-793
Drosophila_melanogaster_Short_Gastrulation_NP_476736.1/100-893
Anopheles_gambiae_Short_Gastrulation_XP_310620.4/2-801
Homo_sapiens_BMP-binding_endothelial_regulator_NP_597725.1/48-325
0.5
d
vWFC1
CHRD1-4
vWFC2-4
Curated sequences
Annelid candidate sequences
Candidate aligned against a vWFC domain?
Domain trimming and concatenation
vWFC domains only
subset MSA 2
Phylogenetic analysis B
If sequence clusters with curated chordin sequences AND Not in phylogenetic analysis A
Putative chordin present
Homo_sapiens_Chordin-like_1_NP_001137454.1/2-197
Danio_rerio_Chordin-like_protein_2_NP_001124431.1/2-194
Capitella_teleta_Chordin-like_224618/35-231
Lottia_gigantea_Chordin-like_2312267/207
Crassostrea_gigas_Chordin-like_EKC43057.1/2-212
Parergodrilus_heideri_TRINITY_DN4161_c0_g1_i1_ORF_(-1)_/43-150
Phascolopsis_gouldii_TRINITY_DN72743_c2_g1_i1_ORF_(2)_/10-97
Crassostrea_gigas_Chordin_EKC31975.1/44-339
Lottia_gigantea_Chordin_164016/42-308
Patella_vulgata_Chordin/2-246
Flabelliderma_ockeri_TRINITY_DN25864_c0_g3_i2_ORF_(-2)_/1-187
Fauvelopsis_sp_TRINITY_DN239867_c0_g2_i1_ORF_(1)_/1-79
Neosabellaria_cementarium_TRINITY_DN88346_c0_g1_i1_ORF_(3)_/11-140
Megalomma_vesiculosum_TRINITY_DN15931_c0_g1_i1_ORF_(1)_/482-674
Sabella_pavonina_TRINITY_DN126253_c0_g2_i2_ORF_(-2)_/177-352
Perinereis_albuhitensis_TRINITY_DN84201_c0_g1_i2_ORF_(2)_/271-443
Hermodice_carunculata_TRINITY_DN184187_c0_g1_i1_ORF_(-2)_/1-34
Hermodice_carunculata_TRINITY_DN62209_c0_g1_i1_ORF_(2)_/1-89
Pareurythoe_californica_TRINITY_DN120484_c0_g1_i1_ORF_(-3)_/1-58
Themiste_lageniformis_MB7P44536_/34-309
Phascolopsis_gouldii_TRINITY_DN75852_c2_g5_i1_ORF_(-2)_/1-62
Phascolosoma_granulatum_TRINITY_DN68179_c0_g1_i1_ORF_(-1)_/11-96
Phascolosoma_granulatum_TRINITY_DN72792_c0_g1_i1_ORF_(2)_/3-64
Siphonosoma_cumanense_TRINITY_DN5935_c0_g1_i1_ORF_(1)_/5-94
Antillesoma_antillarum_TRINITY_DN64897_c0_g1_i3_ORF_(-1)_/1-90
Phascolosoma_granulatum_TRINITY_DN88355_c0_g1_i1_ORF_(-3)_/7-41
Siphonosoma_cumanense_TRINITY_DN19789_c0_g1_i1_ORF_(2)_/15-68
Bonellia_viridis_TRINITY_DN19964_c0_g1_i1_ORF_(-3)_/1-105
Scalibregma_inflatum_TRINITY_DN7369_c0_g1_i1_ORF_(1)_/1-89
Protodriloides_chaetifer_TRINITY_DN79322_c0_g1_i1_ORF_(-3)_/1-55
Lanice_conchilega_TRINITY_DN53809_c0_g1_i1_ORF_(2)_/4-141
Thelepus_sp_TRINITY_DN50587_c0_g1_i1_ORF_(-3)_/1-69
Paralvinella_grasslei_TRINITY_DN18587_c0_g1_i1_ORF_(3)_/105-298
Alvinella_pompejana_TRINITY_DN12334_c0_g1_i1_ORF_(1)_/1-96
Fauvelopsis_sp_TRINITY_DN239867_c0_g1_i1_ORF_(-1)_/1-107
Owenia_fusiformis_Chordin_OFUSG13105.1/41-327
Myriochele_heeri_TRINITY_DN3250_c0_g1_i1_ORF_(1)_/1-185
Myriochele_heeri_TRINITY_DN39733_c0_g1_i1_ORF_(2)_/1-96
Myriochele_heeri_TRINITY_DN54615_c0_g1_i1_ORF_(-3)_/1-81
Pharyngocirrus_tridentiger_TRINITY_DN108532_c0_g2_i1_ORF_(-2)_/11-69
Pareurythoe_californica_TRINITY_DN131450_c0_g3_i1_ORF_(-1)_/47-119
Cirratulus_cirratus_TRINITY_DN75024_c0_g1_i1_ORF_(3)_/1-15
Saccocirrus_burchelli_TRINITY_DN12522_c0_g1_i1_ORF_(-1)_/40-116
Pharyngocirrus_tridentiger_TRINITY_DN119934_c2_g1_i2_ORF_(-1)_/48-127
Protodriloides_chaetifer_TRINITY_DN80836_c0_g1_i1_ORF_(-3)_/1-70
Lanice_conchilega_TRINITY_DN14994_c0_g1_i1_ORF_(-1)_/81-157
Paralvinella_grasslei_TRINITY_DN18587_c0_g2_i1_ORF_(1)_/134-210
Terebratalia_transversa_chordin_AMY99568.1/61-307
Novocrania_anomala_chordin_AMY99573.1/44-312
Lingula_anatina_chordin_XP_013383932.1/34-302
Sabella_pavonina_TRINITY_DN124633_c0_g1_i1_ORF_(-3)_/57-133
Megalomma_vesiculosum_TRINITY_DN44107_c0_g1_i1_ORF_(-1)_/68-117
Magelona_pitelkai_TRINITY_DN117902_c1_g1_i3_ORF_(-2)_/198-389
Magelona_berkeleyi_TRINITY_DN13216_c0_g1_i1_ORF_(2)_/1-66
Magelona_johnstoni_TRINITY_DN31994_c0_g1_i1_ORF_(1)_/4-64
Spiochaetopterus_typicus_TRINITY_DN31920_c0_g1_i1_ORF_(3)_/1-59
Perinereis_albuhitensis_TRINITY_DN81554_c0_g1_i1_ORF_(-1)_/13-90
Magelona_pitelkai_TRINITY_DN113881_c0_g1_i1_ORF_(-1)_/1-69
Pharyngocirrus_tridentiger_TRINITY_DN108532_c0_g1_i1_ORF_(-2)_/1-116
Saccocirrus_burchelli_TRINITY_DN89949_c0_g1_i1_ORF_(3)_/1-114
Scleroilunum_brattstromii_TRINITY_DN26181_c1_g5_i1_ORF_(3)_/4-176
Homo_sapiens_Chordin_NP_003732.2/49-319
Danio_rerio_Chordin_NP571048.1/42-307
Nematostella_vectensis_Chordin_ABC88373.1/36-301
Ophryotrocha_globopalpata_TRINITY_DN19585_c1_g5_i1_ORF_(2)_/9-124
Dinophilus_taeniatus_TRINITY_DN68204_c1_g2_i1_ORF_(-3)_/3-70
Phyllodoce_medipapillata_TRINITY_DN126580_c0_g4_i5_ORF_(3)_/1-70
Alvinella_pompejana_TRINITY_DN20718_c2_g1_i1_ORF_(-3)_/1-90
Thelepus_sp_TRINITY_DN24690_c0_g1_i1_ORF_(2)_/6-75
Neosabellaria_cementarium_TRINITY_DN129042_c1_g2_i3_ORF_(1)_/1-63
Pareurythoe_californica_TRINITY_DN148879_c1_g8_i1_ORF_(3)_/3-83
Drosophila_melanogaster_Short_Gastrulation_NP_476736.1/100-376
Anopheles_gambiae_Short_Gastrulation_XP_310620.4/2-279
Homo_sapiens_BMP-binding_endothelial_regulator_NP_597725.1/48-276
0.5
e
Chordin present
Putative chordin present
Chordin absent
Clade
Species
Sedentaria
Errantia
Dinophiliformia
Sipuncula
Amphinomida
Chaetopteriformia
Oweniida
Alvinellidae
Alvinellidae
Ampharetidae
Arenicolidae
Capitellidae
Cirratulidae
Citellata
Citellata
Citellata
Citellata
Citellata
Citellata
Citellata
Citellata
Diurodrilidae
Echiura
Fauveliopsis sp
Flabelliderdae
Nerillidae
Opheliidae
Orbiniidae
Orbiniidae
Parergodrilidae
Pectinariidae
Sabellariidae
Sabellidae
Sabellidae
Scalibregmatidae
Siboglinidae
Siboglinidae
Siboglinidae
Siboglinidae
Siboglinidae
Siboglinidae
Spionidae
Stemaspidae
Terebellidae
Terebellidae
Dorvilleidae
Dorvilleidae
Dorvilleidae
Eunicidae
Eunicidae
Eunicidae
Glyceridae
Lumbrinereidae
Nephtyidae
Nereididae
Phyllodocidae
Polynoidae
Protodrilidae
Protodrilidae
Saccocirridae
Saccocirridae
Dinophilidae
Dinophilidae
Dinophilidae
Dinophilidae
Lobatocerebridae
Sipuncula
Sipuncula
Sipuncula
Sipuncula
Sipuncula
Sipuncula
Sipuncula
Sipuncula
Amphinomidae
Amphinomidae
Amphinomidae
Apistobranchidae
Chaetopteridae
Chaetopteridae
Chaetopteridae
Chaetopteridae
Chaetopteridae
Chaetopteridae
Psammodrilidae
Psammodrilidae
Magelonidae
Magelonidae
Magelonidae
Oweniidae
Oweniidae
Alvinella pompejana
Paralvinella grasslei
Hypania invalida
Arenicola marina
Capitella teleta
Cirratulus cirratus
Eisenia fetida
Haemadipsa cavatuses
Hirudo nipponia
Hormogaster elisaeg
Hormogaster samnitica
Lumbricus terrestris
Poecilobdella javanica
Diurodrilus subterraneus
Bonellia viridis
Fauveliopsis sp
Flabelliderma ockeri
Mesonerilla fagei
Thoracophelia mucronata
Naineris dendritica
Scoloplos armiger
Parergodrilus heideri
Pectinaria gouldii
Neosabellaria cementarium
Megalomma vesiculosum
Sabella pavonina
Scalibregma inflatum
Osedax frankpressi
Osedax mucofloris
Scoleolunum brattstromi
Spirobrachia sp
Spiochaetopterus typicus
Stemaspis affinis
Lanice conchilega
Thelepus sp
Dorvillea sp
Ophryotrocha globopalpata
Protodorvillea kefersteini
Eunice pennata
Eunice torquata
Marphysa bellii
Glycera tridactyla
Ninoe nigripes
Nephtys caeca
Perinereis albuhitensis
Phyllodoce medipapillata
Harmothoe imbricata
Megadrilus sp
Protodrilus acharens
Protodriloides chaetifer
Pharyngocirrus tridentiger
Saccocirrus burchelli
Apharyngtus punicus
Dimorphilus gyrociliatus
Dinophilus taeniatus
Trilobodrilus axi
Lobatocerebrum sp
Antillesoma antillarum
Aspidosiphon parvulus
Nephasoma pellucidum
Phascolosoma granulatum
Phascolosoma perlucens
Phascolosoma scolopsis
Siphonosoma cumanense
Themiste lageniformis
Hermodice carunculata
Paramphinome jeffreysii
Pareurythoe californica
Apistobranchus tullbergi
Chaetopterus sp
Chaetopterus variopedatus
Mesochaetopterus minutus
Phyllochaetopterus sp
Spiochaetopterus typicus
Psammodrilus aedificator
Psammodrilus balanoglossoides
Magelona berkeleyi
Magelona johnstoni
Magelona pitelkai
Myriochele heeri
Owenia fusiformis

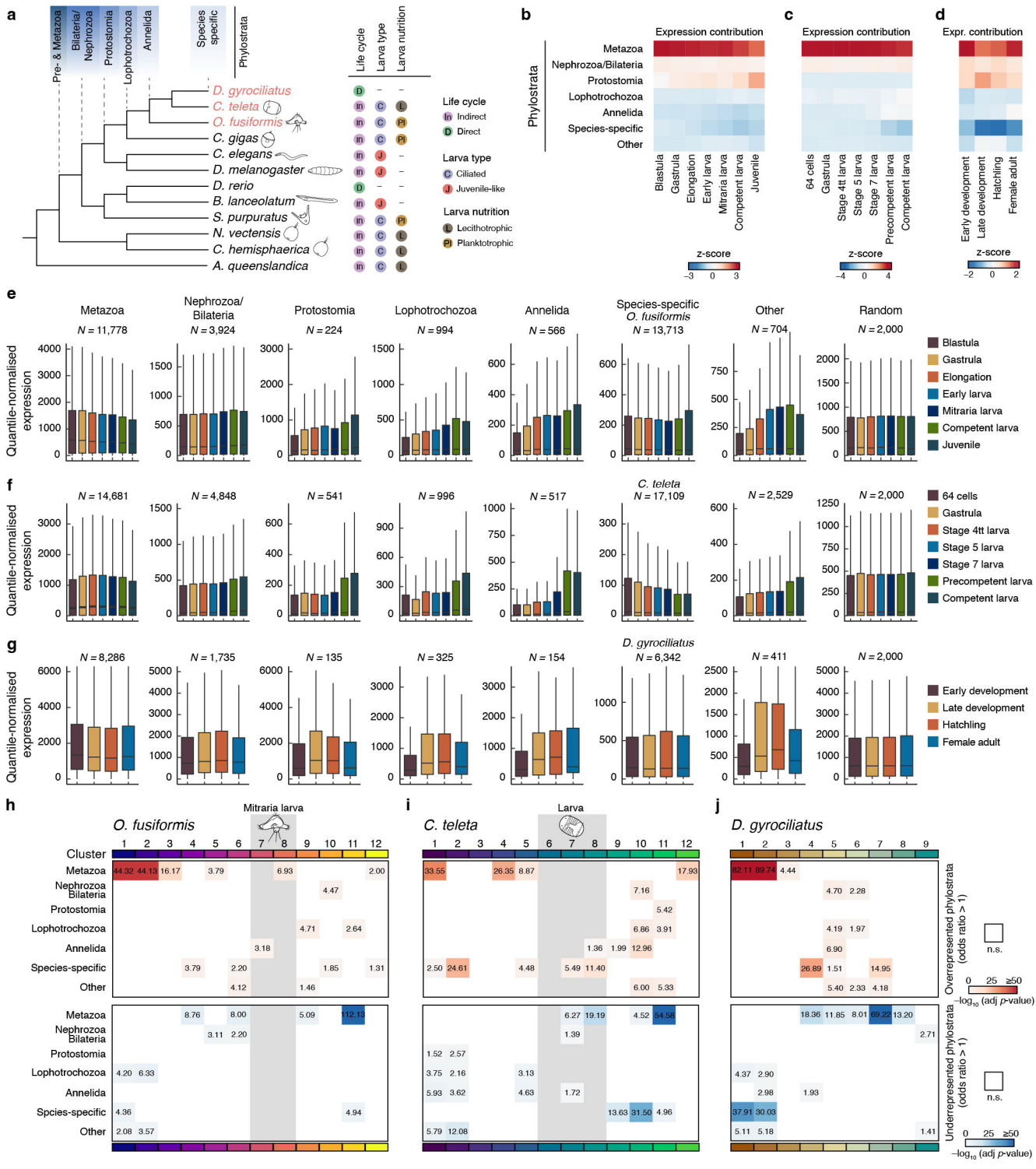

a
O. fusiformis
Planktotrophic larva
z-score
development
Cluster
Blastula
Gastrula
Elongation
Early larva
Mitraria larva
Competent larva
Juvenile
z-score
b
C. teleta
Lecithotrophic larva
64 cells
Gastrula
Stage 4tt larva
Stage 5 larva
Stage 7 larva
Pre-competent
Competent
c
D. gyrociliatus
Direct development
Early embryo
Late embryo
Juvenile
Female adult
d
I: RNA processing and translation
II: RNA biosynthesis
III: Amino acid metabolism
IV: Metabolic regulation
V: Catabolism and proteostasis
VI: Glycoconjugates metabolism
VII: Cellular transport and motility
VIII: OXPHOS metabolism
IX: Complex metabolism
X: Cellular structures assembly I
XI: Cellular structures assembly II
XII: Ion homeostasis
XIII: Cell communication and signaling
XIV: Response to stimuli
XV: Morpho- and organogenesis I
XVI: Morpho- and organogenesis II
XVII: Other processes
−log10(p-value GO term)
e
Bimodal clusters
cluster 3: O. fusiformis
establishment of localization in cell
cellular localization
intracellular transport
cellular catabolic process
catabolic process
vacuolar transport
endosome to lysosome transport
vesicle−mediated transport
organic substance catabolic process
intracellular protein transport
cellular protein catabolic process
ubiquitin−dependent protein catabolic
protein transport
protein catabolic process
modification−dependent protein
−log10(p−value)
cluster 3: C. teleta
respiratory electron transport chain
mitochondrial ATP synthesis coupled
electron transport chain
ATP synthase coupled electron
oxidative phosphorylation
signaling
signal transduction
cell communication
cellular respiration
oxidation−reduction process
acute−phase response
mitochondrial electron transport
response to gravity
positive regulation of endothelial cell
positive regulation of epithelial cell
f
Multi-stage (≥ 3 stages) clusters
cluster 12: O. fusiformis
transport
actin filament organization
ion transmembrane transport
ion transport
transmembrane transport
establishment of localization
localization
actin filament−based process
actin cytoskeleton organization
regulation of biological quality
cell communication
supramolecular fiber organization
cation transport
signaling
inorganic ion transmembrane transport
cluster 12: C. teleta
axoneme assembly
cilium movement
cilium−dependent cell motility
microtubule bundle formation
cilium or flagellum−dependent cell
mitotubule−based movement
cilium assembly
cilium organization
movement of cell or subcellular
cilium movement involved in cell
flagellated sperm motility
sperm motility
cell projection organization
sperm axoneme assembly
plasma membrane bounded cell
g
Mitraria larva and stage 5 larva clusters
cluster 7: O. fusiformis
sulfur compound metabolic process
cilium organization
intraciliary transport
small molecule metabolic process
cellular modified amino acid metabolic
cilium assembly
microtubule−based process
transport
establishment of localization
glutathione metabolic process
smoothened signaling pathway
organic anion transport
carboxylic acid transport
regulation of phospholipid metabolic
ion transport
cluster 6: C. teleta
small molecule metabolic process
small molecule catabolic process
organic acid metabolic process
oxoacid metabolic process
carboxylic acid metabolic process
cellular amino acid metabolic process
cellular modified amino acid metabolic
organonitrogen compound biosynthetic
alpha−amino acid catabolic process
cellular amino acid catabolic process
carboxylic acid catabolic process
organic acid catabolic process
alpha−amino acid metabolic process
purine−containing compound metabolic
nucleobase−containing small molecule
cluster 8: O. fusiformis
microtubule−based movement
cilium organization
cilium assembly
cell projection assembly
plasma membrane bounded cell projection
microtubule−based process
cilium movement
organelle assembly
axoneme assembly
microtubule−based transport
cilium or flagellum−dependent cell
cilium−dependent cell motility
small molecule metabolic process
small molecule catabolic process
organic acid metabolic process
cluster 7: C. teleta
DNA metabolic process
nucleic acid metabolic process
cellular nitrogen compound metabolic
cellular macromolecule metabolic process
nucleobase−containing compound
heterocycle metabolic process
DNA recombination
cellular aromatic compound metabolic
macromolecule metabolic process
organic cyclic compound metabolic
transposition, DNA−mediated
transposition
nitrogen compound metabolic process
primary metabolic process
cellular metabolic process

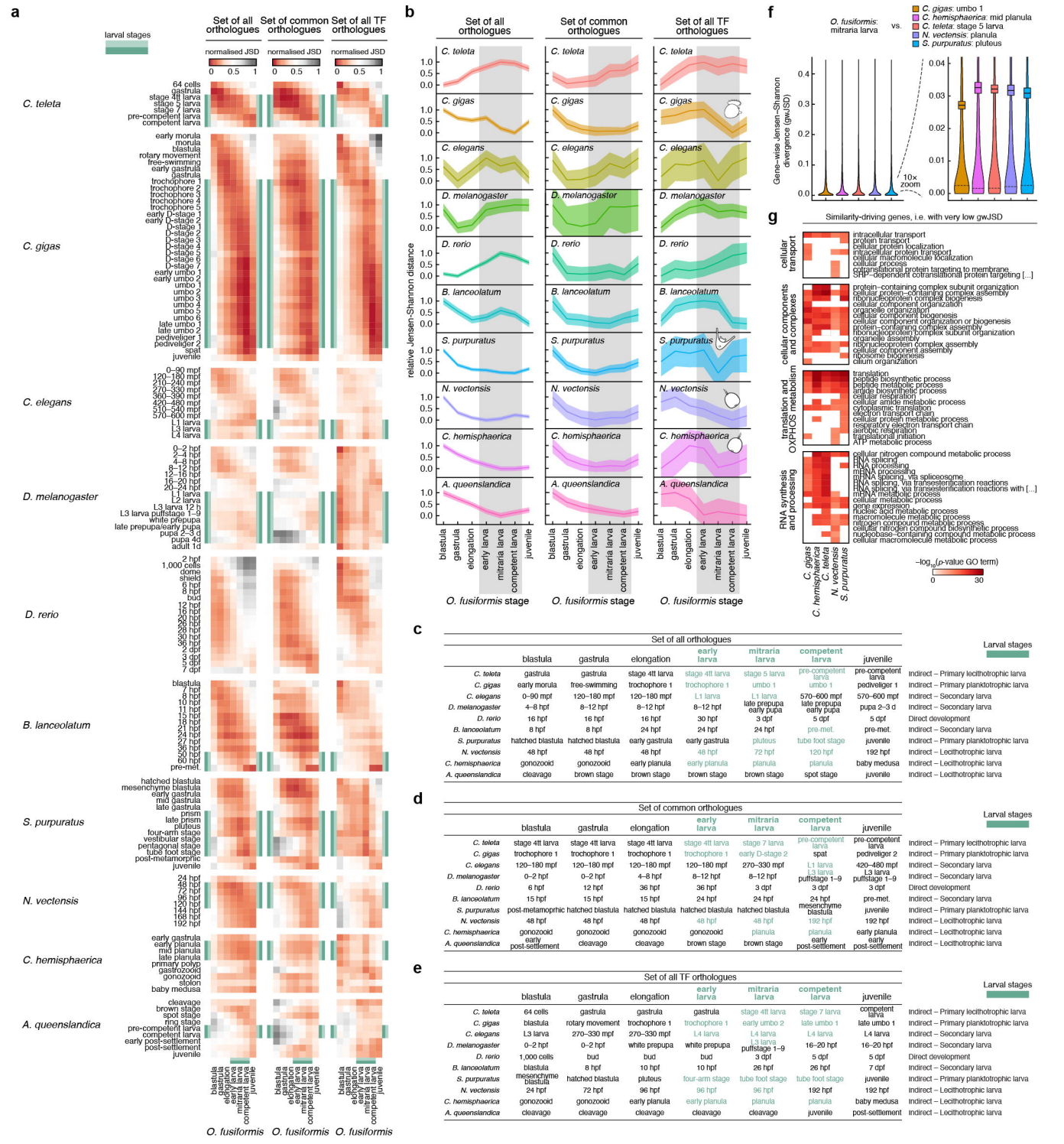

a
O. fusiformis (Planktotrophic larva)
–log10(adj p-value)
0 50 ≥100
Blastula
Gastrula Elongation
Mitraria
Early mitraria to juvenile
C. teleta (Lecithotrophic larva)
Blastula
Gastrula
Competent
b
O. fusiformis (Planktotrophic larva)
C. teleta (Lecithotrophic larva)
Blastula
Gastrula Elongation
Mitraria
Early mitraria to juvenile
Gastrula
Competent
D. gyrociliatus (direct development)
Early embryo
Late embryo & juvenile
c
O. fusiformis
C. teleta
D. gyrociliatus
Mitraria larva
Stage 5 larva
Late embryo
Female adult
Early (pre-larva)
Late (post-larva)
Early (pre-larva)
Late (post-larva)
Early (0 to 3 dpf)
Late (3 dpf to hatchling)
relative similarity
higher than average similarity
lower than average similarity
relative similarity = mean(–log10(adj p-value)quadrant) / mean(–log10(adj p-value)total)
d
Whole genome
C. teleta
early 2.47 0.37
late 0.44 1.02
O. fusiformis
D. gyrociliatus
early 2.95 0.15
late 0.43 1.42
O. fusiformis
D. gyrociliatus
early 3.40 0.32
late 0.47 1.06
C. teleta
e
Transcription factors only
C. teleta
early 2.01 0.79
late 0.61 1.02
O. fusiformis
D. gyrociliatus
early 3.05 0.23
late 1.88 0.93
O. fusiformis
D. gyrociliatus
early 3.03 0.31
late 2.15 0.57
C. teleta
f
C. teleta
ortholog genes
XX (ZZ)
ortholog TFs
early late
3,414 (138)
1,203 (38)
1,380 (67)
1,607 (92)
O. fusiformis
g
KEGG brite
Number of genes
O. fusiformis early – C. teleta late
O. fusiformis late – C. teleta early
Enzymes
Transcription factors
Membrane trafficking
Cilia and their proteins
Cytoskeleton
h
Top enriched GO terms
Organelle organization
Cellular component organization
Intracellular transport
Cellular component organization or biogenesis
Autophagy
Process utilizing autophagic mechanism
–log10(p-value)
O. fusiformis late – C. teleta early
i
KEGG map04140: Animal autophagy pathway
gene O. fusiformis late – C. teleta early
gene O. fusiformis early – C. teleta late
gene No temporal difference
gene No unique OrthoFinder ortholog in both annelids
Autophagosome
Lysosome
Fusion with lysosome (autolysosome)
Degradation of the inner vesicle
Remove LC3 from the lipidated form after closure of the autophagosome
j
Overrepresented
Underrepresented
–log10(adj p-value)
n.s.
O. fusiformis
Competent larva
bZIP
bHLH
zf–C4
zf–C2H2
zf–DM
zf–C2HC
zf–CCCH
Homeodomain
PAX
Forkhead
ETS
IRF
RHD
STAT
CSD
C. teleta
Stage 5 larva to competent
bHLH
AP–2
zf–C4
zf–C2H2
zf–CXXC
zf–CCCH
zf–THAP
Homeodomain
Forkhead
HSF
IRF
HMG–box
CBF
MADS–box
RHD
GRH/LSF
AT–hook
CSD
D. gyrociliatus
zf–THAP
HMG–box
MADS–box
RHD
STAT


a
HOX2
pb
HOX3
Of HOX3
XLOX
Of HOX4
HOX4
Dfd
CDX
Of HOX2
GSX
Of HOX1
HOX1
lab
EVX
0.4
Of LOX4
LOX4
HOX8
Of HOX5
HOX5
LOX2
Of LOX2
Of ANTP
ANTP
Of LOX5
LOX5
HOX6
POST2
Of POST2
Of POST1
POST1
b
LFC > 5
3 < LFC < 5
1 < LFC < 3
-3 < LFC < -1
-5 < LFC < -3
LFC < -5
−log10(p-value)
log2(fold-change)
Antp
Hox4
Hox2
Hox5
Post1
Hox1
Post2
Lox4
Lox2
c
Gene density
low
high
O. fusiformis chromosome 1
O. fusiformis
D. gyrociliatus
C. teleta
L. gigantea
Annelida
d
50 kb
e
Lecithotrophic
Planktotrophic
C. teleta
O. fusiformis
Urechis unicinctus
64 cells
gastrula
stage 4tt larva
stage 5 larva
stage 7 larva
pre-competent larva
competent larva
oocyte
zygote
2 cells
4 cells
8 cells
3 hpf
4 hpf
blastula
gastrula
elongation
early larva
mitraria larva
competent larva
juvenile
blastula
emerged cilia
early trochophore
mid trochophore
late trochophore
early segmentation
segmentation
late segmentation
worm
Relative gene expression
0
max
f
Gastrula
Mitraria
Pre-com-petent
Juvenile
Lateral
Ventral
bp
mo


a
O. fusiformis adult
Head plus anteriormost segments
Head
Tail
Body wall
b
Testis
Retractor muscle
Blood vessel
Ovary
Gut
Tail
Body wall
Head+anteriormost segments
Head
Pearson correlation coefficient
1
0.8
0.6
0.4
0.2
0
c
Head (Prostomium & peristomium)
Head (plus two anteriormost segments)
Tail
Body wall
1640 (7.4%)
1144 (5.1%)
515 (2.3%)
625 (2.8%)
1655 (7.4%)
198 (0.9%)
407 (1.8%)
816 (3.7%)
621 (2.8%)
11987 (53.9%)
324 (1.5%)
583 (2.6%)
286 (1.3%)
565 (2.5%)
860 (3.9%)
Anterior genes
Posterior genes
d
Metazoa
Nephrozoa/Bilateria
Protostomia
Lophotrochozoa
Annelida
O. fusiformis
Other
% of genes
100
75
50
25
0
Anterior
Trunk and posterior
Whole genome
Random (N = 1,000)
e
Anterior genes from adult tissues
Expression dynamics
Metazoa
Nephrozoa/Bilateria
Protostomia
Lophotrochozoa
Annelida
O. fusiformis
Other
Blastula
Gastrula
Elongation
Early larva
Mitraria larva
Competent larva
Juvenile
z-score
3
0
−3
f
Posterior genes from adult tissues
Expression dynamics
Metazoa
Nephrozoa/Bilateria
Protostomia
Lophotrochozoa
Annelida
O. fusiformis
Other
Blastula
Gastrula
Elongation
Early larva
Mitraria larva
Competent larva
Juvenile
z-score
2
0
−2
g
O. fusiformis
Planktotrophic larva
Anterior
Trunk
Posterior
Normalised expression (z-score)
Blastula
Gastrula
Elongation
Early larva
Mitraria larva
Competent larva
Juvenile
h
Mitraria larva
Anterior
Posterior
Trunk
i
C. teleta
Lecithotrophic larva
Anterior
Trunk
Posterior
Blastula
Gastrula
Stage 4tt
Stage 5
Stage 7
Pre−competent
Competent
j
Stage 4tt larva
Anterior
Posterior
Trunk
k
D. gyrociliatus
Direct development
Anterior
Trunk
Early embryo
Late embryo
Hatching
Female adult
l
Anterior
Trunk


a
O. fusiformis
all ATAC-seq peaks
N = 63,726 peaks
12.92%
17.38%
4.01%
6.88%
9.73%
27.33%
21.74%
b
C. teleta
all ATAC-seq peaks
N = 44,368 peaks
20.53%
27.93%
8.41%
8.04%
10.88%
12.80%
11.40%
Promoters
5'-UTR and first exon
First intron
Other exons and 3'-UTR
Other introns
Transcription end sites
Intergenic regions
c
O. fusiformis
Motif accessibility
C. teleta
Motif accessibility
Blastula
Gastrula
Elongation
Mitraria larva
Competent
Stage 4tt
Stage 5
Stage 8
Pearson correlation coefficient
−1 0 1
d
O. fusiformis
Motif binding
C. teleta
Motif binding
e
O. fusiformis
Accessibility
Blastula
Gastrula
Elongation
Mitraria larva
Competent
Pearson correlation coefficient
−0.5 0 0.5 1
Binding score
C. teleta
Stage 4tt
Stage 5
Stage 8
f
Motif binding (z-score)
−1 0 1
Motif accessibility (z-score)
Common annotated annelid motifs
RUNX/RUNT
ZNF189
SCRT
NKX
ZNF692
MEIS/PKNOX/TGIF
SMAD
Owenia/Capitella SOX
PAX
TBX/TBR
E2F
SOX
NF−Y
YY1
ZIC2
EGR
MYOD/TAL
PPAR/NR1C
CEBP
GATA
BATF
Owenia/Capitella AR
HSF
NFKB/REL
RXR/NR2/RAR/NR1B
HOX/CDX/EVX
POU2F
HOX/MEIS/PBX
ETS
ZBTB33/KAISO/BANP
MYB
FOX
PAIRED
Blastula
Gastrula
Elongation
Stage 4tt larva
Mitraria larva
Stage 5 larva
Competent larva
Stage 8 larva
O. fusiformis
C. teleta
Trunk development
g
Overrepresented
0 5 ≥10
−log10(adj p-value)
Underrepresented
0 5 ≥10
−log10(adj p-value)
n.s.
O. fusiformis
Cluster
BATF
CEBP
E2F
EGR
ETS
FOX
GATA
HOX/CDX/EVX
HOX/MEIS/PBX
MEIS/PKNOX/TGIF
MYOD/TAL
NF−Y
NKX
PAIRED
POU2F
PPAR/NR1C
RUNX/RUNT
RXR/NR2/RAR/NR1B
SOX
TBX/TBR
YY1
ZBTB33/KAISO/BANP
ZIC2
ZNF189
O. fusiformis AR
O. fusiformis SOX
Trunk
h
C. teleta
MYB
SCRT
SMAD

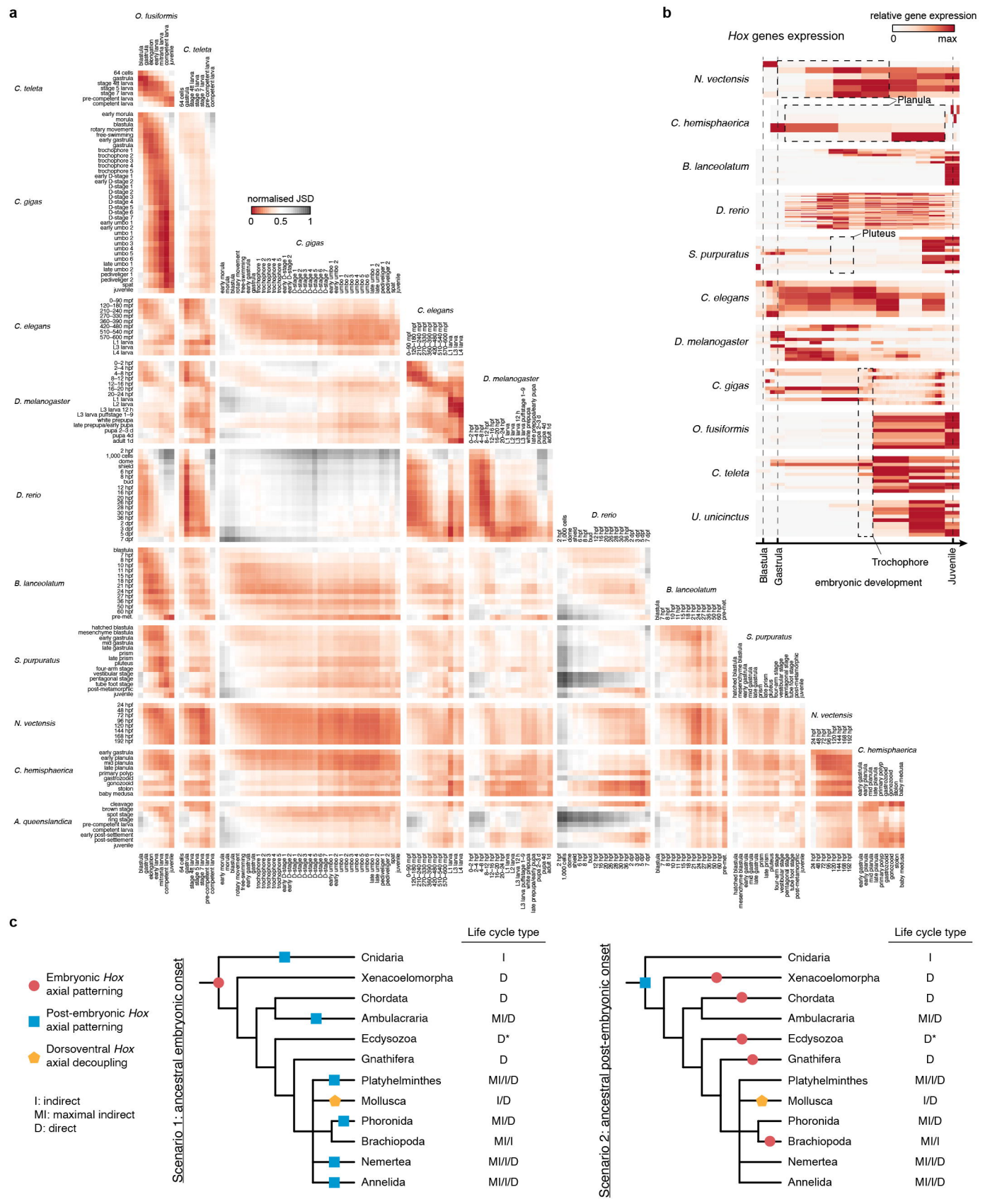

a
Pre- & Metazoa
Bilateria/ Nephrozoa
Protostomia
Lophotrochozoa
Annelida
Species specific
Phylostrata
Life cycle
Larva type
Larva nutrition
D. gyrociliatus
C. teleta
O. fusiformis
C. gigas
C. elegans
D. melanogaster
D. rerio
B. lanceolatum
S. purpuratus
N. vectensis
C. hemisphaerica
A. queenslandica
Life cycle
In Indirect
D Direct
Larva type
C Ciliated
J Juvenile-like
Larva nutrition
L Lecithotrophic
Pl Planktotrophic
b
Expression contribution
Phylostrata
Metazoa
Nephrozoa/Bilateria
Protostomia
Lophotrochozoa
Annelida
Species-specific
Other
Blastula
Gastrula
Elongation
Early larva
Mitraria larva
Competent larva
Juvenile
z-score
c
Expression contribution
64 cells
Gastrula
Stage 4tt larva
Stage 5 larva
Stage 7 larva
Precompetent larva
Competent larva
z-score
d
Expr. contribution
Early development
Late development
Hatchling
Female adult
z-score
e
Metazoa N = 11,778
Nephrozoa/ Bilateria N = 3,924
Protostomia N = 224
Lophotrochozoa N = 994
Annelida N = 566
Species-specific O. fusiformis N = 13,713
Other N = 704
Random N = 2,000
Quantile-normalised expression
Blastula
Gastrula
Elongation
Early larva
Mitraria larva
Competent larva
Juvenile
f
N = 14,681
N = 4,848
N = 541
N = 996
N = 517
C. teleta N = 17,109
N = 2,529
N = 2,000
Quantile-normalised expression
64 cells
Gastrula
Stage 4tt larva
Stage 5 larva
Stage 7 larva
Precompetent larva
Competent larva
g
N = 8,286
N = 1,735
N = 135
N = 325
N = 154
D. gyrociliatus N = 6,342
N = 411
N = 2,000
Quantile-normalised expression
Early development
Late development
Hatchling
Female adult
h
O. fusiformis
Mitraria larva
Cluster
Metazoa
Nephrozoa Bilateria
Protostomia
Lophotrochozoa
Annelida
Species-specific
Other
i
C. teleta
Larva
j
D. gyrociliatus
Overrepresented phylostrata (odds ratio > 1)
Underrepresented phylostrata (odds ratio > 1)
n.s.
0 25 ≥50
−log10 (adj p-value)



	blastula	gastrula	elongation	early larva	mitraria larva	competent larva	juvenile	
C. teleta	gastrula	gastrula	stage 4tt larva	stage 4tt larva	stage 5 larva	pre-competent larva	pre-competent larva	Indirect – Primary lecithotrophic larva
C. gigas	early morula	free-swimming	trochophore 1	trochophore 1	umbo 1	umbo 1	pediveliger 1	Indirect – Primary planktotrophic larva
C. elegans	0–90 mpf	120–180 mpf	120–180 mpf	L1 larva	L1 larva	570–600 mpf	570–600 mpf	Indirect – Secondary larva
D. melanogaster	4–8 hpf	8–12 hpf	8–12 hpf	8–12 hpf	late prepupa early pupa	early pupa	pupa 2–3 d	Indirect – Secondary larva
D. rerio	16 hpf	16 hpf	16 hpf	30 hpf	3 dpf	5 dpf	5 dpf	Direct development
B. lanceolatum	8 hpf	8 hpf	24 hpf	24 hpf	24 hpf	pre-met.	pre-met.	Indirect – Secondary larva
S. purpuratus	hatched blastula	hatched blastula	early gastrula	early gastrula	pluteus	tube foot stage	juvenile	Indirect – Primary planktotrophic larva
N. vectensis	48 hpf	48 hpf	48 hpf	48 hpf	72 hpf	120 hpf	192 hpf	Indirect – Lecithotrophic larva
C. hemisphaerica	gonozooid	gonozooid	early planula	early planula	planula	planula	baby medusa	Indirect – Lecithotrophic larva
A. queenslandica	cleavage	brown stage	brown stage	brown stage	brown stage	spot stage	juvenile	Indirect – Lecithotrophic larva

	blastula	gastrula	elongation	early larva	mitraria larva	competent larva	juvenile	
C. teleta	stage 4tt larva	stage 4tt larva	stage 4tt larva	stage 4tt larva	stage 7 larva	pre-competent larva	pre-competent larva	Indirect – Primary lecithotrophic larva
C. gigas	trochophore 1	trochophore 1	trochophore 1	trochophore 1	early D-stage 2	spat	pediveliger 2	Indirect – Primary planktotrophic larva
C. elegans	120–180 mpf	120–180 mpf	120–180 mpf	120–180 mpf	270–330 mpf	L1 larva	420–480 mpf	Indirect – Secondary larva
D. melanogaster	0–2 hpf	0–2 hpf	4–8 hpf	8–12 hpf	8–12 hpf	L3 larva puffstage 1–9	L3 larva puffstage 1–9	Indirect – Secondary larva
D. rerio	6 hpf	12 hpf	36 hpf	36 hpf	3 dpf	3 dpf	3 dpf	Direct development
B. lanceolatum	15 hpf	15 hpf	15 hpf	24 hpf	24 hpf	24 hpf	pre-met.	Indirect – Secondary larva
S. purpuratus	post-metamorphic	hatched blastula	hatched blastula	hatched blastula	hatched blastula	mesenchyme blastula	juvenile	Indirect – Primary planktotrophic larva
N. vectensis	48 hpf	48 hpf	48 hpf	48 hpf	48 hpf	192 hpf	192 hpf	Indirect – Lecithotrophic larva
C. hemisphaerica	gonozooid	gonozooid	gonozooid	gonozooid	planula	planula	early planula	Indirect – Lecithotrophic larva
A. queenslandica	early post-settlement	cleavage	cleavage	brown stage	brown stage	early post-settlement	early post-settlement	Indirect – Lecithotrophic larva

	blastula	gastrula	elongation	early larva	mitraria larva	competent larva	juvenile	
C. teleta	64 cells	gastrula	gastrula	gastrula	stage 4tt larva	stage 7 larva	competent larva	Indirect – Primary lecithotrophic larva
C. gigas	blastula	rotary movement	trochophore 1	trochophore 1	early umbo 2	late umbo 1	late umbo 1	Indirect – Primary planktotrophic larva
C. elegans	L3 larva	270–330 mpf	270–330 mpf	L4 larva	L4 larva	L4 larva	L4 larva	Indirect – Secondary larva
D. melanogaster	0–2 hpf	0–2 hpf	white prepupa	white prepupa	L3 larva puffstage 1–9	16–20 hpf	16–20 hpf	Indirect – Secondary larva
D. rerio	1,000 cells	bud	bud	bud	3 dpf	5 dpf	5 dpf	Direct development
B. lanceolatum	blastula	8 hpf	10 hpf	10 hpf	26 hpf	26 hpf	7 dpf	Indirect – Secondary larva
S. purpuratus	mesenchyme blastula	hatched blastula	pluteus	four-arm stage	tube foot stage	tube foot stage	juvenile	Indirect – Primary planktotrophic larva
N. vectensis	24 hpf	72 hpf	96 hpf	96 hpf	96 hpf	192 hpf	192 hpf	Indirect – Lecithotrophic larva
C. hemisphaerica	gonozooid	gonozooid	early planula	early planula	planula	planula	baby medusa	Indirect – Lecithotrophic larva
A. queenslandica	cleavage	cleavage	cleavage	cleavage	cleavage	juvenile	post-settlement	Indirect – Lecithotrophic larva


a
O. fusiformis
C. teleta
C. gigas
C. elegans
D. melanogaster
D. rerio
B. lanceolatum
S. purpuratus
N. vectensis
C. hemisphaerica
A. queenslandica
normalised JSD
0
0.5
1
b
relative gene expression
0
max
Hox genes expression
N. vectensis
C. hemisphaerica
B. lanceolatum
D. rerio
S. purpuratus
C. elegans
D. melanogaster
C. gigas
O. fusiformis
C. teleta
U. unicinctus
Planula
Pluteus
Trochophore
Blastula
Gastrula
Juvenile
embryonic development
c
Embryonic Hox axial patterning
Post-embryonic Hox axial patterning
Dorsoventral Hox axial decoupling
I: indirect
MI: maximal indirect
D: direct
Scenario 1: ancestral embryonic onset
Life cycle type
Cnidaria I
Xenacoelomorpha D
Chordata D
Ambulacraria MI/D
Ecdysozoa D*
Gnathifera D
Platyhelminthes MI/I/D
Mollusca I/D
Phoronida MI/D
Brachiopoda MI/I
Nemertea MI/I/D
Annelida MI/I/D
Scenario 2: ancestral post-embryonic onset
Life cycle type
Cnidaria I
Xenacoelomorpha D
Chordata D
Ambulacraria MI/D
Ecdysozoa D*
Gnathifera D
Platyhelminthes MI/I/D
Mollusca I/D
Phoronida MI/D
Brachiopoda MI/I
Nemertea MI/I/D
Annelida MI/I/D