Annelid functional genomics reveal the origins of bilaterian life cycles

Jose Martín-Durán (✉ chema.martin@qmul.ac.uk)
  Queen Mary University of London   https://orcid.org/0000-0002-2572-1061

Yan Liang
  Queen Mary University of London

Francisco M. Martín-Zamora
  Queen Mary University of London   https://orcid.org/0000-0002-2558-3762

Kero Guynes
  Queen Mary University of London

Allan Carrillo-Baltodano
  Queen Mary University of London

Yongkai Tan
  Okinawa Institute of Science and Technology Graduate University

Giacomo Moggioli
  Queen Mary University of London

Oceane Seudre
  Queen Mary University of London

Martin Tran
  Queen Mary University of London

Kate Martimer
  Amgueddfa Cymru – National Museum Wales

Nicholas Luscombe
  Okinawa Institute of Science and Technology Graduate University

Andreas Hejnol
  University of Bergen   https://orcid.org/0000-0003-2196-8507

Ferdinand Marlétaz
  University College London

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)

23

24   **Indirect development with an intermediate larva exists in all major animal lineages[1],**

25   **and thus larvae are central to most scenarios for animal evolution[2-12]. Yet how larvae**

26   **evolved remains disputed. Here we show that changes in the timing of trunk formation**

27   **underpin the diversification of larvae and bilaterian life cycles. Combining**

28   **chromosome-scale genome sequencing with transcriptomic and epigenomic profiling in**

29   **the slow-evolving oweniid *Owenia fusiformis*[13], we found that different genes and**

30   **genomic regulatory elements control the development of its feeding larva and adult**

31   **stage. First, *O. fusiformis* embryos develop into an enlarged anterior domain that forms**

32   **larval tissues and the adult head, as posterior growth and trunk patterning is deferred**

33   **to pre-metamorphic stages. These traits also occur in the so-called "head larvae" of**

34   **other bilaterians[14,15], with whom *O. fusiformis* larva shows extensive transcriptomic**

35   **similarities. Conversely, animals with non-feeding larvae and gradual metamorphoses,**

36   **such as the annelid *Capitella teleta*, start trunk differentiation during embryogenesis,**

37   **like direct developers. Together, our findings suggest that the ancestral temporal**

38   **decoupling of head and trunk formation, as retained in extant "head larvae", allowed**

39   **larval evolution in Bilateria, questioning prevailing scenarios that propose either co-**

40   **option[10,11] or innovation[12] of gene regulatory programmes to explain larva and adult**

41   **origins.**

42

43   Many animal embryos develop into an intermediate, often free-living stage termed larva,

44   which later metamorphoses into the sexually competent adult[1,2]. Larvae are vastly diverse and

45   can display from radically different to more similar morphologies than those of their adults[1,2].

46   Given their broad phylogenetic distribution[2], larvae are central to major scenarios of animal

47   evolution[2-12]. These fundamentally disagree on whether larvae are ancestral[2-7] or secondarily

48   evolved[10,11] life stages and propose different mechanisms to explain the evolution of larva

49   and adult forms (Fig. 1a, b). The "intercalation" hypothesis[10,11] suggests that larval stages

50   were added to animal life cycles multiple times independently, by co-opting genes and

51   genetic programmes originally expressed in the adult (Fig. 1a). Conversely, the "terminal

52   addition" scenario[2,3,12] considers that the ancestral bilaterians resembled existing larvae, and

53   thereby adults convergently evolved through the parallel evolution of adult-specific genetic

54   programmes[12] (Fig. 1b). An assessment of the mechanisms underlying these hypotheses using

55   comparative and functional genomics data is, however, lacking, and thus larval origins—and

56   their importance to explain animal evolution—are still contentious.

57

58   The trochophore[16] is a widespread larval type generally characterised by an apical sensory

59   organ and a pre-oral locomotive ciliary band[17] that is classically assigned to Annelida and

60   Mollusca but also potentially to related clades within Lophotrochozoa[18] (Fig. 1c).

61   Trochophore larvae—traditionally exemplified by that of annelid worms—are pivotal to the

62   "terminal addition" scenario[2,19], which regards this larval type a vestige of the last common

63   adult ancestor to Protostomia[2,3], or even Bilateria[20] (Fig. 1c). Compared to most other annelid

64   lineages, however, the larvae of the annelid groups Oweniidae and Magelonidae—which

65   form Oweniida, the sister group to all other annelids[13]—exhibit unusual characters

66   (Fig. 1c, d). In particular, the idiosyncratic oweniid larva, commonly referred to as

67   "mitraria"[21], has an enlarged pre-oral region and a single tuft of posterior chaetae, as well as a

68   pair of nephridia and a long monociliated ciliary band alike those of phylogenetically distant

69   larvae of deuterostome lineages, such as echinoderms and hemichordates[22-24]. Despite these

70   distinctive larval traits, oweniids exhibit many developmental characters considered ancestral

71   to Annelida, and even Lophotrochozoa as a whole[25,26], as well as similarities in larval

72 molecular patterns with other trochophore and bilaterian larvae[24,25,27,28]. Consequently,

73 whether the mitraria larva is the result of divergent or convergent evolution is unclear, which

74 makes the comparative analysis of this lineage-restricted larva an excellent case study to

75 investigate how larval traits evolve, and thereby formulate and assess scenarios on the origin

76 of animal life cycles.

77

78 Here, we characterise the reference chromosome-scale genome assembly of the oweniid

79 *Owenia fusiformis* (Fig. 1e) and perform a comprehensive study of its developmental

80 transcriptome and regulatory genome that allows us to dissect the gene regulatory events

81 underpinning the formation of the mitraria larva. The comparison of our comprehensive

82 dataset with other lophotrochozoan and bilaterian taxa provides compelling evidence that

83 heterochronic shifts in trunk development, rather than co-option and innovation in genetic

84 programmes—as traditionally proposed—underpin the evolution of bilaterian life cycles.

85

86 **Results**

87 O. fusiformis *has a conservatively evolving genome*

88 To characterise the transcriptomic and genomic regulatory basis for larval development in

89 *O. fusiformis*, we first generated a chromosome-scale reference assembly combining PacBio

90 long-reads, 10x genomics read clouds, optical mapping, and Hi-C scaffolding (Extended Data

91 Fig. 1a). Consistent with flow cytometry and k-mer estimations, the haploid assembly spans

92 505.8 Mb (Extended Data Fig. 1b–d; Supplementary Fig. 1), exhibiting 12 chromosome-scale

93 scaffolds that encompass 89.2% of the assembly (Extended Data Fig. 1e, f). Almost half of

94 the assembly consists of repeats (43.02%; largely DNA transposons) acquired steadily during

95 evolution (Extended Data Fig. 2a–c; Supplementary Table 1). Using transcriptomic data from

4

96   14 developmental stages and 9 adult tissues (Extended Data Fig. 1a), we annotated 26,966

97   protein-coding genes and 31,903 transcripts, representing a nearly complete (97.5%) set of

98   metazoan BUSCO genes (Extended Data Fig. 1g). Gene family reconstruction and principal

99   component analysis on gene content across 22 animal genomes nested *O. fusiformis* within

100   other non-annelid lophotrochozoan species (Fig. 1f), supporting that *O. fusiformis* has fewer

101   gene family gains and losses, and retains more ancestral metazoan orthogroups than other

102   annelid taxa (Fig. 1g; Extended Data Fig. 2d–i; Supplementary Tables 2–7). Therefore,

103   *O. fusiformis* genome contains a more complete gene repertoire that those reported for other

104   annelid lineages, which together with its phylogenetic position and conserved early

105   embryogenesis[25,26] makes it a key lineage to reconstruct the evolution of Annelida and

106   Lophotrochozoa generally.

107

108   Chromosomal linkage of orthologous genes between deuterostomes and protostomes supports

109   the presence of 22 ancestral linkage groups in bilaterians[29,30]. To explore the conservation of

110   the high-order organisation of *O. fusiformis* genome, we compared its syntenic orthology

111   with the cephalochordate *Branchiostoma lanceolatum* (a deuterostome) and the mollusc

112   *Pecten maximus* (a lophotrochozoan). *Owenia fusiformis* has globally retained the ancestral

113   bilaterian linkage groups (Fig. 1h; Extended Data Fig. 2h, i) and exhibits chromosomal

114   fusions that are present in the scallop and even a nemertean (H+Q, J2+L and K+O2), which

115   are thus likely ancestral to Lophotrochozoa. While the nemertean *Lineus longissimus* presents

116   only one more fusion in addition to these three ancestral lophotrochozoan chromosomal

117   rearrangements, *O. fusiformis*, as well as the polychaete *Streblospio benedicti*, show

118   additional fusions of ancestral bilaterian linkage groups, which are notably not shared

119   between the two taxa and therefore none are ancestral to Annelida (Extended Data Fig. 2h, i).

120   However, *O. fusiformis* still retains a more ancestral chromosomal organisation than

121   *S. benedicti*, in which all chromosomes except two are derived from secondary chromosomal

122   fusion events, and other annelids with highly contiguous assemblies[31,32]. Together, our high-

123   quality assembly of *O. fusiformis* supports that phylogenetically distant lophotrochozoan

124   lineages largely share an ancestral syntenic chromosomal organisation, further revealing a

125   dynamic chromosomal evolution in annelids, albeit their generally stable gene complement.

126

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

128   The "intercalation" scenario proposes that larvae evolved secondarily by co-opting genes and

129   gene modules expressed in adult stages[10,11] (Fig. 1a). Larva and adult would thus share

130   transcriptional similarities, while more gradual transcriptional dynamics would be expected

131   during direct development, as in the annelid *Dimorphilus gyrociliatus*[31]. To test this

132   hypothesis, we used bulk stage-specific transcriptomic data from the blastula—the stage at

133   which axial polarity is established in *O. fusiformis*[26]—to juvenile (as a proxy to adult stages),

134   covering larval embryogenesis, growth, and metamorphosis (Fig. 2a; Extended Data Fig. 3a–

135   f; Supplementary Tables 8–10). We observed two main phases of increased gene expression

136   and differential gene expression activity during the life cycle of *O. fusiformis*, with the early

137   mitraria as transitional stage: a first phase leading to larva formation (i.e., embryogenesis),

138   followed by a second phase encompassing larval growth and metamorphosis into the juvenile

139   (Fig. 2b, c; Extended Data Fig. 3g). Soft clustering and weighted gene co-expression network

140   analysis of the 31,678 expressed transcripts generated an optimal number of 12 distinct

141   clusters and 14 modules of temporally co-regulated genes, respectively, which broadly

142   classified as peaking during either the first or the second transcriptional phases (Extended

143   Data Fig. 4; Supplementary Fig. 2–4; Supplementary Tables 11–13). In none of the two

144   approaches, however, were there clusters of genes with peaks of expression at both the

6

145   mitraria and the juvenile stage, and only one cluster (1,426 transcripts; 4.5%) showed a

146   bimodal activation at blastula and juvenile stages (Extended Data Fig. 4a). Therefore,

147   extensive co-option of adult genes into larva embryogenesis—as the "intercalation" scenario

148   posits—does not occur in *O. fusiformis*, and instead larval embryogenesis and juvenile/adult

149   development are two markedly distinct transcriptional phases.

150

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

152   To identify the developmental processes underpinning the two transcriptional phases during

153   *O. fusiformis* life cycle, we performed gene ontology (GO) enrichment analyses on the soft

154   clusters and modules of temporally co-regulated genes (Extended Data Fig. 4b–l). GO terms

155   related to nucleic acid metabolism—i.e., DNA replication and transcription—and complex

156   biosynthetic metabolism predominate in clusters of the first transcriptional phase (Extended

157   Data Fig. 4b–f), consistent with the establishment of the larval body plan and development of

158   the functional digestive system at those stages[21,25]. Among the genes expressed in this first

159   phase, we identified an ortholog of *chordin*, a key bone morphogenetic protein (BMP)

160   inhibitor involved in dorsoventral patterning across Bilateria[33] and thought to be lost in

161   annelids[34] (Fig. 2d; Extended Data Fig. 5a, b). With a complex pattern of retentions and

162   losses in Annelida (Extended Data Fig. 5c–f; Supplementary Tables 14–18), *chordin* is

163   asymmetrically localised around the blastopore lip of the gastrula and later in the larval

164   antero-ventral oral ectoderm (Fig. 2e). This expression resembles that of molluscs[34,35] and

165   brachiopods[28], providing further evidence that conserved developmental mechanisms

166   underpin larval formation in *O. fusiformis*[26-28]. However, differently from non-feeding

167   annelid trochophores[6], anterior ectodermal genes[26-28] pattern most larval tissues in

168   *O. fusiformis* (Fig. 2f), whereas posterior genes are restricted to few circum-anal cells, in line

169   with the vastly enlarged pre-oral region of the mitraria that forms early on in the adult head[21].

170

171   GO terms involved in terminal cell differentiation, morphogenesis, and organogenesis

172   predominate in co-expression clusters of the second transcriptional phase. *Hox* genes, a

173   conserved family of transcription factors involved in anterior-posterior trunk regionalisation

174   in Bilateria[36], are among the most upregulated genes at these stages (Fig. 2g).

175   *Owenia fusiformis* has a conserved complement of 11 *Hox* orthologues arranged as a

176   compact, ordered cluster in chromosome 1, except for *Post1*, which is located downstream on

177   that same chromosome (Extended Data Fig. 6a–c; Supplementary Table 19). This is a similar

178   repertoire and genomic arrangement to that of the annelids *Capitella teleta*[37] and

179   *Platynereis dumerilii*[38]; yet unlike these annelids, which deploy *Hox* genes during or soon

180   after gastrulation[37,39] (Extended Data Fig. 6d), *O. fusiformis* does not express *Hox* genes

181   during embryogenesis to pattern the larval body (Fig. 2h; Extended Data Fig. 6d, e). Instead,

182   *Hox* genes are expressed in the trunk rudiment during larval growth, already in an anterior–

183   posterior staggered pattern that is retained in the juvenile after metamorphosis (Fig. 2h–j;

184   Extended Data Fig. 6e). This late activation of *Hox* genes is not unique to *O. fusiformis* but

185   also occurs in the planktotrophic trochophore of the echiuran annelid *Urechis unicinctus*

186   (Extended Data Fig. 6d; Supplementary Table 20). Therefore, posterior/trunk development is

187   deferred to pre-metamorphic stages in planktotrophic annelid trochophores[14,40] compared to

188   annelids with lecithotrophic larvae[37,38] and direct developers[31] (Extended Data Fig. 6f).

189   Altogether, our findings support that axial and body patterning, as well as anterior and gut

190   differentiation contribute to the first transcriptional phase, while adult trunk formation largely

191   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, Supplementary 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, Supplementary 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, Supplementary 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, Supplementary Tables 25, 26, Supplementary 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

217 neurogenesis (ATOH[41]) are amongst the most differentially accessible in open chromatin

218 regions during embryogenesis, whilst ciliary band genes (OTX[28]), trunk related genes (e.g.,

219 NKX2.1[27]) and most notably *Hox* genes appear in regulatory regions during larval

220 competence and trunk patterning (Fig. 3e; Extended Data Fig. 8c). Therefore, two different

221 sets of *cis*-regulatory elements mirror the major transcriptional dynamics in *O. fusiformis*,

222 reinforcing the view that separate genetic mechanisms that are temporally decoupled

223 underpin larva and adult formation in this annelid.

224

225 *Mitraria development does not rely on novel genes*

226 Novel genes account for a significant proportion of some trochophore and other larval

227 transcriptomes[7,42], as they might be associated with the development of larval-specific

228 characters (e.g., ciliary bands)[7,42,43]. Therefore, recruitment of novel genes at the mitraria

229 stage could also explain the overall transcriptomic differences between this and the adult

230 stage, as well as the distinctive morphological traits of this larval form. To define the

231 contribution of novel genes at each developmental stage in *O. fusiformis*, we classified all

232 predicted transcripts in seven phylostrata according to their time of origin (Fig. 4a,

233 Supplementary Table 27). Older genes (genes of metazoan and pre-metazoan origin;

234 phylostratum 1) represent the largest fraction of genes expressed at all developmental stages

235 except that of the juvenile in *O. fusiformis*, whereas the highest expression levels of

236 novel/younger genes (phylostratum 7) occur in the juvenile stage (Fig. 4b, Extended Data

237 Fig. 9a–d). An enrichment analysis of each phylostrata on each of the 12 recovered gene

238 cluster showed that older genes are significantly enriched in early embryonic stages (clusters

239 1 to 3), while novel genes are significantly overrepresented in gene clusters active at the

240 juvenile stage and most notably, underrepresented in the mitraria larva (clusters 7 and 8;

241   Extended Data Fig. 9e). Therefore, older rather than younger genes contribute to the

242   development of the mitraria larva, suggesting that the increased use of novel genes in other

243   lophotrochozoan larvae[7,42] might be due to lineage-specific traits found on those larvae, such

244   as the shell primordium of molluscan trochophores[7] and perhaps even ciliary bands with

245   multiciliated cells[44], which are absent in oweniid larvae.

246

247   *Bilaterian planktotrophic larvae share maximal transcriptional similarity*

248   The "terminal addition" scenario[2,3,19] posits that larvae are homologous and adults evolved

249   convergently, and thus pre-larval stages should be more conserved than metamorphosis and

250   adult stages across species[11] (Fig. 1b). To test this hypothesis, we first generated bulk stage-

251   specific transcriptomic data for the annelid *C. teleta* (Supplementary Fig. 7, Supplementary

252   Table 28), which exhibits a typical lecithotrophic trochophore (Fig. 1c). Differently from

253   *O. fusiformis*, *C. teleta* does not show two marked transcriptional phases but rather a gradual

254   increase of gene expression during its life cycle (Fig. 4c), as direct developers do[31]. Pairwise

255   transcriptomic comparisons between *O. fusiformis* and *C. teleta* demonstrate overall similar

256   transcriptional dynamics between the development of these annelids, yet the point of

257   maximal transcriptional divergence occurs at larval—and not adult—stages (Fig. 4d). While

258   this observation disagrees with the "terminal addition" scenario, it rather reflects, and is

259   explained by, the different timings of trunk differentiation in *O. fusiformis* and *C. teleta*

260   (Extended Data Fig. 6f).

261

262   We thus extended our comparative transcriptomic approach to six other bilaterian lineages

263   and the cnidarian *Nematostella vectensis* to cover the overall diversity of life cycle strategies

264   found in Metazoa, using both all single copy one-to-one orthologs (Fig. 4a, e; Extended Data

265   Fig. 10a; Supplementary Tables 29, 30) and a reduced set of conserved cross-species single

266   copy orthologs (Extended Data Fig. 10b, c). Transcriptional dynamics between *O. fusiformis*

267   and other major animal groups are more dissimilar (except in *Danio rerio*) at early

268   development and become more similar as development proceeds towards juvenile and adult

269   stages (Fig. 4e, f; Extended Data Fig. 10a–c; Supplementary Tables 31, 32). At the larval

270   stage, *O. fusiformis* shows vast transcriptomic differences with the larvae of *Drosophila*

271   *melanogaster* and *Caenorhabditis elegans*, the two ecdysozoan taxa with indirect

272   development and whose larvae evolved secondarily[10] (Fig. 4f). However, *O. fusiformis* shares

273   maximal transcriptomic similarities at larval stages with bilaterian species with

274   planktotrophic ciliated larvae and even the early planula of the cnidarian *Nematostella*

275   *vectensis* (Fig. 4e, f; Extended Data Fig. 10a, b). Together, these findings question the

276   "terminal addition" scenario, because adult development is generally more conserved than

277   embryogenesis, but also reveal unexpected genome-wide transcriptional similarities between

278   phylogenetically distant bilaterian planktotrophic larvae.

279

280   **Discussion**

281   In this study, we report the chromosome-scale genome assembly of the oweniid *O. fusiformis*

282   which, together with extensive gene expression and genomic regulatory profiling during

283   embryonic and larval development, provides an unprecedented perspective on mainstream

284   scenarios for life cycle evolution in Bilateria. In particular, our results refute predictions that

285   propose either the co-option of adult genes into larval gene regulatory programmes[10,11] (in

286   line with the "intercalation" scenario) or the independent evolution of adult gene regulatory

287   modules[2,3,19] (as it would be expected in the "terminal addition" scenario) as drivers of larva

288   and adult evolution, respectively. Instead, the distinctive mitraria larva of *O. fusiformis*

289   develops from an enlarged head region while trunk differentiation is deferred to late larval

290   stages (Fig. 5a; Extended Data Fig. 6f). Similar developmental traits occur in other feeding

291   annelid larvae[40] (Extended Data Fig. 6e) including that of *Chaetopterus*[45,46], a member of the

292   second earliest branching annelid lineage[47], as well as in other phylogenetically distant clades

293   within Lophotrochozoa (e.g., nemerteans[48] and phoronids[49]), Ecdysozoa (pancrustaceans[15]

294   and pycnogonids[50]), and Deuterostomia (e.g., echinoderms[51,52] and hemichordates[53]), whose

295   larvae are generally referred to as "head larvae"[14,15]. By contrast, non-feeding larvae[37,39] and

296   direct developers[31] in both Annelida (Extended Data Fig. 6f) and other bilaterian taxa[54,55]

297   start trunk patterning soon after gastrulation (Fig. 5b), while the onset of anterior/head

298   patterning always takes place before gastrulation in bilaterians[56]. Regardless of these different

299   timings, major bilaterian clades share head[57,58] and trunk molecular patterns (e.g., a *Hox*

300   code[36]; Fig. 2h), supporting the overall homology of the adult bilaterian body plan. Therefore,

301   our findings indicate that heterochronic shifts of an adult trunk developmental programme

302   correlate with, and may account for, the evolution of the different life cycles in Annelida, and

303   Bilateria generally[53].

304

305   The post-embryonic activation of genetic programmes involved in trunk differentiation

306   observed in "head larvae" has generally been considered a lineage-specific innovation

307   associated with the evolution of maximal indirect development[14,15,23,49]. Consequently, "head

308   larvae" are thought to have evolved convergently and the observed similarities in larval

309   molecular patterns[6,49,53,59] might reflect ancient gene regulatory modules that were

310   independently co-opted to develop analogous larval organs (e.g., apical organs and ciliary

311   bands), in line with the "intercalation" scenario[10,11]. In contrast, our study supports that the

312   late deployment of trunk differentiation programmes is likely ancestral to Annelida (Fig. 5c),

313   and not necessarily related to maximal indirect development. Indeed, an ancestral post-

314   embryonic onset of trunk differentiation is a more parsimonious state for Bilateria, based on

315   our current understanding of the timing of trunk-associated expression of *Hox* genes in major

316   bilaterian clades and cnidaria (Extended Data Fig. 10d; Supplementary Table 33). This

317   temporal decoupling between head and trunk genetic programmes could thus have facilitated

318   the evolution of larvae, which would then be homologous on the grounds of being largely

319   anteriorly derived transitory structures that form intermediate life cycle specialisations.

320   Notably, this scenario does not exclude the possibility that as bilaterians (and thus larvae)

321   diversified, conserved gene regulatory modules were also convergently reused to develop

322   analogous larval organs in different lineages. Regardless of the scenario, however, our study

323   highlights the importance of heterochronic changes in the deployment of ancient genetic

324   programmes for the diversification of bilaterian life cycles, which had profound ecological

325   and evolutionary implications. In the future, comparative genomic studies that thoroughly

326   dissect the regulatory principles underlying head and trunk development in multiple bilaterian

327   lineages will help to decipher how temporal changes in their deployment drove the evolution

328   of larval and adult forms in Bilateria.

329

330   **Materials and Methods**

331   *Adult culture, spawning and in vitro fertilisation*

332   Sexually mature *Owenia fusiformis* Delle Chiaje, 1844 adults were collected from subtidal

333   waters near the Station Biologique de Roscoff and cultured in the lab as described before[25]. *In*

334   *vitro* fertilisations and collections of embryonic and larval stages were performed as

335   previously described[25]. *Capitella teleta* Blake, Grassle & Eckelbarger, 2009 was cultured,

336   grown, and sifted, and its embryos and larvae were collected following established

337   protocols[60]. *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[61] to count and generate a 31-mer histogram

from adaptor-cleaned, short-read Illumina reads (see section below), and GenomeScope 2.0[62]

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[63] assuming 'batOptions="-dg 3 -db 3 -dr 1 -ca 500 -cp 50' and

362 'correctedErrorRate=0.065'. Pacbio reads were remapped using pbalign v.0.3.2 and the

363 assembly polished once using Arrow (genomicconsensus, v2.3.2). Then Illumina paired end

364 reads generated with the 10x Genomics linked reads were extracted, remapped using bwa

365 mem v.0.7.17[64] and used for polishing with Racon v.1.16[65]. Bionano Genomics optical

366 mapping data was used to scaffold the PacBio-based assembly, which was de-haploidised

367 with purge_haplotigs v.1.0.4[66] setting cut-offs at 35, 85 and 70x coverages to reconstruct a

368 high-quality haploid reference assembly. HiC-based chromosome scaffolding was performed

369 as described below. Merqury v.1.1[67] and BUSCO v.5[68] were used to assess genome

370 completeness and evaluate the quality of the assembly.

371

372 *Transcriptome sequencing*

373 Fourteen samples spanning key developmental time points of *O. fusiformis* life cycle,

374 including active oocyte, zygote, 2-cell, 4-cell, and 8-cell stages, 3 hours post-fertilisation

375 (hpf), 4 hpf, coeloblastula (5 hpf), gastrula (9 hpf), axial elongation (13 hpf), early larva

376 (18 hpf), mitraria larva (27 hpf), pre-metamorphic competent larva (3 weeks post-

377 fertilisation, wpf) and post-metamorphic juvenile were collected in duplicates (except for the

378 latter), flash frozen in liquid nitrogen and stored at –80 °C for total RNA extraction. Samples

379 within replicates were paired, with each one containing ~300 embryos or ~150 larvae coming

380 from the same *in vitro* fertilisation. Nine further samples from adult tissues and body regions

381 (blood vessel, body wall, midgut, prostomium, head, ovary, retractor muscle, tail, and testes)

382 were also collected as described above. Likewise, further five samples spanning post-

383 cleavage time points of *C. teleta,* including 64-cell stage, gastrula, stage 4tt, stage 5, and

384 stage 7 larva, were also collected in duplicates. Total RNA was isolated with the Monarch

385 Total RNA Miniprep Kit (New England Biolabs, NEB) following supplier's

16

386   recommendations. Total RNA samples from developmental stages from both *O. fusiformis*

387   and *C. teleta* were used to prep strand-specific mRNA Illumina libraries that were sequenced

388   at the Oxford Genomics Centre (University of Oxford, UK) over three lanes of an Illumina

389   NovaSeq6000 system in 2 × 150 bases mode. Adult tissue samples were sequenced at BGI on

390   a BGISeq-500 platform in 2 × 100 bases mode. All samples were sequenced to a depth of

391   ~50 M reads (Supplementary Tables 8, 28).

392

393   *Annotation of repeats and transposable elements (TEs)*

394   RepeatModeler v.2.0.1[69] and RepBase were used to construct a *de novo* repeat library for

395   *O. fusiformis*, which was then filtered for *bona fide* genes using the predicted proteome of *C.*

396   *teleta* available at Ensembl Metazoa as reference. The filtered consensus repeat predictions

397   were then used to annotate the genome assembly of *O. fusiformis* with RepeatMasker "open-

398   4.0". We next used LTR_finder v.1.07[70], a structural search algorithm, to identify and

399   annotate Long Tandem Repeats (LTR). Finally, we generated a consensus set of repeats by

400   merging RepeatMasker and LTR_finder predictions with RepeatCraft[71], using default

401   parameters but a maximum LTR size of 25 kb (as derived from the LTR_finder annotation).

402   The general feature format (gff) file with the annotation of TEs and repeats is in *O. fusiformis*

403   genome repository (see Data Availability section).

404

405   *Gene prediction and functional annotation*

406   We used SAMtools v.1.9[72] and the annotation of repeats to soft-mask *O. fusiformis* genome

407   assembly before gene prediction. We then mapped all embryonic and adult transcriptomes

408   and a publicly available dataset[47] (SRR1222288) with STAR v. 2.5.3a[73] after removing low-

409   quality read pairs and read pairs containing Illumina sequencing adapters with trimmomatic

410 v.0.39[74]. StringTie v.1.3.6[75] was used to convert STAR alignments into gene transfer format

411 (GTF) files and Portcullis v.1.1.2[76] to generate a curated set of splice junctions. Additionally,

412 we generated *de novo* transcriptome assemblies for all samples with Trinity v.2.5.1[77] with

413 default parameters, which were thereafter mapped to the soft-masked assembly with GMAP

414 v.2020-04-08[78]. We then ran the default Mikado v.2.1 pipeline[79] to merge all transcriptomic

415 evidence and reliable splice junctions into a single set of best-supported transcripts and gene

416 models. From this merged dataset, we filtered full-length, non-redundant transcripts with a

417 BLAST hit on at least 50 % of their length and at least two exons to obtain a gene set that we

418 used to train Augustus v.3.2.3[80]. Simultaneously, we used the Mikado gene annotation and

419 Portcullis splice junctions to generate confident sets of exon and intron hints, respectively.

420 We also ran Exonerate v.2.4.0[81] to generate spliced alignments of the proteome of *C. teleta*

421 proteome on *O. fusiformis* soft-masked genome assembly to obtain further gene hints. We

422 then merged all exon and intron hints into a single dataset which we passed to Augustus

423 v.3.2.3[80] for *ab initio* gene prediction. Finally, PASA v.2.3.3[82] was used to combine RNA-seq

424 and *ab initio* gene models into a final gene set, from which spurious predictions with in-frame

425 STOP codons (228 gene models), predictions that overlapped with repeats (5,779 gene

426 models) and that had high similarity to transposable elements in the RepeatPeps.lib database

427 (2,450 models) were removed. This filtered gene set includes 26,966 genes, encompassing

428 31,903 different transcripts. To assess the completeness of this annotation, we ran BUSCO

429 v.5[68] in proteome mode, resulting in 97.7 % of the core genes present. Protein homologies for

430 the 34,353 filtered transcripts were annotated with BLAST v.2.2.31+[83] on the

431 UniProt/SwissProt database provided with Trinotate v.3.0[84]. We used HMMER v.2.3.2[85] to

432 identify protein domains using Trinotate's PFAM-A database and signalP v.4.1[86] to predict

433 signal peptides. These functional annotations were integrated into a Trinotate database, which

434 retrieved Gene Onthology (GO), eggNOG and KEGG terms for each transcript. In addition,

435   we ran PANTHER HMM scoring tool to assign a PantherDB[87] orthology ID to each

436   transcript. In total, we retrieved a functional annotation for 22,516 transcripts (63.86 %). The

437   functional annotation report is provided in *O. fusiformis* genome repository (see Data

438   Availability section).

439

440   *Chromosome-scale scaffolding*

441   Sperm from a single *O. fusiformis* worm and an entire sexually mature male were used as

442   input material to construct two Omni-C Dovetail libraries following manufacturer's

443   recommendations for marine invertebrates. These libraries were sequenced in an Illumina

444   NovaSeq6000 at the Okinawa Institute of Science and Technology (Okinawa, Japan) to a

445   depth of 229 and 247 million reads. HiC reads were processed using the Juicer pipeline

446   r.e0d1bb7[88] to generate a list of curated contracts ('merged no dups') that was subsequently

447   employed to scaffold the assembly using 3d-dna v.180419[89]. The resulting assembly and

448   contact map was visually inspected and curated using Juicebox v.1.11.08[88] and adjustments

449   submitted for a subsequent run of optimisation using 3d-dna. Finally, repeats and TEs were

450   re-annotated in this chromosome scale assembly as described above, and the annotation

451   obtained for the PacBio-based assembly was lifted over with Liftoff v.1.6.1[90]. All gene

452   models but two were successfully re-annotated in the chromosome-scale assembly.

453

454   *Gene family evolution analyses*

455   We used the AGAT suite of scripts to generate non-redundant proteomes with only the

456   longest isoform for a set of 21 metazoan proteomes (Supplementary Table 2). To reconstruct

457   gene families, we used OrthoFinder v.2.2.7[91] using MMSeqs2[92] to calculate sequence

458   similarity scores and an inflation value of 2. OrthoFinder gene families were parsed and

459   mapped onto a reference species phylogeny to infer gene family gains and losses at different

460   nodes and tips using the ETE 3 library[93], as well as to estimate the node of origin for each

461   gene family. Gene expansions were computed for each species using a hypergeometric test

462   against the median gene number per species for a given family employing previously

463   published code[31]. Principal component analysis was performed on the orthogroups matrix by

464   metazoan lineage, given that orthogroups were present in at least three of the 22 analysed

465   species, to eliminate taxonomically restricted genes.

466

467   *Macrosynteny analyses*

468   Single copy orthologues obtained using the mutual best hit (MBH) approach generated using

469   MMseqs2[92] using the annotations of *Branchiostoma floridae*[94], *Pecten maximus*[95],

470   *Streblospio benedictii*[96], and *Lineus longissimus*[97,98] were used to generate Oxford synteny

471   plots comparing sequentially indexed orthologue positions. Plotting order was determined by

472   hierarchical clustering of the shared orthologue content using the complete linkage method as

473   originally proposed. Comparison of the karyotype of all four species was performed using the

474   Rideogram package by colouring pairwise orthologues according to the ALG assignment in

475   comparisons with *P. maximus* and *B. floridae*.

476

477   *Gene expression profiling*

478   To profile gene expression dynamics from blastula to juvenile stages (*O. fusiformis*) and from

479   64-cell to competent larva stages (*C. teleta*), sequencing adaptors were removed from raw

480   reads using trimmomatic v.0.39[74]. Cleaned reads were pseudo-aligned to the filtered gene

481   models using kallisto v.0.46.2[99] (Supplementary Table 9) and genes with an expression level

482   above an empirically defined threshold of 2 transcripts per million (TPM) were deemed

483　expressed (Extended Data Fig. 3f). DESeq2 v.1.30.1 package[100] was used to normalise read

484　counts across developmental stages (Supplementary Table 10) and to perform pair-wise

485　differential gene expression analyses between consecutive stages. *P*-values were adjusted

486　using the Benjamini-Hochberg method for multiple testing correction. We defined a gene as

487　significantly upregulated for a $\log_2$(fold-change) (LFC) > 1 or downregulated for a LFC < 1,

488　given that adjusted *p*-value < 0.05 (Supplementary Information). Principal component

489　analysis was performed on a variance stabilising-transformed matrix of the normalised

490　DESeq2 matrix. *Hox* gene expression profiling in *Urechis unicinctus*, was performed as

491　described for *O. fusiformis* gene profiling, using an available reference transcriptome[101].

492

493　*Gene clustering and co-expression network analyses*

494　Genes were clustered according to their normalised expression dynamics through soft *k*-

495　means clustering using the mfuzz v.2.52 package[102]. From all 31,903 genes, we discarded 225

496　genes which were not expressed at any stage. We then determined an optimal number of 12

497　clusters for our dataset by applying the elbow method to the minimum centroid distance as a

498　function of the number of clusters. For the construction of the gene co-expression network,

499　we used the WGCNA package v.1.70–3[103]. All 31,678 genes expressed at any developmental

500　stage were used to build a signed network with a minimum module size of 300 genes and an

501　optimised soft-thresholding power of 16. Block-wise network construction returned 15 gene

502　modules, from which one module was dropped due to poor intramodular connectivity. The

503　remaining 14 gene modules (A–N) were labelled with distinct colours with unassigned genes

504　labelled in grey. A random subset consisting of the nodes and edges of 30 % of the genes was

505　fed to Cytoscape v.3.8.2[104] for network visualisation. Module eigengenes were chosen to

506　summarise the gene expression profiles of gene modules. Gene ontology (GO) enrichment

507   analyses of each gene cluster and gene module were performed using the topGO v.2.44

508   package. We performed a Fisher's exact test and listed the top 30 (soft *k*-means clustering) or

509   top 15 (WGCNA modules) significantly enriched GO terms of the class Biological Process.

510   To ease visualisation, all 288 non-redundant enriched GO terms from all 12 gene clusters

511   were clustered through *k*-means by semantic similarity using the simplifyEnrichment v.1.2.0

512   package[105] (Supplementary Fig. 3).

513

514   *Gene orthology assignment*

515   The identification of *chordin* (*chrd*), *chordin-like* (*chrdl*) and *Hox* genes in *O. fusiformis* was

516   based on the genome functional annotation. To mine *chrd* orthologues, 81 annelid

517   transcriptomic datasets were downloaded from SRA (Supplementary Table 14) and

518   assembled with Trinity v.2.5.1[77] to create BLAST local nucleotide databases. We also created

519   a nucleotide database for *C. teleta* using its annotated genome[29] (ENA accession number

520   GCA_000328365.1). Human and *O. fusiformis* CHRD proteins were used as queries to find

521   *chrd* orthologues following the MBH approach (*e*-value $\leq 10^{-3}$), obtaining 104 unique

522   candidate *chrd* transcripts that were then translated (Supplementary Table 15). A single

523   candidate CHRD protein for *Themiste lageniformis* (unpublished data, provided by Michael J

524   Boyle) was included *ad hoc* at this step. In addition, 15 curated CHRD and CHRDL protein

525   sequences (and an outgroup) were fetched from various sources (Supplementary Table 16)

526   and aligned together with *O. fusiformis* CHRD and CHRDL sequences in MAFFT v.7[106] with

527   the G-INS-I iterative refinement method and default scoring parameters. From this mother

528   alignment further daughter alignments were obtained using "mafft --addfragments"[107], the

529   accurate "--multipair" method, and default scoring parameters. For orthology assignment,

530   two phylogenetic analyses were performed on selected candidate sequences, which included

531   the longest isoform for each species-gene combination, given that it included a 10-residue or

532   longer properly aligned fragment in either the CHRD domains or the von Willebrand factor

533   type C (VWFC) domains (Extended Data Figure 5c, d). vWFC and CHRD domains were

534   trimmed and concatenated using domain boundaries defined by ProSITE domain annotation

535   for the human chordin precursor protein (UniProt: Q9H2X0). Either all domains or the

536   VWFC domains only were used for phylogenetic inference with a WAG amino acid

537   replacement matrix[108] to account for transition rates, the FreeRate heterogeneity model

538   (R4)[109] to describe sites evolution rates, and an optimization of amino acid frequencies using

539   maximum likelihood (ML) using IQ-TREE v.2.0.3[110]. 1,000 ultrafast bootstraps (BS)[111] were

540   used to extract branch support values. Bayesian reconstruction in MrBayes v.3.2.7a[112] were

541   also performed using the same WAG matrix but substituting the R4 model for the discrete

542   gamma model[113], with 4 rate categories (G4). For the orthology assignment of *Hox* genes,

543   129 curated Hox sequences were retrieved from various databases (Supplementary Table 19)

544   and aligned with *O. fusiformis* Hox proteins with MAFFT v.7 in automatic mode. Poorly

545   aligned regions were removed with gBlocks v.0.91b[114] yielding the final alignments.

546   Maximum likelihood trees were constructed using RAxML v.8.2.11.9[115] with an LG

547   substitution matrix[116] and 1,000 ultrafast BS. All trees were composed in FigTree v.1.4.4.

548   Alignment files are available in *O. fusiformis* genome repository (see Data Availability

549   section).

550

551   *Whole mount* in situ *hybridisation and immunohistochemistry*

552   Fragments of *chordin* and *Hox* genes were isolated as previously described[26] using gene-

553   specific oligonucleotides and a T7 adaptor. Riboprobes were synthesise with the T7

554   MEGAscript kit (ThermoFisher, AM1334) and stored at a concentration of 50 ng/µl in

555 hybridisation buffer at –20 ºC. Whole mount *in situ* hybridisation in embryonic, larval, and

556 juvenile stages were conducted as described elsewhere[26,28]. Antibody staining in larval stages

557 of *O. fusiformis*, *Magelona* spp. and *C. teleta* was carried out as previously described[25,117].

558 DIC images of the colorimetric *in situs* were obtained with a Leica 560 DMRA2 upright

559 microscope equipped with an Infinity5 camera (Lumenera). Fluorescently stained samples

560 were scanned with a Nikon CSU-W1 Spinning Disk Confocal.

561

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

563 We performed two replicates of ATAC-seq from samples containing ~50,000 cells at the

564 blastula (~900 embryos), gastrula (~500), elongation (~300), mitraria larva (~150) and

565 competent larva (~40) stages following the omniATAC protocol[118], but gently homogenising

566 the samples with a pestle in lysis buffer and incubating them on ice for 3 min. Tagmentation

567 was performed for 30 min at 37°C with an in-house purified Tn5 enzyme[119]. After DNA

568 clean-up, ATAC-seq libraries were amplified as previously described. Primers used for both

569 PCR and qPCR are listed in Supplementary Table 21. Amplified libraries were purified using

570 ClentMag PCR Clean Up Beads as indicated by the supplier and quantified and quality

571 checked on a Qubit 4 Fluorometer (Thermo-Fisher) and an Agilent 2200 TapeStation system

572 before pooling at equal molecular weight. Sequencing was performed on an Illumina

573 HiSeq4000 platform in 2 × 75 bases mode at the Oxford Genomics Centre (University of

574 Oxford, United Kingdom) (blastula, elongation and mitraria larva stages, and one replicate of

575 gastrula) and on an Illumina NovoSeq6000 in 2 × 150 bases mode at Novogene (Cambridge,

576 United Kingdom) (one replicate of gastrula and two replicates of competent larva stages).

577

578 *Chromatin accessibility profiling and motif identification and enrichment analyses*

579   We used cutadapt v.2.5[120] to remove sequencing adaptors and trim reads from libraries

580   sequenced in 2×150 bases mode to 75 bases reads. Quality filtered reads were mapped using

581   NextGenMap v.0.5.5[121] in paired-end mode, duplicates were removed using samtools v.1.9[122]

582   and mapped reads were shifted using deepTools v.3.4.3[123]. Fragment size distribution was

583   estimated from resulting BAM files and transcription start site (TSS) enrichment analysis was

584   computed using computeMatrix and plotHeatmap commands in deepTools v.3.4.3. Peak

585   calling was done with MACS2 v.2.2.7.1[124,125] (-f BAMPE --min-length 100 --max-gap 75 and

586   -q 0.01). Peaks from repetitive regions were filtered with BEDtools v.2.28.0[126] and

587   irreproducible discovery rates (IDR) v.2.0.4.2 was used to identify reproducible peaks

588   (IDR < 0.05) at each developmental stage. Next, we used DiffBind v.3.0.14[127] to generate a

589   final consensus peak set of 47,406 peaks, which were normalised using DESeq2 analysis.

590   Peak clustering according to accessibility dynamics was performed as described above for

591   RNA-seq, using the same number of 12 clusters to make both profiling techniques

592   comparable. Principal component analysis and differential accessibility analyses between

593   consecutive developmental stages were also performed as described above. An LFC > 0 and a

594   LFC < 0 indicates whether a peak opens or closes, respectively, given that the adjusted *p*-

595   value < 0.05. Stage-specific and constitutive peaks were determined using UpSetR v.1.4.0[128]

596   and both the consensus peak set and the stage-specific peak sets were classified by genomic

597   feature/region using HOMER v.4.11[129] and further manually curated. Visualisation of peak

598   tracks and gene structures was conducted with pyGenomeTracks v.2.1[130] and deepTools

599   v.3.4.3[123]. To correlate chromatin accessibility and gene expression, this genomic region

600   annotation was used to assign peaks to their closest gene and the Pearson correlation

601   coefficient between chromatin accessibility and gene expression was computed. Promoter

602   and intergenic peaks were subset, for which we then used GimmeMotifs v.0.16.1[131] to

603   perform *de novo* motif search and differential motif enrichment in our peak sets. Data

604 clustering was performed with mfuzz v.2.52[102] and visualized with ComplexHeatmap package

605 v.2.6.2[132].

606

607 *Phylostratigraphy*

608 To evaluate gene expression dynamics by phylostratum and developmental stage, we used the

609 OrthoFinder gene families and their inferred origins. We deemed all genes originating before

610 and with the Cnidarian-Bilaterian ancestor of metazoan origin (Supplementary Table 27). We

611 then applied a quantile normalisation onto the DESeq2 normalised matrix of gene expression

612 values spanning blastula to juvenile stages. The 75 % percentile of the quantile-normalised

613 gene expression levels was used as the summarising measure of the gene expression

614 distribution by developmental stage. Furthermore, we tested the over- or underrepresentation

615 of the different phylostrata in the gene expression clusters through pair-wise Fisher's exact

616 tests, for which we then adjusted the *p*-values using the Bonferroni correction for multiple

617 testing.

618

619 *Comparative transcriptomics*

620 Publicly available RNA-seq developmental time courses for the development of *Nematostella*

621 *vectensis*, *Strongylocentrotus purpuratus*, *Branchiostoma lanceolatum*, *Danio rerio*,

622 *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Crassostrea gigas*, and two stages of

623 *Capitella teleta* were downloaded from the SRA (Supplementary Table 29), cleaned for

624 adaptors and low-quality reads with trimmomatic v.0.39[74] and pseudo-aligned to their

625 respective non-redundant genome-based gene repertoires using kallisto v.0.46.2[99]. We then

626 performed a quantile transformation of TPM values using scikit-learn v.1.0.2[133] and

627　calculated the Jensen-Shannon divergence (JSD) from single copy orthologs between all

628　possible 1-to-1 species comparisons using the philentropy v.0.5.0 package[134]:

629
$$JSD_{\text{raw}}(P \parallel Q) = \frac{1}{2}\sum_{i=0}^{n} p_i \times \log_2\left(\frac{p_i}{\frac{1}{2}(p_i + q_i)}\right) + \frac{1}{2}\sum_{i=0}^{n} q_i \times \log_2\left(\frac{q_i}{\frac{1}{2}(p_i + q_i)}\right)$$

630　Transcriptomic divergences were calculated based on 250 bootstrap replicates, from which

631　statistically robust mean values and standard deviations were obtained. Raw mean JSD values

632　(JSD_{raw}) were adjusted (JSD_{adj}) by dividing by the number of single copy orthologs of each

633　comparison (Supplementary Table 30), and normalised using the minimum and maximum

634　adjusted JSD values from all 1-to-1 species comparisons as follows:

635
$$JSD_{\text{norm}}(P \parallel Q) = \frac{JSD_{\text{adj}}(P \parallel Q) - \min JSD_{\text{adj}}}{\max JSD_{\text{adj}} - \min JSD_{\text{adj}}}; \; JSD_{\text{norm}} \in (0, 1)$$

636　Relative JSD values were obtained equally, using minimum and maximum adjusted JSD

637　values from each 1-to-1 species comparison instead. For a further analysis with a subset of

638　common orthologues, raw mean JSD values needed no adjustment for comparison purposes.

639

640　**References**

641　1　Hall, B. K. & Wake, M. H. in *The Origin and Evolution of Larval Forms* (eds Brian

642　　　K. Hall & Marvalee H. Wake) Ch. 1, 1-19 (Academic Press, 1999).

643　2　Nielsen, C. Origin and evolution of animal life cycles. *Biol Rev* **73**, 125-155 (1998).

644　3　Nielsen, C. Animal phylogeny in the light of the trochaea theory. *Biological Journal*

645　　　*of the Linnean Society* **25**, 243-299 (2008).

646　4　Garstang, W. The origin and evolution of larval forms. 77-98 (British Association for

647　　　the Advancement of Science, London, 1928).

648   5   Jägersten, G. *Evolution of the Metazoan Life Cycle*. (Academic Press, 1972).

649   6   Marlow, H. *et al.* Larval body patterning and apical organs are conserved in animal

650   evolution. *BMC Biol* **12**, 7 (2014).

651   7   Wang, J. *et al.* Evolutionary transcriptomics of metazoan biphasic life cycle supports

652   a single intercalation origin of metazoan larvae. *Nat Ecol Evol* **4**, 725-736 (2020).

653   8   Salvini-Plawen, L. v. On the origin and evolution of the lower Metazoa. *Z Zool Syst*

654   *Evol Forsch* **16**, 40-87 (1978).

655   9   Hazsprunar, G., Salvini-Plawen, L. v. & Rieger, R. M. Larval Planktotrophy — A

656   Primitive Trait in the Bilateria? *Acta Zoologica* **76**, 141-154 (1995).

657   10   Raff, R. A. Origins of the other metazoan body plans: the evolution of larval forms.

658   *Philos Trans R Soc Lond B Biol Sci* **363**, 1473-1479 (2008).

659   11   Sly, B. J., Snoke, M. S. & Raff, R. A. Who came first - larvae or adults? Origins of

660   bilaterian metazoan larvae. *Int J Dev Biol* **47**, 623-632 (2003).

661   12   Davidson, E. H., Peterson, K. J. & Cameron, R. A. Origin of bilaterian body plans:

662   evolution of developmental regulatory mechanisms. *Science* **270**, 1319-1325 (1995).

663   13   Rouse, G. W., Pleijel, F. & Tilic, E. *Annelida*. (Oxford University Press, 2022).

664   14   Lacalli, T. C. Protochordate body plan and the evolutionary role of larvae: old

665   controversies resolved? *Can. J. Zool.* **83**, 216-224 (2005).

666   15   Strathman, R. Multiple origins of feeding head larvae by the Early Cambrian. *Can. J.*

667   *Zool.* **98**, 761-776 (2020).

668   16   Hatschek, B. Studien über Entwicklungsgeschichte der Anneliden. Ein Beitrag zur

669   Morphologie der Bilaterien. *Arbeiten aus dem Zoologischen Institute der Universität*

670   *Wien und der Zoologischen Station in Triest* **1**, 277-404 (1878).

671   17   Rouse, G. W. Trochophore concepts: ciliary bands and the evolution of larvae in

672   spiralian Metazoa. *Biological Journal of the Linnean Society* **66**, 411-464 (2008).

673   18   Marlétaz, F., Peijnenburg, K., Goto, T., Satoh, N. & Rokhsar, D. S. A New Spiralian

674   Phylogeny Places the Enigmatic Arrow Worms among Gnathiferans. *Curr Biol* **29**,

675   312-318 e313 (2019).

676   19   Nielsen, C. *Animal Evolution. Interrelationships of the Living Phyla*. 3rd edn,

677   (Oxford University Press, 2012).

678   20   Nielsen, C. Evolution of deuterostomy - and origin of the chordates. *Biol Rev Camb*

679   *Philos Soc* **92**, 316-325 (2017).

680   21   Wilson, D. P. IV. On the Mitraria Larva of *Owenia fusiformis* Delle Chiaje.

681   *Philosophical Transactions of the Royal Society of London. Series B, Containing*

682   *Papers of a Biological Character* **221**, 231-334 (1932).

683   22   Nielsen, C. Trochophora larvae: cell-lineages, ciliary bands and body regions. 2.

684   Other groups and general discussion. *J Exp Zool B Mol Dev Evol* **304**, 401-447

685   (2005).

686   23   Smart, T. I. & Von Dassow, G. Unusual development of the mitraria larva in the

687   polychaete *Owenia collaris*. *Biol Bull* **217**, 253-268 (2009).

688   24   Gasiorowski, L. *et al.* Molecular evidence for a single origin of ultrafiltration-based

689   excretory organs. *Curr Biol* **31**, 3629-3638 e3622 (2021).

690   25   Carrillo-Baltodano, A. M., Seudre, O., Guynes, K. & Martin-Duran, J. M. Early

691   embryogenesis and organogenesis in the annelid *Owenia fusiformis*. *Evodevo* **12**, 5

692   (2021).

693   26   Seudre, O., Carrillo-Baltodano, A. M., Liang, Y. & Martín-Durán, J. M. ERK1/2 is an

694   ancestral organising signal in spiral cleavage. *bioRxiv*, 2021.2005.2026.445819

695   (2021).

696   27   Martin-Duran, J. M. *et al.* Convergent evolution of bilaterian nerve cords. *Nature* **553**,

697   45-50 (2018).

698   28   Martin-Duran, J. M., Passamaneck, Y. J., Martindale, M. Q. & Hejnol, A. The

699   developmental basis for the recurrent evolution of deuterostomy and protostomy. *Nat*

700   *Ecol Evol* **1**, 5 (2016).

701   29   Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes.

702   *Nature* **493**, 526-531 (2013).

703   30   Wang, S. *et al.* Scallop genome provides insights into evolution of bilaterian

704   karyotype and development. *Nat Ecol Evol* **1**, 120 (2017).

705   31   Martin-Duran, J. M. *et al.* Conservative route to genome compaction in a miniature

706   annelid. *Nat Ecol Evol* **5**, 231-242 (2021).

707   32   Shao, Y. *et al.* Genome and single-cell RNA-sequencing of the earthworm *Eisenia*

708   *andrei* identifies cellular mechanisms underlying regeneration. *Nat Commun* **11**, 2656

709   (2020).

710   33   Niehrs, C. On growth and form: a Cartesian coordinate system of Wnt and BMP

711   signaling specifies bilaterian body axes. *Development* **137**, 845-857 (2010).

712   34   Tan, S., Huan, P. & Liu, B. Molluscan dorsal-ventral patterning relying on BMP2/4

713   and Chordin provides insights into spiralian development and evolution. *Mol Biol*

714   *Evol* (2021).

715   35   Lyons, D. C., Perry, K. J., Batzel, G. & Henry, J. Q. BMP signaling plays a role in

716   anterior-neural/head development, but not organizer activity, in the gastropod

717   *Crepidula fornicata*. *Dev Biol* **463**, 135-157 (2020).

718   36   Pearson, J. C., Lemons, D. & McGinnis, W. Modulating Hox gene functions during

719   animal body patterning. *Nat Rev Genet* **6**, 893-904, doi:10.1038/nrg1726 (2005).

720   37   Fröbius, A. C., Matus, D. Q. & Seaver, E. C. Genomic organization and expression

721   demonstrate spatial and temporal Hox gene colinearity in the lophotrochozoan

722   *Capitella* sp. I. *PLoS One* **3**, e4004 (2008).

723   38   Hui, J. H. *et al.* Extensive chordate and annelid macrosynteny reveals ancestral

724   homeobox gene organization. *Mol Biol Evol* **29**, 157-165 (2012).

725   39   Kulakova, M. *et al.* Hox gene expression in larval development of the polychaetes

726   *Nereis virens* and *Platynereis dumerilii* (Annelida, Lophotrochozoa). *Dev Genes Evol*

727   **217**, 39-54 (2007).

728   40   Woltereck, R. Trochophora-Studien I. Histogie der larve und die Entstehung des

729   Annelids bei den *Polygordius*-Arten der Nordsee. *Zoologica* **13** (1902).

730   41   Simionato, E. *et al. atonal*- and *achaete-scute*-related genes in the annelid *Platynereis*

731   *dumerilii*: insights into the evolution of neural basic-Helix-Loop-Helix genes. *BMC*

732   *Evol Biol* **8**, 170 (2008).

733   42   Wu, L., Ferger, K. E. & Lambert, J. D. Gene Expression Does Not Support the

734   Developmental Hourglass Model in Three Animals with Spiralian Development. *Mol*

735   *Biol Evol* **36**, 1373-1383 (2019).

736   43   Wu, L. *et al.* Genes with spiralian-specific protein motifs are expressed in spiralian

737   ciliary bands. *Nat Commun* **11**, 4171 (2020).

738   44   Marinković, M., Berger, J. & Jékely, G. Neuronal coordination of motile cilia in

739   locomotion and feeding. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190165 (2020).

740   45   Irvine, S. Q. & Martindale, M. Q. Expression patterns of anterior Hox genes in the

741   polychaete *Chaetopterus*: correlation with morphological boundaries. *Dev Biol* **217**,

742   333-351 (2000).

743   46   Peterson, K. J., Irvine, S. Q., Cameron, R. A. & Davidson, E. H. Quantitative

744   assessment of *Hox* complex expression in the indirect development of the polychaete

745   annelid *Chaetopterus* sp. *Proc Natl Acad Sci U S A* **97**, 4487-4492 (2000).

746   47   Weigert, A. *et al.* Illuminating the base of the annelid tree using transcriptomics. *Mol*

747   *Biol Evol* **31**, 1391-1401 (2014).

748   48   Hiebert, L. S. & Maslakova, S. A. *Hox* genes pattern the anterior-posterior axis of the

749   juvenile but not the larva in a maximally indirect developing invertebrate, *Micrura*

750   *alaskensis* (Nemertea). *BMC Biol* **13**, 23 (2015).

751   49   Gasiorowski, L. & Hejnol, A. Hox gene expression during development of the

752   phoronid *Phoronopsis harmeri*. *Evodevo* **11**, 2 (2020).

753   50   Brenneis, G., Bogomolova, E. V., Arango, C. P. & Krapp, F. From egg to "no-body":

754   an overview and revision of developmental pathways in the ancient arthropod lineage

755   Pycnogonida. *Front Zool* **14**, 6 (2017).

756   51   Arenas-Mena, C., Cameron, A. R. & Davidson, E. H. Spatial expression of *Hox*

757   cluster genes in the ontogeny of a sea urchin. *Development* **127**, 4631-4643 (2000).

758   52   Arenas-Mena, C., Martinez, P., Cameron, R. A. & Davidson, E. H. Expression of the

759   *Hox* gene complex in the indirect development of a sea urchin. *Proc Natl Acad Sci U*

760   *S A* **95**, 13062-13067 (1998).

761   53   Gonzalez, P., Uhlinger, K. R. & Lowe, C. J. The Adult Body Plan of Indirect

762   Developing Hemichordates Develops by Adding a Hox-Patterned Trunk to an

763   Anterior Larval Territory. *Curr Biol* **27**, 87-95 (2017).

764   54   Aronowicz, J. & Lowe, C. J. *Hox* gene expression in the hemichordate *Saccoglossus*

765   *kowalevskii* and the evolution of deuterostome nervous systems. *Integr Comp Biol* **46**,

766   890-901 (2006).

767   55   Pascual-Anaya, J. *et al.* Broken colinearity of the amphioxus *Hox* cluster. *Evodevo* **3**,

768   28 (2012).

769   56   Carrillo-Baltodano, A. M. & Meyer, N. P. Decoupling brain from nerve cord

770   development in the annelid *Capitella teleta*: Insights into the evolution of nervous

771   systems. *Dev Biol* **431**, 134-144 (2017).

772   57   Luo, Y. J. *et al.* Nemertean and phoronid genomes reveal lophotrochozoan evolution

773   and the origin of bilaterian heads. *Nat Ecol Evol* **2**, 141-151 (2018).

774   58   Steinmetz, P. R. *et al. Six3* demarcates the anterior-most developing brain region in

775   bilaterian animals. *Evodevo* **1**, 14 (2010).

776   59   Hiebert, L. S. & Maslakova, S. A. Expression of *Hox*, *Cdx*, and *Six3/6* genes in the

777   hoplonemertean *Pantinonemertes californiensis* offers insight into the evolution of

778   maximally indirect development in the phylum Nemertea. *Evodevo* **6**, 26 (2015).

779   60   Seaver, E. C., Thamm, K. & Hill, S. D. Growth patterns during segmentation in the

780   two polychaete annelids, *Capitella* sp. I and *Hydroides elegans*: comparisons at

781   distinct life history stages. *Evol Dev* **7**, 312-326 (2005).

782   61   Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting

783   of occurrences of *k*-mers. *Bioinformatics* **27**, 764-770 (2011).

784   62   Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and

785   Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1432

786   (2020).

787   63   Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer

788   weighting and repeat separation. *Genome Res* **27**, 722-736,

789   doi:10.1101/gr.215087.116 (2017).

790   64   Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-

791   MEM. *arXiv* **1303.3997** (2013).

792   65   Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome

793   assembly from long uncorrected reads. *Genome Res* **27**, 737-746 (2017).

794   66   Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig

795   reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460

796   (2018).

797   67   Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free

798   quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*

799   **21**, 245 (2020).

800   68   Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO

801   Update: Novel and Streamlined Workflows along with Broader and Deeper

802   Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.

803   *Mol Biol Evol* **38**, 4647-4654 (2021).

804   69   Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable

805   element families. *Proc Natl Acad Sci U S A* **117**, 9451-9457 (2020).

806   70   Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length

807   LTR retrotransposons. *Nucleic Acids Res* **35**, W265-268 (2007).

808   71   Wong, W. Y. & Simakov, O. RepeatCraft: a meta-pipeline for repetitive element de-

809   fragmentation and annotation. *Bioinformatics* **35**, 1051-1052 (2019).

810   72   Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).

811   73   Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21

812   (2013).

813   74   Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina

814   sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

815   75   Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from

816   RNA-seq reads. *Nat Biotechnol* **33**, 290-295 (2015).

817   76   Mapleson, D., Venturini, L., Kaithakottil, G. & Swarbreck, D. Efficient and accurate

818   detection of splice junctions from RNA-seq with Portcullis. *Gigascience* **7** (2018).

819   77   Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data

820   without a reference genome. *Nat Biotechnol* **29**, 644-652 (2011).

821   78   Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for

822   mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875 (2005).

823   79   Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L. & Swarbreck, D.

824   Leveraging multiple transcriptome assembly methods for improved gene structure

825   annotation. *Gigascience* **7** (2018).

826   80   Stanke, M., Schoffmann, O., Morgenstern, B. & Waack, S. Gene prediction in

827   eukaryotes with a generalized hidden Markov model that uses hints from external

828   sources. *BMC Bioinformatics* **7**, 62 (2006).

829   81   Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence

830   comparison. *BMC Bioinformatics* **6**, 31 (2005).

831   82   Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal

832   transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654-5666 (2003).

833   83   Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**,

834   421 (2009).

835   84   Bryant, D. M. *et al.* A Tissue-Mapped Axolotl De Novo Transcriptome Enables

836   Identification of Limb Regeneration Factors. *Cell Rep* **18**, 762-776 (2017).

837   85   Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195

838   (2011).

839   86   Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating

840   signal peptides from transmembrane regions. *Nat Methods* **8**, 785-786 (2011).

841   87   Mi, H. *et al.* PANTHER version 16: a revised family classification, tree-based

842   classification tool, enhancer regions and extensive API. *Nucleic Acids Res* **49**, D394-

843   D403 (2021).

844   88   Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-

845   Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).

846 89 Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C
847   yields chromosome-length scaffolds. *Science* **356**, 92-95 (2017).

848 90 Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations.
849   *Bioinformatics* **37**, 1639-1643 (2021).

850 91 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for
851   comparative genomics. *Genome Biol* **20**, 238 (2019).

852 92 Steinegger, M. & Soding, J. MMseqs2 enables sensitive protein sequence searching
853   for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).

854 93 Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and
855   Visualization of Phylogenomic Data. *Mol Biol Evol* **33**, 1635-1638 (2016).

856 94 Marletaz, F. *et al.* Amphioxus functional genomics and the origins of vertebrate gene
857   regulation. *Nature* **564**, 64-70 (2018).

858 95 Zeng, Q. *et al.* High-quality reannotation of the king scallop genome reveals no 'gene-
859   rich' feature and evolution of toxin resistance. *Comput Struct Biotechnol J* **19**, 4954-
860   4960 (2021).

861 96 Zakas, C., Harry, N. D., Scholl, E. H. & Rockman, M. V. The genome of the
862   poecilogonous annelid *Streblospio benedicti*. *Genome Biology and Evolution* (2022).

863 97 Kwiatkowski, D. *et al.* The genome sequence of the bootlace worm, *Lineus
864   longissimus* (Gunnerus, 1770). *Wellcome Open Res* **6**, 272 (2021).

865 98 Cannon, J. T. *et al.* Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530**,
866   89-93 (2016).

867 99 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-
868   seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).

869 100 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
870   dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

871   101   Park, C. *et al.* The developmental transcriptome atlas of the spoon worm *Urechis*

872   *unicinctus* (Echiurida: Annelida). *Gigascience* **7**, 1-7 (2018).

873   102   Futschik, M. E. & Carlisle, B. Noise-robust soft clustering of gene expression time-

874   course data. *J Bioinform Comput Biol* **3**, 965-988 (2005).

875   103   Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation

876   network analysis. *BMC Bioinformatics* **9**, 559 (2008).

877   104   Shannon, P. *et al.* Cytoscape: a software environment for integrated models of

878   biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).

879   105   Gu, Z. & Hübschmann, D. simplifyEnrichment: an R/Bioconductor package for

880   Clustering and Visualizing Functional Enrichment Results. *bioRxiv*,

881   2020.2010.2027.312116 (2021).

882   106   Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version

883   7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780,

884   doi:10.1093/molbev/mst010 (2013).

885   107   Katoh, K. & Frith, M. C. Adding unaligned sequences into an existing alignment

886   using MAFFT and LAST. *Bioinformatics* **28**, 3144-3146 (2012).

887   108   Whelan, S. & Goldman, N. A general empirical model of protein evolution derived

888   from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*

889   **18**, 691-699 (2001).

890   109   Soubrier, J. *et al.* The influence of rate heterogeneity among sites on the time

891   dependence of molecular rates. *Mol Biol Evol* **29**, 3345-3358 (2012).

892   110   Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and

893   effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol*

894   *Biol Evol* **32**, 268-274 (2015).

895   111   Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2:

896   Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**, 518-522 (2018).

897   112   Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under

898   mixed models. *Bioinformatics* **19**, 1572-1574 (2003).

899   113   Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with

900   variable rates over sites: approximate methods. *J Mol Evol* **39**, 306-314 (1994).

901   114   Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent

902   and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564-

903   577 (2007).

904   115   Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis

905   of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).

906   116   Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol

907   Biol Evol* **25**, 1307-1320 (2008).

908   117   Meyer, N. P., Carrillo-Baltodano, A., Moore, R. E. & Seaver, E. C. Nervous system

909   development in lecithotrophic larval and juvenile stages of the annelid *Capitella

910   teleta*. *Front Zool* **12**, 15 (2015).

911   118   Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and

912   enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962 (2017).

913   119   Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled

914   sequencing projects. *Genome Res* **24**, 2033-2040 (2014).

915   120   Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing

916   reads. *2011* **17**, 3 (2011).

917   121   Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate

918   read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791 (2013).

919   122   Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,

920   2078-2079 (2009).

921   123   Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data

922   analysis. *Nucleic Acids Res* **44**, W160-165 (2016).

923   124   Gaspar, J. M. Improved peak-calling with MACS2. *bioRxiv*, 496521 (2018).

924   125   Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137

925   (2008).

926   126   Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing

927   genomic features. *Bioinformatics* **26**, 841-842 (2010).

928   127   Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with

929   clinical outcome in breast cancer. *Nature* **481**, 389-393 (2012).

930   128   Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization

931   of intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940 (2017).

932   129   Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors

933   prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell*

934   **38**, 576-589 (2010).

935   130   Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic

936   datasets *Bioinformatics* **37**, 422-423 (2020).

937   131   van Heeringen, S. J. & Veenstra, G. J. GimmeMotifs: a *de novo* motif prediction

938   pipeline for ChIP-sequencing experiments. *Bioinformatics* **27**, 270-271 (2011).

939   132   Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in

940   multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).

941   133   Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine*

942   *Learning Research* **12**, 2825−2830 (2011).

943   134   Drost, H. G. Philentropy: Information Theory and Distance Quantification with R.

944   *Journal of Open Source Software* **3**, 765 (2018).

945

961

962   **Author Contributions**

963   JMM-D, FM, YL and FMM-Z conceived and designed the study; YL collected RNA-seq

964   samples for *O. fusiformis* and *C. teleta*, performed ATAC-seq experiments and contributed to

965   all data analyses; FMM-Z performed *chordin* orthology studies and contributed to all data

966   analyses; KG conducted *in situ* hybridisation analyses of *Hox* genes; AC-B collected RNA-

967   seq samples for *C. teleta*, performed immunostainings on larvae and gene expression analyses

968   of *chordin*; YT performed OMNI-C libraries; GM performed repeat annotations and

969   analyses; OS identified and performed *in silico* analyses of *Hox* genes; MT performed

970   genomic extractions and optical mapping; KM collected *Magelona* spp.; AH and NML

971   contributed to sequencing efforts; FM and JMM-D assembled and annotated the genome and

972   contributed to data analyses; YL, FMM-Z and JMM-D drafted the manuscript and all authors

973   critically read and commented on the manuscript.

974

975   **Competing Interests**

976   The authors declare no competing interests.

977

978   **Data availability**

979   All sequence data associated with this project are available at the European Nucleotide

980   Archive (project PRJEB38497) and Gene Expression Omnibus (accession numbers

981   GSE184126 and GSE192478). Genome assemblies, transposable element annotations,

982   genome annotation files used for RNA-seq and ATAC-seq analyses, WGCNA nodes and

983   edges files, alignment files used in orthology assignment, and other additional files are

984   publicly available in https://github.com/ChemaMD/OweniaGenome.

985

986   **Code availability**

987   No custom code was used in this study.

988

Figures

990
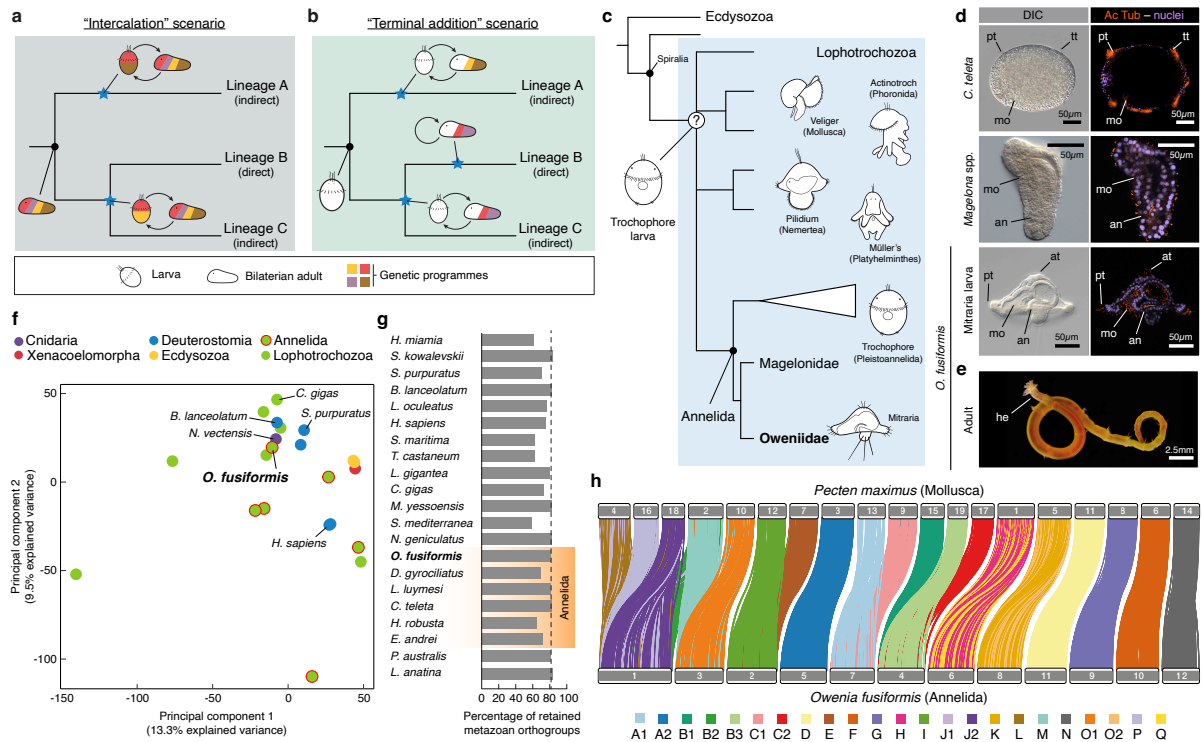
Figure 1 | *Owenia fusiformis* **has a unique larva and a conservatively evolving genome.**

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

1004 *Magelona* spp. larva, and a *O. fusiformis* mitraria stained for DAPI and acetylated α-tubulin.

1005 **e**, Image of an adult of *O. fusiformis*. **f,** Principal component analysis of metazoan gene

1006 complements demonstrates that *Owenia* clusters with other slow-evolving lineages. See

1007 Extended Data Fig. 2g for a fully labelled graph. **g,** Percentage of retained pre-metazoan and

1008 metazoan orthogroups per species. Dotted vertical line represents the value for *O. fusiformis*.

1009 **h,** Karyotypic correspondence between *O. fusiformis* and *Pecten maximus*, which exemplifies

1010 the ancestral spiralian chromosome complement. Each colour represents an ancestral

1011 bilaterian linkage group. Schematic drawings are not to scale. at: apical tuft; an: anus; he:

1012 head; mo: mouth; pt: prototroch; tt: telotroch.
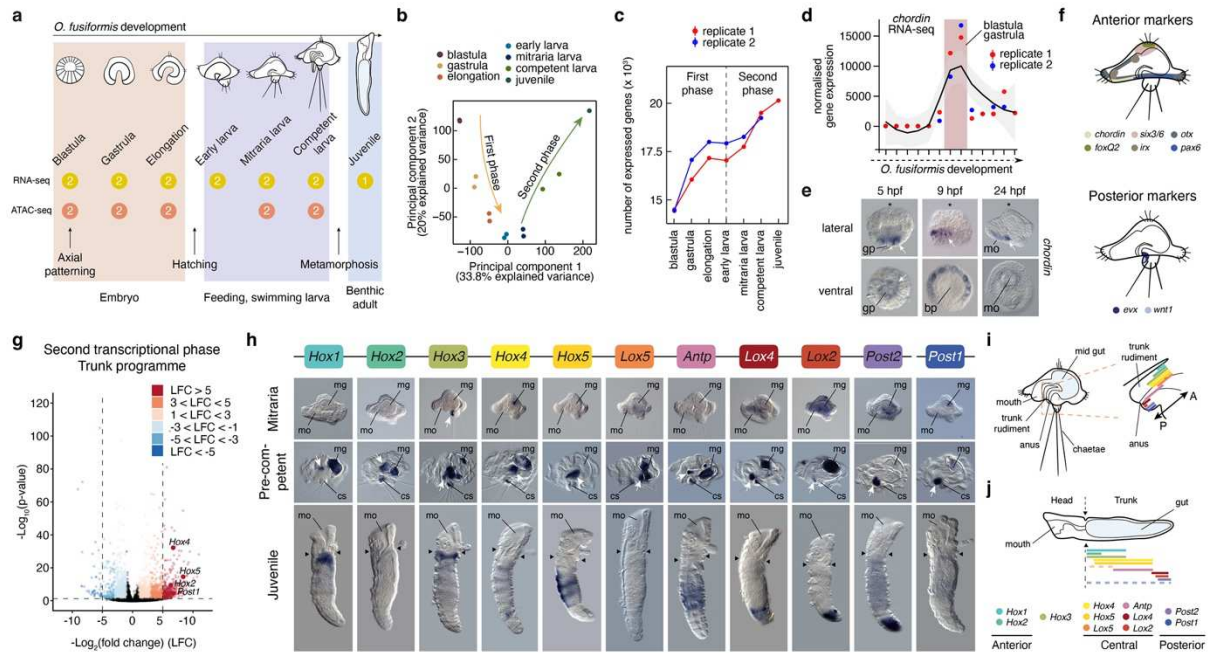
Figure 2 | Two transcriptional phases underpin larva embryogenesis and metamorphosis. a, Schematic representation of *O. fusiformis* development from blastula to juvenile indicating major developmental events and the time points when RNA-seq and ATAC-seq samples were collected. Numbers inside the coloured 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 (27 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

44

1030   anterior-posterior axis at the developing trunk (white arrows in the pre-competent larva) and

1031   juvenile. Dotted lines in the competent larva panels indicate background from the gut content.

1032   Black arrowheads in the juvenile panels indicate head to trunk boundary. cs: chaetal sack;

1033   mg: mid gut; mo: mouth. **i, j,** Schematic representations of the expression of *Hox* genes in the

1034   trunk rudiment of the competent larva (**i**) and juvenile trunk (**j**). A: anterior; P: posterior.

1035   Drawings are not to scale, and schematic expression domains are approximate.
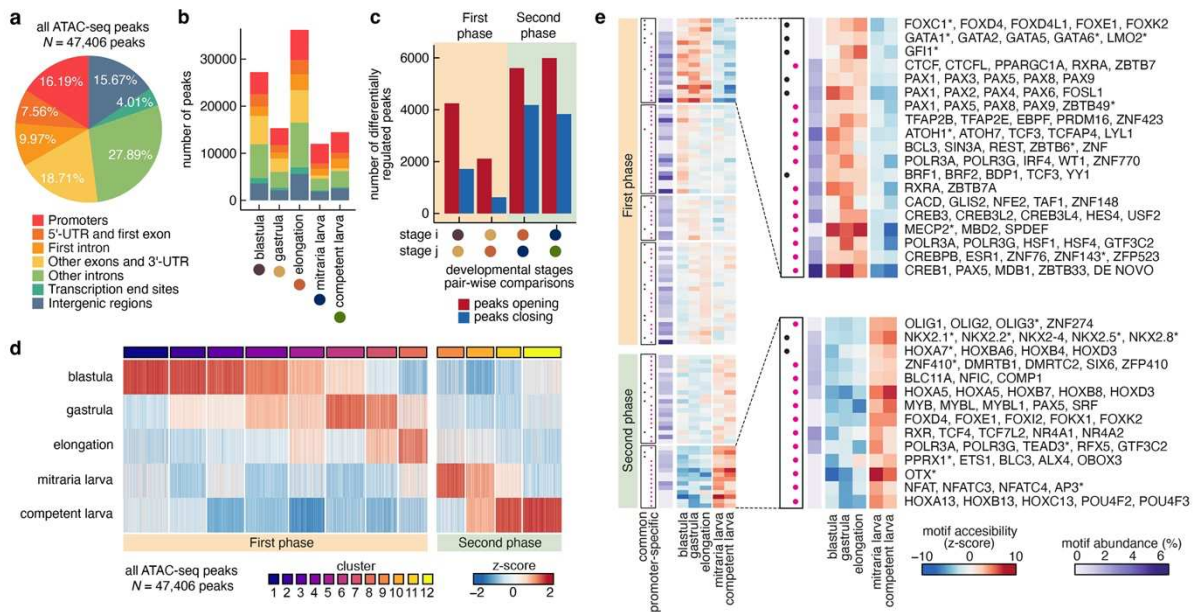
1036

1037 **Figure 3 | Chromatin dynamics support two regulatory modules during *Owenia***

1038 **development. a,** Genomic feature annotation of the consensus ATAC-seq peak set.

1039 **b,** Stacked bar plot showing the number of called peaks per developmental stage, classified

1040 by genomic feature. **c,** Differentially accessible peaks for all four pairwise comparisons

1041 between adjacent stages. Only significant differentially regulated peaks (adjusted *p*–

1042 value < 0.05) are shown. **d,** Heatmap of normalised peak accessibility (z–score) of the soft

1043 clustered consensus ATAC-seq peaks, subdivided into clusters of open chromatin regions that

1044 peak pre- and post-larval stages. **e,** Motif enrichment analysis of promoter peaks. The

1045 clustered heatmap shows normalised motif accessibility (z-score) and the abundance of the

1046 motif in promoter peaks, subdivided in clusters of motifs mostly accessible either pre- or

1047 post-larval stage. Clusters 1 (top right) and 6 (bottom right) are the most accessible of each

1048 regulatory programme, and thus are shown enlarged and with predicted bound transcription

1049 factors to the right. Asterisk denotes predicted factors from GimmeMotif curated databases.

1050
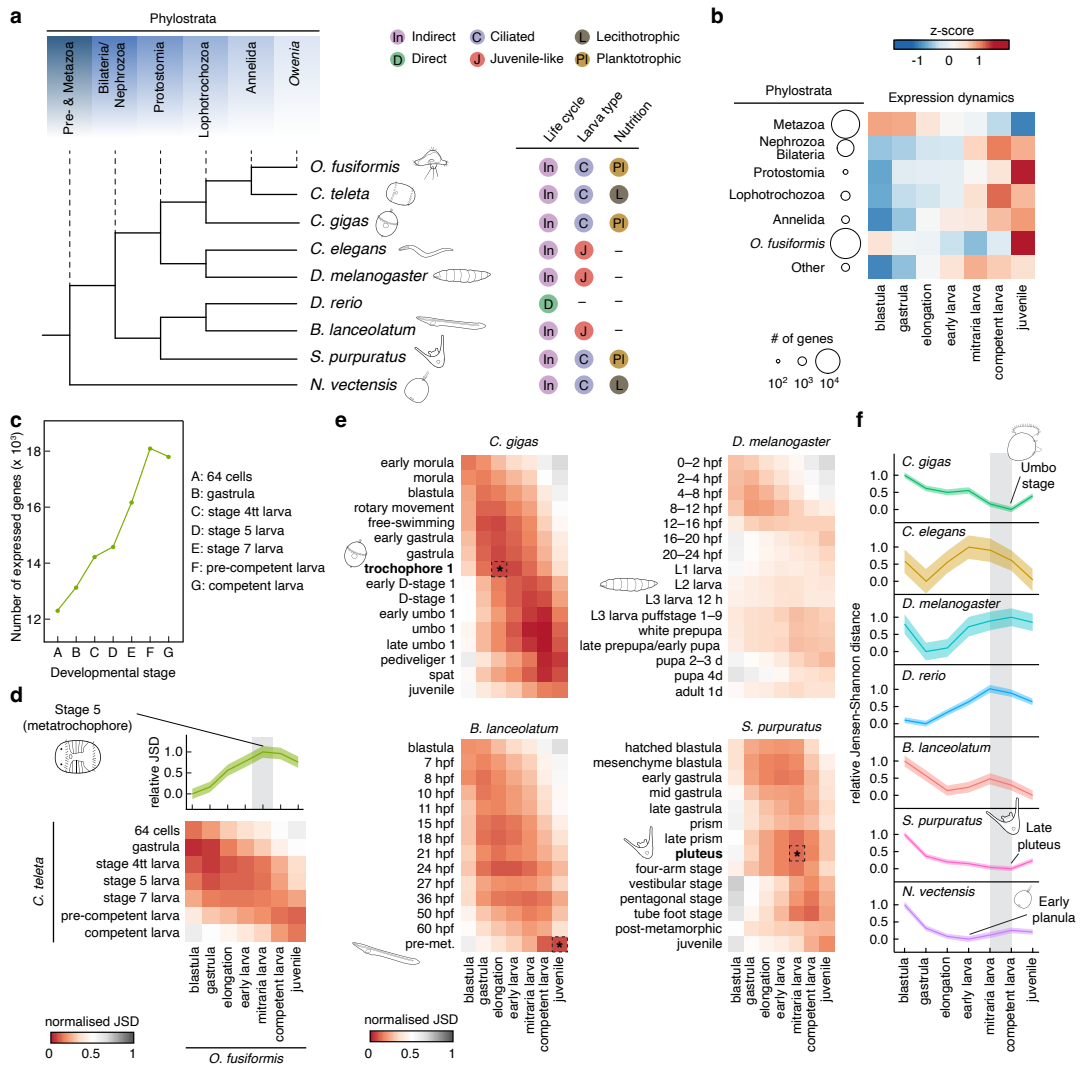
Figure 4 | Novel genes and comparative transcriptional dynamics during larval development. **a,** Cladogram of the species used for comparative transcriptomics analyses, indicating on top the phylogenetic age of each phylostrata considered for phylostratigraphy. For each species, the type of life cycle (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*

1062   and *C. teleta*, with the relative JSD of the *C. teleta* stages of minimal divergence to each

1063   *O. fusiformis* stage on top. **e,** Heatmaps of pairwise normalised JSD between *O. fusiformis*

1064   and *C. gigas*, *D. melanogaster*, *B. lanceolatum* and *S. purpuratus* (the asterisk indicates the

1065   stage of minimal JSD between the larval stage of each species with *O. fusiformis*). **f,** Relative

1066   JSD from stages of minimal divergence of species in (**a**) to each *O. fusiformis* developmental

1067   stage. Confidence intervals in (**b**) and (**f**) represent the standard deviation from 250 bootstrap

1068   resamplings of the one-to-one orthologs.
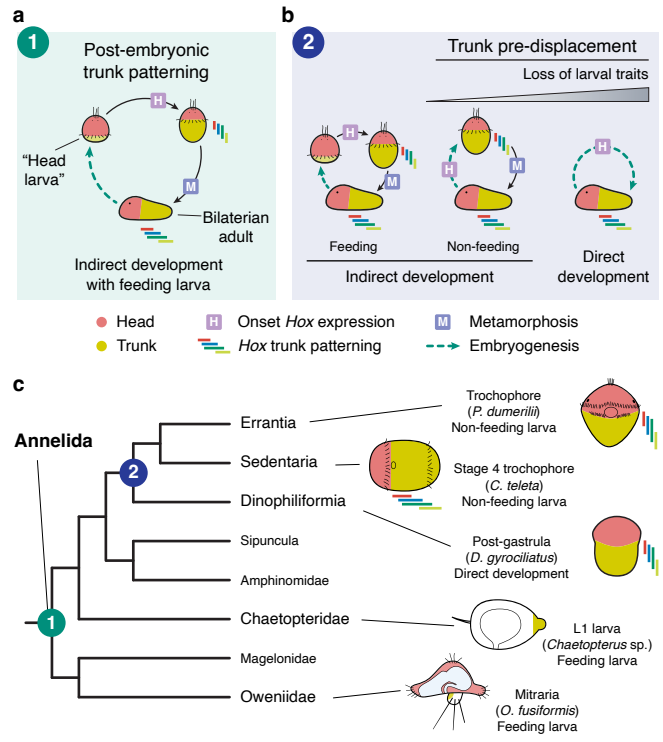
1069

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

1084   established post-gastrulation, as in the direct developer *D. gyrociliatus*. Notably, feeding

1085   larva in Errantia and Sedentaria behave as those of Oweniidae and Chaetopteridae, and thus a

1086   post-embryonic trunk patterning is likely an ancestral condition (green circle with 1) and the

1087   convergent pre-displacement of trunk differentiation to embryogenesis (blue box with 2)

1088   concurred with the evolution of indirect development with feeding larva and direct

1089   development. Drawings are not to scale.

1090
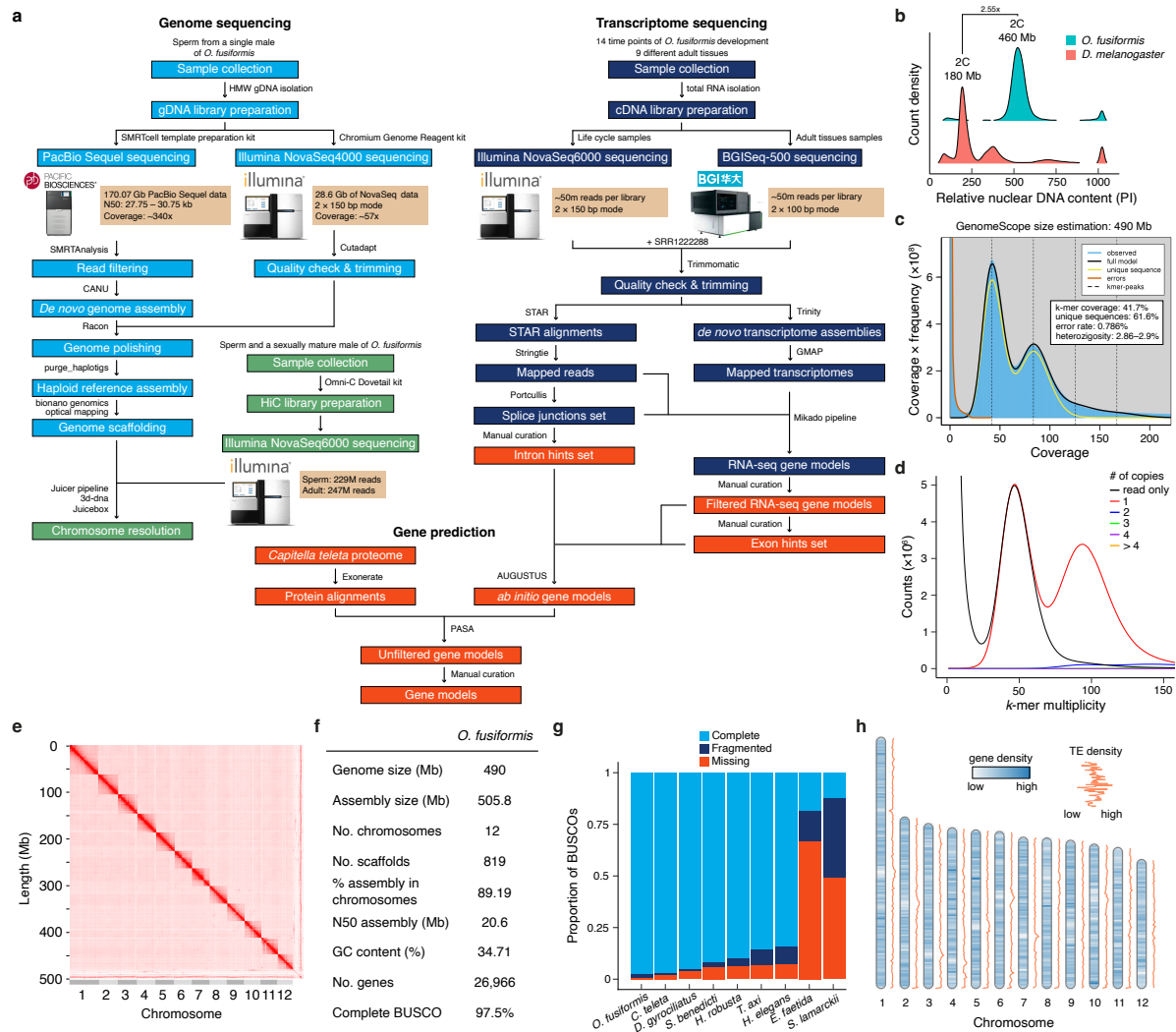
1091

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*.
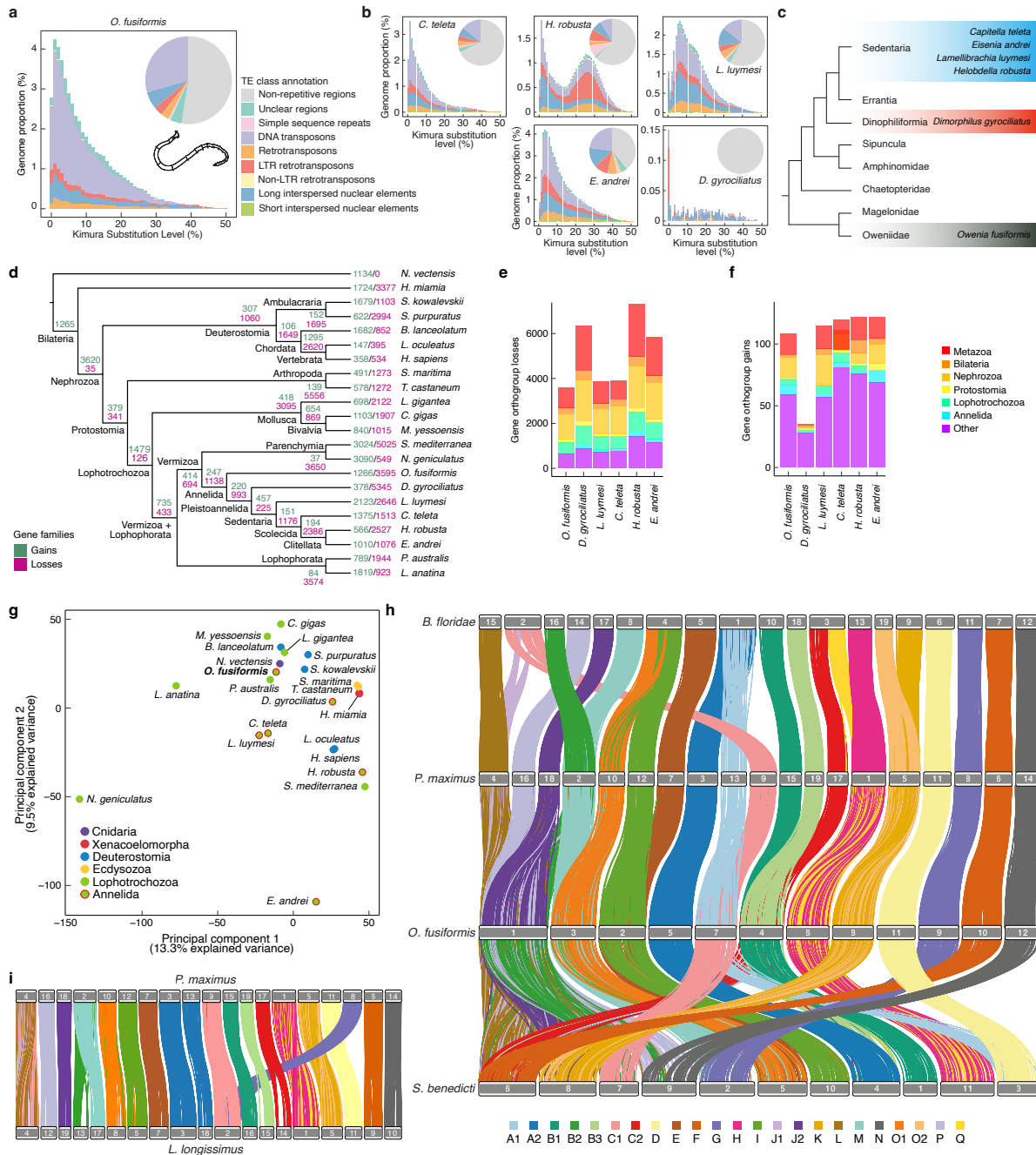
Extended Data Figure 2 | The genome of *Owenia fusiformis* is conservatively evolving.

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

1109   tree topology. Gains are shown in green, losses in violet. Gene family losses in *O. fusiformis*

1110   are like those of slow-evolving lineages. **e, f,** *O. fusiformis* has the lowest number of gene

1111   losses of all sampled annelids (**e**), and the least gene expansions (**f**) after the extremely

1112   compact genome of *D. gyrociliatus*. **g**, Principal component analysis from Fig. 1f, showing

1113   the full set of species. **h**, Macrosynteny analysis between *O. fusiformis*, and from top to

1114   bottom, the cephalochordate *Branchiostoma floridae*, the bivalve *Pecten maximus*, and the

1115   annelid *Streblospio benedicti*. *Owenia fusiformis* retains ancestral linkage groups but also

1116   exhibits annelid- and species-specific chromosomal arrangements. However, the karyotype of

1117   *O. fusiformis* is more conserved than that of the annelid *S. benedicti*. **i**, Macrosynteny analysis

1118   between the bivalve *P. maximus* and the nemeartean worm *L. longissimus*.

1119   *Lineus longissimus* exhibits conserved ancestral bilaterian linkage groups, including three

1120   potential lophotrochozoan-specific chromosomal rearrangements (H+Q, J2+L and K+O2),

1121   plus a nemertean-specific fusion (G+C1).
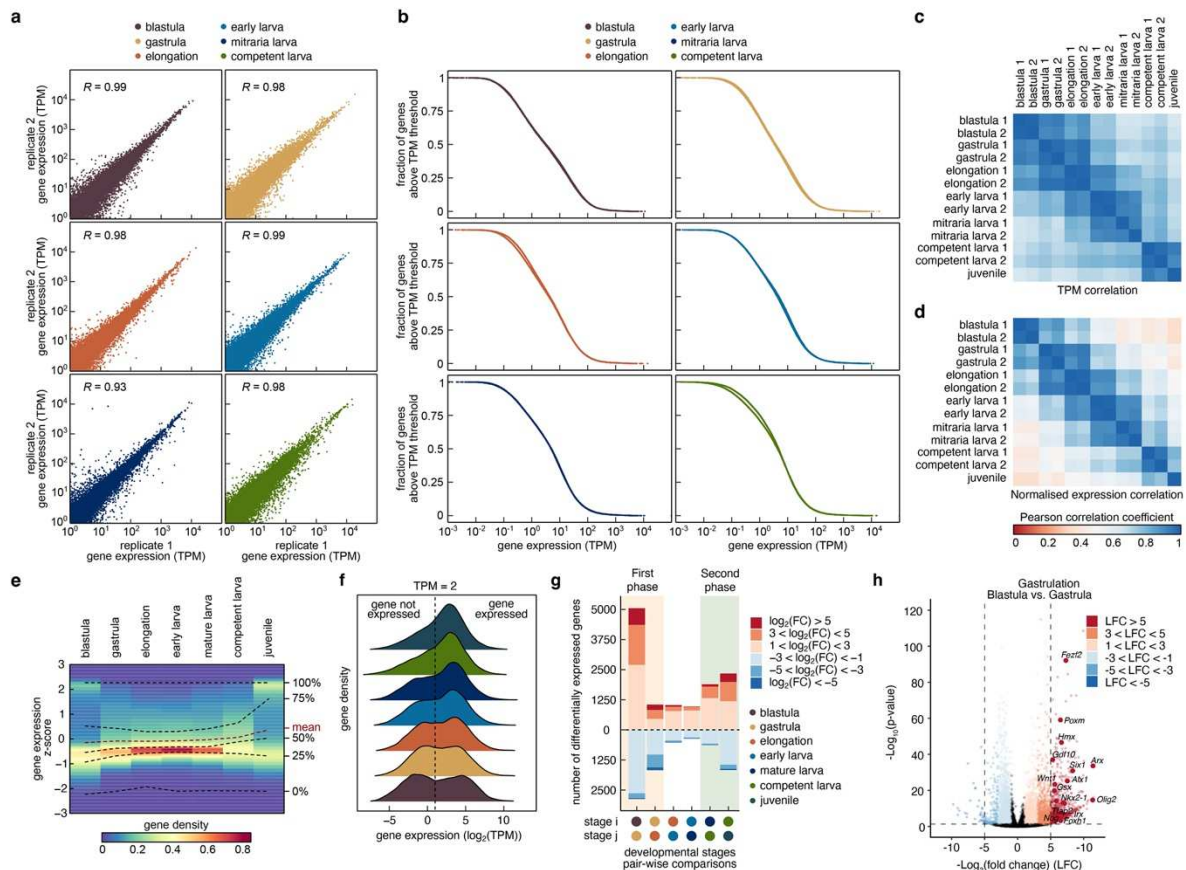
Extended Data Figure 3 | Gene expression dynamics during *O. fusiformis* development.

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

1136   expressed. Colour legend is next to **g**. **g,** Summary of differential gene expression analyses

1137   between consecutive developmental stages. Only significantly differentially expressed (i.e.,

1138   those with an adjusted p–value < 0.05) are shown. $\log_2(FC)$: $\log_2$(fold-change). **h,** Volcano

1139   plot of gastrulation (onset of the first transcriptional phase), highlighting top upregulated

1140   axial patterning genes. Colours of dots represent $\log_2(FC)$ levels as in (**g**).
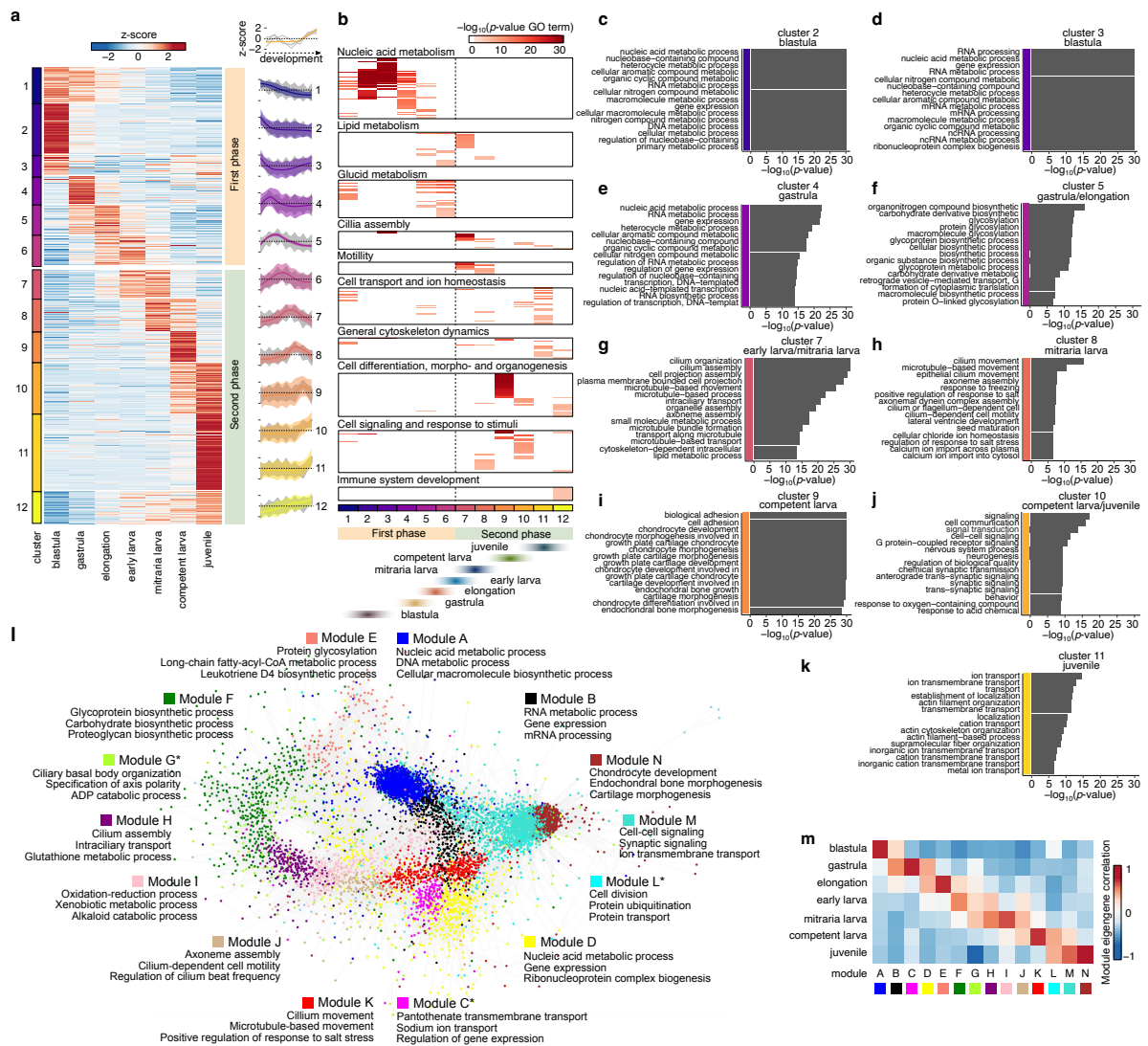
1141

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 $-\log_{10}(p$–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

1152   enriched GO terms in nine representative clusters. For the full list of GO terms and clusters,

1153   see Supplementary Fig. 2. **i**, Force-directed layout representation of the weighted gene co-

1154   expression network analysis (WGCNA). For visualisation purposes, only the nodes and edges

1155   of a random selection of 30% of the transcripts of *O. fusiformis* are depicted here. For each of

1156   the 14 modules, representative enriched GO terms are shown (full lists are in Supplementary

1157   Fig. 4). Unadjusted *p*–values of GO terms from modules flagged with an asterisk (*) were

1158   lower than average. **m,** WGCNA module eigengene correlation with each developmental

1159   stage of *O. fusiformis*.
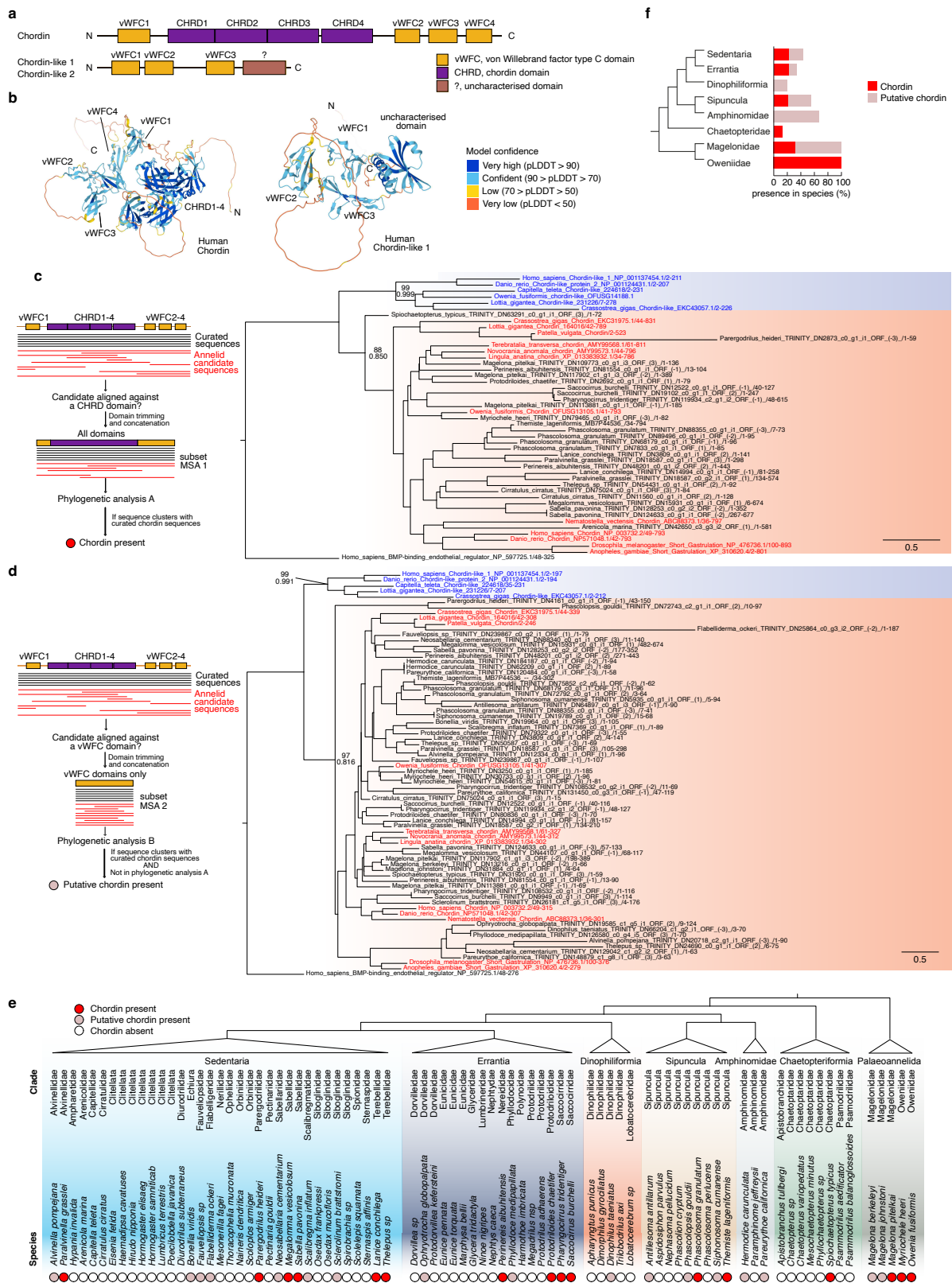
Extended Data Figure 5 | *chordin* **was lost multiple times in annelid lineages. a,** Domain

organisation of Chordin (CHRD) and Chordin-like (CHRDL1/2) proteins, as inferred from

1163   human orthologs. **b,** AlphaFold protein structure prediction for human Chordin (UniProt:

1164   Q9H2X0) and Chordin-like 1 (UniProt: Q9BU40) revealed a previously unknown and

1165   uncharacterised domain in CHRDL1 (also depicted in **a**). **c**, **d**, Orthology assignment of

1166   *chordin* annelid candidates. From the multiple sequence alignment, candidate annelid

1167   sequences with a 10-residue or longer fragment aligned against either the CHRD (**c**; i.e., bona

1168   fide *chordin* genes) or the vWFC domains (**d**; i.e., putative *chordin* genes) were kept for

1169   further analysis. CHRDL cluster is shaded in blue; CHRD cluster, in red. Bootstrap support

1170   values (top) and posterior probabilities (bottom) are shown at key nodes. Sequences in red

1171   and blue are curated CHRD and CHRDL sequences, respectively. **e, f**, Summary

1172   phylogenetic trees of presence/absence of *chordin* (red) or putative chordin (light brown)

1173   across Annelida.
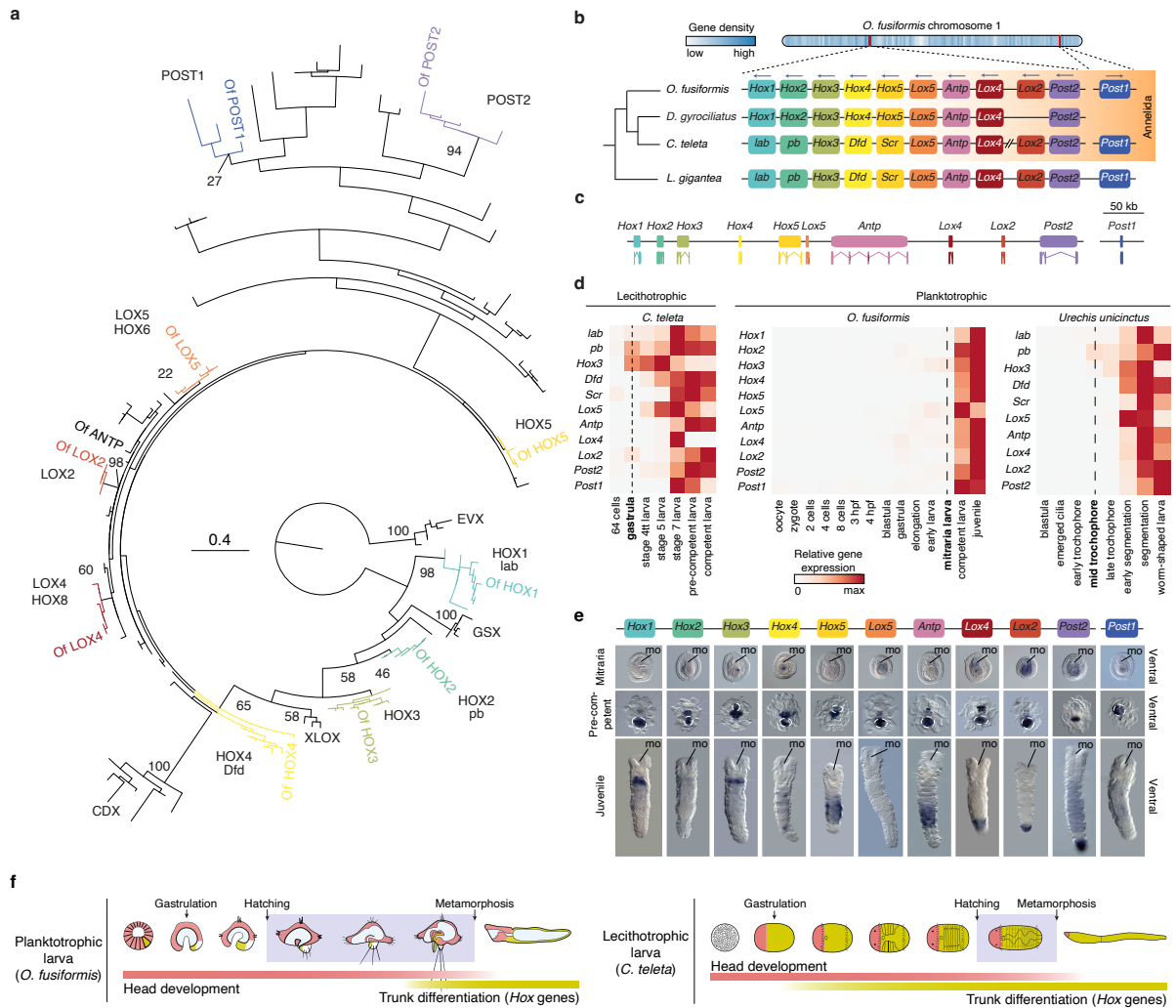
Extended Data Figure 6 | *Hox* **gene complement and expression in** *O. fusiformis* **and**

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.
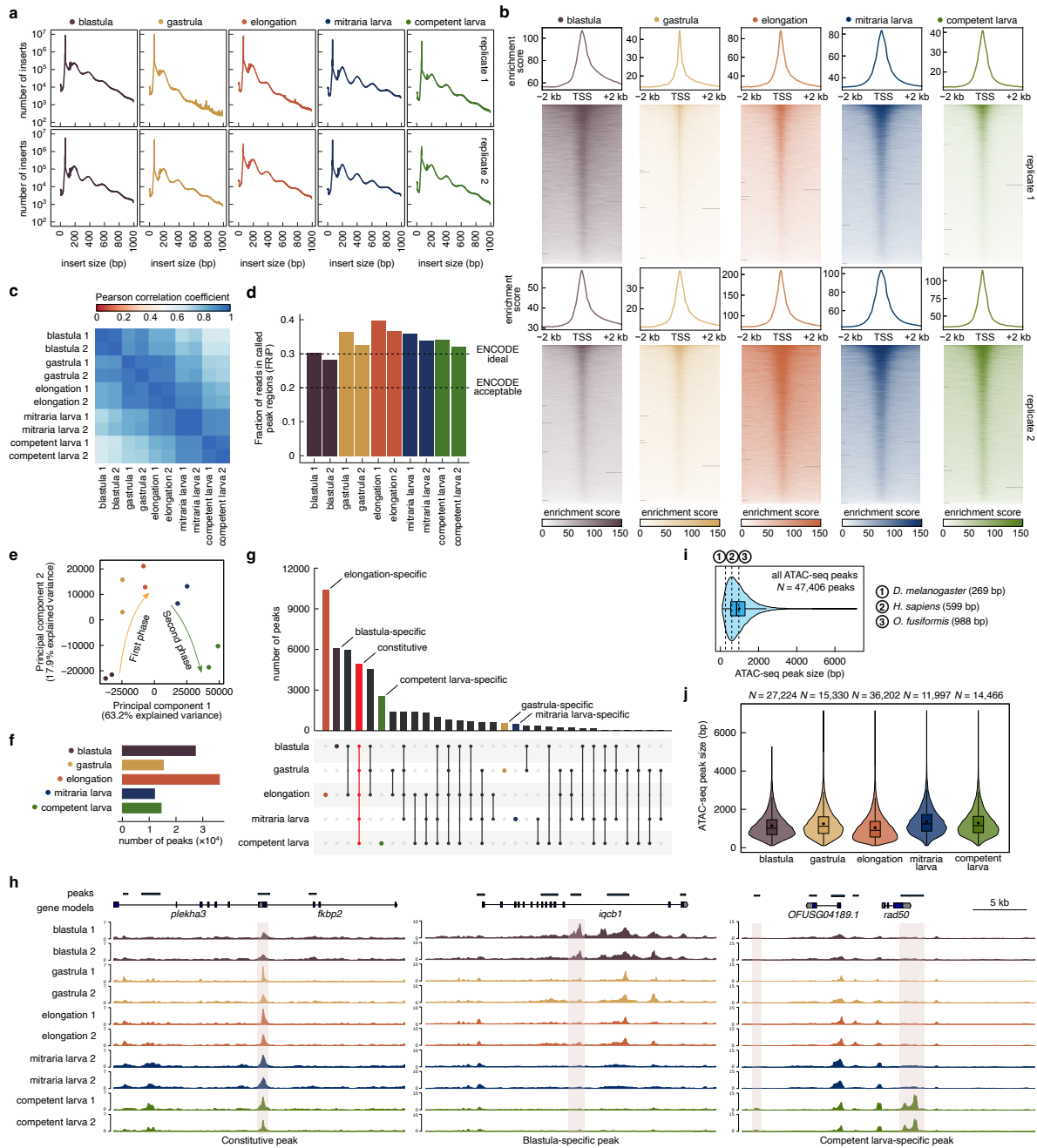
Extended Data Figure 7 | Chromatin accessibility dynamics during *O. fusiformis*

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-

1200   seq samples based on peaks as in (**c**). **f**, Bar plot of ATAC-seq peak number by

1201   developmental stage. **g,** UpSet plot classification of ATAC-seq peaks by developmental

1202   stage. Peaks present in more than one stage are coloured in grey. **h**, Representative views of

1203   constitutive and stage-specific peaks, with the ATAC-seq peak track indicating IDR

1204   consensus peaks. *plekha3:* pleckstrin homology domain containing A3; *fkbp2*: FKBP prolyl

1205   isomerase 2; *iqcb1*: IQ calmodulin-binding motif-containing protein 1; *rad50*: RAD50 double

1206   strand break repair protein. **i**, **j**, Violin plots of ATAC-seq peak width distribution for all peak

1207   set (**i**) and according to developmental stages (**j**). Dotted lines in **i** represent the mean peak

1208   size for fruit fly (1), human (2) and *O. fusiformis* (3). Peaks are narrower during axial

1209   elongation and broader in the mitraria larva.
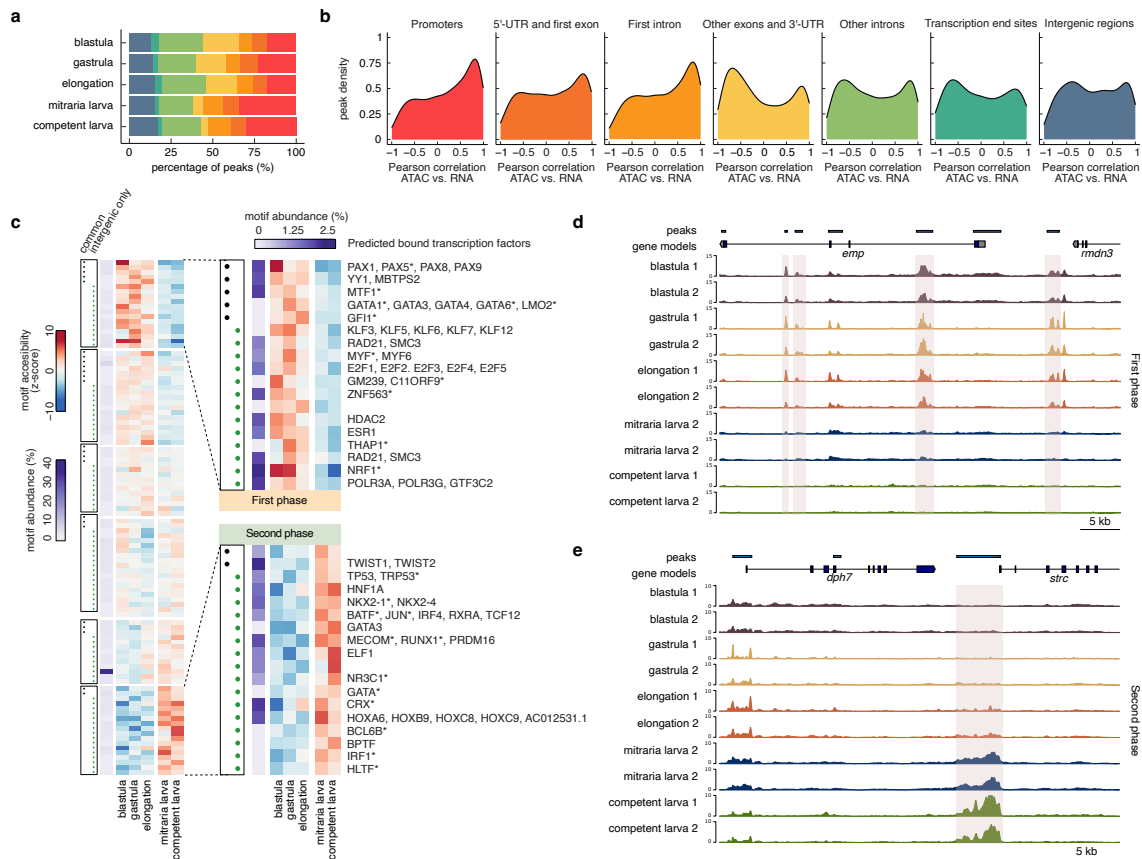
Extended Data Figure 8 | Correlation of ATAC-seq and RNA-seq data. a, Percentage of

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

1223   of the two regulatory programmes highlighted in light red. Peak track corresponds to the

1224   consensus ATAC-seq peak set. *emp*: epithelial membrane protein-related; *rmdn3*: regulator of

1225   microtubule dynamics 3; *dph7*: diphthamide biosynthesis 7; *strc*: stereocilin.
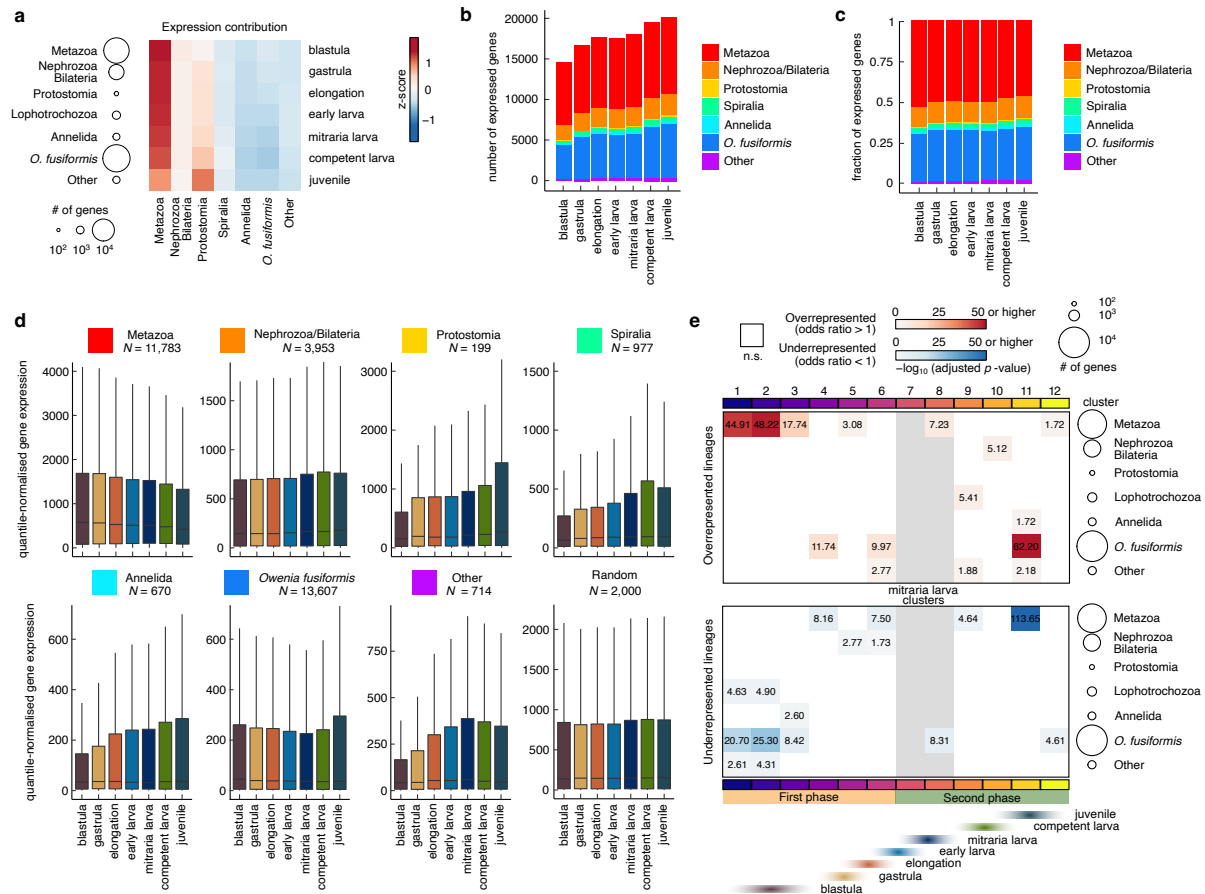
1226

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

1239   underrepresented lineages (OR < 1, adjusted p–value < 0.05). The results match the

1240   expression dynamics displayed in **d** (shaded grey are indicates the mitraria stage). In **c** and **f**,

1241   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

1248 *D. melanogaster*, the bivalve *C. gigas*, and the annelids *O. fusiformis* and *C. teleta*.

1249 **b,** Heatmaps of JSD between *O. fusiformis* and all other eight species using only the subset of

1250 single copy orthologs common to all nine species. This excludes all genes that are not of pre-

1251 metazoan and metazoan origin, as they will not be present in *N. vectensis*. **c,** Relative JSD

1252 from stages of minimal divergence of eight animal species to each *O. fusiformis*

1253 developmental stage calculated from the subset of single copy common orthologs.

1254 Confidence intervals are the standard deviation from 250 bootstrap resamplings of the single

1255 copy ortholog subset ($N = 459$ orthologs for all eight pairwise comparisons). **d,** alternative

1256 evolutionary scenarios for the deployment of *Hox* genes (as proxy for trunk patterning and

1257 assuming the staggered expression along the directive axis of cnidarians and anteroposterior

1258 axis of bilaterians is homologous, which does not necessarily imply homology of the two

1259 axes). Given our current understanding of *Hox* gene deployment in cnidarian and bilaterian

1260 taxa (see Supplementary Table 33), a late post-embryonic *Hox* patterning ancestral to

1261 Bilateria and Cnidaria, as seen in extant lineages with maximal indirect development, is a

1262 more parsimonious scenario (bottom scenario), which could have been favoured the evolution

1263 of bilaterian larvae.
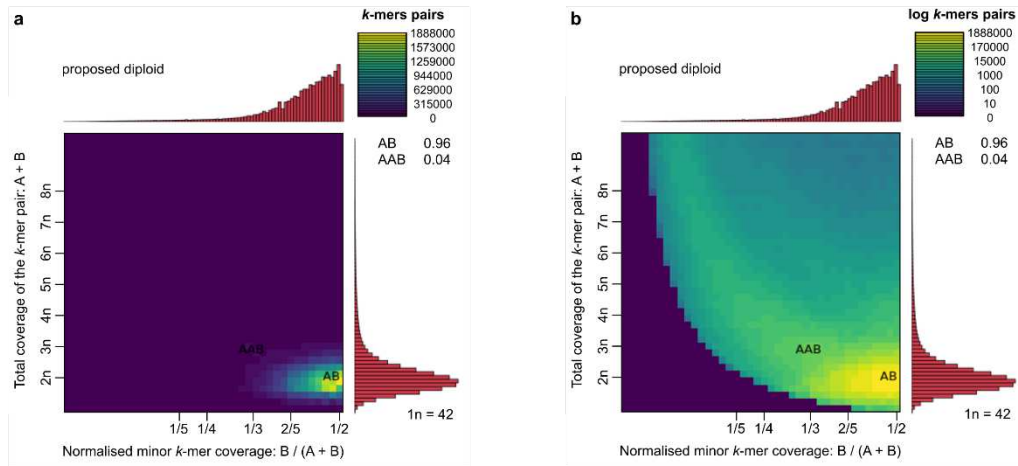
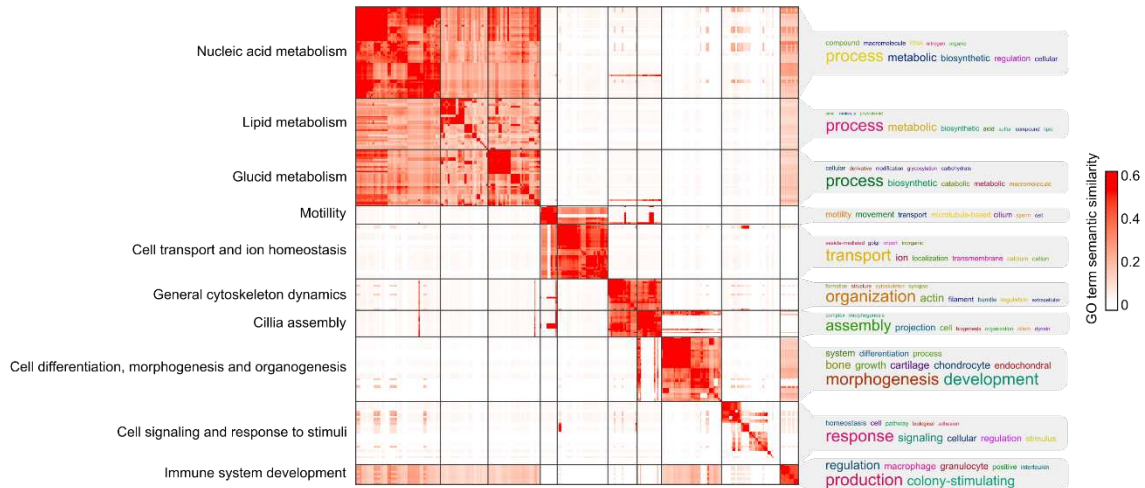1331

Supplementary Figure 1 | Smudgeplot ploidy estimation. a, Linear scale and **b,** base 10

logarithmic scale smudgeplots. Diploid (AB) *k*-mer pairs are predominant (0.96) for

O. fusiformis genome, suggesting diploidy.

1335

Supplementary Figure 2 | Gene ontology terms enrichment for RNA-seq clusters. **a–**

l, Bar plots depicting unadjusted *p*–values of the top 30 gene ontology (GO) terms for

biological process for each of the 12 clusters (1–12, **a–l**) of co-regulated transcripts retrieved

through soft *k*-means clustering.

1340

1341 **Supplementary Figure 3 | Clustering of gene ontology terms enriched in RNA-seq**

1342 **clusters.** Similarity matrix of all 288 non-redundant gene ontology (GO) terms for biological

1343 process from the GO term enrichment analysis performed on RNA-seq clusters. GO terms

1344 were grouped through *k*-means clustering according to their semantic similarity. A word

1345 cloud recapitulates each cluster, with word size being proportional to word frequency in the

1346 GO terms contained within that cluster. Custom umbrella terms were selected to summarise

1347 each cluster of GO terms, here shown on the left of the matrix.

1348
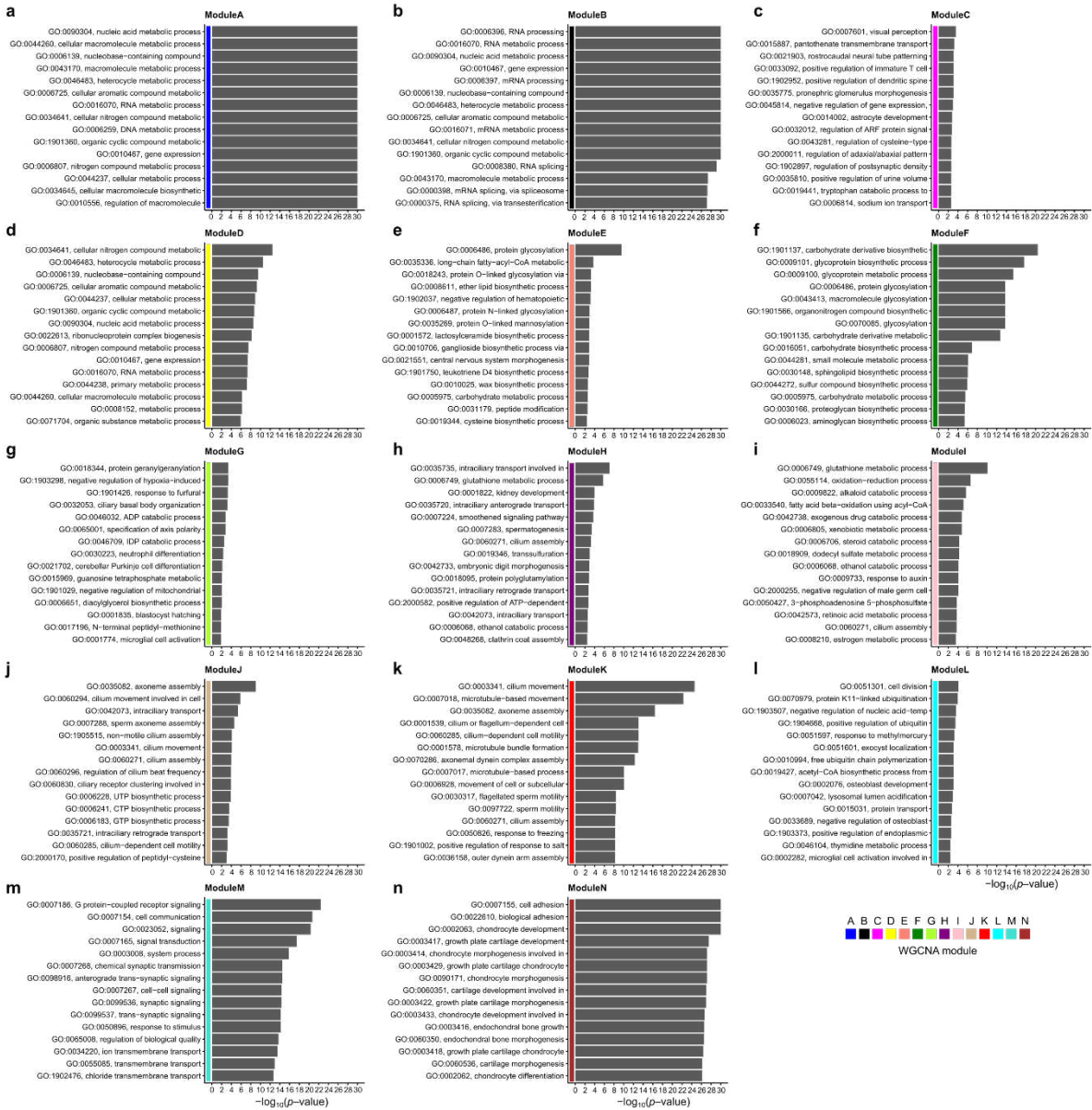
1349

Supplementary Figure 4 | Gene ontology terms enrichment for WGCNA modules. a–n, Barplots depicting unadjusted *p*–values of the top 15 gene ontology (GO) terms for biological process for each of the 14 modules (A–N, **a–n**) of co-expressed transcripts retrieved through our WGCNA approach.
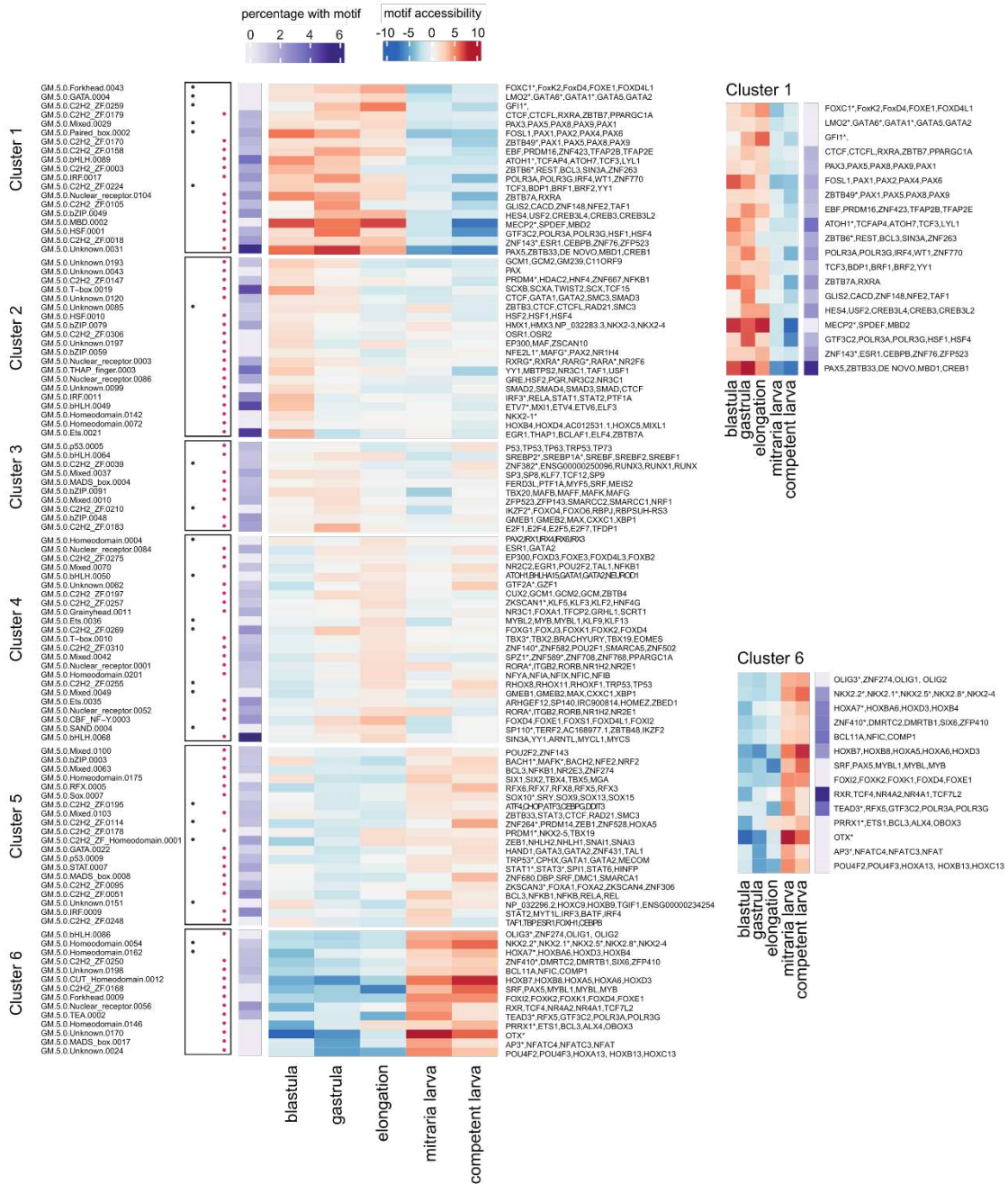
1354

Supplementary Figure 5 | Full motif enrichment analysis of promoter peaks. Full motif

enrichment analysis of promoter 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. The full list of all predicted bound transcription

factors and raw motif IDs is shown here.
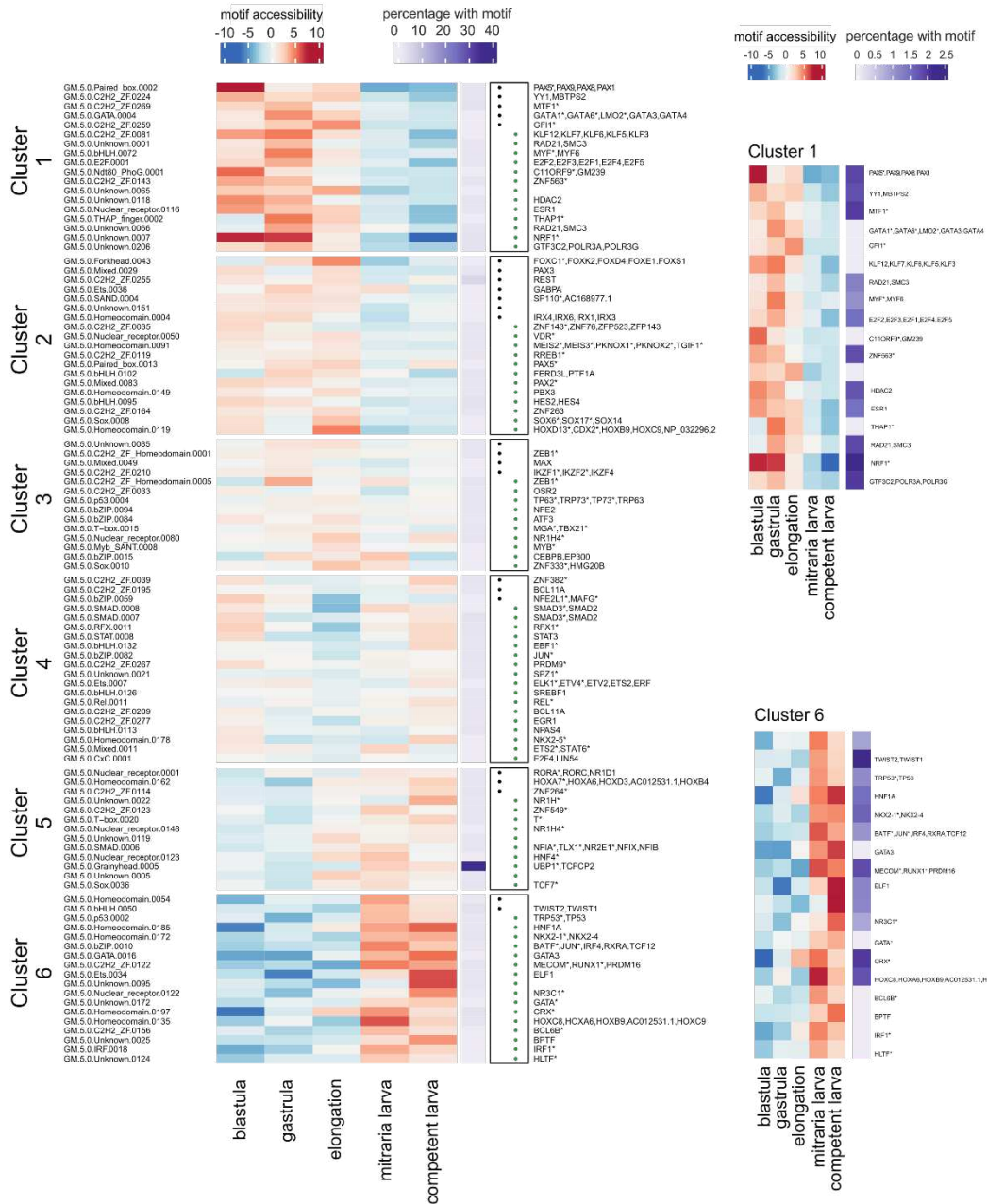
1361

Supplementary Figure 6 | Full motif enrichment analysis of intergenic peaks. Full 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 intergenic peaks. Clusters 1 (top right) and 6 (bottom right) are shown in detail as

representative of each regulatory programme. The full list of all predicted bound transcription

factors and raw motif IDs is shown here. Note the scale for motif abundance is different for

zoomed detailed clusters than for the full set, for comparison purposes.
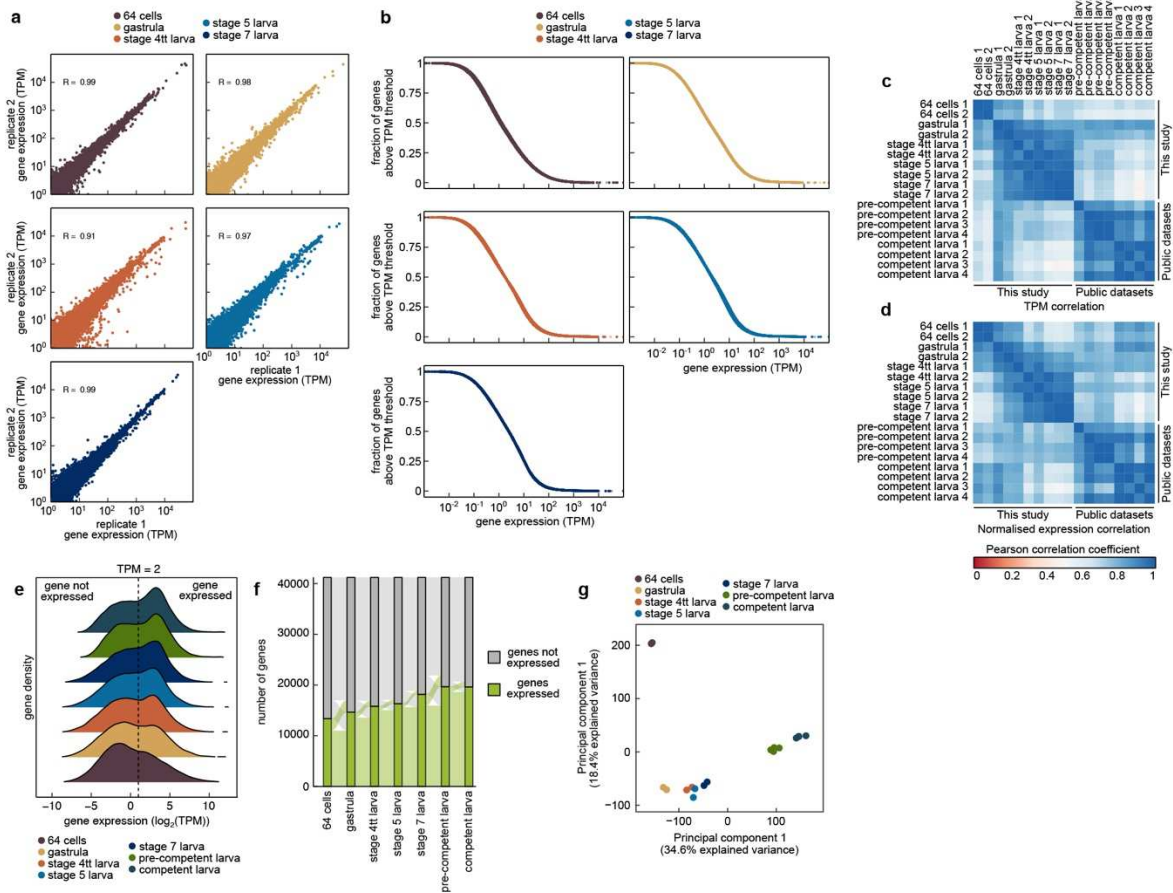
1369

Supplementary Figure 7 | Quality control and sample correlation of stage-specific RNA-seq samples of *C. teleta*. 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**). Adequate agreement between biological replicates can be observed in both matrices. **e,** 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. **f,** Alluvial plot of expressed and not expressed genes throughout the development of *C. teleta*. As in Fig. 4d, the two phases of gene expression observed for

1380   *O. fusiformis* in Fig. 2c cannot be discerned in *C. teleta*. **g,** Principal component analysis of

1381   the developmental RNA-seq time course of *O. fusiformis*.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- LMZSupplementaryTables01.xlsx
- LMZSupplementaryTables02.xlsx