

SCIENTIFIC REPORTS



OPEN

High expression of new genes in trochophore enlightening the ontogeny and evolution of trochozoans

Received: 23 February 2016
Accepted: 19 September 2016
Published: 04 October 2016

Fei Xu^{1,2,3}, Tomislav Domazet-Lošo^{4,5}, Dingding Fan⁶, Thomas L. Dunwell⁷, Li Li^{1,2,3}, Xiaodong Fang⁶ & Guofan Zhang^{1,2,3}

Animals with trochophore larvae belong to Trochozoa, one of the main branches of Bilateria. In addition to exhibiting spiral cleavage and early cell fate determination, trochozoans typically undergo indirect development, which contributes to the most unique characteristics of their ontogeny. The indirect development of trochozoans has provoked discussion regarding the origin and evolution of marine larvae and is interesting from the perspective of phylogeny-ontogeny correspondence. While these phylo-onto correlations have an hourglass shape in Deuterostomia, Ecdysozoa, plants and even fungi, they have seldom been studied in Trochozoa, and even Lophotrochozoa. Here, we compared the ontogenetic transcriptomes of the Pacific oyster, *Crassostrea gigas* (Bivalvia, Mollusca), the Pacific abalone, *Haliotis discus hannai* (Gastropoda, Mollusca), and the sand worm *Perinereis aibuhitensis* (Polychaeta, Annelida) using several complementary phylotranscriptomic methods to examine their evolutionary trajectories. The results revealed the late trochophore stage as the phylotypic phase. However, this basic pattern is accompanied with increased use of new genes in the trochophore stages which marks specific adaptations of the larval body plans.

The reconstruction of the Metazoan ancestor has been challenging due to the existence of indirect developmental life cycles, which are common in the main branches of bilateral animals. Some marine phyla consist of animals that undergo biphasic development characterized by a free-swimming larval stage during ontogenesis¹. For instance, annelids and mollusks are characterized by ciliated trochophore larvae (Fig. 1), and are both placed within the Trochozoa², a group of spiralian animals which account for at least one-third of all marine species³. Trochozoans are often included in studies that discuss the origin of the metazoan body plan^{4,5}. The increasing number of sequenced trochozoan genomes has facilitated the investigation of their developmental patterns^{3,6}. For example, Hox genes are present in all of these animals³ and in some species start to be expressed during mid-embryogenesis, consistent with observations made in other phyla⁷. These findings further suggest the possible existence of common regulatory mechanisms that are ancestral to all animals. Indeed, the phylotypic stage, a phylum specific ontogenetic period, has been assigned to a broad range of phyla⁸, and the common phylogeny-ontogeny correlation has an “hourglass” or “egg timer” shape⁹. In vertebrates, the timing of Hox gene expressions was considered to be an important molecular indicator of phylotype⁹, this has also been well studied on the basis of transcriptomic data that consistently indicate the presence of phylotypic stages^{9,10}. Taking gene regulatory networks into consideration, the hourglass model of embryogenesis can be explained by a hypothesis in which the early stages of embryogenesis are highly diversified as a result of variations in egg construction, which are driven by adult adaptation to the environment. This results in low similarity between species at this

¹Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China. ²Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China. ³National & Local Joint Engineering Laboratory of Ecological Mariculture, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China. ⁴Laboratory of Evolutionary Genetics, Ruđer Bošković Institute, Bijenička cesta 54, P.P. 180, HR-10002, Zagreb, Croatia. ⁵Catholic university of Croatia, Ilica 242, HR-10000, Zagreb, Croatia. ⁶BGI-Shenzhen, Shenzhen, 518083, China. ⁷Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK. Correspondence and requests for materials should be addressed to G.Z. (email: gzhang@qdio.ac.cn)

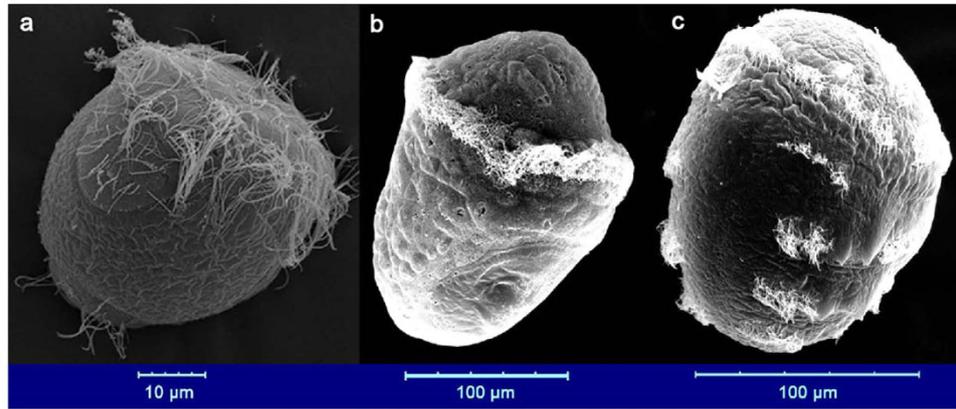


Figure 1. Electron microscopy images showing the trochophores. From left to right: The Pacific oyster, *Crassostrea gigas*, trochophore (14 hpf, in seawater at approximately 26 °C; photo from Zhang *et al.*⁶). The Pacific abalone, *Haliotis discus hannai*, trochophore (15 hpf, in seawater at approximately 22 °C). The sand worm *Perinereis aibuhitensis* trochophore (31 hours post-fertilization (hpf), in seawater at approximately 25 °C).

early developmental stage. As the embryo develops, phylogenetically conserved regulatory mechanisms become fully operative and assert increasing developmental control, resulting in phylotypic stages with high similarity. Finally, the fully developed diversified gene networks begin assert their influence resulting in adult bodies with low similarity¹¹.

While the hourglass profiles in ontogenic transcriptomes has been well described in Ecdysozoa (*Drosophila*, *Caenorhabditis*)^{10,12}, plants (*Arabidopsis*)¹³, and fungi (*Coprinopsis*)¹⁴ it has however been poorly explored whether the convergence of mid-ontogeny is also present in trochozoans^{8,15,16}. Moreover, it is unclear whether the trochophores of different phyla are subject to similar molecular regulatory mechanisms. To help answer these questions, we analyzed the transcriptome characteristics of the main developmental stages in three trochozoans: the Pacific oyster, *Crassostrea gigas* (Bivalvia, Mollusca), the Pacific abalone, *Haliotis discus hannai* (Gastropoda, Mollusca), and the sand worm *Perinereis aibuhitensis* (Polychaeta, Annelida). We also conducted genome re-sequencing in ten individual *C. gigas* oysters to analyze the relationships between genetic variation and the expression patterns observed during embryogenesis.

Results

The transcriptome age index of three trochozoans. The transcriptome age index (TAI) was calculated for different development stages of the three trochozoans. When all genes are considered, the results revealed that the trochophore larvae show the youngest transcriptome of all ontogenic stages (Fig. 2). However, this pattern was mainly affected by species-specific genes, *i.e.* the youngest phylostratum (ps11 oyster, ps11 abalone and ps10 sand worm, Fig. 2 and Supplementary Fig. S1). We found that only about 56 genes (51 from ps11 and 5 from ps10) contribute to increased TAI in trochophores in oyster, and oyster trochophores show the lowest TAI when these genes were excluded from the analysis (Supplementary Fig. S2, S3 and Supplementary Table S1). The overall pattern was robust when we calculated TAI for the oyster dataset with all the 38 developmental samples and all the annotated genes included (see Supplementary Fig. S4). Thus, the selected development stages used in comparative analyses of the three trochozoans reflect well the basic transcriptome pattern.

To get a more detailed picture of TAI differences we have depicted the relative expression levels (REL) for each phylostratum across the oyster ontogeny. It was evident that two groups of phylostrata exhibited opposing trends in expression after the trochophore stage (Fig. 3). The first group of phylostrata (ps1, 3, 6 and 8) show a rapid increase in relative expression values immediately following the trochophore stage. By contrast, the relative expression values of the other group (ps4, 5, 10 and 11) show a striking decrease. These differences reach their maximum at the late trochophore stage. Genes in phylum phylostratum (ps9) also has a quick increased REL during trochophore. In contrast, expression of ps2 genes shows decreasing pattern along the development, while ps7 has the lowest REL at the trochophore stage.

The transcriptome similarities between trochozoans. As the TAI catches only one aspect of evo-devo correlations it is good to complement it with other phylo-transcriptomic measures^{17,18}. Therefore, we further analyzed the similarities between the transcriptomes of the three species over the course of embryogenesis. According to one view the adaptive pressures for changes in the phylotypic stage of development are minimized¹⁵. For this reason, this stage is most likely to retain features found in the last common ancestor; thus, similarities in the transcriptomes of the three trochozoan species should be highest during the phylotypic stage⁹. We identified orthologous genes between the species pairs and compared the transcriptomes of the different embryonic stages in an all-to-all manner. This analysis shows a pretty good correspondence between developmental stages of the three species (Fig. 4). We generally found the highest similarity during the trochophore period between the two mollusks. The high similarities were also seen between the metatrochophore of sand worm and veliger of abalone. For sand worm and oyster, we detected the highest similarity in the gastrula stage that precedes the trochophore period (Fig. 4).

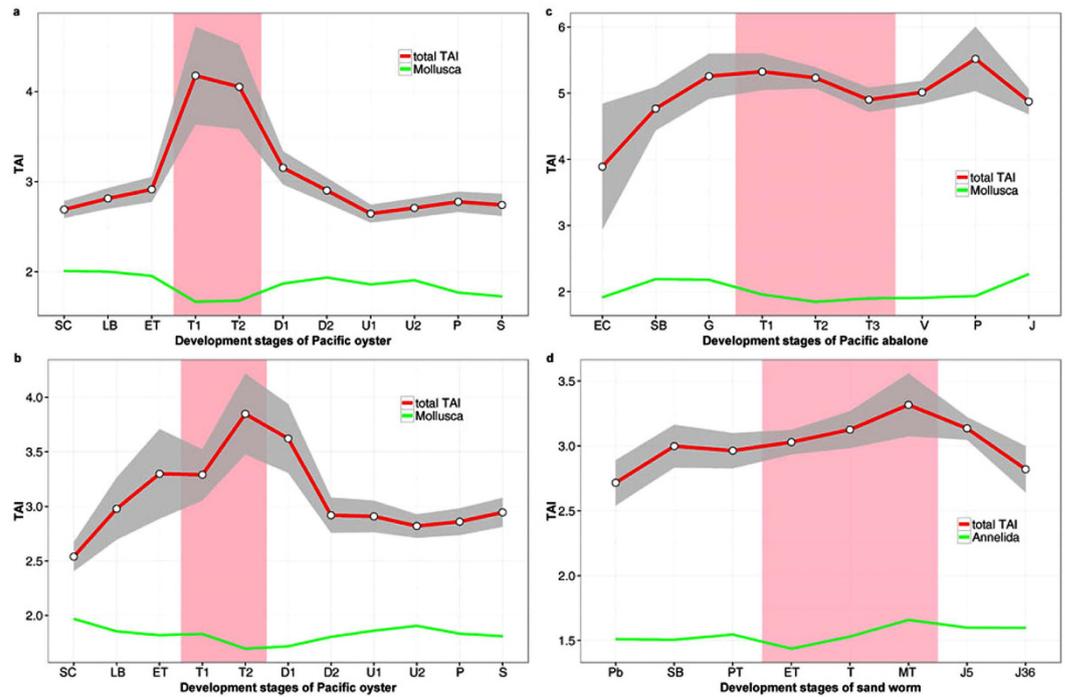


Figure 2. Transcriptome age indices across the ontogeny of the Pacific oyster, *Crassostrea gigas* (**a,b** indicate the two experimental replicates), the Pacific abalone, *Haliotis discus hannai* (**c**), and the sand worm *Perinereis aibuhitensis* (**d**). The red lines indicate the total cumulative value of the TAI. The green lines in panels a, b, and c show TAI for phylostrata older than Mollusca (ps1–ps9), while the green line in panel d represents TAI for phylostrata older than Annelida (ps1–ps9). The developmental stages and their timelines are defined in Supplementary Table S2–S5. The pink area designates the trochophore stage of each species. See the supplementary information for more details. The grey area designates the 95% confidence interval for TAI with a bootstrapping strategy. The overall patterns of the TAI profiles are significant as measured by the permutation tests (**a**. $P = 2.1 \times 10^{-11}$; **b**. $P = 1.2 \times 10^{-10}$; **c**. $P = 3.3 \times 10^{-16}$; **d**. $P = 2.6 \times 10^{-6}$).

The transcriptome polymorphism index of oyster. So far we tested profiles at the macroevolutionary scale. To further explore the conservation level within the population we also developed the novel transcriptome polymorphism index (TPI). This measure estimates, akin of TAI and TDI¹³, the mean sequence diversity of a transcriptome by weighting the nucleotide diversity (θ) of each gene by its expression level. By obtaining genome re-sequencing data for 10 oyster individuals we found that the TPI was lowest during the trochophore stage (Fig. 5a), which indicated that the transcriptome at this stage is more conserved.

The expression pattern of oyster Hox family genes. The conservation of genes expressed during trochophore stage can also be seen in Hox gene expression pattern. As molecular characteristics of vertebrate phylotype, the timing of Hox gene expression coincides with the pharyngula stage^{9,19,20}. Here we surveyed the expression pattern of the ten Hox genes in oyster (Fig. 5b). The expression timing for oyster Hox genes follows a similar temporal pattern between each group of Hox genes; the Hox1-Hox2-Hox3 and Hox4-Hox5 anterior group, Lox5-Lox4-Lox2 central group, and Post2-Post1 posterior group. In total, nine of the ten Hox genes are expressed during trochophore stage, indicating the full activation of this gene family, which is essential for the proper organization of the body plan.

Discussion

Our study marks the trochophore as the phylotypic stage in both Mollusca and Annelida. In the three assayed species the lowest phylum-level TAI value was found during the trochophore stage, indicating this stage having the oldest transcriptome (Fig. 2 and Supplementary Fig. S1). During this stage, the shell field has formed in mollusks and the three body segments can be observed in sand worm (Fig. 1). As few studies have focused on phylotypic stage analyses in trochozoans, speculations have been primarily made based on the main body plan characteristics of the taxon. For instance, it has been proposed that the segmentation stage should be considered the phylotype for annelids, whereas the veliger is a candidate for the phylotype stage in mollusks⁸. However, other authors employing phenotypic analysis have proposed the trochophore as the phylotypic stage in mollusks^{15,16}. In that line, high-density sampling and electron microscopy studies confirm that the molluscan shell field starts forming by trochophore stages⁶. Additionally, some annelid species form segments during the trochophore stage²¹. The timing at which these main phylum body plan characteristics arise is coincident with the observation that the lowest phylum-level TAIs occurred during the trochophore stage in the three trochozoans examined in this work. The highlighted similarities and low transcriptome polymorphism observed between the transcriptomes during

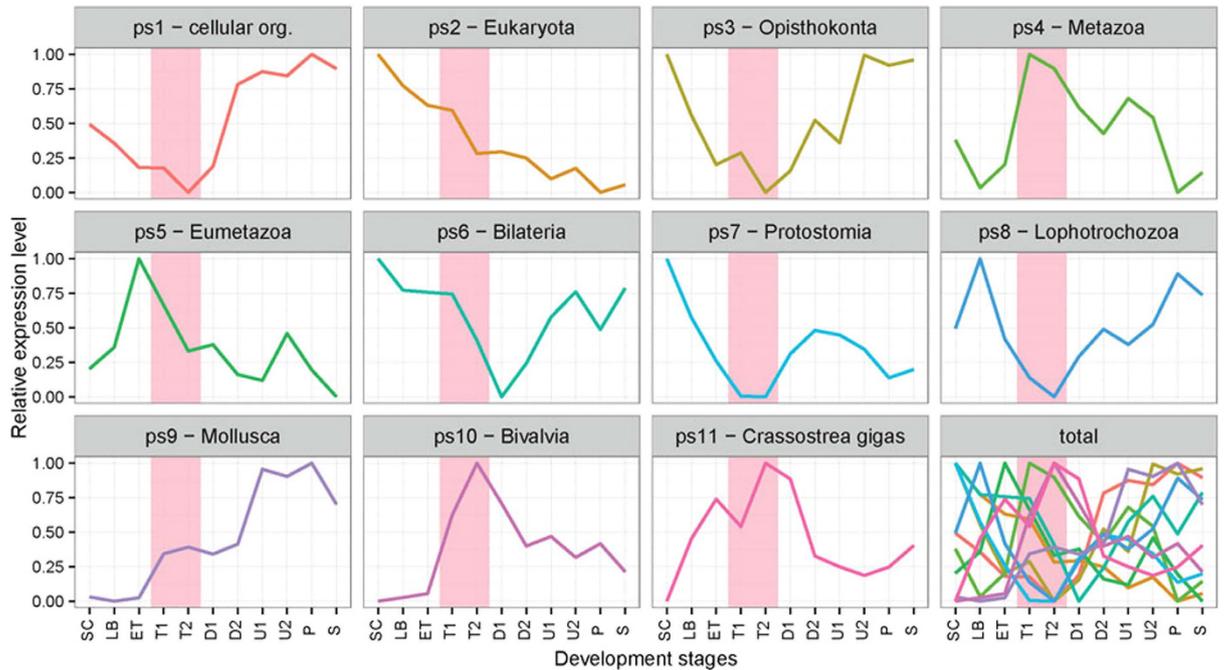


Figure 3. The mean relative expression levels (REL) of genes in each phylostratum. Phylostratum name is shown at the top label of each panel. The phylostrata (ps) 1, 3, 6, 8 relatively highly expressed at late development, while ps 4, 5, 10, 11 relatively lowly expressed at late development. Pink shaded area designates trochophore stage when the two types of phylostrata have the largest difference (the panel marked with “total”).

the molluscs trochophore stage resembled the molecular hourglass model reported in vertebrates⁹, *Drosophila*²², *Caenorhabditis*¹², plants¹³, and fungi¹⁴. A previous study on the expression pattern of oyster specific Homeobox genes, especially the observation that nine of the ten Hox cluster genes in oyster are expressed during trochophore stage, also suggested that the trochophore stage should be considered the phylotype⁷. However, the pattern for trochozoans we report here is less clear than in other organisms given that we found strong influence of some very young genes on the transcriptome composition in trochophores.

A recent report has indicated that transcriptome differences were minimal at trochophore stage in annelid *Platynereis dumerilii*, and thus this should be the phylotypic stage¹⁶. In this study, the highest levels of transcriptome similarity was also identified in the trochophore stage of the two mollusks. Among the common traits between mollusks and annelids, the trochophore larvae shows plesiomorphic characters such as the prototroch^{23,24}. However, trochophore larvae can never the less be highly diverse²³. In this study, the transcriptomes between the Annelida and the two mollusks did not show the highest similarity during the trochophore stage, which indicated substantial divergence between the two types of trochophores (Fig. 4). Indeed, the trochozoans are a group of animals showing wide body plan diversification, and some studies have suggested that many of the assembled features, such as the metatroch, have different embryonic origins or evolutionary histories²⁴. The complex genomic context of trochozoans includes high rates of genome polymorphism and large amounts of repeated content, making the molecular evolution of trochozoans intricate²⁵. The transcriptome divergence of trochophore between the two phylum (Mollusca and Annelida) may be a result of integrating new genes into regulatory networks through long-term evolution. The differences in trochophore transcriptomes between these two closely related phyla resembles the inverse hourglass pattern observed between other more distant animal phyla¹⁶. In this regard it was suggested that ‘phylum’ could be defined as a set of species sharing the phylogenetically conserved molecular regulatory network during the mid-development¹⁶.

While phylum level TAI and transcriptome similarity analyses reveal the transcriptome characteristics of genes widely present in multiple species, new genes may also play key roles in developmental plasticity and evolution of novel body plans²⁶. In this study, “new genes” were defined as those that are mapped to the shallowest node on the phylogeny used in the phylostratigraphic analyses^{26,27}. For TAI analysis, these genes were assigned to ps11 in oyster and abalone, and ps 10 in sand worm. These new genes were found to have the highest expression levels in trochophore stage (Fig. 3), indicating the high degree of regulation of the activity of new genes during this stage. A phylostratigraphic analysis of oyster miRNAs also showed that new miRNAs were also highly expressed during the trochophore stage²⁸. One possible explanation is that new genes arise as a result of the adaptive change of the species in response to the environment²⁹. As a link between the gastrula and adult-like stages (the veliger in mollusks and the nectochaete in annelids), the trochophore may be an important stage determining the development of the adult body plan. Furthermore, the population dispersion of benthic marine animals mainly relies on swimming larvae³⁰. Thus, the larvae will encounter more complex environments, and the high expression of new genes during trochophore could be the result of both the need of environmental adaption and full development of adult-like organs at the larvae stage. Indeed, the development of juvenile structures in larvae is believed to be

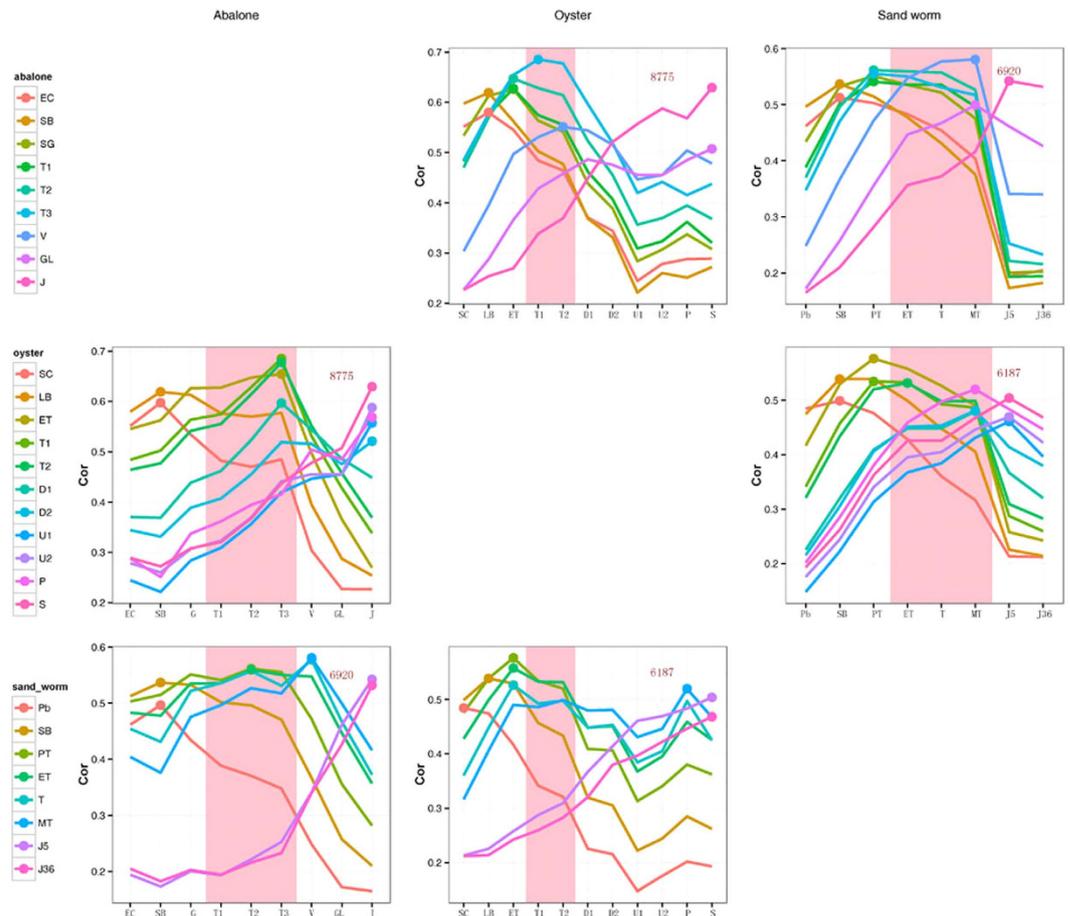


Figure 4. Transcriptome similarities of the different development stages between species. The Spearman correlation coefficients (ρ) of the transcriptome profiles of the different development stages were calculated in a pairwise manner between species. The x-axis indicates the development stages stated in the corresponding column header. The differently colored lines indicate the different development stages of the species stated in the corresponding row header. The pink area designates trochophore stage. The highest ρ value throughout the developmental stages is indicated by a point on the corresponding line. The numbers in each panel show the gene numbers used for the analysis.

one of the most important adaptive mechanisms for the rapid metamorphosis of marine invertebrate larvae³¹. These findings support the view that evolutionary processes in the stem of a phylum should be investigated by assessing the wide range of extant species within the phylum³², as expression patterns of individual species could give a biased picture.

Phylo-transcriptomic index used in this study (TAI) rely on the phylostratigraphic approach which maps genes on the consensus phylogeny based on the sequence similarity inference using BLAST^{10,27}. Two recent papers questioned reliability of the BLAST algorithm in phylostratigraphic analyses by claiming that in tracing remote homologs false negative BLAST error rate, which they estimated at 15%, could impact phylostratigraphic inference^{33,34}. Interestingly, a reanalysis of these simulation studies found unrealistic parameter choices, irreproducibility, statistical flaws and partial representation of results which all together suggest that BLAST is an appropriate and sufficiently sensitive tool in phylostratigraphic analysis³⁵. However, it is important to understand that phylostratigraphic approach is based on the model of punctuated evolution of protein families and is therefore designed to trace significant shifts in protein sequence, which are relevant for biological adaptations, and not remote homologs^{35–37}. Therefore, remote homolog BLAST error rates, regardless of their magnitude, are not directly relevant for TAI calculations which rely on gene age estimates according to the model of punctuated evolution of protein families³⁵.

The larval transition from the trochophore to the D-shaped veliger stage is probably the most dramatic event to occur during the development of oysters. During this transition, most of the organs form in approximately 10 hours. The footprint of this transition could be clearly seen at the expression levels when the detailed contributions of genes from all phylostrata were simultaneously depicted. In this context, it was also indicative that the expression of Bivalvia- (ps10) and oyster-specific genes (ps11) peaked at the trochophore stage, as well as the relative expression levels (REL) for each phylostratum experienced significant change during the trochophore

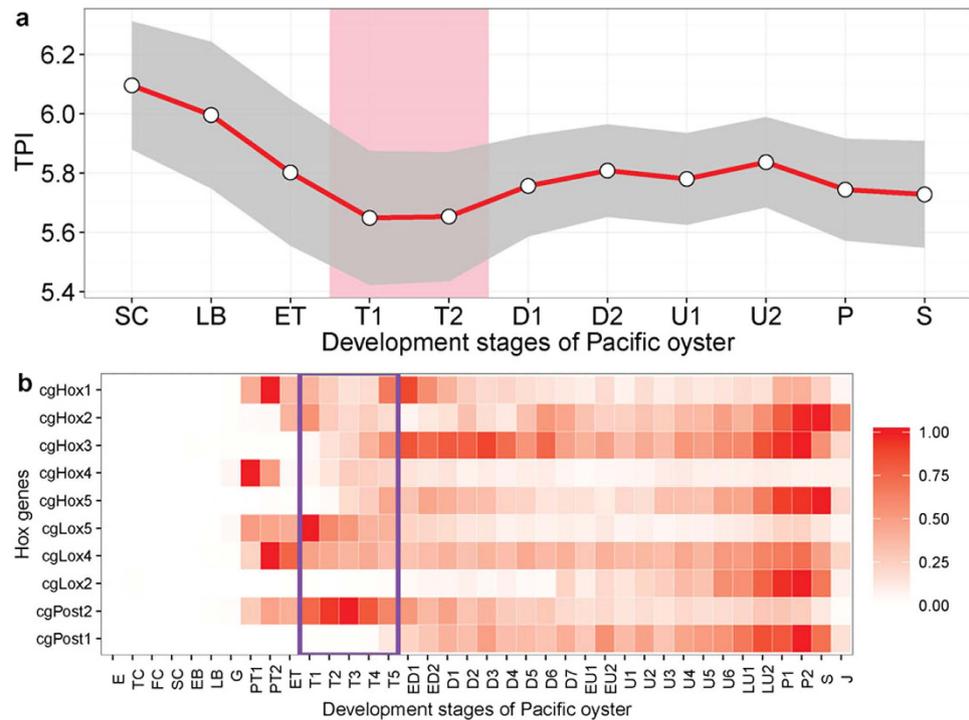


Figure 5. The transcriptome diversity indices over embryo stages (a), and the expression pattern of oyster Hox family genes (b). (a) The transcriptome diversity indices (TPI) were calculated based on the expression and the nucleotide diversity of the genes among 10 individuals. The area between pink lines designates the trochophore stage. The grey area designates the 95% confidence interval for TPI with a bootstrapping strategy. The overall pattern of the TPI profile is significant as measured by the permutation tests ($P = 5.3 \times 10^{-38}$). Pink shaded area designates trochophore stage. (b) RPKM Expression values were normalized between 0 and 1 for each gene across all stages. A value of 0 indicates no/lowest express and 1 indicates the stage where each gene was maximally expressed. Purple rectangle designates the trochophore stages.

stage (Fig. 3). However, more detailed studies on the gene functions are needed to bridge this general trend of the transcriptome and the real developmental mechanisms and morphological outcomes.

The expression waves of phylostrata-specific genes are less conclusive in relation to a long-standing debate about the origin of indirect development in marine invertebrates. According to the ‘larvae-first’ hypothesis, the original metazoan life cycle included pelagic adults, which were similar to the modern trochophore larvae⁴. An opposing view is presented by the ‘intercalation’ hypothesis, which proposes that the pelagic larvae evolved secondarily in different phyla and became intercalated multiple times into a life cycle that was originally direct^{5,38}. Our analysis shows elements of both scenarios. For instance, the exclusive wave of Metazoa-specific genes (ps4) during the trochophore stage suggests that this larvae relies on ancient processes in accordance with ‘larvae-first’ scenario. However, the high expression of phylum-, class- and species-specific genes (ps9, ps10 & 11) during the trochophore stage suggests massive recent adaptive elaborations that go in favor of the ‘intercalation’ idea. These elements may be the main cause of phylum level difference (Annelida vs. Mollusca) between the trochophore larvae. As marine larvae bear both development and population dispersal function their evolutionary history might be far more complex than the simple ‘intercalation’ vs. ‘early larvae’ hypotheses dichotomy could explain.

Methods

Sampling and developmental staging. The Pacific oyster *Crassostrea gigas* was sampled in two parallel experiments. The RNA-seq data from one replicate have previously been reported⁶. We designed the second replicate according to the main stages sampled previously³⁹. Briefly, the oysters used for the first experiment were the F₂ mass-mating progenies of 51 females and 1 male from the “G3” family, which was constructed from 1 farm-cultured female and 1 farm-cultured male that were collected from Yantai, China. As the sampling time intervals were short in this experiment, the transcriptome age indices of some adjacent stages were similar. We thus only chose a few typical stages in this study, which were the sixteen cells (SC), late blastula (LB), early trochophore (ET), two trochophore stages (T1 & T2), two D-shape stages (D1 & D2), two umbo larvae stages (U1 & U2), the pediveliger (P) and the spat (S) (see Supplementary Table S2). The oysters used for the second experiment were F₁ progenies produced by individuals from a wild population around Zhangzidao Island, Liaoning, China. Twenty females and one male were used. The embryos and larvae were cultured at 22 °C in seawater. The same developmental stages were sampled as in the first repeat (see Supplementary Table S3). The gamete collection, insemination, larval culturing, and sampling methods used for RNA extraction and electronic microscopy examination were the same as those employed in a previous report⁶.

The gonads of mature Pacific abalone *Haliotis discus hannai* were induced at the Laodong Aquaculture Breeding Company, Qingdao, China, via air exposure and using sand-filtered ultraviolet-irradiated seawater, as described previously⁴⁰. Each induced individual was then placed separately into a 10 L bucket for spawning. The eggs from ten females were inseminated with the sperm from ten males. The embryos and larvae were cultured at 22 °C in seawater using the previously described methods⁴⁰. Staging was conducted on-site under light microscopy according to a previous report⁴¹ and was confirmed through electronic microscopy examination (see Supplementary Fig. S5). The details for each of the sampled developmental stages are listed in Supplementary Table S4.

The gonads of mature and spawn-ready sand worms *Perinereis aibuhitensis* were collected from a sand worm culture farm at Rushan, Weihai, China. The matured worms were dry-transported at a low temperature to the aquarium of the Institute of Oceanology, Chinese Academy of Sciences. Approximately 30 worms were then placed into a bucket containing 30 L of 25 °C seawater. The worms were spawned soon after they were placed into the seawater. The mixture of eggs and sperm was collected every 10 min by transferring the worms into a new bucket. Thus, a total of three buckets of zygotes were produced with developmental age intervals of approximately 10 min. Developmental staging was conducted on-site through light microscopy according to a previous report^{42,43} and was confirmed via electronic microscopy examination of representative embryos (see Supplementary Fig. S6). The details for each of the sampled developmental stages are listed in Supplementary Table S5. As sand worm eggs have thick envelopes, the volume of the zygote is quite large. As a result, a normal 1.5-mL centrifuge tube can only be used to collect a few zygotes, and the thick envelope has a negative influence on RNA extraction. To solve this problem, we first removed the egg envelopes and then extracted RNA from fresh samples. Briefly, approximately 1.2 mL of an embryonic or larval sample was first mixed with 0.2 mL of guanidinium thiocyanate-phenol-chloroform (TRIzol, Invitrogen, Beijing, China). The samples were then collected immediately by centrifugation (8,000 × g) for 30 s. Four tubes of samples were treated at the same time to ensure sufficient quantity of RNA. The embryos or larvae were then released from the envelope and precipitated to the bottom of the centrifuge tube. After the envelope was removed, a portion of the retained embryos or larvae was sampled for electronic microscopy examination. The RNA of the remaining embryos or larvae was extracted using TRIzol according to the manufacturer's protocol.

To analyze the DNA polymorphism and variance of the oyster genes, we conducted genome re-sequencing in ten individuals. The oysters were collected from different sites in northern China, and DNA was isolated from the gills of each individual using standard molecular biology techniques.

Illumina-based RNA and DNA sequencing. Because there are no reference transcriptomes for Pacific abalone and sand worms, we conducted paired-end RNA-seq on the two species and assembled their transcriptomes. The transcriptome of the Pacific abalone was sequenced using the Illumina Genome Analyzer Iix platform at Biomarker Technologies CO., LTD, while that of the sand worm was sequenced using the Illumina HiSeq 2000 platform at BGI-Shenzhen. RNAs from each developmental stage were first mixed in equal quantities. Then, sequencing libraries were constructed according to the Illumina's recommended protocol.

Single-end RNA-seq was conducted to obtain the number of gene reads for each developmental stage of the three species. Single-end RNA sequencing data for the oyster and abalone samples was generated using the Illumina Cluster Station and Illumina Genome Analyzer Iix, while the sand worm was sequenced using the Illumina HiSeq 2000 platform. The sequencing libraries were constructed according to Illumina's recommended protocol.

A short insert library with 300-bp fragments was constructed for resequencing of the oyster genome using a standard protocol recommended by Illumina. The DNA was sequenced using the Illumina HiSeq 2000 platform.

Transcriptomic analysis and SNP calling. RNA-seq reads can be from any site within the total mRNA, in contrast to the fixed tag detection employed in microarrays. Thus, we needed to normalize the number of reads by dividing the reads by the length of the count region. This allows the normalized number of reads to be regarded as representative of the absolute number of gene transcript molecules. At the same time, the trimmed mean of M values (TMM) normalization was applied to adjust for the RNA composition effect⁴⁴. To date, most global gene expression analyses based on RNA-seq have assumed that the total mRNA abundance does not differ between different experimental samples. The problem with this assumption is that mRNA abundance is not equal, as expected. This assumption is irrelevant to TAI analysis, which only investigates how the partial concentrations of individual genes contribute to the overall expression levels in the transcriptome. However, the TAI value is sensitive to a few highly abundant mRNAs. The partial concentrations of transcripts of certain genes can account for more than 1% of the assayed transcriptome. When the expression level fluctuates across developmental stages, the TAI pattern is highly influenced by these extreme genes. We found that in the oyster, some genes showed very similar expression patterns (with Pearson's correlation coefficients greater than 0.99) and fluctuating expression curves during the course of ontogeny. These genes had short nucleotide sequences (186–465 bp) and extremely high expression levels in most samples (RPKM was as high as 120,000). These genes were therefore considered to be special genomic elements such as retrotransposons. In this case, we only retained one gene among each cluster. As a result, a total of 13 genes were discarded in this step (see Supplementary Fig. S7). Furthermore, some genes may be duplicated and showed similar nucleotide sequences, we thus filtered the redundant genes out of version 9 of the oyster genome (released in 2012)⁶. To construct the non-redundant gene set, BLAT⁴⁵ alignment searches were performed on the CDS and protein sets using the sets themselves as separate queries. The gene clusters were then constructed based on the similar sequences and were visualized using Cytosape⁴⁶. The following criteria were used to determine the similar genes: 1. the protein sequences shared at least 97% identity in the alignment region, while the DNA sequences shared at least 95% identity; and 2. there were no gaps in the aligned protein regions. As a result, 505 genes were discarded because their sequences were redundant (see Supplementary Fig. S8).

The filtering of these genes did not change the main TAI pattern compared with that using the whole predicted gene set and all the sampled development stages (see Supplementary Fig. S4), but reduced the high impact of few abundant short questionable genes.

For the Pacific abalone and sand worm, we assembled the transcriptomes with the Trinity program⁴⁷. The transcript abundances were estimated using RSEM software⁴⁸ based on the single-end read alignment obtained in Bowtie2⁴⁹. The transcript abundances for the oyster genes were calculated using a previously described method⁶. Briefly, the reads were mapped using Tophat⁵⁰, and the mapped reads for every gene were calculated using a Perl script developed in-house. The data for the hit reads can be found in Supplementary Table S6.

The sequenced reads from the genomes of ten individual oysters were mapped onto the oyster genome using Bowtie2⁴⁹. A combination of SAMtools and BCFtools⁵¹ was used in SNP calling and allelic genotyping. The diploid genome sequences from each individual were inferred using the genomic allele site information. Then the gene sequences for each individual were extracted according to the oyster gene model annotation, and were used as the input for the DnaSP software⁵² to calculate the value of theta per nucleotide site (θ , see Supplementary Table S7).

TAI and TPI. A database containing the complete genomes of 3,915 species was constructed (see Supplementary Table S8). The species were sorted into 11 phylostrata for the TAI analysis of the oysters and abalone and 10 phylostrata for the TAI analysis of the sand worm (see Supplementary Fig. S9). Blastp searches were conducted against the database using oyster proteins, while Blastx searches were conducted using the unigenes of abalones and sand worms.

The transcriptome age index of developmental stage s (TAIs) was calculated according to the method reported by Domazet-Lošo and Tautz¹⁰:

$$\text{TAIs} = \frac{\sum_{i=1}^n ps_i e_{is}}{\sum_{i=1}^n e_{is}} \quad (1)$$

where ps_i is the evolutionary age (phylostratum) of gene i ; e_{is} is the weighted read number of gene i at stage s , which is weighted by the CDS length of the gene; and n is the total number of genes analyzed.

To calculate the relative expression levels (REL) of the genes for a given phylostratum (ps) and developmental stage (s), we first measured the partial concentration (f_{is}) according to the formula $f_{is} = e_{is}/(e_1 + e_2 + \dots + e_n)$. Then, RE was calculated according to the following equation:

$$\text{RE}(ps)_s = \frac{\bar{f} - \bar{f}_{\min}}{\bar{f}_{\max} - \bar{f}_{\min}} \quad (2)$$

where f is the average partial concentration of genes from phylostratum ps for a given stage, s ; and \bar{f}_{\max} and \bar{f}_{\min} are the maximal and minimal average partial concentrations from phylostratum ps across all stages considered, respectively.

As oysters show a high degree of polymorphism, the SNP content varied for the different oyster genes. Some genes rarely exhibited SNP variations, indicating purifying selection during evolution. However, there was a high proportion of segregating (polymorphic) sites among some other genes. We calculated the θ for each gene and introduced the transcriptome polymorphism index (TPI), which was calculated according to the following equation:

$$\text{TPIs} = \frac{\sum_{i=1}^n \theta_i e_{is}}{\sum_{i=1}^n e_{is}} \times 10^3 \quad (3)$$

where θ_i is the theta per nucleotide site of gene i . When multiplying by 10^3 , the value is theta per kilo nucleotide sites.

Other analysis methods. The orthologs were determined between each pair of species based on a reciprocal best-hit Blast with an E-value cutoff of $1e-5$. Spearman's correlation coefficient (ρ) was used to evaluate transcriptome similarity. The statistical analyses and plots were performed and generated in R (<http://www.R-project.org/>)⁵³. A 1000 replication bootstrapping strategy was used to create a 95% confidence interval for TAI and TPI values with the “boot” R package. To determine the statistical significance of TAI and TPI pattern, permutation test was performed. The three components (phylostratum in TAI or theta value in TPI, gene length, and RPKM) of each gene were randomly permuted 1000 times within each developmental stage s . In this way, 1000 TAIs surrogate values were calculated for each development stage. These values were approximated by gamma distribution, and two parameters (shape and scale) were estimated by the method of moments. The fitted gamma distribution was treated as null distribution, and the P -value of observed TAIs was computed from the null distribution. Then the combined P -value for the TAI pattern was computed using the meta-analytic approach⁵⁴. TPI followed the same procedure.

References

1. Arenas-Mena, C. Indirect development, transdifferentiation and the macroregulatory evolution of metazoans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 653–669 (2010).
2. Dunn, C. W., Giribet, G., Edgecombe, G. D. & Hejnol, A. Animal phylogeny and its evolutionary implications. *Annu. Rev. Ecol. Evol. Syst.* **45**, 371–395 (2014).
3. Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526–531 (2013).

4. Nielsen, C. Life cycle evolution: was the eumetazoan ancestor a holopelagic, planktotrophic gastraea? *BMC Evol. Biol.* **13**, 1–18 (2013).
5. Raff, R. A. Origins of the other metazoan body plans: the evolution of larval forms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 1473–1479 (2008).
6. Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).
7. Paps, J., Xu, F., Zhang, G. & Holland, P. W. H. Reinforcing the egg-timer: Recruitment of novel Lophotrochozoa homeobox genes to early and late development in the Pacific oyster. *Genome Biol. Evol.* **7**, 677–688 (2015).
8. Slack, J. M. In *Keywords and Concepts in Evolutionary Developmental Biology* (eds Hall, B. K. & Olson, W. M.) 309–318 (Harvard Univ. Press, 2003).
9. Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* **2**, 248 (2011).
10. Domazet-Lošo, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010).
11. Kirschner, M. & Gerhart, J. Evolvability. *Proc. Nat. Acad. Sci. USA* **95**, 8420–8427 (1998).
12. Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo. *Dev. Cell* **22**, 1101–1108 (2012).
13. Quint, M. *et al.* A transcriptomic hourglass in plant embryogenesis. *Nature* **490**, 98–101 (2012).
14. Cheng, X., Hui, J. H. L., Lee, Y. Y., Wan Law, P. T. & Kwan, H. S. A “Developmental Hourglass” in Fungi. *Mol. Biol. Evol.* **32**, 1556–1566 (2015).
15. Shigeno, S., Takenori, S. & von Boletzky, S. In *Cephalopods - Present and Past* (eds Tanabe, K., Shigeta, Y., Sasaki, T. & Hirano, H.) 23–34 (Tokai University Press, 2010).
16. Levin, M. *et al.* The mid-developmental transition and the evolution of animal body plans. *Nature* **531**, 637–641 (2016).
17. Drost, H. G., Gabel, A., Grosse, I. & Quint, M. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol. Biol. Evol.* **32**, 1221–1231 (2015).
18. Drost, H. G., Gabel, A., Domazetlošo, T., Quint, M. & Grosse, I. Capturing Evolutionary Signatures in Transcriptomes with myTAI. *bioRxiv*, 051565 (2016).
19. Duboule, D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev. Suppl.* **120**, 135–142 (1994).
20. Slack, J., Holland, P. & Graham, C. Development of the zootype. *Nature* **363**, 307–308 (1993).
21. Irvine, S. M. & Martindale, M. Q. Cellular and molecular mechanisms of segmentation in annelids. *Semin. Cell Dev. Biol.* **7**, 593–604 (1996).
22. Kalinka, A. T. *et al.* Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010).
23. Henry, J. Q., Hejnol, A., Perry, K. J. & Martindale, M. Q. Homology of ciliary bands in Spiralian Trochophores. *Integr. Comp. Biol.* **47**, 865–871 (2007).
24. Nielsen, C. Trochophora larvae: Cell-lineages, ciliary bands, and body regions. 1. Annelida and Mollusca. *J. Exp. Zool. B Mol. Dev. Evol.* **302**, 35–68 (2004).
25. Romiguier, J. *et al.* Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261–263 (2014).
26. Kemkemer, C. & Long, M. New genes important for development. *EMBO Rep.* **15**, 460–461 (2014).
27. Tautz, D. & Domazet-Lošo, T. The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **12**, 692–702 (2011).
28. Xu, F. *et al.* Identification of Conserved and Novel MicroRNAs in the Pacific Oyster *Crassostrea gigas* by Deep Sequencing. *PLoS ONE* **9**, e104371 (2014).
29. Long, M., Betran, E., Thornton, K. & Wang, W. The origin of new genes: glimpses from the young and old. *Nature Rev. Genet.* **4**, 865–875 (2003).
30. Oyarzun, F. X. & Strathmann, R. R. Plasticity of Hatching and the Duration of Planktonic Development in Marine Invertebrates. *Integr. Comp. Biol.* **51**, 81–90 (2011).
31. Hadfield, M. G. Why and how marine-invertebrate larvae metamorphose so fast. *Semin. Cell Dev. Biol.* **11**, 437–443 (2000).
32. Hejnol, A. & Dunn, C. W. Animal Evolution: Are Phyla Real? *Curr Biol* **26**, R424–R426 (2016).
33. Moyers, B. A. & Zhang, J. Evaluating phylostratigraphic evidence for widespread de novo gene birth in genome evolution. *Mol. Biol. Evol.* **33**, 1245–1256 (2016).
34. Moyers, B. A. & Zhang, J. Phylostratigraphic bias creates spurious patterns of genome evolution. *Mol. Biol. Evol.* **32**, 258–267 (2015).
35. Domazet-Lošo, T. *et al.* No evidence for phylostratigraphic bias impacting inferences on patterns of gene emergence and evolution. *bioRxiv*, 060756 (2016).
36. Domazet-Lošo, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
37. Domazet-Lošo, T. & Tautz, D. Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC Bio.* **8**, 66 (2010).
38. Page, L. R. Molluscan larvae: Pelagic juveniles or slowly metamorphosing larvae? *Biol. Bull.* **216**, 216–225 (2009).
39. Du, Y. *et al.* Validation of housekeeping genes as internal controls for studying gene expression during Pacific oyster (*Crassostrea gigas*) development by quantitative real-time PCR. *Fish Shellfish Immun.* **34**, 939–945 (2013).
40. Deng, Y., Liu, X., Zhang, G. & Guo, X. Inbreeding depression and maternal effects on early performance of Pacific abalone. *N. Am. J. Aquac.* **67**, 231–236 (2005).
41. Seki, T. Biological studies on the seed production of the northern Japanese abalone, *Haliotis discus hannai* Ino. *Bull. Tohoku Natl. Fish. Res. Inst.* **59**, 1–71 (1997).
42. Liu, X., Jiang, M. & Zhang, Y. Study on the embryonic and larval development of *Perinereis aibuhitensis*. *Marine Fish. Res.* **26**, 13–17 (2005).
43. Fischer, A. H., Henrich, T. & Arendt, D. The normal development of *Platyneris dumerilii* (Nereididae, Annelida). *Front. Zool.* **7**, 1–39 (2010).
44. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, 79–82 (2010).
45. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
46. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431 (2011).
47. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
48. Li, B. & Dewey, C. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
50. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
51. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

52. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
53. Ihaka, R. & Gentleman, R. R: A language for data analysis and graphics. *J Comput. Graph. Stat.* **5**, 299–314 (1996).
54. Zaykin, D. V. Optimally weighted Z -test is a powerful method for combining probabilities in meta-analysis. *J. Evol. Biol.* **24**, 1836–1841 (2011).

Acknowledgements

The authors thank X. Guo, P. Yang, P. Holland, H. Qi and K. Song for discussions. We also thank Y. Zhou, D. Yang, F. Wu, M. Yang, W. Huang, and W. Wang for help with the sampling and RNA extraction. We acknowledge the Laodong Aquaculture Breeding Company for providing the facility for culturing the abalone. Most of the calculations were performed using the supercomputer cluster of the High Performance Computing Center (HPCC) at the Institute of Oceanology, Chinese Academy of Sciences. We acknowledge grant support from the National Natural Science Foundation of China (31402285), the National Basic Research Program of China (2010CB126401), the National Natural Science Foundation of China (31530079), The Scientific and Technological Innovation Project Financially Supported by Qingdao National Laboratory for Marine Science and Technology (No. 2015ASKJ02), the earmarked fund for Modern Agro-industry Technology Research System (CARS-48) and the Taishan Scholars Climbing Program of Shandong. T.D. funding from the City of Zagreb and Adris Foundation grants.

Author Contributions

F.X., X.F. and G.Z. conceived the study. F.X., T.D.-L., D.F. and X.F. analyzed the data. F.X. cultured the oysters, abalone and sand worms. L.L. provided the oysters and the genome re-sequencing data. F.X., T.D.-L. and T.L.D. wrote the paper.

Additional Information

Accession codes: The transcriptome, RNA-seq and genome re-sequencing raw data have been deposited with GenBank under BioProject SRP056090.

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Xu, F. *et al.* High expression of new genes in trochophore enlightening the ontogeny and evolution of trochozoans. *Sci. Rep.* **6**, 34664; doi: 10.1038/srep34664 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016