The Mediterranean mussel, *Mytilus galloprovincialis*, a novel model for developmental studies of mollusks

Angelica Miglioli[1$†], Marion Tredez[1$], Manon Boosten[1,2], Camille Sant[1,2], João E. Carvalho[1], Philippe Dru[1], Laura Canesi[3], Michael Schubert[1*], and Rémi Dumollard[1*]

1- Laboratoire de Biologie du Développement de Villefranche-sur-Mer (LBDV), Institut de la Mer de Villefranche (IMEV), Sorbonne Université, CNRS, Villefranche-sur-Mer, France

2- Laboratoire d'Océanologie de Villefranche (LOV), Institut de la Mer de Villefranche (IMEV), Sorbonne Université, CNRS, Villefranche-sur-Mer, France

3- Università degli Studi di Genova, Dipartimento di Scienze della Terra dell'Ambiente e della Vita (DISTAV), Genova, Italy

[$] Co-First authors

[*] Co-Last authors

[†] Corresponding author: angelica.miglioli@imev-mer.fr

Keywords: bivalve mollusks, embryonic and larval development, trochophore, shell field, transcriptome, hybridization chain reaction

Summary statement: Resources and techniques are described for the Mediterranean mussel *Mytilus galloprovincialis*, which, together, establish a novel model system for studying mollusk development and animal evolution.

Running title: A novel mollusk model system: the Mediterranean mussel

Abstract

A model organism in developmental biology is defined by its experimental amenability as well as by resources created for the model system by the scientific community. For the most powerful models, the combination of both has already yielded a thorough understanding of development. However, the number of developmental model systems is still very limited, and their phylogenetic distribution is heavily biased. Members of one of the largest animal phyla, the mollusks, for example, have long been neglected as developmental model organisms. To remedy this shortcoming, we produced a detailed developmental transcriptome for the Mediterranean mussel *Mytilus galloprovincialis*, a bivalve mollusk, and expanded the list of experimental protocols available for this species. Our high-quality transcriptome allowed us to identify transcriptomic signatures of developmental transitions and to perform a first comparison with the Pacific oyster *Crassostrea gigas* that can be used in future multi-species analyses. To allow co-labelling studies, we optimized protocols for immunohistochemistry and hybridization chain reaction and combined both techniques to create high-resolution co-expression maps of developmental genes. The resources and protocols we describe here thus represent an enormous boost for the establishment of the Mediterranean mussel as a laboratory model in developmental biology.

1. Introduction

The development of aquatic invertebrates has inspired the curiosity of biologists for well over a century (De Robertis and Tejeda-Muñoz, 2022; Hunt and Maslakova, 2017; Nielsen, 2005, 2009; Wang et al., 2020). Morphological similarities between larvae have, for example, been used as taxonomic characters and comparisons between larval types have yielded highly influential hypotheses on animal evolution (Carrier et al., 2018). One such larval type, the trochophore, is widely distributed amongst the lophotrochozoans and may have already been present in the last common ancestor of this large clade of bilaterian animals (De Robertis and Tejeda-Muñoz, 2022). However, despite decades of studies (Carrier et al., 2018; Carrillo-Baltodano et al., 2021; Nielsen, 2005; Page, 2009; Rawlinson, 2010; Wada et al., 2020; Xu et al., 2016; Yang et al., 2020), our understanding of the evolutionary origin of the trochophore larva as well as of the biological processes underpinning its ontogeny and transition to the adult body remains extremely fragmentary (Carrier et al., 2018; De Robertis and Tejeda-Muñoz, 2022; Wada et al., 2020; Wang et al., 2020; Wu et al., 2019; Xu et al., 2016; Yang et al., 2020). Although multiple genome and transcriptome sequencing projects have provided abundant data from diverse organisms, most of these datasets have yet to be exploited to provide insights into the overall conservation of trochophore development between different taxa (Liu et al., 2014; Sayers et al., 2021; Wada et al., 2020). A major limitation for comparisons seems to be that lophotrochozoan development is characterized by the deployment of species-specific gene sets. This fact makes the identification of comparable larval phases, even between closely related species of the same class, quite difficult (Wada et al., 2020; Wu et al., 2019; Xu et al., 2016; Yang et al., 2020).

Within the lophotrochozoans, the mollusks are the largest and anatomically most diverse phylum (Yang et al., 2020). Yet, as of today, the scientific community still lacks an amenable laboratory model for studying the mechanisms of mollusk development. Reasons include the limited availability of genomic and transcriptomic resources and the complexity of establishing even basic techniques, such as immunocytochemistry and *in situ* hybridization (Yang et al., 2020). It has been proposed that the overall fragility of mollusk larvae as well as the presence of a shell, at least in conchiferan (i.e., shell-bearing) mollusks (Kocot et al., 2016; Miglioli et al., 2019; Miglioli et al., 2021a; Sayers et al., 2021; Wada et al., 2020; Yang et al., 2020), have hindered efforts to devise more elaborate resources and techniques for representatives

of this phylum (Carrier et al., 2018; Sayers et al., 2021; Wada et al., 2020). Altogether, these issues have slowed research into the molecular aspects of trochophore development in mollusks, hence limiting our capacity for comparisons of trochophore development between mollusks and other taxa (Wada et al., 2020; Yang et al., 2020).

The Mediterranean mussel, *Mytilus galloprovincialis*, is an established model in environmental studies, and a species of great ecological and economic interest with a worldwide distribution (Brzozowska et al., 2012; Daguin and Borsa, 2000; FAO, 2022). Large-scale larval cultures of *M. galloprovincialis* are easily set up in the laboratory, following standardized culture protocols by the International Standardization Organization (ISO) and the American Society for Testing Materials (ASTM) (ASTM E724-21, 2021; ISO 17244:2015, 2020). *M. galloprovincialis* is also the first metazoan with a published pangenome (Gerdol et al., 2020), a feature of significant importance for researchers interested in genome evolution. The presence of detailed culture protocols and of a well-annotated genome make *M. galloprovincialis* a promising model organism for developmental studies. In addition, basic protocols for pharmacological treatments, quantification of shell biogenesis, quantitative polymerase chain reaction (qPCR) as well as immunohistochemistry and *in situ* hybridization have recently been described for developing *M. galloprovincialis* embryos and larvae (Miglioli et al., 2019, 2021a, 2021b).

In this work, we provide a compendium of updated protocols and novel resources to significantly facilitate access to and work with *M. galloprovincialis* embryos and larvae. We thus compiled a detailed developmental transcriptome and validated it with a standardized qPCR protocol. We further established a novel protocol for multi-color, fluorescent whole-mount *in situ* hybridization, using the hybridization chain reaction (HCR) approach (Choi et al., 2014), and successfully coupled this *in situ* hybridization approach with immunohistochemistry assays. The developmental transcriptome of *M. galloprovincialis* allowed us to define major developmental periods and transitions, results that were further confirmed in other bivalve mollusks and that provide a basis for future comparative developmental studies of lophotrochozoans. The novel *in situ* hybridization and immunohistochemistry protocols yielded detailed information on the expression of genes marking developmental transitions, for example, in the early shell field, apical sensory cells, peripheral mantle cells, and hindgut. Taken together, this work represents a valuable resource for future work with *M. galloprovincialis* embryos and larvae, setting the stage for

establishing this animal as a mollusk model system. As such, this work will have a significant impact on developmental biology and associated research fields.

2. Results

2.1. Morphological staging of *Mytilus galloprovincialis* early development and description of transcriptome libraries

The developmental transcriptome of *M. galloprovincialis* embryos and larvae consists of 15 developmental stages (*i.e.*, a total of 30 libraries from two independent biological replicates). Samples were obtained every 4 hours under standard culture conditions at 16°C (ASTM E724-21, 2021; ISO 17244:2015, 2020), from unfertilized egg to 52 hours post fertilization (hpf) with an additional sampling timepoint at 72 hpf (Fig. 1A). As previously described in other Mytilids, the *M. galloprovincialis* zygote undergoes holoblastic, unequal cleavage (Kurita et al., 2009). The micromeres form at the animal pole of the embryo at 4 hpf, and subsequent cleavage follows a spiral pattern (Kurita et al., 2009). By 12 hpf, the macromeres at the vegetal pole start to invaginate, indicating the start of gastrulation. At 16 hpf, the stomodeum (or presumptive mouth) has moved anteriorly, and a rosette of dorsal ectodermal cells forms an invagination, indicating the onset of shell field ontogeny (Kniprath, 1980, 1981; Miglioli et al., 2019). At 20 hpf, the columnar cells of the shell field thicken, creating the impression of a deeper invagination. At 24 hpf, the larvae have the stereotypical morphology of a trochophore, with a ventral stomodeum and a dorsal shell field that is now completing the evagination process and secreting the organic matrix of the larval shell (Kniprath, 1980; Miglioli et al., 2019). By 28 hpf, the hinge region of the shell field is completely flat. At 32 hpf, a cavity appears in the hinge region of the larva, and, at 36 hpf, it expands and seems to connect to the former stomodeum (that has now moved anteriorly) and to the anal cavity on the posterior side. At 40 hpf, the D-margin of the shell becomes clearly visible and only the velum (also called the prototroch) of the larva is protruding. By 44 hpf, the velum is enclosed in the D-shaped shell. Thereafter, no major morphological changes are observable, excepting the progressive outgrowth of the shell with respect to the mantle tissue along the D-border (Fig. 1A). In summary, our transcriptome is representative of all developmental stages of early bivalve development.

2.2. Transcriptome assembly and gene expression dynamics

More than 31 million paired-end clean reads per library were mapped on the published *M. galloprovincialis* genome (Gerdol et al., 2020), with an average input length of 200 bp (Supplementary Table 1). More than 85% of the reads were mapped, with 71% of them assigned to a unique locus (Supplementary Table 1). The output gene count matrix included 52,422 expressed genes (sum of gene raw counts in all libraries > 0) of a total of 60,302 entries in all libraries (Supplementary Table 2). An analysis of the number of genes expressed during development is presented in Fig. 1B. The minimum number of expressed genes (EGs) was found in unfertilized eggs (at 0 hpf), with around 25,000, while the highest was found at the late D-veliger stage (at 72 hpf), with around 40,000, indicating that at least 15,000 genes (that is about 25% of the genome) are transcriptionally modulated during early development of *M. galloprovincialis* (Fig. 1B). The number of EGs increased dramatically during the first 12 hours of development, with the two replicates yielding similar results. Thereafter, a slow but steady increase of EGs was observed until the last sampling point. A heatmap of the Euclidean distances between stages and a principal component analysis (PCA) of the normalized libraries validated sample reproducibility (Fig. 1C and 1D). The heatmap further revealed that the libraries clustered by developmental time and not by biological replicate (Fig. 1C). Library clustering yielded two main clusters, one before and one after 32 hpf, indicating a potential mid-development transcriptomic shift within the trochophore stage (Fig. 1C). The PCA showed a continuous transcriptomic progression from 0 to 72 hpf along the axis of the first principal component, indicating that the maximum variation is found between the 0 to 24 hpf and the 32 to 72 hpf libraries (Fig. 1D), which is consistent with a potential transcriptomic shift during early trochophore development. A majority of the variance was explained by the first principal component (78%), which is mainly correlated with developmental timing, lending support to the notion that morphological changes during development are driven by transcriptomic changes.

2.3. Differential expression analysis and developmental clustering

Differential expression analysis was performed by pairwise comparison of samples of consecutive stages. The analysis identified 11,996 differentially expressed genes (DEGs) during *M. galloprovincialis* development (Supplementary File 2 and Supplementary Table 3). Hierarchical clustering of the libraries based on DEG expression values identified statistically distinct developmental phases (adjusted-unbiased p-values, au > 0.95) (Fig. 2A) and divided

the 30 libraries in two main clusters: one regrouping the libraries from 0 to 16 hpf (embryonic phase) and one regrouping the libraries from 20 to 72 hpf (larval phase). In the embryonic phase of development, the 0 and 4 hpf libraries were clearly distinct from the ones covering 8 to 16 hpf, corresponding to, respectively, zygote and gastrulation (Fig. 2A). In the larval phase of development, three sub-clusters were observable. The first one was formed by libraries corresponding to three trochophore stages, from 20 to 28 hpf, with three older trochophore-like larvae (advanced trochophore), from 32 to 40 hpf, establishing a second cluster. The third cluster grouped the libraries of the four stages with D-veliger morphology, from 44 to 72 hpf. The number of DEGs varied as development proceeded, with an overall decreasing trend (Fig. 2B). About 55% of DEGs were found to be exclusively up-regulated, 25% exclusively down-regulated, and the remaining 20% alternated between up- and down-regulation. Of note, 40% of DEGs were differentially expressed more than once during development (Fig. 2B, insert). More than 1,500 DEGs were already detected at 4 hpf, indicating an early initiation of zygotic gene expression. The highest number of DEGs was reached at 8 hpf, and it was followed by a progressive decrease as development proceeded. Following the differential expression analysis, DEGs were grouped as a function of their most recurrent expression profiles (Fig. 2C and 2D, Supplementary Tables 4 and 5). The analysis identified 56 clusters with more than 15 genes and of the top 6 clusters, based on the number of DEGs, we found 5 to consist of genes with low maternal (0 hpf) expression levels. This indicates that the DEGs we identified are predominantly derived from zygotic expression (Fig. 2C). An expression level-based association of gene clusters with developmental periods (Supplementary Table 4) showed that 31% of the clusters reached their highest expression levels at the zygote, 32% at the gastrula, 14% at the trochophore, 7% at the advanced trochophore, and 25% at the D-veliger stage (Fig. 2D). These results indicate that zygote and D-veliger stages are characterized by genes with a higher diversity of expression profiles than the trochophore and advanced trochophore stages. Moreover, the analysis indicated that expression levels of individual clusters are generally skewed towards earlier developmental periods and that cluster expression levels might be correlated with key morphological events during development of both embryo and larva.

2.4. qPCR and *in situ* hybridization of selected genes

The differential expression analyses were corroborated by qPCR experiments carried out with material from both biological replicates. Four target genes were selected based on their transcriptomic expression profiles. The *Foxb2* (MGAL_10B093191) and *Wnt8a* (MGAL_10B085403) genes were selected as members of one of the top 6 clusters (Fig. 2C, Group 6), characterized, respectively, by high and low mean expression levels. *SLC46A3* (MGAL_10B020966) and *SYTL4* (MGAL_10B044489) were selected because they showed a highly dynamic expression profile (Supplementary Tables 3 and 4). For each one of the four genes, the qPCR experiments validated both the levels and the temporal patterns of their expression (Fig. 2E). Given their shared temporal expression patterns, we further assessed the spatiotemporal expression of *Foxb2* and *Wnt8a* in relation to each other and relative to the immunoreactivity of serotonin (5-HT), a marker of mussel nervous system cells. To do so, we established a fluorescent double *in situ* hybridization protocol based on the hybridization chain reaction (HCR) protocol for *M. galloprovincialis* and combined it with an immunohistochemistry assay for serotonin (Fig. 2E and Supplementary Figure 1). Co-labeling with serotonin allowed a standardized orientation of embryos and larvae, which significantly facilitated the interpretation of the *in situ* hybridization signals. We thus found that, while *Wnt8a* was detectable in a discrete group of cells that will give rise to one of the posterior neural ganglions of the larva, *Foxb2* was detectable in the stomodeal and trochal area. These observations were consistent with the expression patterns described for these two genes in other lophotrochozoan species (Lartillot et al., 2002; Marlow et al., 2014; Tomer et al., 2010). In sum, qPCR analysis and *in situ* hybridization confirmed the developmental trends identified by our transcriptome analyses.

2.5. Qualitative analysis of differentially expressed genes

We next performed a qualitative analysis of the DEGs based on functional domains and gene ontologies (GOs) (Fig. 3). DEGs were screened for transcription factor domains (Fig. 3A), domains known to be involved in shell formation (Fig. 3B) as well as for receptors, ion channels, and neurotransmitter transporters (Fig. 3C). In total, 711 transcription factors with 65 different domains were annotated, with the most recurrent (> 20 genes) being: zinc fingers, homeodomains, helix-loop-helix and fork head domains (Fig. 3A, Supplementary Table 6). With respect to shell formation, 779 genes were found to contain domains associated with shell matrix proteins and calcification (Fig. 3B) (Ramesh et al., 2019). The most recurrent

domains included calcium binding proteins (EF-hand and calponin), von Willebrand factors (WKA), extracellular matrix proteins (fibronectin), carbohydrate binding proteins (lectin and concanavalin), chitin binding proteins (CB), ion channels, and transporters (Fig. 3B, Supplementary Table 5). Regarding receptors, 225 DEGs were annotated as G protein-coupled receptors of the rhodopsin and secretin families, 92 as ephrin-A receptors, 23 belonged to the nuclear receptor superfamily, and other DEGs included scavenger, low density (LD) lipoprotein, and TGFβII receptors (Fig. 3C). As many as 125 DEGs were predicted to be involved in neurotransmitter transport and reception (Fig. 3C, Supplementary Table 3). A gene ontology enrichment analysis was subsequently performed on the DEGs to identify stage-enriched biological processes (Fig. 3D). The GO enrichment analysis for biological processes identified 6 GO clusters: transmembrane transport, actin polymerization, cell adhesion, receptor-mediated signal transduction, biosynthesis, and transcription-coupled repair. While most biological processes did not exhibit a clear association with a specific developmental phase, transmembrane transport and receptor-mediated signal transduction were enriched in the trochophore and advanced trochophore stages (20 to 28 hpf and 32 to 40 hpf). Actin polymerization and transcription-coupled repair were weakly enriched in embryonic development (zygote and gastrula), and biosynthesis in the larval phase (from the early trochophore onward). Altogether, the results of the quantitative analysis confirmed the relevance of the identified DEGs with respect to the major regulatory (*i.e.*, transcription factors) and ontogenetic (*i.e.*, shell matrix proteins) programs controlling early development of bivalve mollusks.

2.6. Gene clustering and transcriptomic signatures of early developmental transitions

Soft clustering of expressed gene was performed to reveal potential transcriptomic dynamics during development. The analysis identified nine soft clusters (Fig. 4A), which were named according to the developmental staging system described above (Fig. 1A). While the embryonic gene clusters nicely recovered the hierarchical clusters of the embryonic phase, the larval soft clusters predominantly highlighted stages between the developmental phases identified by hierarchical clustering, thereby revealing potential developmental transitions. The early trochophore gene cluster matched the hierarchical clusters transitioning from the gastrula to the trochophore (16 to 20 hpf, Transition 1), the advanced trochophore gene clusters matched the hierarchical clusters transitioning from trochophore to advanced

trochophore (28 to 32 hpf, Transition 2), and the veliger gene cluster matched the hierarchical clusters transitioning from advanced trochophore to D-veliger (40 to 44 hpf, Transition 3) (Fig. 4A, Supplementary Table 7). To identify representative genes for developmental Transitions 1, 2, and 3, we isolated the genes of the three soft clusters that were differentially expressed at the corresponding transition stages (Supplementary Table 8). We found 98, 11, and 8 genes for, respectively, Transition 1, 2, and 3 (Fig. 4B, Supplementary Table 3). To assess whether similar transition signatures exist in other bivalves, we searched for homologous genes in the Pacific oyster *Crassostrea gigas* and analyzed their developmental expression using publicly available transcriptomic data (Fig. 4C, Supplementary Table 8) (Liu et al., 2021). The results of these comparisons indicated that, of the 98 Transition 1 genes in *M. galloprovincialis*, 22 were characterized by conspicuous expression at gastrula and trochophore 1 stages in *C. gigas*. However, most of these genes were also highly expressed at earlier or later developmental stages (Fig. 4C). Of the 11 *M. galloprovincialis* Transition 2 genes, we identified two characterized by high expression between the corresponding stages in *C. gigas*, *i.e.*, trochophore 3 and trochophore 4 (Fig. 4C). For Transition 3, marked by 8 genes in *M. galloprovincialis*, we found that three genes in *C. gigas* were characterized by a peak of expression at the passage from advanced trochophore (trochophore 5) to early veliger (early D-veliger 1) (Fig. 4C). Our unbiased cluster analysis thus revealed three developmental transitions during early *M. galloprovincialis* development that are impossible to identify morphologically and that are at least partially conserved in other bivalves.

2.7. Tissue localization of conserved transition genes by *in situ* hybridization

To shed light on the potential developmental functions of the conserved transition genes, we assessed the expression of a set of these genes by *in situ* hybridization using our HCR protocol. Genes were selected based on their differential expression in *M. galloprovincialis* (log fold change > 2) and their expression profiles in *C. gigas* (z-score value becoming positive) at the timepoints of the respective transition (Table 1, Supplementary Figure 2, Supplementary File 2). The genes were named according to the results of BLAST searches with their predicted proteins of the Uniprot and NCBI databases (Table 1, Supplementary Table 3). As Transition 1 contained a significantly higher number of conserved genes, we selected 8 genes for Transition 1 and two genes each for Transitions 2 and 3. Expression of these genes was analyzed at their corresponding transition stages. At 20 hpf, the transcription

factor genes *Tbx2b* and *Erg* were detectable in cells surrounding the stomodeum and in the apical tuft, with *Tbx2b* also being expressed in the hindgut (Fig. 5A). The other 6 assayed genes of Transition 1 were all expressed in or in association with the shell field: *HPP*, *TYRO2*, and *Pxdn* were detectable in groups of cells surrounding the invaginating shell field at 16 hpf, with *TYRO2* maintaining this expression at 20 hpf, *Dmbt1* in column-like cells at 16 and 20 hpf, *NA03* in islets of cells that elongated from invaginated tissue at 20 hpf, and *SLC4A4* in layers of cells covering the invaginated tissue at 20 hpf (Fig. 5A). For the conserved Transition 2 genes, *SLC26A5* was detectable in two symmetric groups of ectodermal cells at 28 hpf (Fig. 5B). At that stage, *Grin3a* was weakly expressed throughout the post-trochal larva, with a small group of anterior cells beneath the apical tuft expressing the gene more conspicuously (Fig. 5B). The two selected genes of Transition 3 were both expressed symmetrically with respect to the hinge axis at 44hpf. Thus, while *EF-hand* was expressed in the velum, the esophagus, and in a dense group of cells close to the anus (Fig. 5C), *COLVI* was detectable in individual cells along the D-border of the mantle (Fig. 5C).

	***Mg*ID**	***Cg*ID**	***Mg* annotation**
Transition 1	MGAL_10B029885	CGI_10016326	T-box transcription factor 2b (*Tbx2b*)
	MGAL_10B094294	CGI_10003121	ETS-related gene (*Erg*)
	MGAL_10B037868	CGI_10025712	Not annotated 03 (*NA03)*
	MGAL_10B039707	CGI_10009468	Tyrosinase 2 (*TYRO2*)
	MGAL_10B009273	CGI_10027254	Solute carrier family 4 member 4 (*SLC4A4*)
	MGAL_10B027895	CGI_10026985	Deleted in malignant brain tumors 1 (*Dmbt1*)
	MGAL_10B078802	CGI_10013987	Peroxidasin (*Pxdn*)
	MGAL_10B078617	CGI_10021520	Hypothetical predicted protein (*HPP)*
Transition 2	MGAL_10B055887	CGI_10015018	Solute carrier family 26 member 5 (*SLC26A5*)
	MGAL_10B039344	CGI_10021966	Glutamate ionotropic receptor NMDA type subunit 3A (*Grin3a*)
Transition 3	MGAL_10B028954	CGI_10015228	*EF-hand*
	MGAL_10B070218	CGI_10024466	Collagen VI-like (*COLVI*)

Table 1. List of conserved transition genes in *Mytilus galloprovincialis* (*Mg*ID) and *Crassostrea gigas* (*Cg*ID) selected for *in situ* hybridization in *M. galloprovincialis*.

3. Discussion

3.1. Fine-grained transcriptomic approaches combined with thorough co-labelling techniques provide valuable insights into bivalve mollusk development

The developmental transcriptome of *M. galloprovincialis* presented in this study is currently unmatched in terms of sampling depth and detail with respect to the embryonic and larval stages represented in the dataset (Fig. 1). The sampling schedule in 4-hour intervals prevented biased sampling and overrepresentation of stereotypical stages (such as gastrula, trochophore or veliger), leading to the generation of a highly consistent dataset between biological replicates whose gene expression dynamics are dictated by developmental timing (Fig. 1B). The results of our differentially expression analysis, indicating a decrease in the number of DEGs as development proceeds, is seemingly in contrast with the findings of previous studies (Foulon et al., 2019; Moreira et al., 2018; Niu et al., 2016; Núñez-Acuña et al., 2022; Song et al., 2016a). However, it is important to note that differences in sampling strategies (*i.e.*, based on morphological criteria *versus* fixed time intervals) and settings of contrasts and thresholds in pairwise comparisons can yield different results. Our qPCR experiments corroborated the results obtained by our computational analyses, hence demonstrating the usefulness of our dataset for illustrating gene expression dynamics during early *M. galloprovincialis* development.

The representativeness of our transcriptome in the context of early bivalve development is further supported by the results of the qualitative analysis of the functional domains of DEGs. Transcription factors known to act in the regulation of metazoan development (*e.g.*, homeobox and fork head genes) and cohorts of genes involved in early shell formation were highly represented among the differentially expressed genes (Guo et al., 2023; Lartillot et al., 2002; Morino et al., 2013; Paps et al., 2015; Ramesh et al., 2019; Setiamarga et al., 2013; Zhao et al., 2018; Zheng et al., 2019). In addition, the analysis revealed a strong enrichment of receptors, transporters, and channels for hormones and neurotransmitters, which is consistent with previous analyses in bivalve mollusks implicating these mediators of intercellular signaling in the regulation of larval shell biogenesis, neurogenesis, and larval morphogenesis (Guo et al., 2023; Liu et al., 2020; Miglioli, 2022; Miglioli et al., 2021b; Song et al., 2016a). The fact that transmembrane transport and receptor-mediated signal transduction were highly significant biological processes in the *M. galloprovincialis* trochophore and

advanced trochophore lends further support to this notion (Guo et al., 2023; Ramesh et al., 2019; Zhao et al., 2018; Zheng et al., 2019).

To correlate the transcriptional and developmental programs controlling early bivalve development, we set up a high-resolution, pan-larval protocol for multicolor *in situ* hybridization coupled with immunostaining in *M. galloprovincialis*. In combination with the developmental transcriptome, this protocol allows the parallel mapping of the expression dynamics of multiple genes in the context of cell type-specific markers. Using this approach, we were able to demonstrate the developmental expression and co-localization, in *M. galloprovincialis* embryos and larvae, of *Wnt8a* and *Foxb2,* which are key mediators of animal development, using immunoreactivity for the neurotransmitter serotonin (5HT) as a marker of the apical organ (Lartillot et al., 2002; Marlow et al., 2014; Miglioli et al., 2021a; Tomer et al., 2010). The combination of transcriptomic data and elaborate fluorescent whole-mount staining protocols has so far not been used for studying early development of bivalve mollusks and thus constitutes a major step towards the establishment of *M. galloprovincialis* as a model system.

3.2. Gene expression analyses suggest that the trochophore is a biphasic larval stage

Hierarchical clustering of libraries was carried out to identify developmental phases based on DEG dynamics (Leclère et al., 2019; Suzuki and Shimodaira, 2006). The results identified the main developmental phases of bivalve development (*i.e.*, zygote, gastrula, trochophore, and D-veliger) plus an additional phase, the advanced trochophore (Fig. 2A). In addition, a principal component analysis and a heatmap of Euclidean distances revealed a transcriptomic shift during the *M. galloprovincialis* trochophore stages, which is consistent with the notion that the trochophore is composed of two distinct developmental phases (Fig. 1). The existence of an advanced (or late) larval stage intercalated between the trochophore and the D-veliger has previously been proposed, but our results are the first to demonstrate the existence of this developmental phase on a transcriptomic level (Kamenev et al., 2018; Miglioli, 2022; Vogeler et al., 2016). Given that the morphotypes of the trochophore and of the advanced trochophore are very similar, our results indicate that developmental staging of bivalve mollusks by visual observation needs to be carried out with great care to avoid erroneous stage assignments. The existence of an additional larval stage linking the trochophore and the competent larva in mollusks, and possibly other lophotrochozoans, is an

important discovery for evolutionary studies into the origin of the trochophore larva that sounds a note of caution for generalized comparisons of trochophores from distantly related species.

3.3. Identification of transcriptional transitions during bivalve mollusk development, with potential implications for shell formation and larval morphogenesis

Soft clustering revealed transcriptomic signatures overlapping the main developmental phases identified by hierarchical clustering. These signatures marked three developmental transitions: from gastrula to trochophore (Transition 1), from trochophore to advanced trochophore (Transition 2), and from advanced trochophore to D-veliger (Transition 3). Taking advantage of our fine-grained developmental sampling, the combined DEG and soft clustering analyses identified highly representative genes for each of the three transitions: 98 genes for Transition 1, 11 genes for Transition 2, and 8 genes for Transitions 3. The high number of Transition 1 genes marks the initiation of gene regulatory networks responsible for the ontogeny of the trochophore larva that includes developmental processes, such as neurogenesis, myogenesis, hematopoiesis, gut formation, and shell secretion (Arenas-Mena, 2013; Dyachuk and Odintsova, 2009; Dyrynda et al., 1997; Kniazkina and Dyachuk, 2022; Miglioli et al., 2019, 2021a; Song et al., 2016b; Yurchenko et al., 2018). In contrast, Transition 2 and Transition 3 are characterized by significantly lower numbers of genes, indicating that the transitions, respectively, of the trochophore to the advanced trochophore and of the advanced trochophore to the D-veliger transitions might require less dramatic transcriptomic changes.

To evaluate the potential evolutionary conservation of these three transitions within mollusks, we carried out a cross-species comparison with the Pacific oyster *C. gigas* (Liu et al., 2021). This comparison revealed that, while homologous sets of genes clearly marked Transition 1 (gastrula to trochophore) in the *C. gigas* developmental transcriptome, the conservation was much less apparent for Transition 2 (trochophore to advanced trochophore) and Transition 3 (advanced trochophore to D-veliger). It might be that inconsistent sampling designs of the two datasets blurred the transcriptional signatures of the transitions in *C. gigas*. Alternatively, the relative lack of conserved genes for Transition 2 and Transition 3 might be symptomatic of a global decrease of homologous transcriptomic traits during trochophore development of bivalve mollusks. In fact, phylostratigraphic analyses have revealed that the late trochophore

stages of bivalves are characterized by the youngest transcriptome, which is largely derived from a class-specific phylostratum (Wu et al., 2019; Xu et al., 2016). Accordingly, specific adaptations of the larval body plan of bivalves, such as the shell field, originate from developmental programs initiated during the trochophore stage (Kemkemer and Long, 2014; Salamanca-Díaz et al., 2022; Xu et al., 2016).

To infer, which developmental processes might be controlled by conserved transition genes, we studied the expression of a subset of these genes in *M. galloprovincialis* embryos and larvae by *in situ* hybridization. For Transition 1 (16 to 20 hpf), we characterized the expression of 8 genes conserved between *M. galloprovincialis* and *C. gigas* (*Erg*, *Tbx2b*, *Dmbt1*, *Pxdn*, *SLC4A4*, *TYRO2*, *NA03*, and *HPP*) and found evidence for their involvement in developmental processes defining the transition from embryo to larva: ectomesoderm formation, dorsal specification, and shell field ontogeny. *Erg* and *Tbx2b* have respectively been described as pivotal regulators of ectomesoderm formation and dorsal specification in other lophotrochozoans, and the tissue localization in *M. galloprovincialis* early larvae is suggestive for a similar developmental role (Arenas-Mena, 2013; Osborne et al., 2018). With respect to shell field ontogeny, *Pxdn* and *TYRO2* are well known regulators of molluscan shell biogenesis, *SLC4A4* is an important player in shell calcification, and *Dmbt1* might be involved in the terminal differentiation of shell field columnar cells (Miglioli et al., 2019; Ramesh et al., 2019; Takito and Al-Awqati, 2004). The remaining two conserved genes of Transition 1 (*NA03* and *HPP*) could not be annotated, although similar sequences exist in other mollusks. However, we found expression of both genes in *M. galloprovincialis* to be associated with invaginating shell field tissue, which is suggestive of a potential role in larval shell biogenesis.

For Transition 2 (28 to 32 hpf), we selected two genes conserved between *M. galloprovincialis* and *C. gigas*. Consistent with the notion that the period between 28 and 32 hpf marks the completion of a fully calcified larval shell in *M. galloprovincialis* (Miglioli et al., 2019), we found *SLC26A5*, a gene encoding a bicarbonate transporter that provides substrate for shell calcification (Ramesh et al., 2019), expressed in two symmetric groups of ectodermal cells at Transition 2. The second gene, *Grin3a*, encodes the NMDA-type receptor subunit NR3, and, at Transition 2, it was most conspicuously expressed in a group of cells close to the apical tuft. NMDA subunits NR1 and NR2 have recently been characterized in nerves and ganglions of *C. gigas,* but the functions of NMDA-type receptors during bivalve trochophore development have yet to be assessed (Vogeler et al., 2021).

Transition 3 (40 to 44 hpf) marks the internalization of the velum into the D-shaped shell, which requires a functional muscular system (Dyachuk and Odintsova, 2009). The first of the two conserved transition genes we selected for Transition 3, *EF-hand*, has been linked to myosin regulation in scallop muscle fibers, and the expression in *M. galloprovincialis* D-veligers shows similarities to the ventral distribution of paramyosin immunoreactivity in D-veliger larvae of the mussel *M. trossulus* (Dyachuk and Odintsova, 2009; Fromherz and Szent-Györgyi, 1995). The second gene, *COLVI*, is expressed in mantle cells of *M. galloprovincialis*, on both sides of the D-border of the larval shell, which is consistent with a function in nacre formation, as has previously been suggested for the *COLVI* homolog of the pearl oyster *Pinctada martensii* (Zheng et al., 2017).

Taken together, the *in situ* hybridization-based expression analysis of conserved transition genes highlighted developmental processes that might be regulated by homologous sets of genes in *M. galloprovincialis* and *C. gigas*. We thus identified genes with possible functions in ectomesoderm formation, dorsal specification, muscle development as well as shell ontogeny, secretion, and calcification, whose expression is likely characteristic of bivalve mollusk development. We hypothesize that the transcriptomic staging approach we used for the characterization of *M. galloprovincialis* development can be applied to other species (bivalves, mollusks, lophotrochozoans) and used in cross-species comparisons for (1) the definition of developmental stages at a transcriptomic level, (2) the identification of major developmental transitions in transcriptome datasets, and (3) the comparison of homologous gene sets characteristic of a given developmental phase or transition.

3.4. The Mediterranean mussel, *Mytilus galloprovincialis*, a novel model for developmental studies of mollusks

The results reported in this study represent the first thorough transcriptomic analysis of the development of the Mediterranean mussel *M. galloprovincialis*. It further presents a collection of protocols essential for the validation and exploitation of transcriptomic data, including experimental tools, such as qPCR, immunohistochemistry, and *in situ* hybridization. The combination of fluorescent immunohistochemistry and *in situ* hybridization is particularly powerful for assessing the co-localization of crucial developmental factors in mussel embryos and larvae. Potential applications for this transcriptome, and associated experimental protocols, include developmental studies, cross-species comparisons, and ecotoxicological

surveys. Taking advantage of the developmental transcriptome and the novel protocols, we are the first to document the expression of a suite of genes in *M. galloprovincialis* that characterize bivalve mollusk development, revealing the full potential of this animal system for becoming a reference for developmental studies using mollusk models.

4. Materials and Methods

4.1. Spawning, fertilization, and rearing of embryos and larvae

Sexually mature *M. galloprovincialis* adults were harvested from the natural population in the bay of Villefranche-sur-Mer, France (43.682°N, 7.319°E) during the spawning seasons of 2021, 2022, and 2023 (January through April). Animals were acclimated to and maintained in laboratory conditions, as previously described (Miglioli et al., 2021b). Thirty adult mussels were exposed to a heat shock to induce gamete release. Briefly, epibionts, dirt, and algae were scrubbed off the shells under running filtered natural seawater. Byssus and threads were carefully cut off. Clean mussels were placed on ice for 10 minutes and then immersed, in individual 200 mL containers and on a rocking shaker, in filtered natural seawater kept at 28°C until gamete release. Spawning mussels were immediately moved to filtered natural seawater at 16°C. Prior to fertilization, gametes were collected, washed, and maintained in Millipore-filtered natural seawater (using a 0.2 µm Millipore filter). Fertilization was carried out as previously described with a 1:10 ratio of eggs to sperm (Miglioli et al., 2021a, 2021b). After 30 minutes, the fertilization success rate (*i.e.*, the ratio of the number of fertilized eggs over the total number eggs x 100) was established by microscopic observation. Parental pairs with a high fertilization success rate (> 90%) were selected for the following experimental manipulations. Two experimental replicates were set up, each consisting of a homogeneous pool of fertilized eggs from two independent parental pairs. Embryos from each replicate were distributed into 200 mL flasks for suspended cultures and brought to a density of 100 embryos per mL with Millipore-filtered natural seawater. A total of 14 flasks per replicate were prepared, corresponding to the pre-defined sampling strategy: in addition to sampling before fertilization, sample collections were carried out every 4 hours from 4 hpf to 52 hpf, with an additional sampling timepoint at 72 hpf. Embryo cultures were maintained at 16°C throughout the experiment. Samples for immunohistochemistry and *in situ* hybridization were produced following the same culture protocol. After filtration, larvae were fixed in 4% paraformaldehyde

in 1x PBS overnight at 4°C, washed 3 times for 15 minutes in 1x PBS with 0.01% Tween 20, and then stored in 100% methanol at -20°C (Miglioli et al., 2021b).

4.2. Sampling, total RNA extraction, and high-throughput sequencing

RNA extraction and purification were carried out with the RNAqueous-Micro Kit (Invitrogen) and the RNeasy MinElute Cleanup Kit (Qiagen) following published protocols (Leclère et al., 2019). Unfertilized eggs were directly collected from spawning females, and embryos and larvae were sampled with a 60 µm mesh filter and concentrated in RNAse free 1.5 mL tubes. The tubes were then placed on ice and gently centrifuged multiple times to remove the seawater from the samples. The pellet was immediately resuspended in lysis buffer, homogenized by energetic pipetting, and flash-frozen in liquid nitrogen. For each sample, 10 µL of the 1.5 mL filtrate were placed in a separate tube and fixed overnight in 4% paraformaldehyde in 1x PBS at 4°C (Miglioli et al., 2021a). The fixed larvae were stained with Hoechst (Invitrogen) and imaged with a SP8 confocal microscope (Leica Microsystems) to validate proper developmental progression (Miglioli et al., 2021a). Lysed samples were stored at -80°C until RNA extraction and purification (Leclère et al., 2019). RNA concentration and quality was assessed with a Nanodrop (Thermo Scientific) and an Agilent 2100 Bioanalyzer. Samples were shipped on dry ice to BGI (Hong Kong) for high-throughput sequencing. At least thirty million clean paired-end reads of 200 bp each were produced using the BGISEQ-500 Platform. Read cleaning and trimming was performed by BGI (Hong Kong). After sequencing, the raw reads were filtered. Data filtering included removal of adapter sequences, contaminations, and low-quality reads from the raw data. The entire read was deleted, if more than 28% matched the adapter sequence, if more than 40% had a quality value lower than 20 or if there were more than 3% of unidentifiable nucleotides in the read. This step was completed by the SOAP nuke software developed by BGI. Libraries were not subjected to further trimming (Dobin et al., 2013).

4.3. Genome-guided transcriptome assembly and differential expression analysis

Sequencing quality of the libraries was assessed with FastQC (v. 0.11.7) (Andrews, 2010). The libraries were mapped on the indexed *M. galloprovincialis* reference genome with STAR (v. 2.7.10a) using default parameters (Dobin et al., 2013; Gerdol et al., 2020). Read counts were obtained using StringTie (v2.2.1) in counting mode with the genome annotation file (Pertea et al., 2015, 2016). The resulting gene count matrix was exported using prepDE.py

and processed in Rstudio (v. 4.2.2) with the DESeq2 (v. 1.24.0) package using default parameters and a design file indicating replicate number and developmental time for each library (Love et al., 2014). The dataset was then pre-filtered by keeping all genes with at least 10 reads in at least two replicated libraries to reduce the potential batch effect derived from the variability in gene presence-absence characterizing the species (Gerdol et al., 2020). The resulting matrix was normalized using the regularized logarithms of counts (DESeq2::rlog) and used to perform principal component analyses and heatmaps (Kolde, 2019). Stage-by-stage differential expression analysis was performed with DESeq2 using the pairwise comparison function, with the reference stage for comparisons relevelled for each contrast (Love et al., 2014). Pairwise comparisons were performed on samples of consecutive stages (0 hpf *versus* 4 hpf, 4 hpf *versus* 8 hpf, etc.). DEGs were filtered by adjusted p-value (padj.cutoff < 0.05) and log2 fold change (lfc.cutoff > 0.58).

4.4. Developmental staging and gene clustering

To discriminate stages of *M. galloprovincialis* development from a transcriptomic perspective, bootstrapped hierarchical clustering with pvclust (v. 2.2-0) default parameters was performed on regularized logarithms (rlog) of DEG counts as previously described (Leclère et al., 2019; Love et al., 2014; Suzuki and Shimodaira, 2006). DEGs (rlog normalized counts) were then clustered with DegReport (v. 1.36.0) using default parameters to identify groups of DEGs with similar expression dynamics during *M. galloprovincialis* development (Pantano, 2023). Soft clustering with the Mfuzz (v. 2.56.0) package was performed on the transcripts per million normalized matrix to identify representative gene clusters of different developmental phases and potential transcriptomic waves (Futschik and Carlisle, 2005). The number of soft clusters was determined using the elbow method to the minimum centroid distance (Futschik and Carlisle, 2005). Genes were assigned to the cluster with the highest membership probability.

4.5. Functional annotation, gene ontologies, and enrichment analyses

The predicted protein sequences of the DEGs were annotated using BLASTp (v. 2.11.0) against the UniProt/SwissProt database, conserved domains were identified with hmmerscan from HMMER 3.3 against the Pfam database, and transcription factor domains were screened with AnimalTFDB (v2.0) (Finn et al., 2015, 2017; Jones et al., 2014; Mahram and Herbordt, 2015; Sayers et al., 2021; Zhang et al., 2015; Mistry et al., 2021). Ontology enrichment analysis was performed on the list of DEGs per stage and soft cluster. The protein sequence

corresponding to genes of interest were extracted from the *M. galloprovincialis* proteome and annotated with Gene Ontology (GO) identifiers by Interproscan (v. 95.0) analysis (Finn et al., 2017; Gerdol et al., 2020; Jones et al., 2014). Enrichment analysis of Biological Processes (BP) was conducted with topGO R (v3.17). Significance of the enrichment was computed with Fisher's exact test (Alexa and Rahnenführer, 2023), and significant lists of GOs were clustered by k-means using semantic similarity of SimplifyEnrichment (v. 1.2.0) to identify potential sets of genes responsible for stage-enriched BPs (Gu and Hübschmann, 2022).

4.6. Transcriptome validation by real time quantitative polymerase chain reaction

Complementary first strand DNAs (cDNAs) were synthesized from total RNA extracts of both biological replicates at all sample timepoints using the SuperScript VILO Kit (Invitrogen). cDNA synthesis was performed using random primers following the manufacturer's instructions: 10 minutes at 25°C, 60 minutes at 42°C, and 5 minutes at 85°C. To validate the differential gene expression analysis and the cluster analysis, four target genes were selected: *Wnt8a* (MGAL_10B085403), *Foxb2* (MGAL_10B093191), *SLC46A3* (MGAL_10B020966), and *SYTL4* (MGAL_10B044489). *EF-α1* was used for data normalization as one of the best-performing reference gene products (Balbi et al., 2016). Primer pairs were designed within the open reading frame of each gene using Primer3 (Koressaar and Rem, 2007). Sequences, efficiency, and amplicon sizes are available in the supplementary material (Supplementary Table 9). qPCR reactions were carried out for each set of primers, in experimental triplicates. The amplification was followed in real time using LightCycler 480 SYBR Green 1 master mix (Roche) in 45 cycles (initial 5 minutes at 95ºC, followed by 45 cycles of 20 seconds at 95ºC, 20 seconds at 60ºC, and 30 seconds at 72ºC) in a LightCycler 480 (Roche). Quantification was carried out in parallel for each of the target genes and the reference gene. The quantification of the expression at a given stage was determined in relation to the expression of the reference gene using the comparative CT method (Schmittgen and Livak, 2008). To compare qPCR and transcriptome results, expression of the target genes from both datasets were normalized with respect to their maximum value. Statistical differences in qPCR results were assessed using a Student's t-test (de Winter, 2013).

4.7. *In situ* hybridization using hybridization chain reaction and co-labeling by immunohistochemistry

Hybridization probe sets were designed using a user-friendly python interface (https://github.com/rwnull/insitu_probe_generator) and synthesized by OligoPools (Twist Bioscience). Amplifiers in red (B2: 546 nm) and far red (B1: 647 nm) were purchased from Molecular Instruments (https://www.molecularinstruments.com/). All probe sets with relative amplifier pairings are reported in Supplementary Table 10. The larval samples previously collected and stored in methanol were re-hydrated with four washes of 5 minutes followed by a 10-minute incubation in 1x PBS with 0.01% Triton X-100 at room temperature. Permeabilization was carried out by a 30-minute incubation in the detergent solution (1% SDS, 0.005% Tween 20, 50 mM HCl pH 7.5, 1 mM EDTA, pH 8, 0.15 M NaCl) at room temperature. After two washes of 1 minute and two washes of 5 minutes in 1x PBS with 0.01% Triton X-100 at room temperature, the specimens were placed in pre-hybridization for 30 minutes at 37°C with gentle shaking in hybridization buffer (HB) (30% formamide, 5x SSC, 9 mM citric acid, pH 6, 0.01% Tween 20, 2 mg heparin, 1x Denhardt's solution, 10% dextran sulphate). After addition of the probes (at a final concentration of 4 mM), hybridization was carried out overnight. Subsequent washes were performed at 37°C: starting with an initial addition of 1.5 volumes of washing solution (30% formamide, 5x SSC, 9 mM citric acid, pH 6, 0.01% Tween 20, 2 mg Heparin) with respect to the volume of the hybridization solution, four washes of 15 minutes were performed with washing solution. Specimens were then washed, at room temperature, twice for 5 minutes in 5x SSC with 0.01% Tween 20, once for 5 minutes in 1x SSC with 0.01% Tween 20, and 30 minutes in amplification buffer (5x SSC, 0.01% Tween 20, 10% dextran sulphate), before amplifiers were added (at a final concentration of 60 mM) to start the amplification chain reaction. Amplifiers were prepared as follows: the required volume of hairpins 1 and 2 from each amplifier were placed in a 0.6 ml tube and incubated for 90 seconds at 95°C. After the heat shock, amplifiers were placed in the dark at room temperature for 30 minutes. The amplifiers were then resuspended in amplification buffer and distributed to the samples. Amplification was carried out at room temperature in the dark overnight. Excess hairpins were removed in two washes of 5 minutes and two washes of 30 minutes in 1x SSC with 0.01% Tween 20, followed by two washes of 5 minutes in 1x PBS with 0.01% Tween 20. Co-labeling by immunocytochemistry with serotonin (5-HT) anti-rabbit antibody (1:10,000; #20080 Immunostar, Hudson, WI, USA) and goat anti-rabbit Alexa Fluor 488 conjugated secondary antibody (1:500; #4412S Cell Signaling Technology, Leiden, The Netherlands) was performed immediately after HCR amplification using previously described

conditions (Miglioli et al., 2021a, 2021b). Specimens were then prepared for imaging by incubation in 1 µg/L Hoechst (Invitrogen) in 1x PBS for 30 minutes at room temperature. Following three washes of 5 minutes in 1x PBS with 0.01% Tween 20, specimens were mounted with the antifading agent CitiFluor AF1 (Agar Scientific) and imaged on a SP8 confocal microscope (Leica Microsystems), as previously described (Miglioli et al., 2021b). Images were analyzed with ImageJ (Schneider et al., 2012) and fluorescent signals were corrected as follows: autofluorescence from tissue and shell was subtracted with “math”, the signal was homogenized with the “despeckle” or “smooth” functions, and z-stacks were merged by either “maximum intensity” or “sum slices”.

4.8. Developmental transition gene selection and comparisons with the developmental transcriptome of the Pacific oyster *Crassostrea gigas*

Representative genes of the developmental transitions were identified in *M. galloprovincialis* following this strategy. Developmental transitions were defined based on the hierarchical clustering of the stages. Three transitions were identified: from 16 to 20 hpf, 28 to 32 hpf, and 40 to 44 hpf. Genes were subsequently clustered according to their expression profile by soft clustering (see section 4.4), allowing the identification of three gene clusters corresponding to the three transitions. From these clusters, only genes showing a differential expression at the time of transition were selected. For Transition 1, genes were selected, if they were differentially expressed (log fold change ≥ 2, approximated by excess) between 16 and 20 hpf. For Transition 2, genes were selected, if they were differentially expressed (log fold change ≥ 2, approximated by excess) between 28 and 32 hpf. For Transition 3, genes were selected, if they were differentially expressed (log fold change ≥ 2, approximated by excess) between 40 and 44 hpf. To identify potentially conserved genes, a BLASTp search (Mahram and Herbordt, 2015) of the *M. galloprovincialis* transition genes was performed against the proteome of the Pacific oyster *C. gigas* (Liu et al., 2021). Their expression profiles in *C. gigas* was then visualized with a heatmap (Kolde, 2019) to investigate the presence of similar developmental transitions in *M. galloprovincialis* and *C. gigas*. *In situ* hybridization experiments were subsequently carried out on a selection of the conserved transition genes in *M. galloprovincialis* (see section 4.7). All selected *M. galloprovincialis* genes had a homolog in *C. gigas*, whose z-score value turned positive (> 0) at the time of the transition.

5. Acknowledgments

The authors are indebted to Guy Lhomond and Lucas Leclère for technical assistance and fruitful discussions. The authors would like to thank Sébastien Schaub of the Plateforme d'Imagerie par Microscopie (PIM) as well as Laurent Gilletta and Axel Duchene of the Service Aquariologie (SA) of the Institut de la Mer de Villefranche (IMEV), France, that is supported by EMBRC-France and funded by the Agence Nationale de la Recherche (ANR) (ANR-10-INBS-02). We thank the members of the Ascidian BioCell and EvoInSiDe teams at the Laboratoire de Biologie du Développement de Villefranche-sur-Mer, France, as well as the members of the team of Environmental Physiology at the Università Degli Studi Di Genova, Italy, for their support of this work.

6. Competing Interests

The authors declare that they have no competing financial interests or personal relationships that could have influenced the work reported in this paper.

7. Funding

This work was funded by grants from the CNRS and the ANR (ANR-21-CE34-0006-02), both of which are French granting agencies, to Michael Schubert and Rémi Dumollard.

8. Data Availability

All libraries analyzed in this article have been deposited in the NCBI SRA database (Accession: PRJNA996031; ID: 996031).

9. References

Alexa, A. and Rahnenführer, J. (2007). Gene set enrichment analysis with topGO. *Bioconductor Improv.* 27.

Andrews, S. (2016). Data, FASTQC. A quality control tool for high throughput sequence.

Arenas-Mena, C. (2013). Brachyury, Tbx2/3 and sall expression during embryogenesis of the indirectly developing polychaete Hydroides elegans. *Int. J. Dev. Biol.* **57**, 73–83.

ASTM E724-21, A. S. for T. and M. (2021). ASTM E724-21. Standard Guide for Conducting Static Short-Term Chronic Toxicity Tests Starting with Embryos of Four Species of Saltwater Bivalve Molluscs.

Balbi, T., Franzellitti, S., Fabbri, R., Montagna, M., Fabbri, E. and Canesi, L. (2016). Impact of bisphenol A (BPA) on early embryo development in the marine mussel Mytilus galloprovincialis: Effects on gene transcription. *Environ. Pollut.* **218**, 996–1004.

Brzozowska, R., Sui, Z. and Kang, K. H. (2012). Testing the usability of sea mussel (Mytilus sp.) for the improvement of seawater quality-An experimental study. *Ecol. Eng.* **39**, 133–137.

Choi, H. M. T., Beck, V. A. and Pierce, N. A. (2014). Next-generation in situ hybridization chain reaction: Higher gain, lower cost, greater durability. *ACS Nano* **8**, 4284–4294.

Daguin Claire, B. P. (2000). Genetic relationships of Mytilus galloprovincialis Lamarck populations worldwide: evidence from nuclear-DNA markers. *Geol. Soc. London, Spec. Publ.*

De Robertis, E. M. and Tejeda-Muñoz, N. (2022). Evo-Devo of Urbilateria and its larval forms. *Dev. Biol.* **487**, 10–20.

de Winter, J. C. F. (2013). Using the student's t-test with extremely small sample sizes. *Pract. Assessment, Res. Eval.* **18**, 1–12.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.

Dyachuk, V. and Odintsova, N. (2009). Development of the larval muscle system in the mussel Mytilus trossulus (Mollusca, Bivalvia): Original Article. *Dev. Growth Differ.* **51**, 69–79.

Dyrynda, E. A., Pipe, R. K. and Ratcliffe, N. A. (1997). Sub-populations of haemocytes in the adult and developing marine mussel, Mytilus edulis, identified by use of monoclonal antibodies. *Cell Tissue Res.* **289**, 527–536.

FAO (2022). *The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation*.

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A. and Eddy, S. R. (2015). HMMER web server: 2015 Update. *Nucleic Acids Res.* **43**, W30–W38.

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H. Y., Dosztanyi, Z., El-Gebali, S., Fraser, M., et al. (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199.

Foulon, V., Boudry, P., Artigaud, S., Guérard, F. and Hellio, C. (2019). In silico analysis of pacific oyster (Crassostrea gigas) transcriptome over developmental stages reveals candidate genes for larval settlement. *Int. J. Mol. Sci.* **20**,.

Fromherz, S. and Szent-Györgyi, A. G. (1995). Role of essential light chain EF hand domains in calcium binding and regulation of scallop myosin. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7652–7656.

Gerdol, M., Moreira, R., Cruz, F., Gómez-Garrido, J., Vlasova, A., Rosani, U., Venier, P., Naranjo-Ortiz, M. A., Murgarella, M., Greco, S., et al. (2020). Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biol.* **21**, 275.

Gu, Z. and Hübschmann, D. (2022). Simplify enrichment: A bioconductor package for clustering and visualizing functional enrichment results. *Genomics. Proteomics Bioinformatics*.

Guo, X., Li, X., Zhao, F., Liu, D., Yang, Z., Li, M., Li, Y., Wei, H., Wang, H., Qin, Z., et al. (2023). Full-length transcriptome analysis provides insights into larval shell formation in Mulinia lateralis. *Front. Mar. Sci.* **9**, 1–14.

Hunt, M. K. and Maslakova, S. A. (2017). Development of a lecithotrophic pilidium larva illustrates convergent evolution of trochophore-like morphology. *Front. Zool.* **14**, 1–18.

ISO 17244:2015, I. O. for S. (2020). ISO 17244:2015. Water quality — Determination of the toxicity of water samples on the embryo-larval development of Japanese oyster (Crassostrea gigas) and mussel (Mytilus edulis or Mytilus galloprovincialis).

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240.

Kamenev, Y. O., Eliseikina, M. G., Maslennikov, S. I. and Dolmatov, I. Y. (2018). The ultrastructural features of embryonic and early larval development in Yesso scallop, Mizuhopecten yessoensis. *Tissue Cell* **53**, 76–86.

Kemkemer, C. and Long, M. (2014). New genes important for development. *EMBO Rep.* **15**, 460–461.

Kniazkina, M. and Dyachuk, V. (2022). Neurogenesis of the scallop Azumapecten farreri: from the first larval sensory neurons to the definitive nervous system of juveniles. *Front. Zool.* **19**, 1–19.

Kniprath, E. (1980a). Roux's Archives of Developmental Biology. **204**, 201–204.

Kniprath, E. (1980b). Larval development of the shell and the shell gland in Mytilus (Bivalvia). *Wilhelm Roux's Arch. Dev. Biol.* **188**, 201–204.

KNIPRATH, E. (1981). Ontogeny of the Molluscan Shell Field: a Review. *Zool. Scr.* **10**, 61–79.

Kocot, K. M., Aguilera, F., McDougall, C., Jackson, D. J. and Degnan, B. M. (2016). Sea shell diversity and rapidly evolving secretomes: Insights into the evolution of biomineralization. *Front. Zool.* **13**, 1–10.

Kolde, R. (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12.

Kurita, A., Kurita, Y., Deguchi, R. and Wada, H. Early Development and Cleavage Pattern of the Japanese Purple Mussel , Septifer virgatus Early Development and Cleavage Pattern of the Japanese Purple Mussel , Septifer virgatus. **26**, 814–820.

Kurita, Y., Deguchi, R. and Wada, H. (2009). Early development and cleavage pattern of the Japanese purple mussel, septifer virgatus. *Zoolog. Sci.* **26**, 814–820.

Lartilot, N., Le Gouar, M. and Adoutte, A. (2002). Expression pattern of fork head and goosecoid homologues in the mollusc Patella vulgata supports the ancestry of the anterior mesendoderm across Bilateria. *Dev. Genes Evol.* **212**, 551–561.

Leclère, L., Horin, C., Chevalier, S., Lapébie, P., Dru, P., Peron, S., Jager, M., Condamine, T., Pottin, K., Romano, S., et al. (2019). The genome of the jellyfish Clytia hemisphaerica and the evolution of the cnidarian life-cycle. *Nat. Ecol. Evol.* **3**, 801–810.

Liu, Z., Zhou, Z., Zhang, Y., Wang, L., Song, X., Wang, W., Zheng, Y., Zong, Y., Lv, Z. and Song, L. (2020). Ocean acidification inhibits initial shell formation of oyster larvae by suppressing the biosynthesis of serotonin and dopamine. *Sci. Total Environ.* **735**, 139469.

Liu, F., Li, Y., Yu, H., Zhang, L., Hu, J., Bao, Z. and Wang, S. (2021). MolluscDB: An integrated functional and evolutionary genomics database for the hyper-diverse animal phylum Mollusca. *Nucleic Acids Res.* **49**, D988–D997.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21.

M. Futschik, B. C. (2005). Noise robust clustering of gene expression time-course data. 965–988.

Mahram, A. and Herbordt, M. C. (2015). NCBI BLASTP on high-performance reconfigurable computing systems. *ACM Trans. Reconfigurable Technol. Syst.* **7**,.

Marlow, H., Tosches, M. A., Tomer, R., Steinmetz, P. R., Lauri, A., Larsson, T. and Arendt, D. (2014). Larval body patterning and apical organs are conserved in animal evolution. *BMC Biol.* **12**, 1–17.

Martín, A., Baltodano, C., Seudre, O., Guynes, K., María, J. and Durán, M. (2021). Early embryogenesis and organogenesis in the annelid Owenia fusiformis. *Evodevo* 1–18.

Maureen Liu, M., Davey, J. W., Jackson, D. J., Blaxter, M. L. and Davison, A. (2014). A conserved set of maternal genes? Insights from a molluscan transcriptome. *Int. J. Dev. Biol.* **58**, 501–511.

Miglioli, A. (2022). Pathways of endocrine disruption in the larval development of the mediterranean mussel Mytilus galloprovincialis To cite this version□: HAL Id□: tel-03696706 PhD thesis of the Università degli studi di Genova and Sorbonne Université Presented by PhD candi. 203.

Miglioli, A., Dumollard, R., Balbi, T., Besnardeau, L. and Canesi, L. (2019). Characterization of the main steps in first shell formation in Mytilus galloprovincialis: Possible role of tyrosinase. *Proc. R. Soc. B Biol. Sci.* **286**,.

Miglioli, A., Balbi, T., Montagna, M., Dumollard, R. and Canesi, L. (2021a). Tetrabromobisphenol A acts a neurodevelopmental disruptor in early larval stages of Mytilus galloprovincialis. *Sci. Total Environ.* **793**, 148596.

Miglioli, A., Balbi, T., Besnardeau, L., Dumollard, R. and Canesi, L. (2021b). Bisphenol A interferes with fi rst shell formation and development of the serotoninergic system in early larval stages of Mytilus galloprovincialis. *Sci. Total Environ.* **758**, 144003.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2020). Pfam: The protein families database in 2021. Nucleic Acids Research 49, D412–D419. 10.1093/nar/gkaa913.

Moreira, R., Pereiro, P., Balseiro, P., Milan, M., Pauletto, M., Bargelloni, L., Novoa, B. and Figueras, A. (2018). Revealing Mytilus galloprovincialis transcriptomic profiles during ontogeny. *Dev. Comp. Immunol.* **84**, 292–306.

Morino, Y., Okada, K., Niikura, M., Honda, M., Satoh, N., & Wada, H. (2013). Genome-Wide Survey of Genes Encoding Transcription Factors in the Japanese Pearl Oyster, Pinctada fucata: I. Homeobox Genes. *Zoolog. Sci.* **30**, 851–857.

Nayshool, O., Kol, N., Javaski, E., Amariglio, N. and Rechavi, G. (2022). SurviveAI: Long Term Survival Prediction of Cancer Patients Based on Somatic RNA-Seq Expression. *Cancer Inform.*

21,.

Nielsen, C. (2005). Trochophora larvae: Cell-lineages, ciliary bands and body regions. 2. Other groups and general discussion. *J. Exp. Zool. Part B Mol. Dev. Evol.* **304**, 401–447.

Nielsen, C. (2009). How did indirect development with planktotrophic larvae evolve? *Biol. Bull.* **216**, 203–215.

Niu, D., Wang, F., Xie, S., Sun, F., Wang, Z., Peng, M. and Li, J. (2016). Developmental Transcriptome Analysis and Identification of Genes Involved in Larval Metamorphosis of the Razor Clam, Sinonovacula constricta. *Mar. Biotechnol.* **18**, 168–175.

Núñez-Acuña, G., Fernandez, C., Sanhueza-Guevara, S. and Gallardo-Escárate, C. (2022). Transcriptome profiling of the early developmental stages in the giant mussel Choromytilus chorus exposed to delousing drugs. *Mar. Genomics* **65**,.

Osborne, C. C., Perry, K. J., Shankland, M. and Henry, J. Q. (2018). Ectomesoderm and epithelial–mesenchymal transition-related genes in spiralian development. *Dev. Dyn.* **247**, 1097–1120.

Page, L. R. (2009). Molluscan larvae: Pelagic juveniles or slowly metamorphosing larvae? *Biol. Bull.* **216**, 216–225.

Pantano, L. (2023). DEGreport: Report of DEG analysis. R package version.

Paps, J., Xu, F., Zhang, G. and Holland, P. W. H. (2015). 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.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T. and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667.

Ramesh, K., Yarra, T., Clark, M. S., John, U. and Melzner, F. (2019). Expression of calcification-related ion transporters during blue mussel larval development. *Ecol. Evol.* **9**, 7157–7172.

Rawlinson, K. A. (2010). Embryonic and post-embryonic development of the polyclad flatworm Maritigrella crozieri; implications for the evolution of spiralian life history traits. *Front. Zool.* **7**, 1–25.

Robertis, E. M. De and Tejeda-mu, N. (2022). Evo-Devo of Urbilateria and its larval forms ~ oz. **487**, 10–20.

Salamanca-Díaz, D. A., Ritschard, E. A., Schmidbaur, H. and Wanninger, A. (2022). Comparative Single-Cell Transcriptomics Reveals Novel Genes Involved in Bivalve Embryonic Shell Formation and Questions Ontogenetic Homology of Molluscan Shell Types. *Front. Cell Dev. Biol.* **10**, 1–10.

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., Comeau, D. C., Funk, K., Kim, S., Klimke, W., et al. (2021). Database resources of the National Center for

Biotechnology Information. *Nucleic Acids Res.* **49**, D10–D17.

Schmittgen, T. D. and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3**, 1101–1108.

Setiamarga, D. H. E., Shimizu, K., Kuroda, J., Inamura, K., Sato, K., Isowa, Y., Ishikawa, M., Maeda, R., Nakano, T., Yamakawa, T., Hatori, R., Ishio, A., Kaneko, K., Matsumoto, K., Sarashina, I., Teruya, S., Zhao, R., Satoh, N., Sasaki, T., … Endo, K. (2013). An In-silico Genomic Survey to Annotate Genes Coding for Early Development-Relevant Signaling Molecules in the Pearl Oyster, Pinctada fucata. *Zoolog. Sci.* **30**, 877–888.

Song, H., Yu, Z. L., Sun, L. N., Gao, Y., Zhang, T. and Wang, H. Y. (2016a). De novo transcriptome sequencing and analysis of Rapana venosa from six different developmental stages using Hi-seq 2500. *Comp. Biochem. Physiol. - Part D Genomics Proteomics* **17**, 48–57.

Song, X., Wang, H., Chen, H., Sun, M., Liang, Z., Wang, L. and Song, L. (2016b). Conserved hemopoietic transcription factor Cg-SCL delineates hematopoiesis of Pacific oyster Crassostrea gigas. *Fish Shellfish Immunol.* **51**, 180–188.

Suzuki, R. and Shimodaira, H. (2006). Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542.

Takito, J. and Al-Awqati, Q. (2004). Conversion of ES cells to columnar epithelia by hensin and to squamous epithelia by laminin. *J. Cell Biol.* **166**, 1093–1102.

Tomer, R., Denes, A. S., Tessmar-Raible, K. and Arendt, D. (2010). Profiling by Image Registration Reveals Common Origin of Annelid Mushroom Bodies and Vertebrate Pallium. *Cell* **142**, 800–809.

Tyler J. Carrier, Adam M. Reitzel, A. H. (2018). *Evolutionary Ecology of Marine Invertebrate Larvae*. Oxford University Press.

Vogeler, S., Bean, T. P., Lyons, B. P. and Galloway, T. S. (2016). Dynamics of nuclear receptor gene expression during Pacific oyster development. *BMC Dev. Biol.* **16**, 1–13.

Vogeler, S., Carboni, S., Li, X., Ireland, J. H., Miller-Ezzy, P. and Joyce, A. (2021). Cloning and characterisation of NMDA receptors in the Pacific oyster, Crassostrea gigas (Thunberg, 1793) in relation to metamorphosis and catecholamine synthesis. *Dev. Biol.* **469**, 144–159.

Wada, H., Phuangphong, S., Hashimoto, N. and Nagai, K. (2020). Establishment of the novel bivalve body plan through modification of early developmental events in mollusks. *Evol. Dev.* **22**, 463–470.

Wang, J., Zhang, L., Lian, S., Qin, Z., Zhu, X., Dai, X., Huang, Z., Ke, C., Zhou, Z., Wei, J., et al. (2020). Evolutionary transcriptomics of metazoan biphasic life cycle supports a single intercalation origin of metazoan larvae. *Nat. Ecol. Evol.* **4**, 725–736.

Wu, L., Ferger, K. E. and David Lambert, J. (2019). Gene expression does not support the developmental hourglass model in three animals with Spiralian development. *Mol. Biol. Evol.* **36**, 1373–1383.

Xu, F., Domazet-Lošo, T., Fan, D., Dunwell, T. L., Li, L., Fang, X. and Zhang, G. (2016). High

expression of new genes in trochophore enlightening the ontogeny and evolution of trochozoans. *Sci. Rep.* **6**, 1–10.

Yang, Z., Zhang, L., Hu, J., Wang, J., Bao, Z. and Wang, S. (2020). The evo-devo of molluscs: Insights from a genomic perspective. *Evol. Dev.* **22**, 409–424.

Yurchenko, O. V., Skiteva, O. I., Voronezhskaya, E. E. and Dyachuk, V. A. (2018). Nervous system development in the Pacific oyster, Crassostrea gigas (Mollusca: Bivalvia). *Front. Zool.* **15**, 1–21.

Zhang, H. M., Liu, T., Liu, C. J., Song, S., Zhang, X., Liu, W., Jia, H., Xue, Y. and Guo, A. Y. (2015). AnimalTFDB 2.0: A resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* **43**, D76–D81.

Zhao, R., Takeuchi, T., Luo, Y. J., Ishikawa, A., Kobayashi, T., Koyanagi, R., Villar-Briones, A., Yamada, L., Sawada, H., Iwanaga, S., et al. (2018). Dual gene repertoires for larval and adult shells reveal molecules essential for molluscan shell formation. *Mol. Biol. Evol.* **35**, 2751–2761.

Zheng, Z., Du, X., Xiong, X., Jiao, Y., Deng, Y., Wang, Q. and Huang, R. (2017). PmRunt regulated by Pm-miR-183 participates in nacre formation possibly through promoting the expression of collagen VI-like and Nacrein in pearl oyster Pinctada martensii. *PLoS One* **12**, 1–14.

Zheng, Z., Hao, R., Xiong, X., Jiao, Y., Deng, Y. and Du, X. (2019). Developmental characteristics of pearl oyster Pinctada fucata martensii: Insight into key molecular events related to shell formation, settlement and metamorphosis. *BMC Genomics* **20**, 1–11.

10. Figure Captions

Figure 1: **Morphological and transcriptomic staging of *Mytilus galloprovincialis* early development**. A) Development of *M. galloprovincialis* under standard culture conditions. Orientation is as shown by arrows in the top right of each picture, hours post fertilization (hpf) (at 16°C) is indicated on the top left. Red and black arrowheads highlight stomodeum/presumptive mouth and anus, respectively, blue dotted line indicates the shell field outline (16 to 24 hpf) and the hinge region (28 hpf). At 40 hpf, the green line indicates the D-margin of the calcified shell. Abbreviations: An: animal; Ve: vegetal; A: anterior; P: posterior; D: dorsal; V: ventral; a: anus; at: apical tuft; h: hinge; m: mantle; sf: shell field; st: stomodeum/presumptive mouth; ve: velum. Scale bar: 20 µm. B) Dot plot of the number of genes expressed in each library before DESeq2 normalization. The black dotted line represents the mean trend throughout development. C) Heatmap of the Euclidean distance matrix of the libraries clustered with default hclust parameters. D) Principal component analysis (PCA) of DESeq2 normalized libraries, dotted line indicates the axis of principal component 2.

Figure 2: **Differential expression analysis highlights stage-specific transcriptomic signatures.** A) Hierarchical clustering of differentially expressed gene expression values in the transcriptome libraries. Bootstrap values are shown in red on each branch. Names for each cluster are assigned according to morphology and developmental timepoint. B) Histogram showing the number of up- (red) and down- (blue) regulated genes at each assayed stage. Color code is according to log fold change values. Insert: percentage (%) of differentially expressed genes that are up-, down- or up-/down-regulated in the dataset, either once (1x) or multiple times. C) Top six gene expression profiles by number of genes. D) Pie graph depicting the percentage (%) of profiles reaching their highest expression levels at different developmental stages. E) Normalized expression of *Wnt8a*, *Foxb2*, *SLC46A3*, and *SYTL4* from transcriptomic (left) and qPCR (right) data. F) Maximum projection superimpositions of *Foxb2* (yellow) and *Wnt8a* (red) fluorescent *in situ* hybridization with serotonin (5-HT) immunohistochemistry (green) and Hoechst (grey) or in bright field in trochophore (24 hpf) and D-veliger (48 hpf) larvae. Abbreviations: an: anus; ao/cg: apical organ/cerebral ganglion; at: apical tuft; aso: apical sensory organ; eso: esophagus; h: hinge; sf: shell field; st: stomodeum/presumptive mouth; ve: velum. Scale bar: 20 µm.

Figure 3: **Qualitative analysis of functional domains and gene ontologies (GOs) of differentially expressed genes (DEGs).** A) Transcription factors. B) Domains associated with shell biogenesis: shell matrix proteins in blue and calcification proteins in green. C) Receptors and neurotransmitter transporters. D) Composite image showing the heatmap of significant GO terms paired with GO cluster analysis at different stages of development. Colors indicate the $-\log_{10}$(p-value). Column breaks in the heatmap divide the different developmental stages identified by hierarchical clustering, and row breaks separate the GO clusters. Row color annotation (on the left) highlights different GO clusters, and the most significant terms for each cluster are shown in their respective box (on the right).

Figure 4: **Soft clustering reveals transcriptomic signatures of early developmental transitions.** A) Left panel: expression patterns of single soft clusters. The green lines depict the expression pattern in scaled transcripts per million (z-scores) of each gene belonging to a soft cluster, the red line indicates the mean trend of the expression data, vertical grey lines indicate developmental transitions. Right panel: heatmap of all soft clusters in scaled transcripts per million (z-scores) arranged by developmental time of appearance. White lines

identify developmental transitions identified by hierarchical clustering. B) Expression of developmental transition genes in scaled transcripts per million (z-scores) in the mussel (*Mytilus galloprovincialis*) transcriptome. C) Expression of developmental transitions genes in scaled transcripts per million (z-scores) in the Pacific oyster (*Crassostrea gigas*) developmental transcriptome. Column breaks indicate the three transitions. Color annotations on the left, in B) and C), highlight the three developmental transitions: blue, gold, and red for Transition 1, 2, and 3, respectively.

Figure 5: **Developmental expression of homologous transition genes in *Mytilus galloprovincialis* larvae*.*** A) Genes selected for Transition 1 (gastrula to trochophore). B) Genes selected for Transition 2 (trochophore to advanced trochophore). C) Genes selected for Transition 3 (advanced trochophore to D-veliger). Left and right valves are presented next to each other, relative to the hinge region. The *in situ* hybridization signal is shown in red, Hoechst-stained nuclei are shown in grey. Gene names are indicated in red in the top left corner. Developmental timing in hours post fertilization (hpf) (at 16°C) is indicated in white under the gene name. Abbreviations: D: dorsal view; S: sagittal view; V: ventral view; a: anus; at: apical tuft; eso: esophagus; h: hinge; sf: shell field; st: stomodeum/presumptive mouth. Scale bar: 20 µm.

11. Supplementary Figure Legends

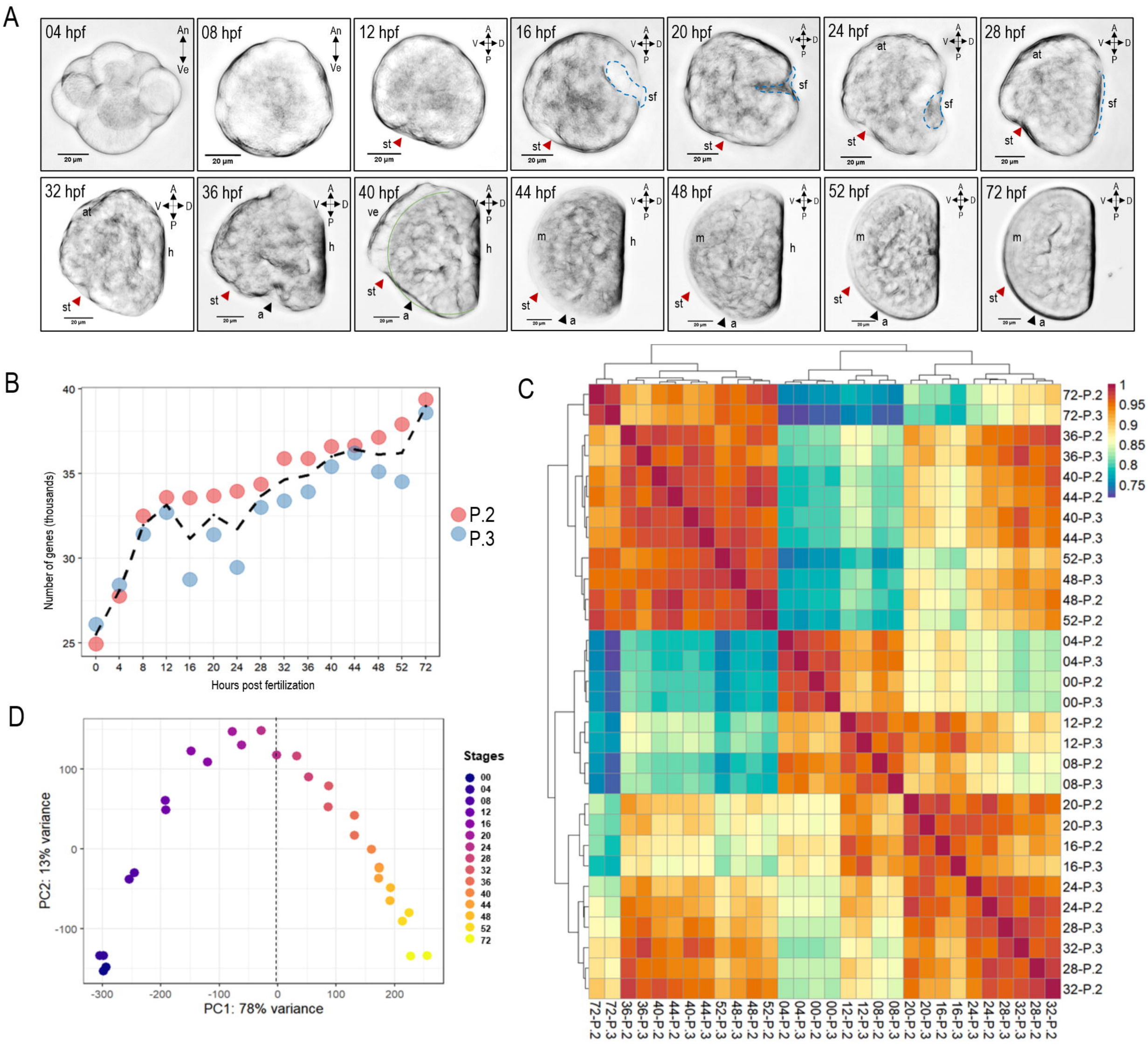
Supplementary Figure 1: **Fluorescent *in situ* hybridization protocol validation with *Foxb2* and *Wnt8a* probes at different stages of *Mytilus galloprovincialis* development.** A) Expression pattern of *Foxb2*. Upper panel shows maximum z-projections of the *in situ* hybridization signal (red) with Hoechst nuclear staining (grey). Lower panel shows the *in situ* hybridization signal (red) in brightfield. B) Expression pattern of *Wnt8a*. Upper panel shows maximum z-projections of the *in situ* hybridization signal (red) with Hoechst nuclear staining (grey). Lower panel shows the *in situ* hybridization signal (red) in brightfield. Abbreviations: at: apical tuft; h: hinge; sf: shell field; st: stomodeum/presumptive mouth; ve: velum. Scale bar: 20µm.

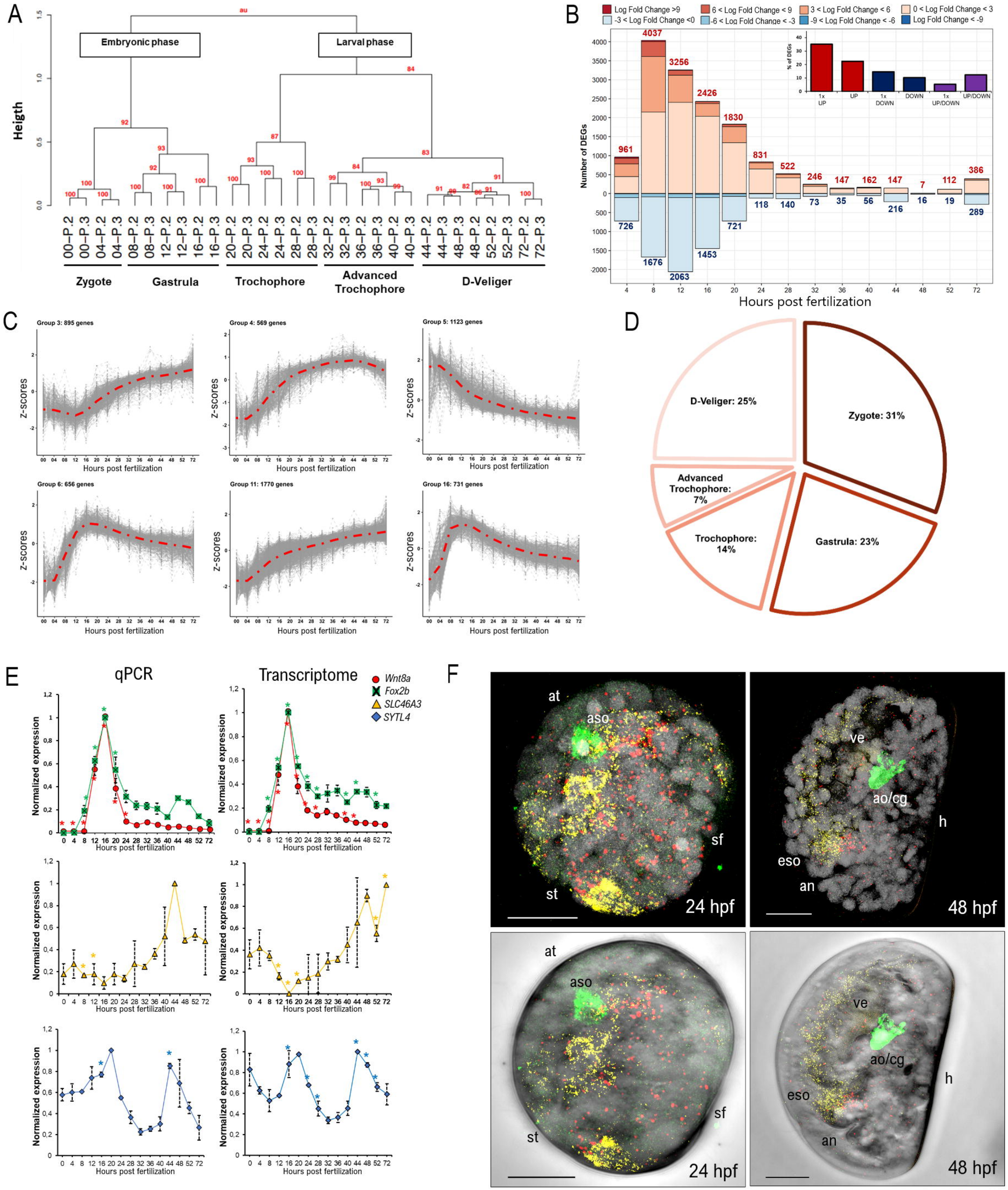
Supplementary Figure 2: **Developmental expression patterns of transition genes in a developmental transcriptome of the Pacific oyster *Crassostrea gigas*.** A) Transition1. B) Transition 2. C) Transition 3.

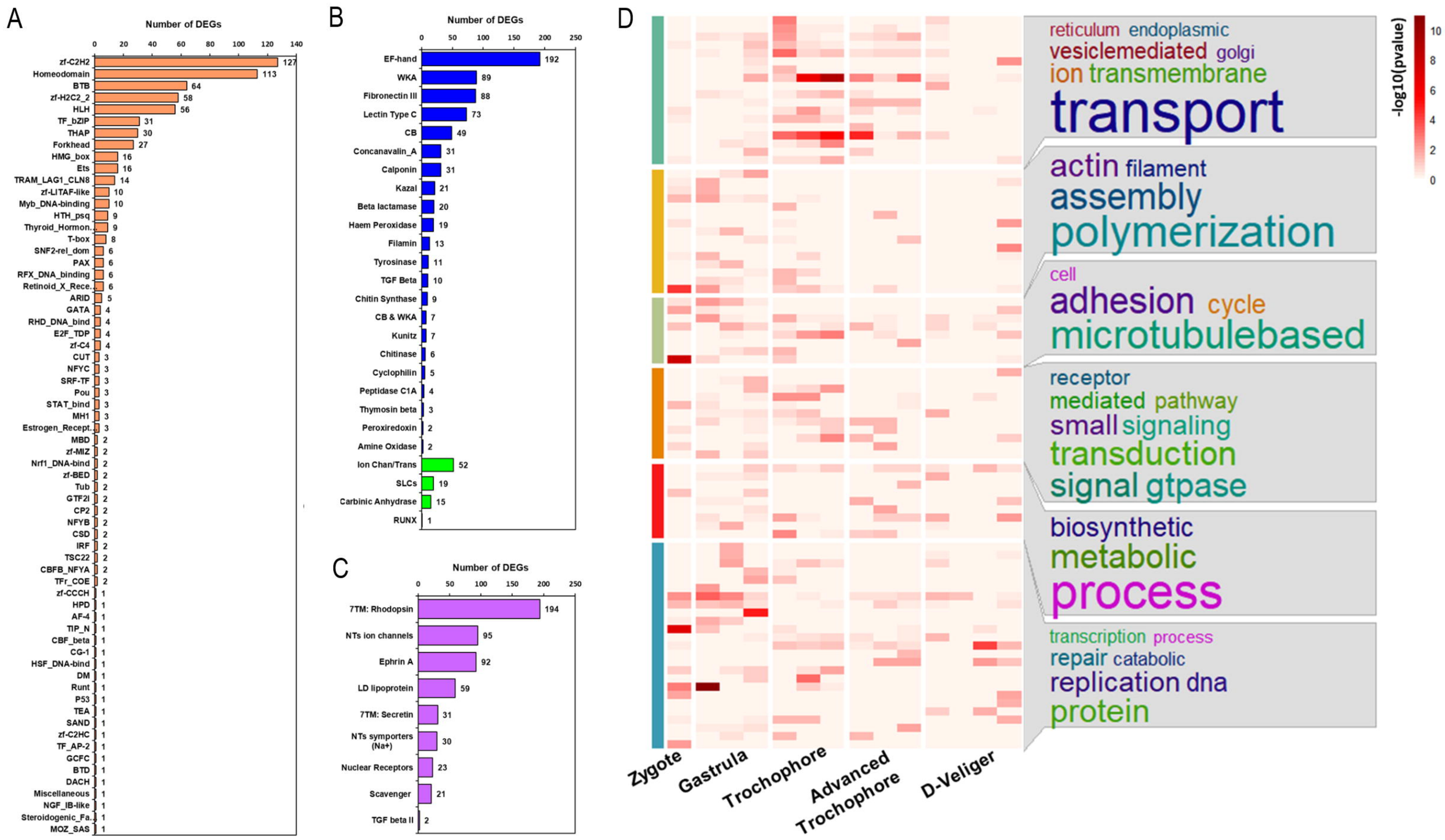
12. Supplementary Files and Tables

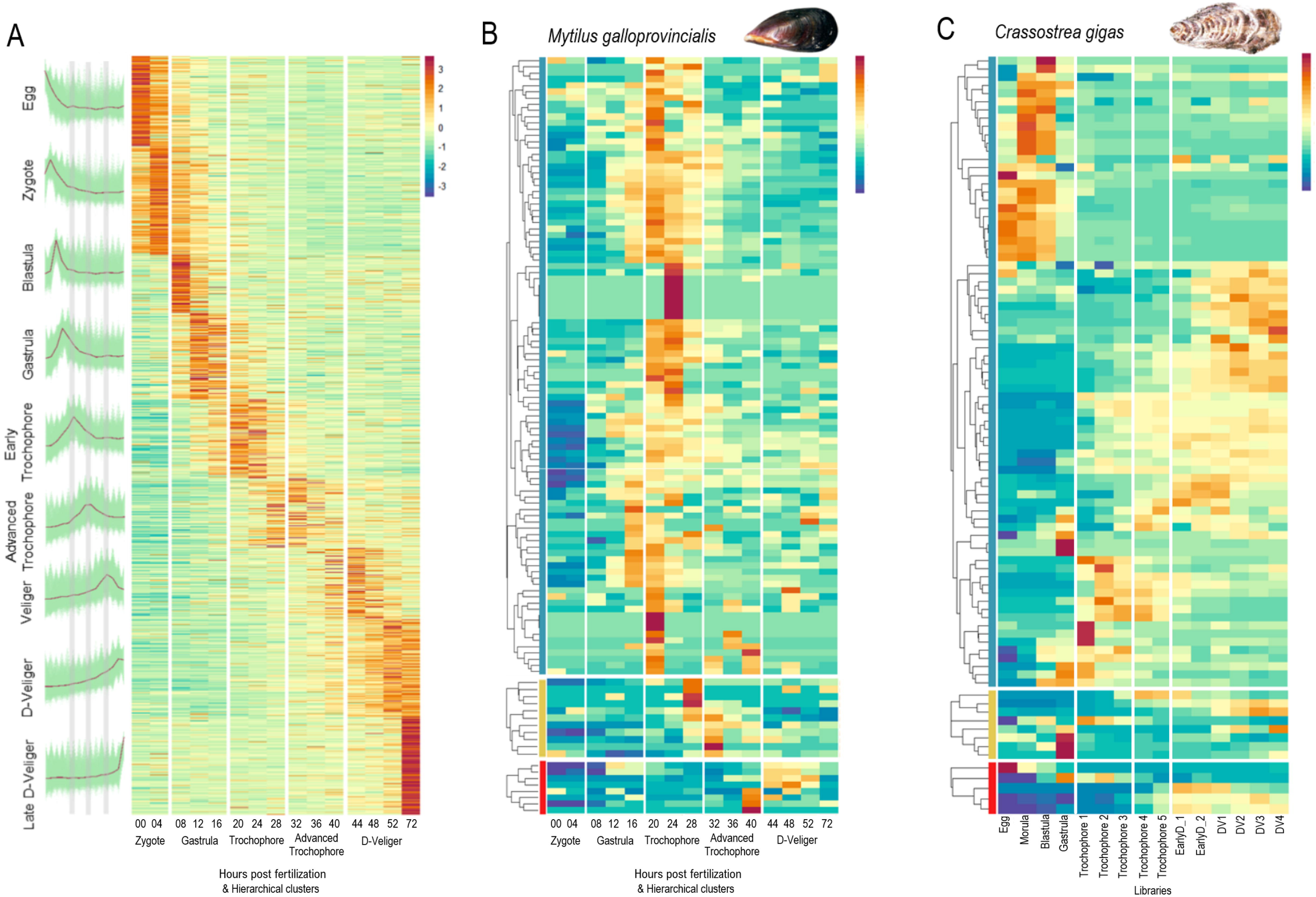
Supplementary File 1: Supplementary Table 1. Summary of mapping results for each library. Supplementary Table 2. Matrix of raw gene counts. Supplementary Table 3. Dynamics of differentially expressed genes with Uniprot protein annotation, Pfam domain annotation, DEG cluster annotation, developmental transition membership, and soft clustering membership. Supplementary Table 4. DegReport analysis. The mean expression value of each cluster is reported for each timepoint, and the expression trend is presented as a sparkline histogram. The green bar indicates the highest and the red one the lowest value. Supplementary Table 5. Binary matrix of the DegReport output. Supplementary Table 6. List of differentially expressed transcription factors identified with AnimalTFDB. Supplementary Table 7. Binary matrix of the Mfuzz soft clustering analysis. Supplementary Table 8. Best blastp hits of *Mytilus galloprovincialis* transition genes in the *Crassostrea gigas* genome. Supplementary Table 9. Primer pairs used for qPCR analyses. Supplementary Table 10. Probe sets for *in situ* hybridization by hybridization chain reaction used in the study.

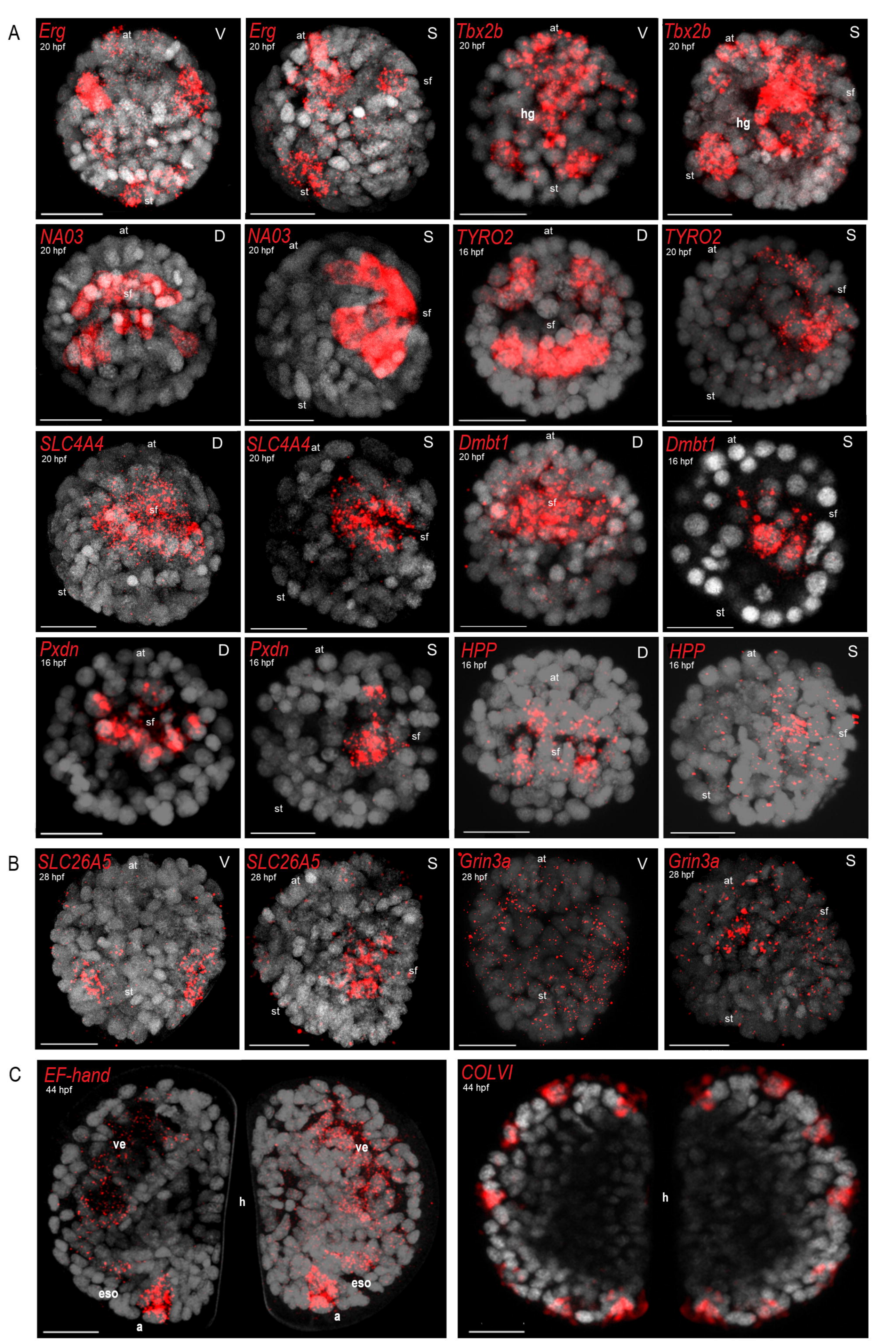
Supplementary File 2: DESeq2 output of differentially expressed genes by pairwise comparison. Every tab corresponds to a pairwise comparison.

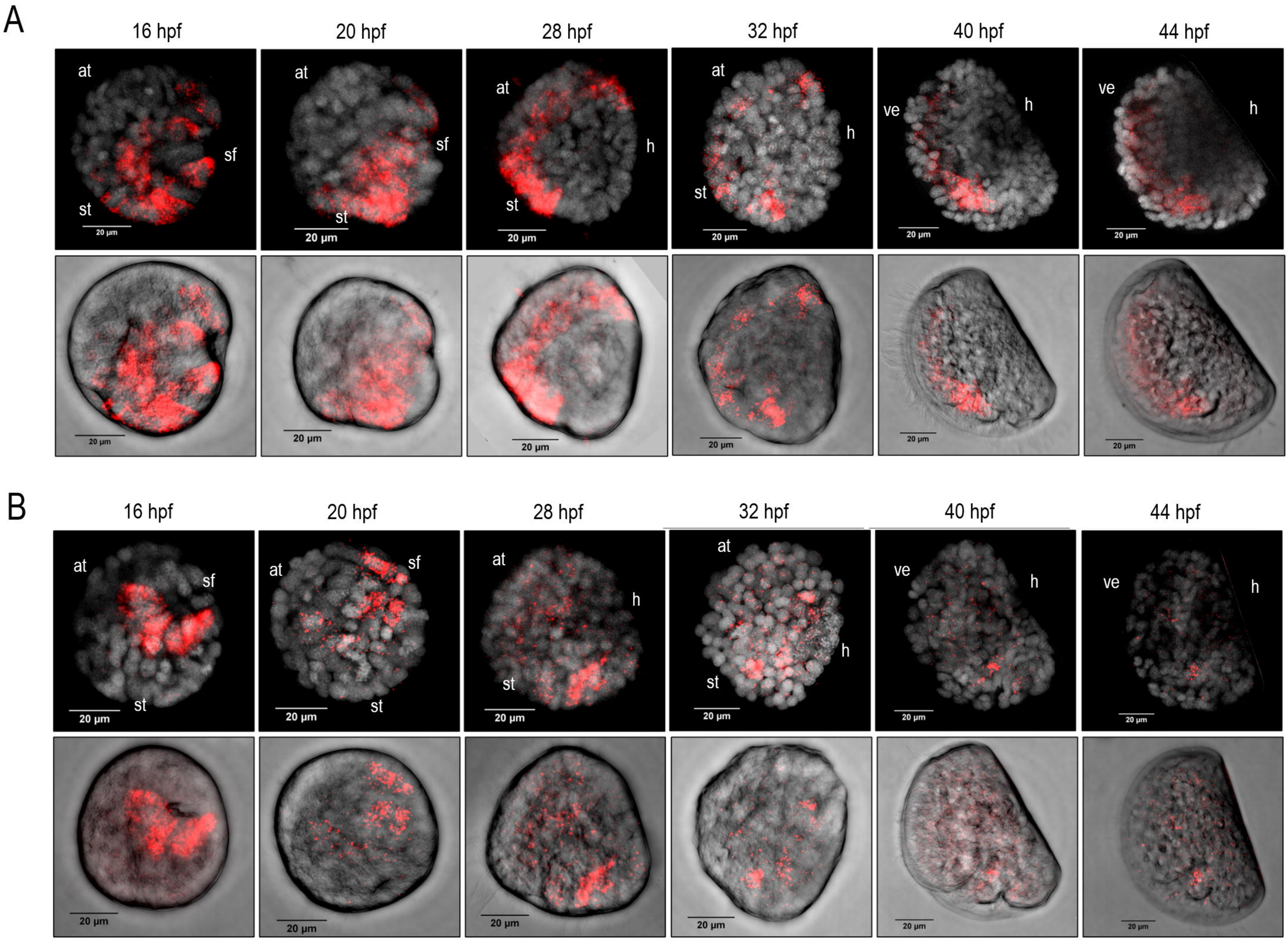

A
04 hpf
08 hpf
12 hpf
16 hpf
20 hpf
24 hpf
28 hpf
32 hpf
36 hpf
40 hpf
44 hpf
48 hpf
52 hpf
72 hpf
An
Ve
A
V
D
P
at
sf
st
ve
m
h
a
20 µm
B
Number of genes (thousands)
Hours post fertilization
P.2
P.3
C
D
PC2: 13% variance
PC1: 78% variance
Stages
00
04
08
12
16
20
24
28
32
36
40
44
48
52
72

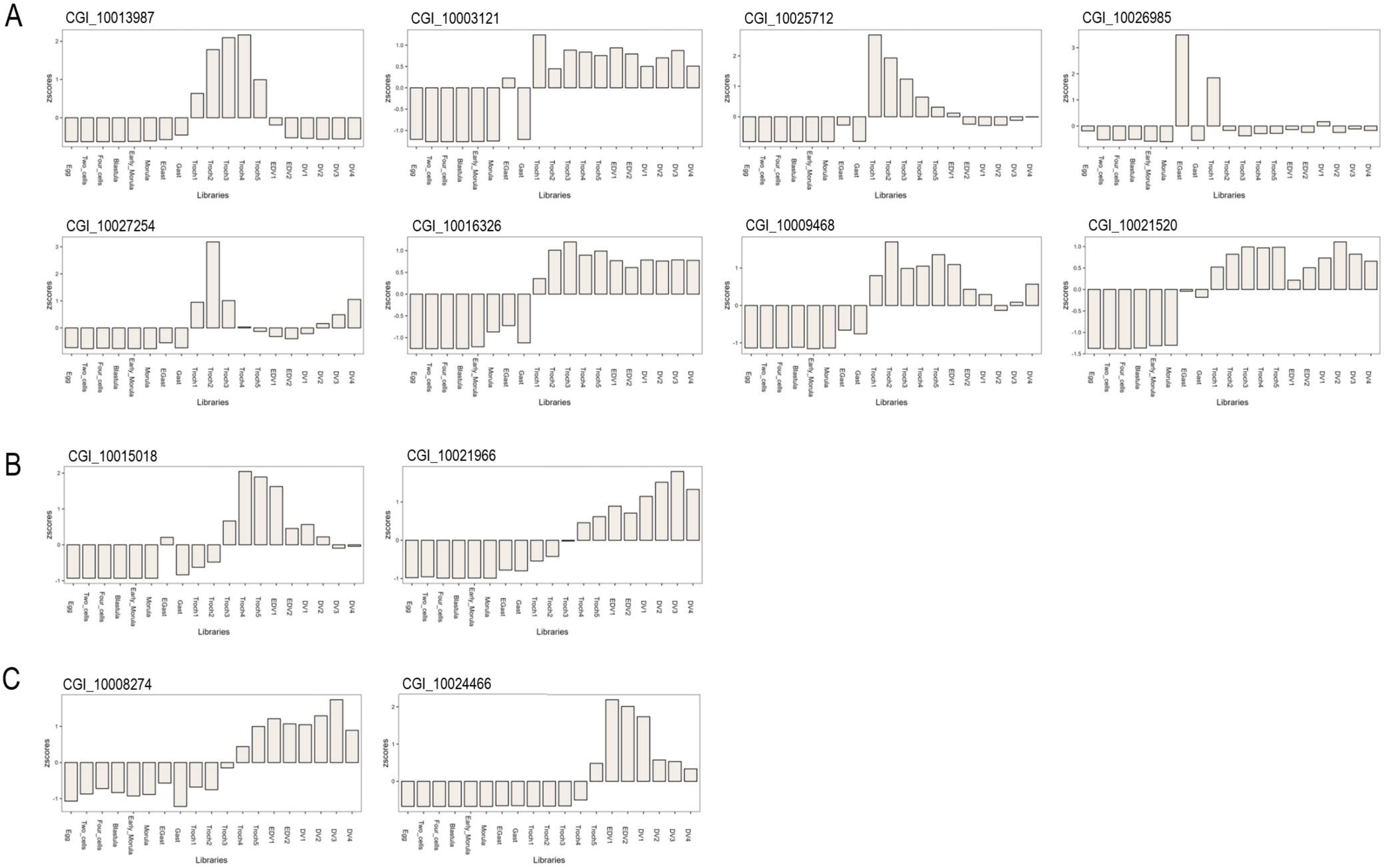

A
Embryonic phase
Larval phase
Heigth
Zygote
Gastrula
Trochophore
Advanced Trochophore
D-Veliger
B
Log Fold Change >9
6 < Log Fold Change < 9
3 < Log Fold Change < 6
0 < Log Fold Change < 3
-3 < Log Fold Change <0
-6 < Log Fold Change < -3
-9 < Log Fold Change < -6
Log Fold Change < -9
Number of DEGs
Hours post fertilization
% of DEGs
1x UP
UP
1x DOWN
DOWN
1x UP/DOWN
UP/DOWN
C
Group 3: 895 genes
Group 4: 569 genes
Group 5: 1123 genes
Group 6: 656 genes
Group 11: 1770 genes
Group 16: 731 genes
Z-scores
Hours post fertilization
D
D-Veliger: 25%
Zygote: 31%
Advanced Trochophore: 7%
Trochophore: 14%
Gastrula: 23%
E
qPCR
Transcriptome
Wnt8a
Fox2b
SLC46A3
SYTL4
Normalized expression
Hours post fertilization
F
at
aso
sf
st
24 hpf
ve
ao/cg
h
eso
an
48 hpf


A
Number of DEGs
0 20 40 60 80 100 120 140
zf-C2H2 127
Homeodomain 113
BTB 64
zf-H2C2_2 58
HLH 56
TF_bZIP 31
THAP 30
Forkhead 27
HMG_box 16
Ets 16
TRAM_LAG1_CLN8 14
zf-LITAF-like 10
Myb_DNA-binding 10
HTH_psq 9
Thyroid_Hormon.. 9
T-box 8
SNF2-rel_dom 6
PAX 6
RFX_DNA_binding 6
Retinoid_X_Rece.. 6
ARID 5
GATA 4
RHD_DNA_bind 4
E2F_TDP 4
zf-C4 4
CUT 3
NFYC 3
SRF-TF 3
Pou 3
STAT_bind 3
MH1 3
Estrogen_Recept.. 3
MBD 2
zf-MIZ 2
Nrf1_DNA-bind 2
zf-BED 2
Tub 2
GTF2I 2
CP2 2
NFYB 2
CSD 2
IRF 2
TSC22 2
CBFB_NFYA 2
TFr_COE 2
zf-CCCH 1
HPD 1
AF-4 1
TIP_N 1
CBF_beta 1
CG-1 1
HSF_DNA-bind 1
DM 1
Runt 1
P53 1
TEA 1
SAND 1
zf-C2HC 1
TF_AP-2 1
GCFC 1
BTD 1
DACH 1
Miscellaneous 1
NGF_IB-like 1
Steroidogenic_Fa.. 1
MOZ_SAS 1
B
Number of DEGs
0 50 100 150 200 250
EF-hand 192
WKA 89
Fibronectin III 88
Lectin Type C 73
CB 49
Concanavalin_A 31
Calponin 31
Kazal 21
Beta lactamase 20
Haem Peroxidase 19
Filamin 13
Tyrosinase 11
TGF Beta 10
Chitin Synthase 9
CB & WKA 7
Kunitz 7
Chitinase 6
Cyclophilin 5
Peptidase C1A 4
Thymosin beta 3
Peroxiredoxin 2
Amine Oxidase 2
Ion Chan/Trans 52
SLCs 19
Carbinic Anhydrase 15
RUNX 1
C
Number of DEGs
0 50 100 150 200 250
7TM: Rhodopsin 194
NTs ion channels 95
Ephrin A 92
LD lipoprotein 59
7TM: Secretin 31
NTs symporters (Na+) 30
Nuclear Receptors 23
Scavenger 21
TGF beta II 2
D
reticulum endoplasmic
vesiclemediated golgi
ion transmembrane
transport
actin filament
assembly
polymerization
cell
adhesion cycle
microtubulebased
receptor
mediated pathway
small signaling
transduction
signal gtpase
biosynthetic
metabolic
process
transcription process
repair catabolic
replication dna
protein
Zygote
Gastrula
Trochophore
Advanced Trochophore
D-Veliger
-log10(pvalue)
10
8
6
4
2
0


A
Egg
Zygote
Blastula
Gastrula
Early Trochophore
Advanced Trochophore
Veliger
D-Veliger
Late D-Veliger
3
2
1
0
-1
-2
-3
00 04 08 12 16 20 24 28 32 36 40 44 48 52 72
Zygote
Gastrula
Trochophore
Advanced Trochophore
D-Veliger
Hours post fertilization
& Hierarchical clusters
B
Mytilus galloprovincialis
00 04 08 12 16 20 24 28 32 36 40 44 48 52 72
Zygote
Gastrula
Trochophore
Advanced Trochophore
D-Veliger
Hours post fertilization
& Hierarchical clusters
C
Crassostrea gigas
Egg
Morula
Blastula
Gastrula
Trochophore 1
Trochophore 2
Trochophore 3
Trochophore 4
Trochophore 5
EarlyD_1
EarlyD_2
DV1
DV2
DV3
DV4
Libraries


A
Erg
20 hpf
at
st
V
Erg
20 hpf
at
sf
st
S
Tbx2b
20 hpf
at
hg
st
V
Tbx2b
20 hpf
at
sf
hg
st
S
NA03
20 hpf
at
sf
D
NA03
20 hpf
at
sf
st
S
TYRO2
16 hpf
at
sf
D
TYRO2
20 hpf
at
sf
st
S
SLC4A4
20 hpf
at
sf
st
D
SLC4A4
20 hpf
at
sf
st
S
Dmbt1
20 hpf
at
sf
D
Dmbt1
16 hpf
at
sf
st
S
Pxdn
16 hpf
at
sf
D
Pxdn
16 hpf
at
sf
st
S
HPP
16 hpf
at
sf
D
HPP
16 hpf
at
sf
st
S
B
SLC26A5
28 hpf
at
st
V
SLC26A5
28 hpf
at
sf
st
S
Grin3a
28 hpf
at
st
V
Grin3a
28 hpf
at
sf
st
S
C
EF-hand
44 hpf
ve
h
eso
a
ve
eso
a
COLVI
44 hpf
h


A
16 hpf
20 hpf
28 hpf
32 hpf
40 hpf
44 hpf
at
sf
st
h
ve
20 µm
B
16 hpf
20 hpf
28 hpf
32 hpf
40 hpf
44 hpf
at
sf
st
h
ve
20 µm


A
CGI_10013987
CGI_10003121
CGI_10025712
CGI_10026985
CGI_10027254
CGI_10016326
CGI_10009468
CGI_10021520
B
CGI_10015018
CGI_10021966
C
CGI_10008274
CGI_10024466
zscores
Libraries
Egg
Two_cells
Four_cells
Blastula
Early_Morula
Morula
EGast
Gast
Troch1
Troch2
Troch3
Troch4
Troch5
EDV1
EDV2
DV1
DV2
DV3
DV4