Annelid functional genomics reveal the origins of bilaterian life cycles

Yan Liang[1,§], Francisco M. Martín-Zamora[1,§], Kero Guynes[1], Allan M. Carrillo-Baltodano[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,*]

[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

§ These authors contributed equally.

* 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], and thus larvae are central to most scenarios for animal evolution[2-12]. Yet how larvae evolved remains disputed. Here we show that changes in the timing of trunk formation underpin the diversification of larvae and bilaterian life cycles. Combining chromosome-scale genome sequencing with transcriptomic and epigenomic profiling in the slow-evolving oweniid *Owenia fusiformis*[13], we found that different genes and genomic regulatory elements control the development of its feeding larva and adult stage. First, *O. fusiformis* embryos develop into an enlarged anterior domain that forms larval tissues and the adult head, as posterior growth and trunk patterning is deferred to pre-metamorphic stages. These traits also occur in the so-called "head larvae" of other bilaterians[14,15], with whom *O. fusiformis* larva shows extensive transcriptomic similarities. Conversely, animals with non-feeding larvae and gradual metamorphoses, such as the annelid *Capitella teleta*, start trunk differentiation during embryogenesis, like direct developers. Together, our findings suggest that the ancestral temporal decoupling of head and trunk formation, as retained in extant "head larvae", allowed larval evolution in Bilateria, questioning prevailing scenarios that propose either co-option[10,11] or innovation[12] of gene regulatory programmes to explain larva and adult origins.

INTRODUCTION

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 vastly diverse and can display from radically different to more similar morphologies than those of their adults[1,2]. Given their broad phylogenetic distribution[2], larvae are central to major scenarios of animal evolution[2-12]. These fundamentally disagree on whether larvae are ancestral[2-7] or secondarily evolved[10,11] life stages and propose different mechanisms to explain the evolution of larva and adult forms (Fig. 1a, b). 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 (Fig. 1a). 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] (Fig. 1b). An assessment of the mechanisms underlying these hypotheses using comparative and functional genomics data is, however, lacking, 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 classically assigned to Annelida and Mollusca but also potentially to related clades within Lophotrochozoa[18] (Fig. 1c). Trochophore larvae—traditionally exemplified by that of annelid worms—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). Compared to most other annelid lineages, however, the larvae of the annelid groups Oweniidae and Magelonidae—which form

Oweniida, the sister group to all other annelids[13]—exhibit unusual characters (Fig. 1c, d). In particular, the idiosyncratic oweniid larva, commonly referred to as "mitraria"[21], has an enlarged pre-oral region and a single tuft of posterior chaetae, as well as a pair of nephridia and a long monociliated ciliary band alike those of phylogenetically distant larvae of deuterostome lineages, such as echinoderms and hemichordates[22-24]. Despite these distinctive larval traits, oweniids exhibit many developmental characters considered ancestral to Annelida, and even Lophotrochozoa as a whole[25,26], as well as similarities in larval molecular patterns with other trochophore and bilaterian larvae[24,25,27,28]. Consequently, whether the mitraria larva is the result of divergent or convergent evolution is unclear, which makes the comparative analysis of this lineage-restricted larva an excellent case study to investigate how larval traits evolve, and thereby formulate and assess scenarios on the origin of animal life cycles.

Here, we characterise the reference chromosome-scale genome assembly of the oweniid *Owenia fusiformis* (Fig. 1e) and perform a comprehensive study of its developmental transcriptome and regulatory genome that allows us to dissect the gene regulatory events underpinning the formation of the mitraria larva. The comparison of our comprehensive dataset with other lophotrochozoan and bilaterian taxa provides compelling evidence that heterochronic shifts in trunk development, rather than co-option and innovation in genetic programmes—as traditionally proposed—underpin the evolution of bilaterian life cycles.

RESULTS

O. fusiformis has a conservatively evolving genome

To characterise the transcriptomic and genomic regulatory basis for larval development in *O. fusiformis*, we first generated a chromosome-scale reference assembly combining PacBio long-reads, 10x genomics read clouds, optical mapping, and Hi-C scaffolding (Extended Data Fig. 1a). Consistent with flow cytometry and *k*-mer estimations, the haploid assembly spans 505.8 Mb (Extended Data Fig. 1b–d; Supp. Fig. 1), exhibiting 12 chromosome-scale scaffolds that encompass 89.2% of the assembly (Extended Data Fig. 1e, f). Almost half of the assembly consists of repeats (43.02%; largely DNA transposons) acquired steadily during evolution (Extended Data Fig. 2a–c; Supp. Table 1). Using transcriptomic data from 14 developmental stages and 9 adult tissues (Extended Data Fig. 1a), we annotated 26,966 protein-coding genes and 31,903 transcripts, representing a nearly complete (97.5%) set of metazoan BUSCO genes (Extended Data 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. 2d–i; Supp. Tables 2–7). Therefore, *O. fusiformis* genome contains a more complete gene repertoire that those reported for other annelid lineages, which together with its phylogenetic position and conserved early embryogenesis[25,26] makes it a key lineage to reconstruct the evolution of Annelida and Lophotrochozoa generally.

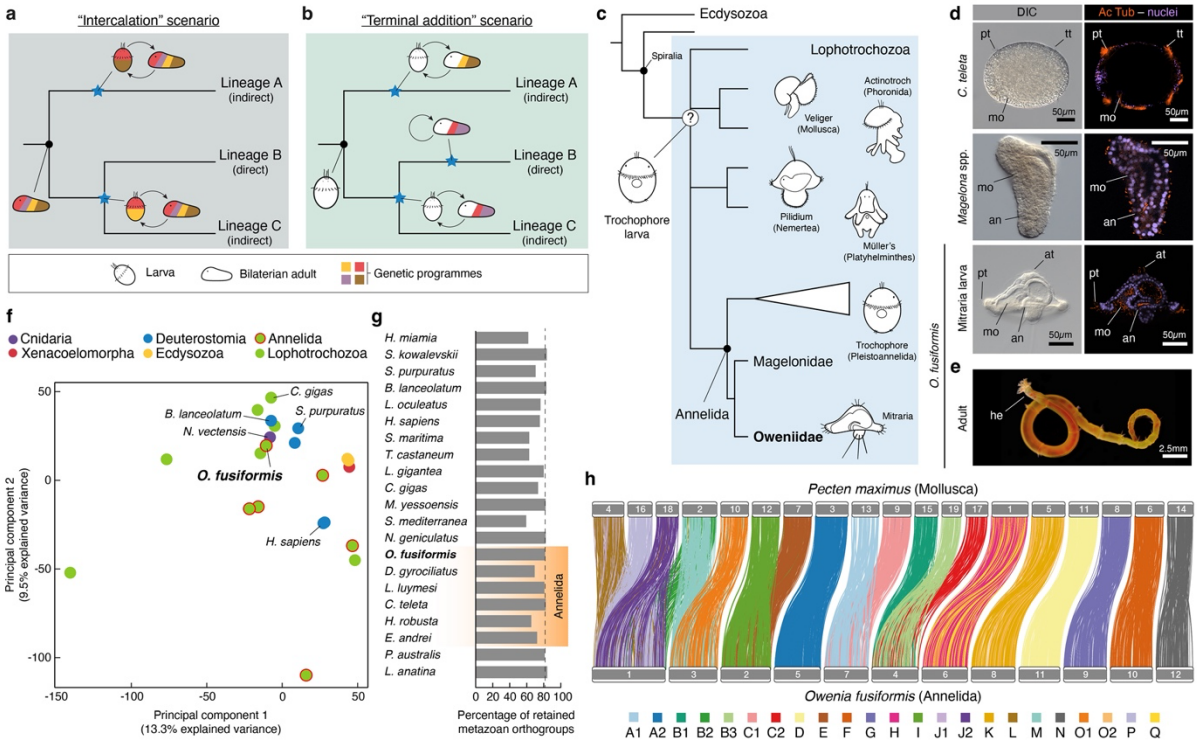


Figure 1 | *Owenia fusiformis* has a unique larva and a conservatively evolving genome. **a**, **b**, Major scenarios for the evolution of bilaterian life cycles and larval forms. The "intercalation" scenario (**a**) deems bilaterian adults ancestral and larvae secondary specialisations that evolved independently in certain lineages by co-opting genetic programmes (highlighted with different colours) originally expressed in the adult. The "terminal addition" scenario (**b**) considers that bilaterian larvae are ancestral and that adult forms evolved secondarily by incorporating new genetic programmes (indicated with different colours) after the larval stage. Direct development (as in lineage B) would have then evolved by losing the ancestral larval stage. **c**, A trochophore larval type has been proposed to be ancestral to Spiralia or even Protostomia (Ecdysozoan + 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. 2g 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; he: head; mo: mouth; pt: prototroch; tt: telotroch.

Chromosomal linkage of orthologous genes between deuterostomes and protostomes supports the presence of 22 ancestral linkage groups in bilaterians[29,30]. To explore the conservation of the high-order organisation of *O. fusiformis* genome, we compared its syntenic orthology with the cephalochordate *Branchiostoma lanceolatum* (a deuterostome) and the mollusc *Pecten maximus* (a lophotrochozoan). *Owenia fusiformis* has globally retained the ancestral bilaterian linkage groups (Fig. 1h; Extended Data Fig. 2h, i) and exhibits chromosomal fusions that are present in the scallop and even a nemertean (H+Q, J2+L and K+O2), which are thus likely ancestral to Lophotrochozoa. While the nemertean *Lineus longissimus* presents only one more fusion in addition to these three ancestral lophotrochozoan chromosomal rearrangements,

O. fusiformis, as well as the pleistoannelid *Streblospio benedicti*, show additional fusions of ancestral bilaterian linkage groups, which are notably not shared between the two taxa and therefore none are ancestral to Annelida (Extended Data Fig. 2h, i). However, *O. fusiformis* still retains a more ancestral chromosomal organisation than *S. benedicti*, in which all chromosomes except two are derived from secondary chromosomal fusion events, and other annelids with highly contiguous assemblies[31,32]. Together, our high-quality assembly of *O. fusiformis* supports that phylogenetically distant lophotrochozoan lineages largely share an ancestral syntenic chromosomal organisation, further revealing a dynamic chromosomal evolution in annelids, albeit their generally stable gene complement.

O. fusiformis larva does not co-opt genes expressed in the adult

The "intercalation" scenario proposes that larvae evolved secondarily by co-opting genes and gene modules expressed in adult stages[10,11] (Fig. 1a). Larva and adult would thus share transcriptional similarities, while more gradual transcriptional dynamics would be expected during direct development, as in the annelid *Dimorphilus gyrociliatus*[31]. To test this hypothesis, we used bulk stage-specific transcriptomic data from the blastula—the stage at which axial polarity is established in *O. fusiformis*[26]—to juvenile (as a proxy to adult stages), covering larval embryogenesis, growth, and metamorphosis (Fig. 2a; Extended Data Fig. 3a–f; Supp. Tables 8–10). We observed two main phases of increased gene expression and differential gene expression activity during the life cycle of *O. fusiformis*, with the early mitraria as transitional stage: a first phase leading to larva formation (i.e., embryogenesis), followed by a second phase encompassing larval growth and metamorphosis into the juvenile (Fig. 2b, c; Extended Data Fig. 3g). Soft clustering and weighted gene co-expression network analysis of the 31,678 expressed transcripts generated an optimal number of 12 distinct clusters and 14 modules of temporally co-regulated genes, respectively, which broadly classified as peaking during either the first or the second transcriptional phases (Extended Data Fig. 4; Supp. Fig. 2–4; Supp. Tables 11–13). In none of the two approaches, however, were there clusters of genes with peaks of expression at both the mitraria and the juvenile stage, and only one cluster (1,426 transcripts; 4.5%) showed a bimodal activation at blastula and juvenile stages (Extended Data Fig. 4a). Therefore, extensive co-option of adult genes into larva embryogenesis—as the "intercalation" scenario posits—does not occur in *O. fusiformis*, and instead larval embryogenesis and juvenile/adult development are two markedly distinct transcriptional phases.

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

To identify the developmental processes underpinning the two transcriptional phases during *O. fusiformis* life cycle, we performed gene ontology (GO) enrichment analyses on the soft clusters and modules of temporally co-regulated genes (Extended Data Fig. 4b–l). GO terms related to nucleic acid metabolism—i.e., DNA replication and transcription—and complex biosynthetic metabolism predominate in clusters of the first transcriptional phase (Extended Data Fig. 4b–f), consistent with the establishment of the larval body plan and development of the functional digestive system at those stages[21,25]. Among the genes expressed in this first phase, we identified an ortholog of *chordin*, a key bone morphogenetic protein (BMP) inhibitor involved in dorsoventral patterning across Bilateria[33] and thought to be lost in annelids[34]

(Fig. 2d; Extended Data Fig. 5a, b). With a complex pattern of retentions and losses in Annelida (Extended Data Fig. 5c–f; Supp. Tables 14–18), *chordin* is asymmetrically localised around the blastopore lip of the gastrula and later in the larval antero-ventral oral ectoderm (Fig. 2e). This expression resembles that of molluscs[34,35] and brachiopods[28], providing further evidence that conserved developmental mechanisms underpin larval formation in *O. fusiformis*[26-28]. However, differently from non-feeding annelid trochophores[6], anterior ectodermal genes[26-28] pattern most larval tissues in *O. fusiformis* (Fig. 2f), whereas posterior genes are restricted to few circum-anal cells, in line with the vastly enlarged pre-oral region of the mitraria that forms early on in the adult head[21].

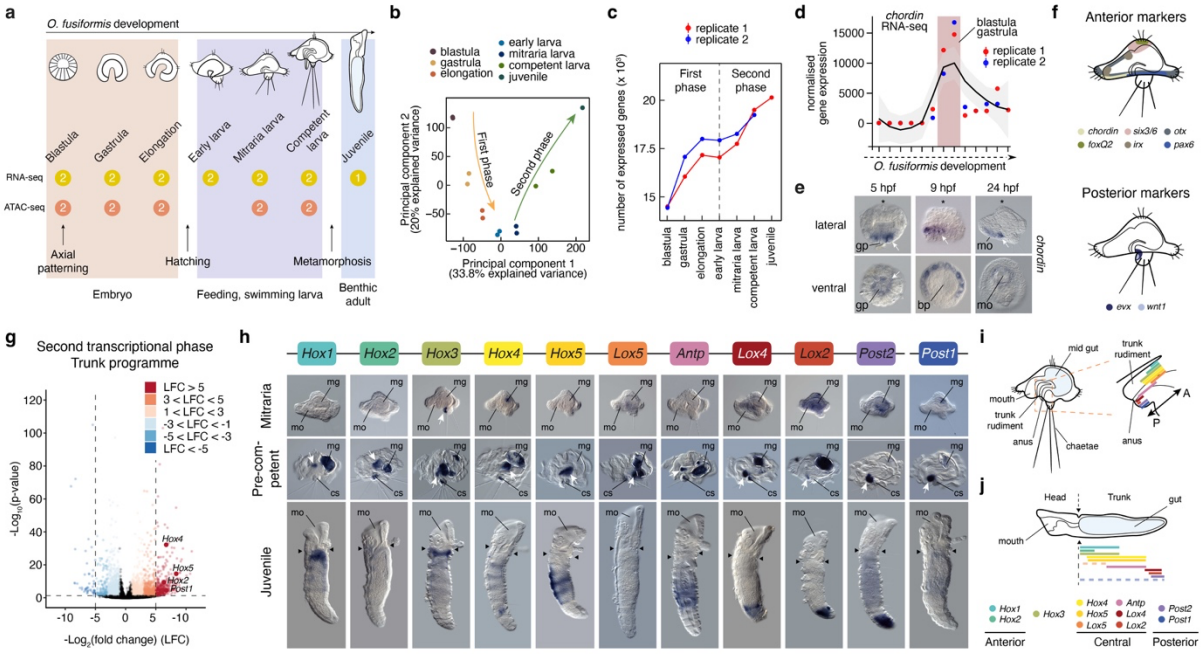


Figure 2 | Two transcriptional waves underpin larva embryogenesis and metamorphosis. **a,** Schematic representation of *O. fusiformis* development from blastula to juvenile indicating major developmental events and the time points when RNA-seq and ATAC-seq samples were collected. Numbers inside the coloured circles indicate the number of biological replicates. **b,** Principal component analysis of the developmental RNA-seq time course. **c,** Number of expressed genes during *O. fusiformis* development. **d**, Expression levels of *chordin*, which peaks at the blastula and gastrula stages, after the specification and inductive activity of the embryonic organiser. **e,** Whole mount *in situ* hybridisation of *chordin* at the blastula (5 hours post fertilisation, hpf), gastrula (9 hpf), and mitraria larva (24 hpf) stages. Asterisks mark the animal/anterior pole. gp: gastral plate; bp: blastopore, mo: mouth. **f,** Schematic representation of the expression of anterior and posterior ectodermal gene markers at the mitraria stage, demonstrating how anterior territories expand throughout most of the larva. **g,** Volcano plot of the mitraria to competent larva transition, highlighting the marked upregulation of certain *Hox* genes. **h**, Whole mount *in situ* hybridisation of *Hox* genes at the mitraria, pre-competent, and juvenile stages. Only *Hox3* appears to be expressed at the mitraria stage (white arrow), while *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. **i, j,** Schematic representations of the expression of *Hox* genes in the trunk rudiment of the competent larva (**i**) and juvenile trunk (**j**). A: anterior; P: posterior. Drawings are not to scale, and schematic expression domains are approximate.

GO terms involved in terminal cell differentiation, morphogenesis, and organogenesis predominate in co-expression clusters of the second transcriptional phase. *Hox* genes, a conserved family of transcription factors involved in anterior-posterior trunk regionalisation in Bilateria[36], are among the most upregulated genes at these stages (Fig. 2g). *Owenia 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. 6a–c; Supp. Table 19). This is a similar repertoire and genomic arrangement to that of the annelids *Capitella teleta*[37] and *Platynereis dumerilii*[38]; yet unlike these annelids, which deploy *Hox* genes during or soon after gastrulation[37,39] (Extended Data Fig. 6d), *O. fusiformis* does not express *Hox* genes during embryogenesis to pattern the larval body (Fig. 2h; Extended Data Fig. 6d, e). 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. 2h–j; Extended Data Fig. 6e). 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. 6d; Supp. Table 20). Therefore, posterior/trunk development is deferred to pre-metamorphic stages in planktotrophic annelid trochophores[14,40] compared to annelids with lecithotrophic larvae[37,38] and direct developers[31] (Extended Data Fig. 6f). Altogether, our findings support that axial and body patterning, as well as anterior and gut differentiation contribute to the first transcriptional phase, while adult trunk formation largely drives the second post-embryonic transcriptional phase during *O. fusiformis* life cycle.

Chromatin dynamics support two regulatory modules during *Owenia* development

To investigate the genomic regulatory basis for the observed temporal decoupling of head and trunk development in *O. fusiformis*, we profiled open chromatin regions with ATAC-seq at five developmental stages, from blastula to competent larva (Fig. 2a; Extended Data Fig. 7a–e, Supp. Tables 21–24). In total, we identified 47,406 consensus regulatory regions (Extended Data Fig. 7f, g), mostly abundant within gene bodies (68.14%) rather than in promoters (16.19%) and distal intergenic regions (15.67%) (Fig. 3a, Supp. Table 23). While the total number of accessible regions is greater during embryogenesis than in larval stages (Fig. 3b; Extended Data Fig. 7f–h), the largest changes in peak accessibility occur in the mitraria and competent larva (Fig. 3c, Supp. Table 24). At these stages, there is an increase in ATAC-seq peak width and proportion of accessible chromatin in promoters and upstream gene body elements (i.e., 5' UTRs and first exons; Extended Data Fig. 7i, j; Extended Data Fig. 8a), which are regulatory regions that correlate positively with gene expression (Extended Data Fig. 8b). Soft clustering revealed that most regulatory regions act during either embryogenesis (29,611; 63.99%) or larval growth (13,457; 29.08%), with just 3,210 peaks (6.93%; cluster 8) being accessible from gastrula to mitraria (Fig. 3d). To explore the interplay between regulatory and transcriptional programmes, we predicted *de novo* DNA binding motifs on ATAC-seq peaks located in promoters and intergenic regions. Soft clustering in six optimal groups according to their differential use during *O. fusiformis* life cycle recovered two main regulatory modules, one mostly active during embryogenesis and a second mostly used during larval growth (Fig. 3e; Extended Data Fig. 8c–e, Supp. Tables 25, 26, Supp. Fig. 5, 6). Consistent with our

transcriptomic dataset, motifs related to transcription factors involved in patterning anterior territories (e.g., PAX2/5/8 and PAX4/6[27]), muscle and gut development (GATA[28], FOXC[28]) and early neurogenesis (ATOH[41]) are amongst the most differentially accessible in open chromatin regions during embryogenesis, whilst ciliary band genes (OTX[28]), trunk related genes (e.g., NKX2.1[27]) and most notably *Hox* genes appear in regulatory regions during larval competence and trunk patterning (Fig. 3e; Extended Data Fig. 8c). Therefore, two different sets of *cis*-regulatory elements mirror the major transcriptional dynamics in *O. fusiformis*, reinforcing the view that separate genetic mechanisms that are temporally decoupled underpin larva and adult formation in this annelid.

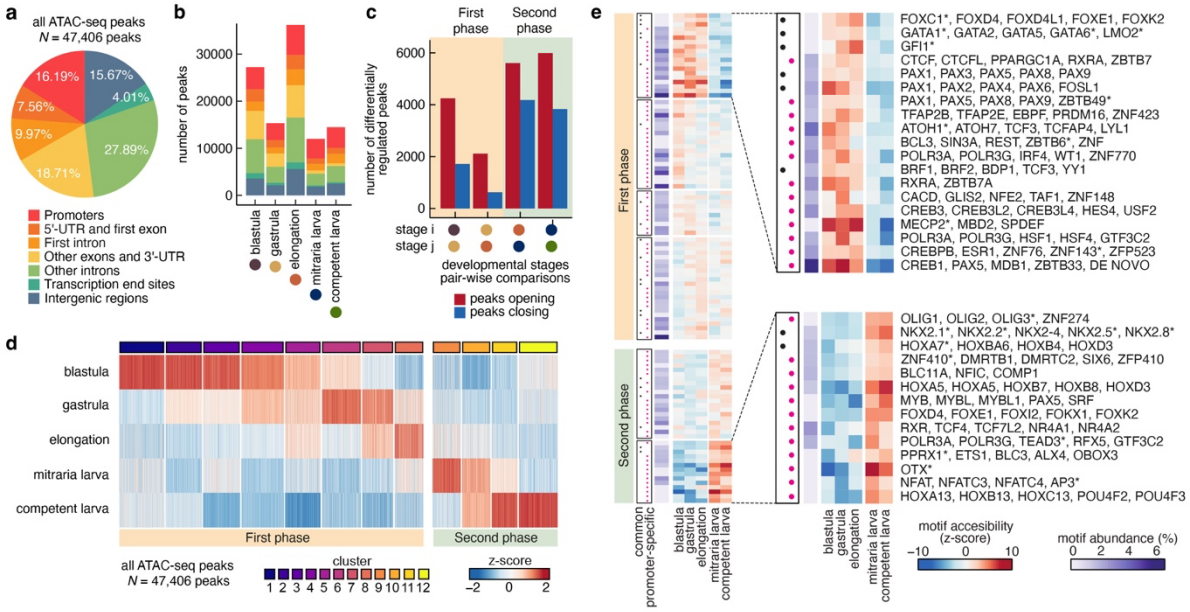


Figure 3 | Chromatin dynamics support two regulatory modules during *Owenia* development. **a,** Genomic annotation by functional region of the consensus ATAC-seq peak set. **b,** Stacked bar plot showing the number of called peaks per developmental stage, classified by genomic location. **c,** Differentially accessible peaks for all four pairwise comparisons between adjacent stages. Only significant differentially regulated peaks (adjusted p–value < 0.05) are shown. **d,** Heatmap of normalised peak accessibility (z–score) of the soft clustered consensus ATAC-seq peaks, subdivided into clusters of open chromatin regions that peak pre- and post-larval stages. **e,** Motif enrichment analysis of promoter peaks. The clustered heatmap shows normalised motif accessibility (z-score) and the abundance of the motif in promoter peaks, subdivided in clusters of motifs mostly accessible either pre- or post-larval stage. Clusters 1 (top right) and 6 (bottom right) are the most accessible of each regulatory programme, and thus are shown enlarged with predicted bound transcription factors to the right. Asterisk denotes predicted factors from GimmeMotif curated databases.

Mitraria development does not rely on novel genes

Novel genes account for a significant proportion of some trochophore and other larval transcriptomes[7,42], as they might be associated with the development of larval-specific characters (e.g., ciliary bands)[7,42,43]. Therefore, recruitment of novel genes at the mitraria stage could also explain the overall transcriptomic differences between this and the adult stage, as well as the distinctive morphological traits of this larval form. To define the contribution of novel genes at each developmental stage in *O. fusiformis*, we classified all predicted transcripts in seven phylostrata according to their time of origin (Fig. 4a, Supp. Table 27). Older genes

(genes of metazoan and pre-metazoan origin; phylostratum 1) represent the largest fraction of genes expressed at all developmental stages except that of the juvenile in *O. fusiformis*, whereas the highest expression levels of novel/younger genes (phylostratum 7) occur in the juvenile stage (Fig. 4b, Extended Data Fig. 9a–d). An enrichment analysis of each phylostrata on each of the 12 recovered gene cluster showed that older genes are significantly enriched in early embryonic stages (clusters 1 to 3), while novel genes are significantly overrepresented in gene clusters active at the juvenile stage and most notably, underrepresented in the mitraria larva (clusters 7 and 8; Extended Data Fig. 9e). Therefore, older rather than younger genes contribute to the development of the mitraria larva, suggesting that the increased use of novel genes in other lophotrochozoan larvae[7,42] might be due to lineage-specific traits found on those larvae, such as the shell primordium of molluscan trochophores[7] and perhaps even ciliary bands with multiciliated cells[44], which are absent in oweniid larvae.

Bilaterian planktotrophic larvae share maximal transcriptional similarity

The "terminal addition" scenario[2,3,19] posits that larvae are homologous and adults evolved convergently, and thus pre-larval stages should be more conserved than metamorphosis and adult stages across species[11] (Fig. 1b). To test this hypothesis, we first generated bulk stage-specific transcriptomic data for the annelid *C. teleta* (Supp. Fig. 7, Supp. Table 28), which exhibits a typical lecithotrophic trochophore (Fig. 1c). Differently from *O. fusiformis*, *C. teleta* does not show two marked transcriptional phases but rather a gradual increase of gene expression during its life cycle (Fig. 4c), as direct developers do[31]. Pairwise transcriptomic comparisons between *O. fusiformis* and *C. teleta* demonstrate overall similar transcriptional dynamics between the development of these annelids, yet the point of maximal transcriptional divergence occurs at larval—and not adult—stages (Fig. 4d). While this observation disagrees with the "terminal addition" scenario, it rather reflects, and is explained by, the different timings of trunk differentiation in *O. fusiformis* and *C. teleta* (Extended Data Fig. 6f).

We thus extended our comparative transcriptomic approach to six other bilaterian lineages and the cnidarian *Nematostella vectensis* to cover the overall diversity of life cycle strategies found in Metazoa, using both all single copy one-to-one orthologs (Fig. 4a, e; Extended Data Fig. 10a; Supp. Tables 29, 30) and a reduced set of conserved cross-species single copy orthologs (Extended Data Fig. 10b, c). Transcriptional dynamics between *O. fusiformis* and other major animal groups are more dissimilar (except in *Danio rerio*) at early development and become more similar as development proceeds towards juvenile and adult stages (Fig. 4e, f; Extended Data Fig. 10a–c; Supp. Tables 31, 32). At the larval stage, *O. fusiformis* shows vast transcriptomic differences with the larvae of *Drosophila melanogaster* and *Caenorhabditis elegans*, the two ecdysozoan taxa with indirect development and whose larvae evolved secondarily[10] (Fig. 4f). However, *O. fusiformis* shares maximal transcriptomic similarities at larval stages with bilaterian species with planktotrophic ciliated larvae and even the early planula of the cnidarian *Nematostella vectensis* (Fig. 4e, f; Extended Data Fig. 10a, b). Together, these findings question the "terminal addition" scenario, because adult development is generally more conserved than embryogenesis, but also reveal unexpected genome-wide transcriptional similarities between phylogenetically distant bilaterian planktotrophic larvae.

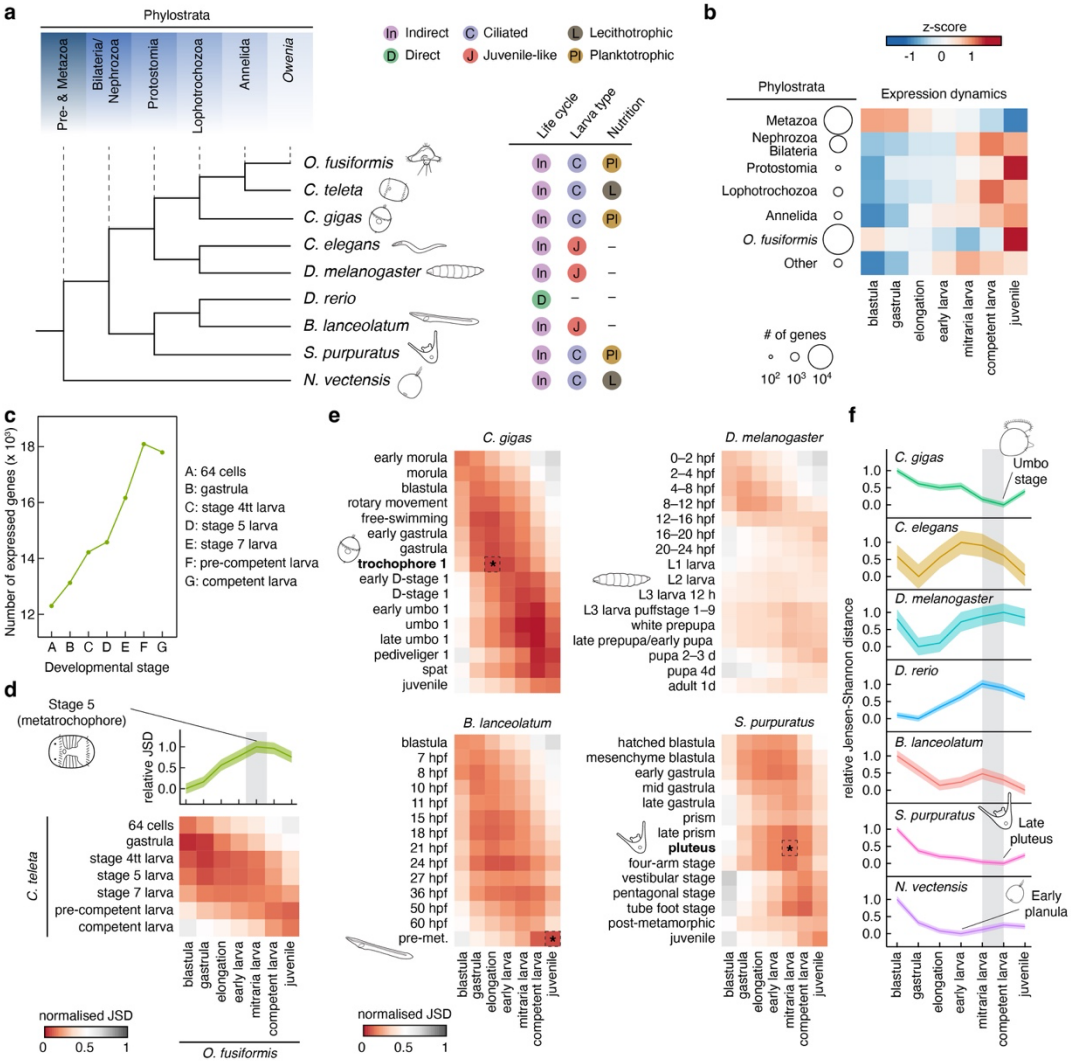


Figure 4 | Novel genes and comparative transcriptional dynamics during larval development. **a,** Cladogram of the species used for comparative transcriptomics analyses, indicating on top the phylogenetic age of each phylostrata considered for phylostratigraphy. For each species, the type of life cycle (direct/indirect), larval type (ciliated/juvenile-like) and larval nutritional mode (planktotrophic/lecithotrophy) are shown on the right. **b**, Expression dynamics of each phylostratum by developmental stage, calculated from the 75 % percentile of a quantile-normalised matrix of gene expression levels. Earlier stages are enriched in metazoan genes, and *O. fusiformis*-specific genes reach their maximum expression levels at the juvenile stage. **c,** Number of expressed genes across *C. teleta* development. **d,** Heatmap of pairwise normalised transcriptomic Jensen–Shannon Divergence (JSD) between *O. fusiformis* and *C. teleta*, with the relative JSD of the *C. teleta* stages of minimal divergence to each *O. fusiformis* stage on top. **e,** Heatmaps of pairwise normalised JSD between *O. fusiformis* and *C. gigas*, *D. melanogaster*, *B. lanceolatum* and *S. purpuratus* (the asterisk indicates the stage of minimal JSD between the larval stage of each species with *O. fusiformis*). **f,** Relative JSD from stages of minimal divergence of species in (**a**) to each *O. fusiformis* developmental stage. Confidence intervals in (**b**) and (**f**) represent the standard deviation from 250 bootstrap resamplings of the one-to-one orthologs.

DISCUSSION

In this study, we report the chromosome-scale genome assembly of the oweniid *O. fusiformis* which, together with extensive gene expression and genomic regulatory profiling during embryonic and larval development, provides an unprecedented perspective on mainstream scenarios for life cycle evolution in Bilateria. In particular, our results refute predictions that propose either the co-option of adult genes into larval 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 drivers of larva and adult evolution, respectively. Instead, the distinctive mitraria larva of *O. fusiformis* develops from an enlarged head region while trunk differentiation is deferred to late larval stages (Fig. 5a; Extended Data Fig. 6f). Similar developmental traits occur in other feeding annelid larvae[40] (Extended Data Fig. 6e) including that of *Chaetopterus*[45,46], a member of the second earliest branching annelid lineage[13], as well as in other phylogenetically distant clades within Lophotrochozoa (e.g., nemerteans[47] and phoronids[48]), Ecdysozoa (pancrustaceans[15] and pycnogonids[49]), and Deuterostomia (e.g., echinoderms[50,51] and hemichordates[52]), whose larvae are generally referred to as "head larvae"[14,15]. By contrast, non-feeding larvae[37,39] and direct developers[31] in both Annelida (Extended Data Fig. 6f) and other bilaterian taxa[53,54] start trunk patterning soon after gastrulation (Fig. 5b), while the onset of anterior/head patterning always takes place before gastrulation in bilaterians[55]. Regardless of these different timings, major bilaterian clades share head[56,57] and trunk molecular patterns (e.g., a *Hox* code[36]; Fig. 2h), supporting the overall homology of the adult bilaterian body plan. Therefore, our findings indicate that heterochronic shifts of an adult trunk developmental programme correlate with, and may account for, the evolution of the different life cycles in Annelida, and Bilateria generally[52].

The post-embryonic activation of genetic programmes involved in trunk differentiation observed in "head larvae" has generally been considered a lineage-specific innovation associated with the evolution of maximal indirect development[14,15,23,48]. Consequently, "head larvae" are thought to have evolved convergently and the observed similarities in larval molecular patterns[6,48,52,58] might reflect ancient gene regulatory modules that were independently co-opted to develop analogous larval organs (e.g., apical organs and ciliary bands), in line with the "intercalation" scenario[10,11]. In contrast, our study supports that the late deployment of trunk differentiation programmes is likely ancestral to Annelida (Fig. 5c), and not necessarily related to maximal indirect development. Indeed, an ancestral post-embryonic onset of trunk differentiation is a more parsimonious state for Bilateria, based on our current understanding of the timing of trunk-associated expression of *Hox* genes in major bilaterian clades (Extended Data Fig. 10d; Supp. Table 33). This temporal decoupling between head and trunk genetic programmes could thus have facilitated the evolution of larvae, which would then be homologous on the grounds of being largely anteriorly derived transitory structures that form intermediate life cycle specialisations. Notably, this scenario does not exclude the possibility that as bilaterians (and thus larvae) diversified, conserved gene regulatory modules were also convergently reused to develop analogous larval organs in different lineages. Regardless of the scenario, however, our study highlights the importance of heterochronic

changes in the deployment of ancient genetic programmes for the diversification of bilaterian life cycles, which had profound ecological and evolutionary implications. In the future, comparative genomic studies that thoroughly dissect the regulatory principles underlying head and trunk development in multiple bilaterian lineages will help to decipher how temporal changes in their deployment drove the evolution of larval and adult forms in Bilateria.

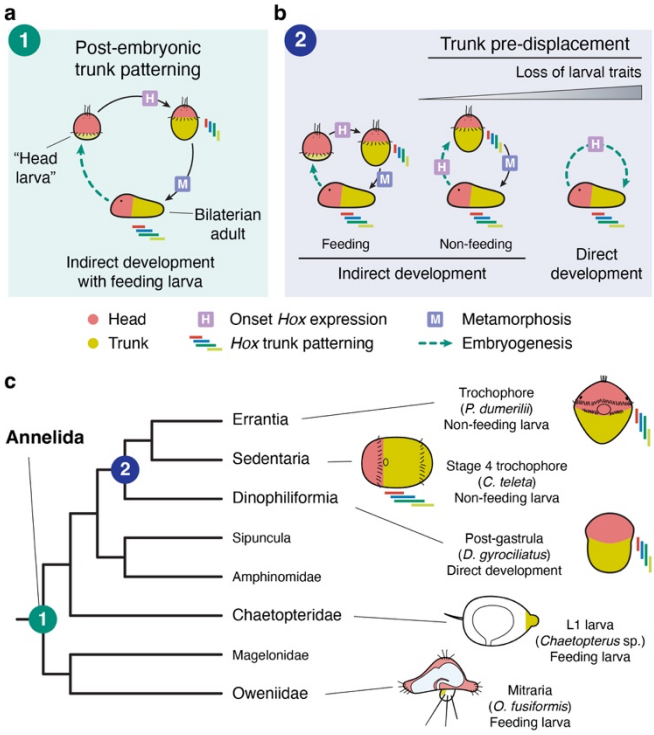


Figure 5 | Heterochronic shifts in trunk patterning underpin larval evolution in Annelida. a, Schematic drawing of the life cycle and patterning events in a bilaterian with indirect development with a feeding larva like the annelid *O. fusiformis*. Embryogenesis results in a larva with mostly anterior 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. **b**, 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. **c**, Proposed evolutionary scenario for larval and life cycle evolution in Annelida. The early branching lineages Oweniid and Chaetopteridae exhibit feeding larvae that originally develop with reduced trunk regions. However, non-feeding larvae in Errantia and Sedentaria (e.g., those of *P. dumerilii* and *C. teleta*, respectively) already exhibit a patterned trunk region, which is established post-gastrulation, as in the direct developer *D. gyrociliatus*. Notably, feeding larva in Errantia and Sedentaria behave as those of Oweniidae and Chaetopteridae, and thus a post-embryonic trunk patterning is likely an ancestral condition (green circle with 1) and the convergent pre-displacement of trunk differentiation to embryogenesis (blue box with 2) concurred with the evolution of indirect development with feeding larva and direct development. Drawings are not to scale.

METHODS

Adult culture, spawning and *in vitro* fertilisation

Sexually mature *Owenia fusiformis* Delle Chiaje, 1844 adults were collected from subtidal waters near the Station Biologique de Roscoff and cultured in the lab as described before[25]. *In vitro* fertilisations and collections of embryonic and larval stages were performed as previously described[25]. *Capitella teleta* Blake, Grassle & Eckelbarger, 2009 was cultured, grown, and sifted, and its embryos and larvae were collected following established protocols[59]. *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*[25] and adult individuals of *Drosophila melanogaster* as reference. Additionally, we used Jellyfish v.2.3[60] to count and generate a 31-mer histogram from adaptor-cleaned, short-read Illumina reads (see section below), and GenomeScope 2.0[61] 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[62] 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[63] and used for polishing with Racon v.1.16[64]. Bionano Genomics optical mapping data was used to scaffold the PacBio-based assembly, which was de-haploidised with purge_haplotigs v.1.0.4[65] 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[66] and BUSCO v.5[67] 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 (17 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-cell stage, gastrula, stage 4tt, stage 5, and stage 7 larva, 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 8, 28).

Annotation of repeats and transposable elements (TEs)

RepeatModeler v.2.0.1[68] 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* available at Ensembl Metazoa as reference. 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[69], 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[70], using default parameters but a maximum LTR size of 25 kb (as derived from the LTR_finder annotation). The general feature format (gff) file with the annotation of TEs and repeats is in *O. fusiformis* genome repository (see Data Availability section).

Gene prediction and functional annotation

We used SAMtools v.1.9[71] 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[13] (SRR1222288) with STAR v. 2.5.3a[72] after removing low-quality read pairs and read pairs containing Illumina sequencing adapters with trimmomatic v.0.39[73]. StringTie v.1.3.6[74] was used to convert STAR alignments into gene transfer format (GTF) files and Portcullis v.1.1.2[75] to generate a curated set of splice junctions. Additionally, we generated *de novo* transcriptome assemblies for all samples with Trinity v.2.5.1[76] with default parameters, which were thereafter mapped to the soft-masked assembly with GMAP v.2020-04-08[77]. We then ran the default Mikado v.2.1 pipeline[78] 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[79]. 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[80] to generate spliced alignments of the proteome of *C. teleta* 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[79] for *ab initio* gene prediction. Finally, PASA v.2.3.3[81] 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[67] in proteome mode, resulting in 97.7 % of the core genes present. Protein homologies for the 34,353 filtered transcripts were annotated with BLAST v.2.2.31+[82] on the UniProt/SwissProt database provided with Trinotate v.3.0[83]. We used HMMER v.2.3.2[84] to identify protein domains using Trinotate's PFAM-A database and signalP v.4.1[85] 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[86] orthology ID to each transcript. In total, we retrieved a functional annotation for 22,516 transcripts (63.86 %). The functional annotation report is provided in *O. fusiformis* genome 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[87] to generate a list of curated contracts ('merged no dups') that was subsequently employed to scaffold the assembly using 3d-dna v.180419[88]. The resulting assembly and contact map was visually inspected and curated using Juicebox v.1.11.08[87] 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[89]. 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 (Supp. Table 2). To reconstruct gene families, we used OrthoFinder v.2.2.7[90] using MMSeqs2[91] 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[92], 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[31]. 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.

Macrosynteny analyses

Single copy orthologues obtained using the mutual best hit (MBH) approach generated using MMseqs2[91] using the annotations of *Branchiostoma floridae*[93], *Pecten maximus*[94], *Streblospio benedictii*[95], and *Lineus longissimus*[96,97] 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*.

Gene expression profiling

To profile gene expression dynamics from blastula to juvenile stages (*O. fusiformis*) and from 64-cell to competent larva stages (*C. teleta*), sequencing adaptors were removed from raw reads using trimmomatic v.0.39[73]. Cleaned reads were pseudo-aligned to the filtered gene models using kallisto v.0.46.2[98] (Supp. Table 9) and genes with an expression level above an empirically defined threshold of 2 transcripts per million (TPM) were deemed expressed (Extended Data Fig. 3f). DESeq2 v.1.30.1 package[99] was used to normalise read counts across developmental stages (Supp. Table 10) and to perform pair-wise differential gene expression analyses between consecutive 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 (Supplementary Information). Principal component analysis was performed on a variance stabilising-transformed matrix of the normalised DESeq2 matrix. *Hox* gene expression profiling in *Urechis unicinctus*, was performed as described for *O. fusiformis* gene profiling, using an available reference transcriptome[100].

Gene clustering and co-expression network analyses

Genes were clustered according to their normalised expression dynamics through soft *k*-means clustering using the mfuzz v.2.52 package[101]. From all 31,903 genes, we discarded 225 genes which were not expressed at any stage. We then determined an optimal number of 12 clusters for our dataset 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 network, we used the WGCNA package v.1.70–3[102]. All 31,678 genes 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. Block-wise network construction returned 15 gene modules, from which one module was dropped due to poor intramodular connectivity. The remaining 14 gene modules (A–N) were labelled with distinct colours with unassigned genes labelled in grey. A random subset consisting of the nodes and edges of 30 % of the genes was fed to Cytoscape v.3.8.2[103] for network visualisation. Module eigengenes were chosen to summarise the gene expression profiles of gene modules. Gene ontology (GO) enrichment analyses of each gene cluster and gene module were performed using the topGO v.2.44 package. We performed a Fisher's exact test and listed the top 30 (soft *k*-means clustering) or top 15 (WGCNA modules) significantly enriched GO terms of the class Biological Process. To ease visualisation, all 288 non-redundant enriched GO terms from all 12 gene clusters were clustered through *k*-means by semantic similarity using the simplifyEnrichment v.1.2.0 package[104] (Supp. Fig. 3).

Gene orthology assignment

The identification of *chordin* (*chrd*), *chordin-like* (*chrdl*) and *Hox* genes in *O. fusiformis* was based on the genome functional annotation. To mine *chrd* orthologues, 81 annelid transcriptomic datasets were downloaded from SRA (Supp. Table 14) and assembled with Trinity v.2.5.1[76] to create BLAST local nucleotide databases. We also created a nucleotide database for *C. teleta* using its annotated genome[29] (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 104 unique candidate *chrd* transcripts that were then translated (Supp. Table 15). 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 (Supp. Table 16) and aligned together with *O. fusiformis* CHRD and CHRDL sequences in MAFFT v.7[105] with the G-INS-I iterative refinement method and default scoring parameters. From this mother alignment further daughter alignments were obtained using "mafft --addfragments"[106], 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 (Extended Data Figure 5c, d). 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 with a WAG amino acid replacement matrix[107] to account for transition rates, the FreeRate heterogeneity model (R4)[108] to describe sites evolution rates, and an optimization of amino acid frequencies using maximum likelihood (ML) using IQ-TREE v.2.0.3[109]. 1,000 ultrafast bootstraps (BS)[110] were used to extract branch support values. Bayesian reconstruction in MrBayes v.3.2.7a[111] were also performed using the same WAG matrix but substituting the R4 model for the discrete gamma model[112], with 4 rate categories (G4). For the orthology assignment of *Hox* genes, 129 curated Hox sequences were retrieved from various databases (Supp. Table 19) and aligned with *O. fusiformis* Hox proteins with MAFFT v.7 in automatic mode. Poorly aligned regions were removed with gBlocks v.0.91b[113] yielding the final alignments. Maximum likelihood trees were constructed using RAxML v.8.2.11.9[114] with an LG substitution matrix[115] and 1,000 ultrafast BS. All trees were composed in FigTree v.1.4.4. Alignment files are available in *O. fusiformis* genome repository (see Data Availability section).

Whole mount *in situ* hybridisation and immunohistochemistry

Fragments of *chordin* and *Hox* genes were isolated as previously described[26] 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[26,28]. Antibody staining in larval stages of *O. fusiformis*, *Magelona* spp. and *C. teleta* was carried out as previously described[25,116]. 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), mature larva (~150) and competent larva (~40) stages following the omniATAC protocol[117], 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[118]. After DNA clean-up, ATAC-seq libraries were amplified as previously described. Primers used for both PCR and qPCR are listed in Supp. Table 21. 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 gastrula) and on an Illumina NovoSeq6000 in 2 × 150 bases mode at Novogene (Cambridge, United Kingdom) (one replicate of gastrula and two replicates of competent larva stages).

Chromatin accessibility profiling and motif identification and enrichment analyses

We used cutadapt v.2.5[119] 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[120] in paired-end mode, duplicates were removed using samtools v.1.9[121] and mapped reads were shifted using deepTools v.3.4.3[122]. 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[123,124] (-f BAMPE --min-length 100 --max-gap 75 and -q 0.01). Peaks from repetitive regions were filtered with BEDtools v.2.28.0[125] and irreproducible discovery rates (IDR) v.2.0.4.2 was used to identify reproducible peaks (IDR < 0.05) at each developmental stage. Next, we used DiffBind v.3.0.14[126] to generate a final consensus peak set of 47,406 peaks, which were normalised using DESeq2. 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. PCA 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[127] and both the consensus peak set and the stage-specific peak sets were classified by genomic feature/region using HOMER v.4.11[128] and further manually curated. Visualisation of peak tracks and gene structures was conducted with pyGenomeTracks v.2.1[129] and deepTools v.3.4.3[122]. To correlate chromatin accessibility and gene expression, this genomic region annotation was used to assign peaks to their closest gene and the Pearson correlation coefficient between chromatin accessibility and gene expression was computed. Promoter and intergenic peaks were subset, for which we then used GimmeMotifs v.0.16.1[130] to perform *de novo* motif search and differential motif enrichment in

our peak sets. Data clustering was performed with mfuzz v.2.52[101]and visualized with ComplexHeatmap package v.2.6.2[131].

Phylostratigraphy

To evaluate gene expression dynamics by phylostratum and developmental stage, we used the OrthoFinder gene families and their inferred origins. We deemed all genes originating before and with the Cnidarian-Bilaterian ancestor of metazoan origin (Supp. Table 27). We then applied a quantile normalisation onto the DESeq2 normalised matrix of gene expression values spanning blastula to juvenile stages. The 75 % percentile of the quantile-normalised gene expression levels was used as the summarising measure of the gene expression distribution by developmental stage. Furthermore, we tested the over- or underrepresentation of the different phylostrata in the gene expression clusters through pair-wise Fisher's exact tests, for which we then adjusted the *p*-values using the Bonferroni correction for multiple testing.

Comparative transcriptomics

Publicly available RNA-seq developmental time courses for the development of *Nematostella vectensis*, *Strongylocentrotus purpuratus*, *Branchiostoma lanceolatum*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Crassostrea gigas*, and two stages of *Capitella teleta* were downloaded from the SRA (Supp. Table 29), cleaned for adaptors and low-quality reads with trimmomatic v.0.39[73] and pseudo-aligned to their respective non-redundant genome-based gene repertoires using kallisto v.0.46.2[98]. We then performed a quantile transformation of TPM values and calculated the Jensen-Shannon divergence (JSD) from single copy orthologs between all possible 1-to-1 species comparisons using the philentropy v.0.5.0 package[132]:

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

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 of each comparison (Supp. Table 30), and normalised using the minimum and maximum adjusted JSD values from all 1-to-1 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 1-to-1 species comparison instead. For a further analysis with a subset of common orthologues, raw mean JSD values needed no adjustment for comparison purposes.

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

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 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).
22 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).
23 Smart, T. I. & Von Dassow, G. Unusual development of the mitraria larva in the polychaete *Owenia collaris*. *Biol Bull* **217**, 253-268 (2009).
24 Gasiorowski, L. *et al.* Molecular evidence for a single origin of ultrafiltration-based excretory organs. *Curr Biol* **31**, 3629-3638 e3622 (2021).
25 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).
26 Seudre, O., Carrillo-Baltodano, A. M., Liang, Y. & Martín-Durán, J. M. ERK1/2 is an ancestral organising signal in spiral cleavage. *bioRxiv*, 2021.2005.2026.445819 (2021).
27 Martin-Duran, J. M. *et al.* Convergent evolution of bilaterian nerve cords. *Nature* **553**, 45-50 (2018).
28 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).
29 Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526-531 (2013).
30 Wang, S. *et al.* Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol* **1**, 120 (2017).
31 Martin-Duran, J. M. *et al.* Conservative route to genome compaction in a miniature annelid. *Nat Ecol Evol* **5**, 231-242 (2021).
32 Shao, Y. *et al.* Genome and single-cell RNA-sequencing of the earthworm *Eisenia andrei* identifies cellular mechanisms underlying regeneration. *Nat Commun* **11**, 2656 (2020).
33 Niehrs, C. On growth and form: a Cartesian coordinate system of Wnt and BMP signaling specifies bilaterian body axes. *Development* **137**, 845-857 (2010).
34 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).
35 Lyons, D. C., Perry, K. J., Batzel, G. & Henry, J. Q. BMP signaling plays a role in anterior-neural/head development, but not organizer activity, in the gastropod *Crepidula fornicata*. *Dev Biol* **463**, 135-157 (2020).
36 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).
37 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).

38 Hui, J. H. *et al.* Extensive chordate and annelid macrosynteny reveals ancestral homeobox gene organization. *Mol Biol Evol* **29**, 157-165 (2012).
39 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).
40 Woltereck, R. Trochophora-Studien I. Histogie der larve und die Entstehung des Annelids bei den *Polygordius*-Arten der Nordsee. *Zoologica* **13** (1902).
41 Simionato, E. *et al. atonal*- and *achaete-scute*-related genes in the annelid *Platynereis dumerilii*: insights into the evolution of neural basic-Helix-Loop-Helix genes. *BMC Evol Biol* **8**, 170 (2008).
42 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).
43 Wu, L. *et al.* Genes with spiralian-specific protein motifs are expressed in spiralian ciliary bands. *Nat Commun* **11**, 4171 (2020).
44 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).
45 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).
46 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).
47 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).
48 Gasiorowski, L. & Hejnol, A. Hox gene expression during development of the phoronid *Phoronopsis harmeri*. *Evodevo* **11**, 2 (2020).
49 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).
50 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).
51 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).
52 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).
53 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).
54 Pascual-Anaya, J. *et al.* Broken colinearity of the amphioxus *Hox* cluster. *Evodevo* **3**, 28 (2012).

55 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).
56 Luo, Y. J. *et al.* Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. *Nat Ecol Evol* **2**, 141-151 (2018).
57 Steinmetz, P. R. *et al. Six3* demarcates the anterior-most developing brain region in bilaterian animals. *Evodevo* **1**, 14 (2010).
58 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).
59 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).
60 Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764-770 (2011).
61 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).
62 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).
63 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **1303.3997** (2013).
64 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).
65 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).
66 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).
67 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).
68 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).
69 Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265-268 (2007).
70 Wong, W. Y. & Simakov, O. RepeatCraft: a meta-pipeline for repetitive element de-fragmentation and annotation. *Bioinformatics* **35**, 1051-1052 (2019).
71 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).
72 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

73 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
74 Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295 (2015).
75 Mapleson, D., Venturini, L., Kaithakottil, G. & Swarbreck, D. Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *Gigascience* **7** (2018).
76 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652 (2011).
77 Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875 (2005).
78 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).
79 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).
80 Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
81 Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654-5666 (2003).
82 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
83 Bryant, D. M. *et al.* A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep* **18**, 762-776 (2017).
84 Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195 (2011).
85 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).
86 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).
87 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).
88 Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95 (2017).
89 Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639-1643 (2021).
90 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).
91 Steinegger, M. & Soding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).
92 Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol* **33**, 1635-1638 (2016).

93 Marletaz, F. *et al.* Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64-70 (2018).
94 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).
95 Zakas, C., Harry, N. D., Scholl, E. H. & Rockman, M. V. The genome of the poecilogonous annelid *Streblospio benedicti*. *Genome Biology and Evolution* (2022).
96 Kwiatkowski, D. *et al.* The genome sequence of the bootlace worm, *Lineus longissimus* (Gunnerus, 1770). *Wellcome Open Res* **6**, 272 (2021).
97 Cannon, J. T. *et al.* Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530**, 89-93 (2016).
98 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).
99 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).
100 Park, C. *et al.* The developmental transcriptome atlas of the spoon worm *Urechis unicinctus* (Echiurida: Annelida). *Gigascience* **7**, 1-7 (2018).
101 Futschik, M. E. & Carlisle, B. Noise-robust soft clustering of gene expression time-course data. *J Bioinform Comput Biol* **3**, 965-988 (2005).
102 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
103 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
104 Gu, Z. & Hübschmann, D. simplifyEnrichment: an R/Bioconductor package for Clustering and Visualizing Functional Enrichment Results. *bioRxiv*, 2020.2010.2027.312116 (2021).
105 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).
106 Katoh, K. & Frith, M. C. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**, 3144-3146 (2012).
107 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).
108 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).
109 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).
110 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).
111 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).

112 Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* **39**, 306-314 (1994).
113 Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564-577 (2007).
114 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
115 Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol Biol Evol* **25**, 1307-1320 (2008).
116 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).
117 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962 (2017).
118 Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**, 2033-2040 (2014).
119 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
120 Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791 (2013).
121 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
122 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).
123 Gaspar, J. M. Improved peak-calling with MACS2. *bioRxiv*, 496521 (2018).
124 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
125 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
126 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393 (2012).
127 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).
128 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).
129 Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets *Bioinformatics* **37**, 422-423 (2020).
130 van Heeringen, S. J. & Veenstra, G. J. GimmeMotifs: a *de novo* motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics* **27**, 270-271 (2011).
131 Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
132 Drost, H. G. Philentropy: Information Theory and Distance Quantification with R. *Journal of Open Source Software* **3**, 765 (2018).

ACKNOWLEDGEMENTS

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 *Themiste lageniformis chordin* sequence, 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*; 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

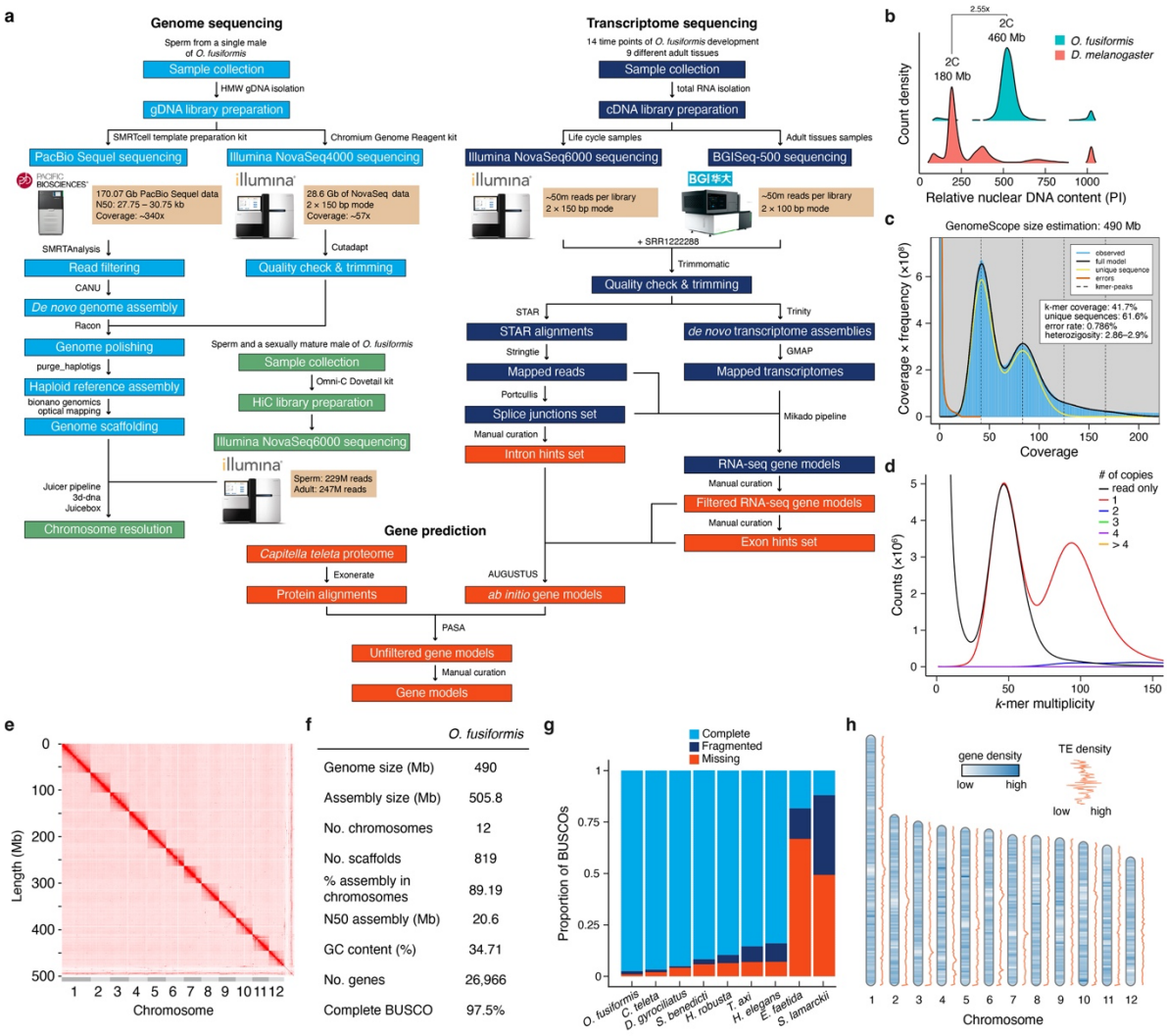
The authors declare no competing interests.

DATA AVAILABILITY

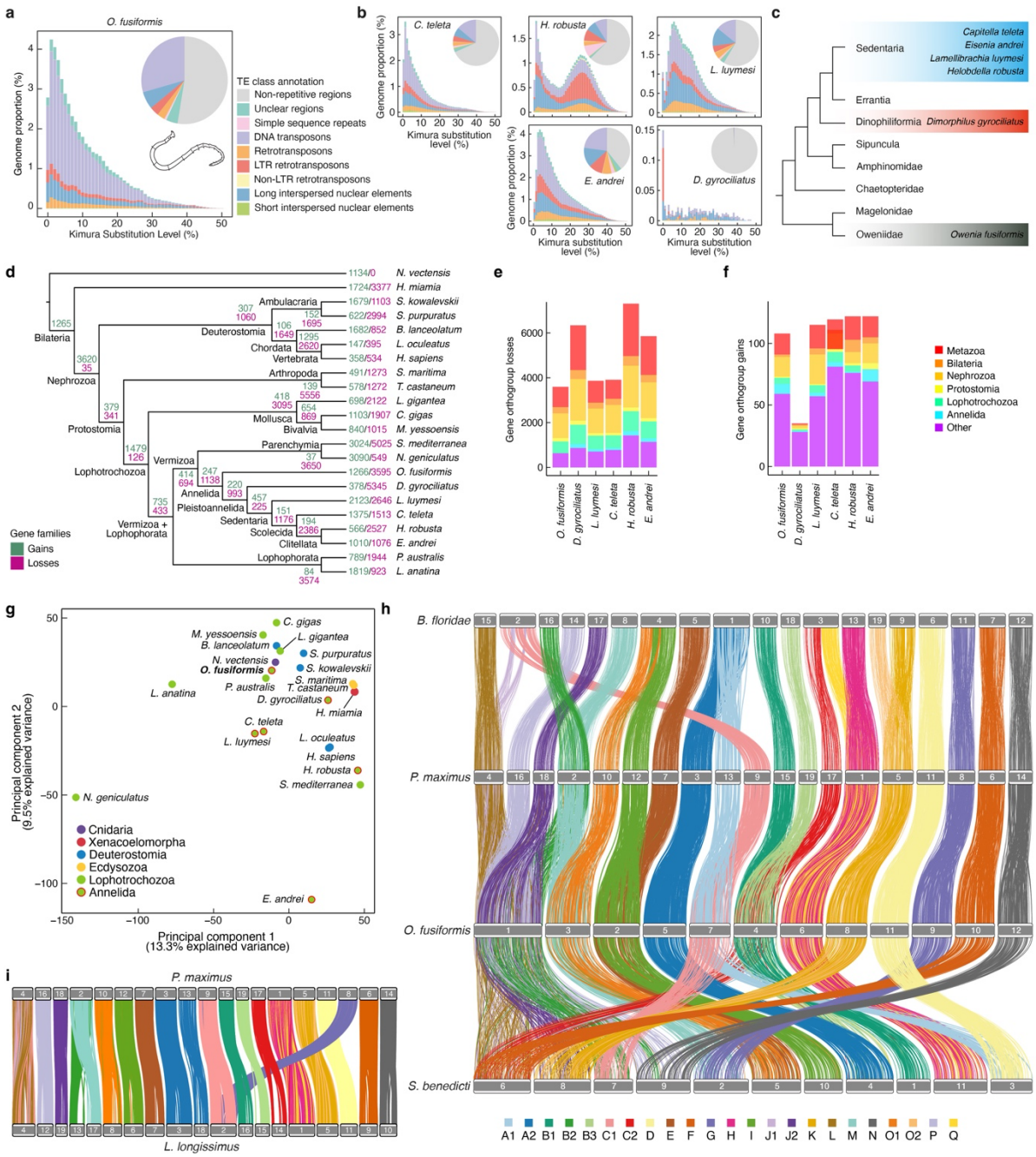
All sequence data associated with this project are available at the European Nucleotide Archive (project PRJEB38497) and Gene Expression Omnibus (accession numbers GSE184126 and GSE192478). Genome assemblies, transposable element annotations, genome annotation files used for RNA-seq and ATAC-seq analyses, 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

No custom code was used in this study.

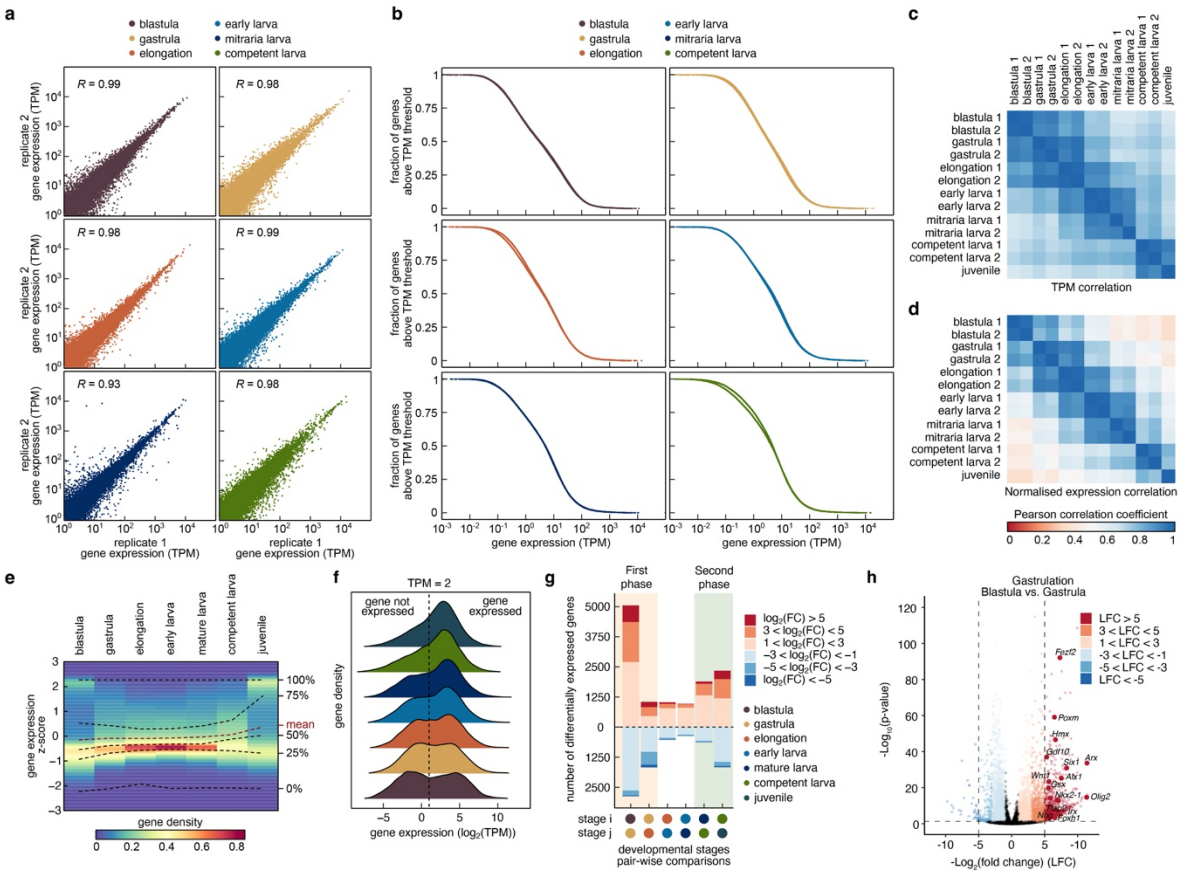


Extended Data Figure 1 | Genome sequencing, assembly, and annotation. a, Flow charts summarising genome (light blue) and transcriptome sequencing (dark blue), chromosome-level scaffolding (green) and gene prediction (red). **b,** Flow cytometry estimation of *O. fusiformis* genome size by comparison of its propidium iodide (PI)-stained nuclear DNA content against *Drosophila melanogaster*. **c,** GenomeScope 2.0 profile and *k*-mer based genome size estimation. **d,** *k*-mer distribution plot indicates the nearly complete de-haploidisation of the reference genome assembly. **e,** Heatmap of HiC contacts showing the inferred twelve chromosomes of *O. fusiformis*. **f,** Genome assembly and annotation statistics. **g,** Comparison of metazoan BUSCO values of selected annelid genomes. **h,** Gene and transposable element (TE) density over the inferred karyotype of *O. fusiformis*.

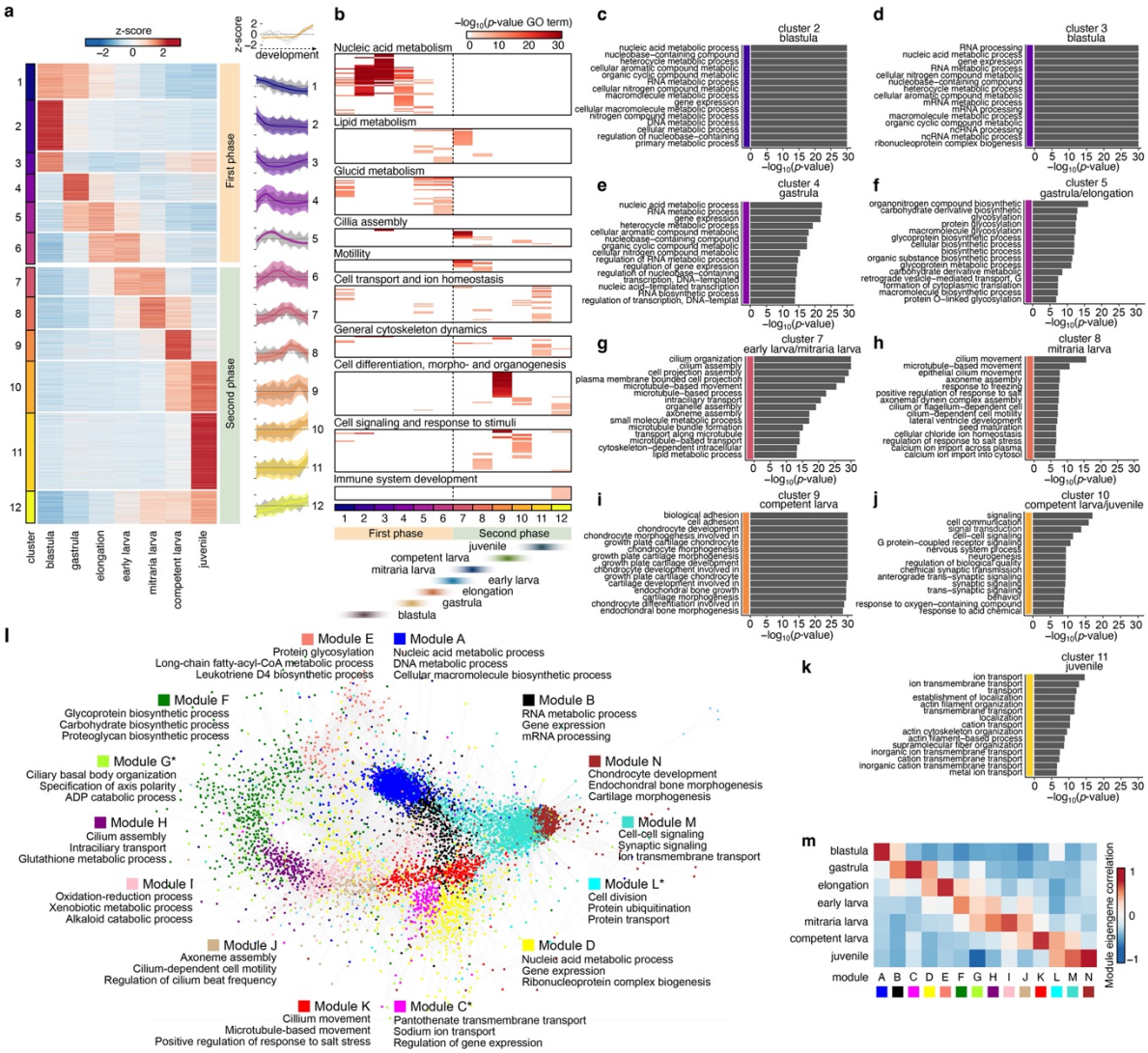


Extended Data Figure 2 | The genome of *Owenia fusiformis* is conservatively evolving. a, b, Pie charts of the transposable element content and Kimura substitution plots of transposable element divergence for *O. fusiformis* and other selected annelid species belonging to different annelid clades as 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, *O. fusiformis*' karyotype is more conserved than that of the annelid *S. benedicti*. **i,** Macrosynteny analysis between the bivalve

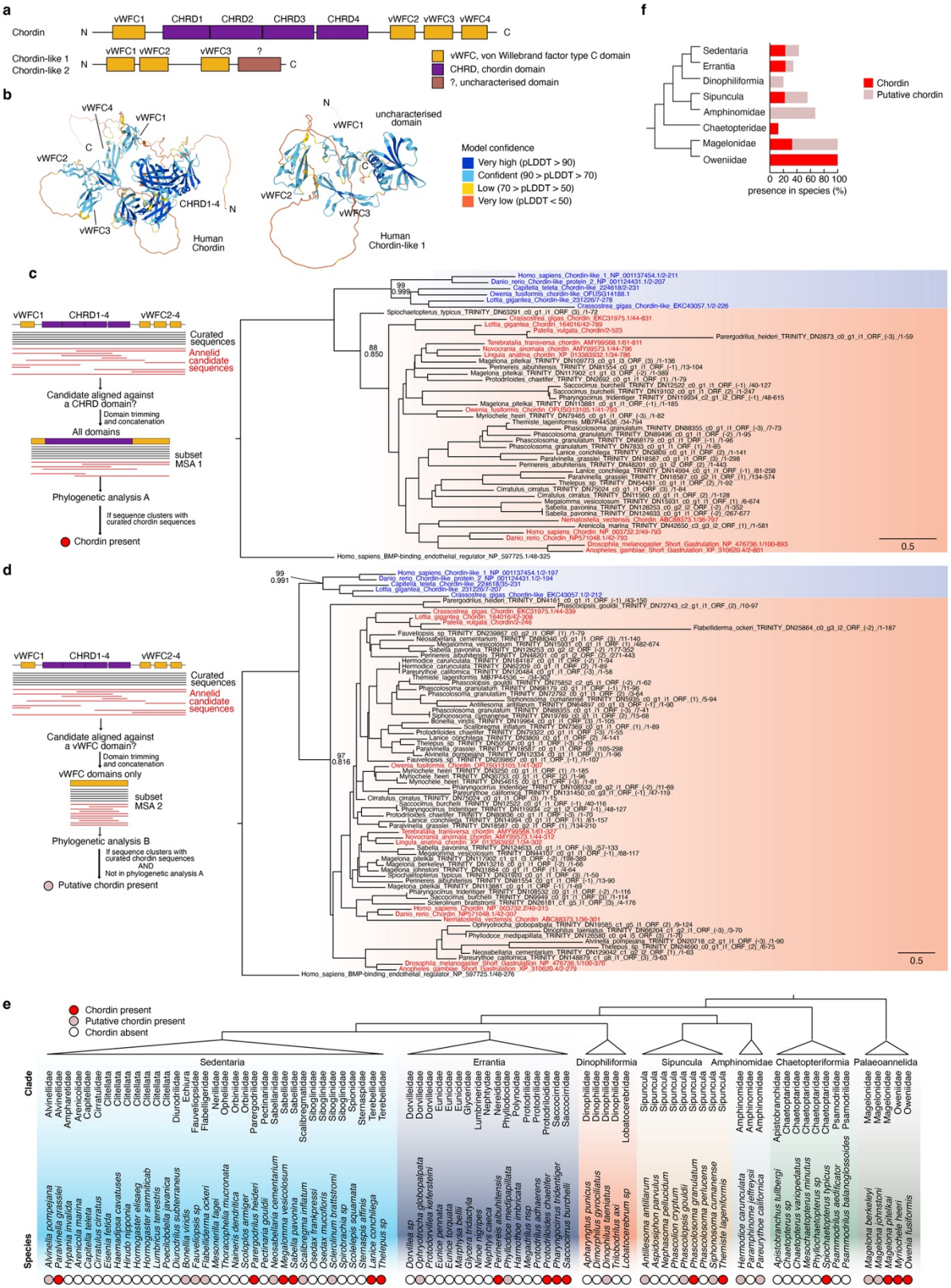
P. maximus and the nemeartean worm *L. longissimus*. *Lineus longissimus* exhibits conserved ancestral bilaterian linkage groups, including three potential lophotrochozoan-specific chromosomal rearrangements (H+Q, J2+L and K+O2), plus a nemertean-specific fusion (G+C1).



Extended Data Figure 3 | Gene expression dynamics during *O. fusiformis* development. Pairwise scatterplots (**a**) and pairwise cumulative frequency plots (**b**) of gene expression levels in transcripts per million (TPM) between biological replicates. *R*: Pearson correlation coefficient. **c, d,** Correlation matrices between RNA-seq experiments, calculated from a matrix of gene expression levels in TPM (**c**) and a variance stabilising-transformed matrix of the normalised DESeq2 matrix (**d**). The satisfactory agreement between biological replicates and an apparent continuum of the developmental transcriptional dynamics can be noted in both matrices. **e,** Gene expression density heatmap, portraying the normalised gene expression (z-score) during *Owenia* development. Dotted black lines represent 0%, 25%, 50%, 75% and 100% quantiles of gene expression. Dotted red line denotes the mean gene expression. As development progresses, gene expression levels increase, with an inflexion point around the early larva stage. **f**, Ridgeline plot of the distribution of genes by gene expression levels used to experimentally defined a cut-off value of TPM = 2 to deem a gene expressed. Colour legend is next to **g**. **g,** Summary of differential gene expression analyses between consecutive developmental stages. Only significantly differentially expressed (i.e., those with an adjusted *p*–value < 0.05) are shown. $\log_2$(FC): $\log_2$(fold-change). **h,** Volcano plot of gastrulation (onset of the first transcriptional phase), highlighting top upregulated axial patterning genes. Colours of dots represent $\log_2$(FC) levels as in (**g**).

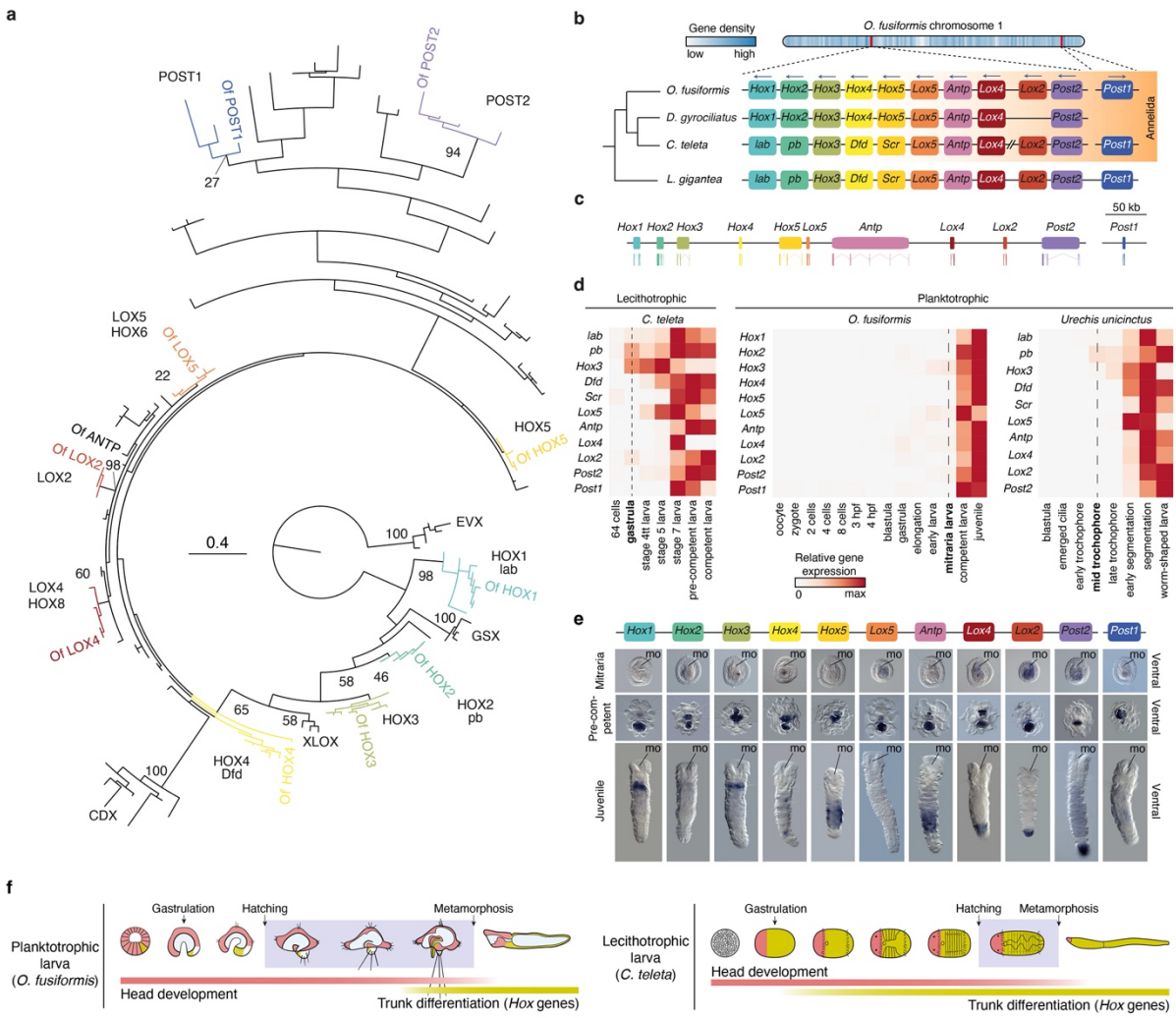


Extended Data Figure 4 | Gene clustering and co-expression network analyses. a, Soft clustered heatmap of all 31,678 transcripts whose expression was not null in at least one developmental stage into an optimal number of 12 clusters. On the right, gene-wise expression dynamics (grey lines) and locally estimated scatterplot smoothing (coloured lines) for each cluster. Coloured shaded areas represent standard error of the mean. **b,** 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. Dotted vertical line depicts the inflexion point between the first and second transcriptional phases and the approximate developmental stage corresponding to each cluster is to the bottom. **c–k,** Bar plots representing *p*–values of top 15 enriched GO terms in nine representative clusters. For the full list of GO terms and clusters, see Supp. Fig. 2. **i**, Force-directed layout representation of the weighted gene co-expression network analysis (WGCNA). For visualisation purposes, only the nodes and edges of a random selection of 30% of the transcripts of *O. fusiformis* are depicted here. For each of the 14 modules, representative enriched GO terms are shown (full lists are in Supp. Fig. 4). Unadjusted *p*–values of GO terms from modules flagged with an asterisk (*) were lower than average. **m,** WGCNA module eigengene correlation with each developmental stage of *O. fusiformis*.

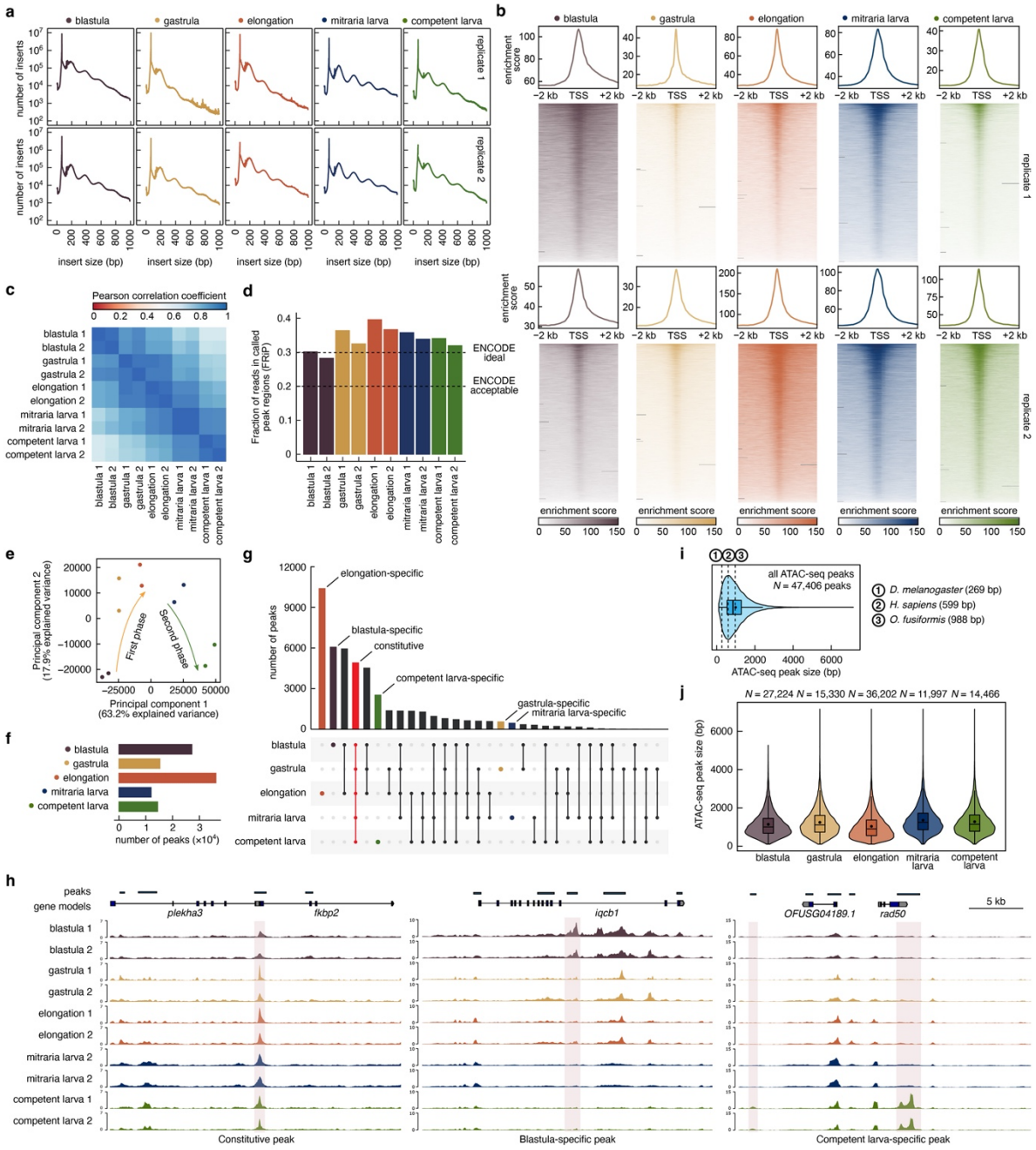


Extended Data Figure 5 | *chordin* was lost multiple times in annelid lineages. a, Domain organisation of Chordin (CHRD) and Chordin-like (CHRDL1/2) proteins, as inferred from human orthologs. **b,** AlphaFold protein structure prediction for human Chordin (UniProt: Q9H2X0) and Chordin-like 1 (UniProt: 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.

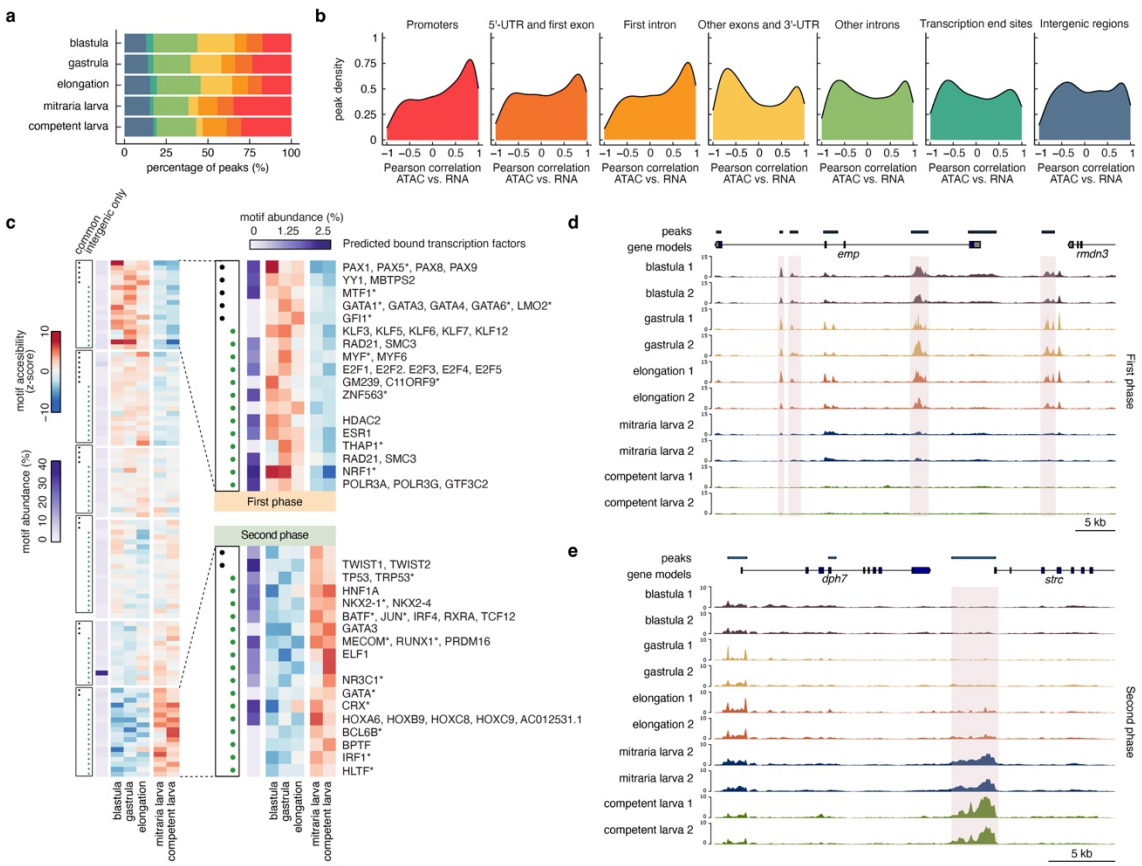


Extended Data Figure 6 | *Hox* gene complement and expression in *O. fusiformis* and other annelid with maximally indirect development. a, Orthology assignment of *O. fusiformis Hox* genes through maximum likelihood phylogenetic inference. Bootstrap support values are shown for major gene groups. Of: *O. fusiformis*. **b,** 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). **c,** Schematic representation to scale of the genomic loci and intron–exon composition of *Hox* genes in *O. fusiformis*. **d,** 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*. **e**, Whole mount *in situ* hybridisation (ventral views) of *Hox* genes in the mitraria, pre-competent, and juvenile stages. The area encircled by a dotted white line at the pre-competent stage highlights a region of probe trapping from ingested food content. mo: mouth. **f,** Heterochrony model depicting the pre-displacement of trunk differentiation in lecithotrophic larvae (as in *C. teleta*) and direct developers, compared to *O. fusiformis* and other planktotrophic annelid larvae.

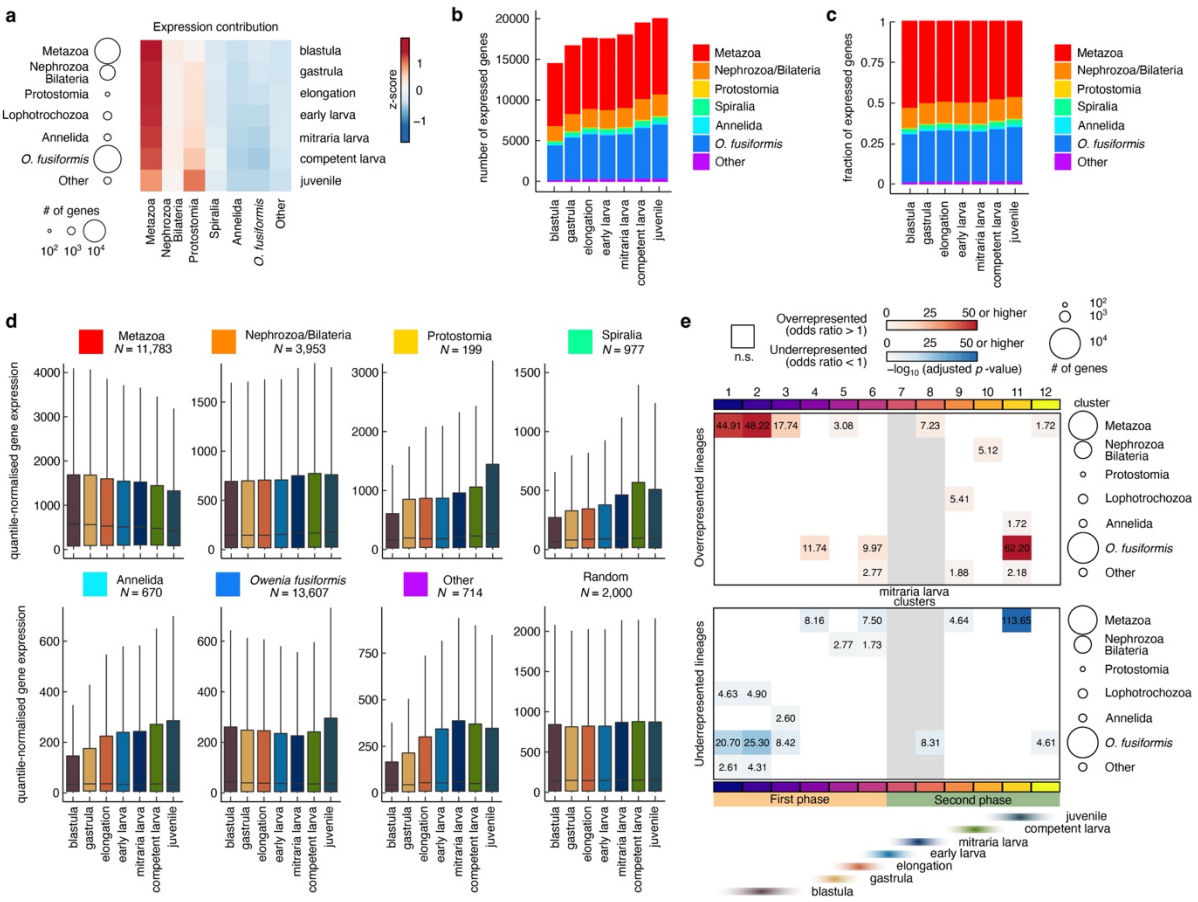


Extended Data Figure 7 | Chromatin accessibility dynamics during *O. fusiformis* development. **a,** Insert size distribution of ATAC-seq libraries show nucleosomal ladders for all samples. **b,** Heatmaps and summary plots of ATAC-seq enrichment around transcription start sites (TSS; ±2 kb) demonstrate at least two-fold enrichment for all samples. **c,** Correlation matrix based on peak accessibility of the called consensus peak set. **d,** FRiP scores for ATAC-seq samples indicating acceptable (FRiP = 0.2) and ideal (FRiP = 0.3) values as per ENCODE standards (dotted lines). **e,** Principal component analysis of ATAC-seq samples based on peaks as in (**c**). **f**, Bar plot of ATAC-seq peak number by developmental stage. **g,** UpSet plot classification of ATAC-seq peaks by developmental stage. Peaks present in more than one stage are coloured in grey. **h**, Representative views of constitutive and stage-specific peaks, with the ATAC-seq peak track indicating IDR consensus peaks. *plekha3:* pleckstrin homology domain containing A3; *fkbp2*: FKBP prolyl isomerase 2; *iqcb1*: IQ calmodulin-binding motif-containing protein 1; *rad50*: RAD50 double strand break repair protein. **i**, **j**, Violin plots of ATAC-seq peak width distribution for all peak set (**i**) and according to developmental stages (**j**). Dotted lines in **i** represent the mean peak size for fruit fly (1),

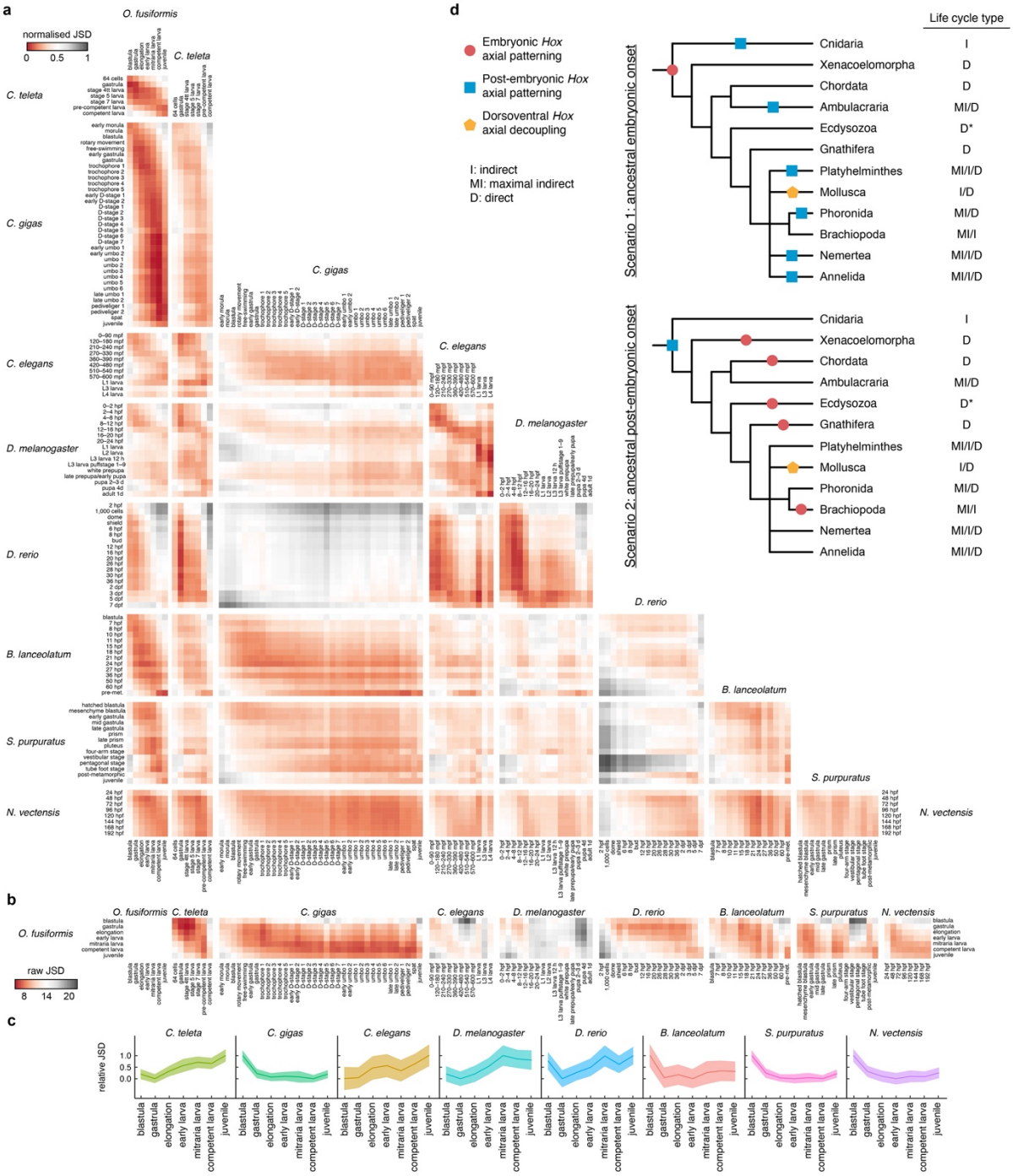
human (2) and *O. fusiformis* (3). Peaks are narrower during axial elongation and broader in the mitraria larva.



Extended Data Figure 8 | Correlation of ATAC-seq and RNA-seq data. a, Percentage of ATAC-seq peaks by genomic feature annotation across *O. fusiformis* development. As development progresses, peaks present in non-first exons and 3'-untranslated regions (3'-UTRs) become less abundant in favour of peaks located in promoters, first exons and 5'-UTRs. **b,** Density plots based on the Pearson correlation coefficient between ATAC-seq peak accessibility and RNA-seq expression level of the nearest gene/transcript to the ATAC-seq peak. Peaks in promoters and first introns are the most positively correlated with gene expression, with most other genomic regions displaying a bimodal distribution. **c**, Motif enrichment analysis of intergenic peaks. Clustered heatmap shows the normalised motif accessibility (z-score), as inferred from chromatin accessibility, and the abundance of the motif in promoter peaks. Clusters 1 (top right) and 6 (bottom right) are shown in detail as representative of each regulatory programme. **d, e,** Representative peaks associated with each of the two regulatory programmes highlighted in light red. Peak track corresponds to the consensus ATAC-seq peak set. *emp*: epithelial membrane protein-related; *rmdn3*: regulator of microtubule dynamics 3; *dph7*: diphthamide biosynthesis 7; *strc*: stereocilin.



Extended Data Figure 9 | The development of the mitraria larva does not rely on gene novelties. **a**, Expression contribution of each phylostratum by developmental stage, calculated from the 75% percentile of a quantile-normalised matrix of gene expression levels. Metazoan genes are expressed at the highest levels at all stages but the juvenile, at which genes of protostomian origin take over. **b,** Number and **c,** percentage of expressed genes throughout development classified by phylostratum, as per Fig. 4a. **e,** Boxplots of quantile-normalised expression levels of genes classified by phylostratum across *O. fusiformis* development. **f,** Enrichment analysis of the number of genes per phylostratum in clusters of co-transcribed genes as inferred through soft clustering and shown in Extended Data Fig. 4a. For each cluster and lineage combination, the Bonferroni-adjusted *p*–value of the Fisher's exact test is shown. The upper table includes significantly overrepresented lineages (odds ratio, OR > 1, adjusted *p*–value < 0.05) and the lower table includes significantly underrepresented lineages (OR < 1, adjusted *p*–value < 0.05). The results match the expression dynamics displayed in **d** (shaded grey are indicates the mitraria stage). In **c** and **f**, bubble size is proportional to the number of genes in each phylostratum.



Extended Data Figure 10 | Comparative transcriptomic analysis of bilaterian life cycles. a, Matrix of heatmaps of normalised transcriptomic Jensen–Shannon Divergence (JSD) from pairwise comparisons of all single copy, one-to-one orthologs between nine animal species with different life cycles: the cnidarian *N. vectensis*, the sea urchin *S. purpuratus*, the cephalochordate *B. lanceolatum*, the vertebrate *D. rerio*, the nematode *C. elegans*, the insect *D. melanogaster*, the bivalve *C. gigas*, and the annelids *O. fusiformis* and *C. teleta*. **b,** Heatmaps of JSD between *O. fusiformis* and all other eight species using only the subset of single copy orthologs common to all nine species. This excludes all genes that are not of pre-metazoan and metazoan origin, as they will not be present in *N. vectensis*. **c,** Relative JSD from stages of minimal divergence of eight animal species to each *O. fusiformis* developmental stage calculated from the subset of single copy common orthologs. Confidence intervals are the standard deviation from 250 bootstrap resamplings of the single copy ortholog subset ($N = 459$ orthologs for all eight pairwise comparisons). **d**, 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 (see Supp. Table 33), 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 (bottom scenario), which could have been favoured the evolution of bilaterian larvae.