

SCIENTIFIC REPORTS



OPEN

Unravelling paralogous gene expression dynamics during three-spined stickleback embryogenesis

Elisavet Kaitetzidou^{1,2}, Ioanna Katsiadaki³, Jacques Lagnel^{2,4}, Efthimia Antonopoulou¹ & Elena Sarropoulou²

Development requires the implementation of a plethora of molecular mechanisms, involving a large set of genes to ensure proper cell differentiation, morphogenesis of tissues and organs as well as the growth of the organism. Genome duplication and resulting paralogs are considered to provide the raw genetic materials important for new adaptation opportunities and boosting evolutionary innovation. The present study investigated paralogous genes, involved in three-spined stickleback (*Gasterosteus aculeatus*) development. Therefore, the transcriptomes of five early stages comprising developmental leaps were explored. Obtained expression profiles reflected the embryo's needs at different stages. Early stages, such as the morula stage comprised transcripts mainly involved in energy requirements while later stages were mostly associated with GO terms relevant to organ development and morphogenesis. The generated transcriptome profiles were further explored for differential expression of known and new paralogous genes. Special attention was given to *hox* genes, with *hoxa13a* being of particular interest and to pigmentation genes where *itgb1*, involved in the melanophore development, displayed a complementary expression pattern throughout studied stages. Knowledge obtained by untangling specific paralogous gene functions during development might not only significantly contribute to the understanding of teleost ontogenesis but might also shed light on paralogous gene evolution.

The three-spined stickleback (*Gasterosteus aculeatus*), a small fresh-water teleost species, has been used for many years as a fish model species. Studies regarding three-spined stickleback developmental stages were significantly enhanced by the ability to manipulate spawning in captivity¹. To date and to the best of our knowledge, gene expression analysis during ontogenesis was performed only in the late developmental stages of the three-spined stickleback (3 days post-hatching, dph) after maternal exposure to predation risk². Investigations of the molecular background during the early development of the three-spined stickleback involved only the expression of specific genes, such as *hox* genes, in relation to the axial formation of the embryos³, as well as of *fgf8* co-orthologs⁴. In non-mammalian vertebrates, a considerable number of developmental studies were performed using the model fish, zebrafish (*Danio rerio*)^{5,6}. Nevertheless, zebrafish has been mainly used as a model species in human medical research, and may not be sufficient to unravel all differences in the developmental processes among vertebrates and especially among teleosts^{7,8}. In addition, the phylogenetic position of the three-spined stickleback⁹ may be of advantage for knowledge transfer to other species belonging to the Eupercaria which comprise most of the prevalent species in Mediterranean aquaculture.

With regard to the molecular toolbox of the three-spined stickleback, an excess of genome, as well as transcriptome data, has been continuously accumulated. Yet, mis- or un-annotated genes may be present, especially considering the teleost-specific whole genome duplication (TGD) event. The TGD has been estimated to have occurred 320–400 Ma ago^{10,11} and about 15% of the duplicated genes, namely paralogs, have been retained in the genome of teleosts¹². In addition, duplicated genes may undergo functional divergence and altered selective constraints^{13,14}. Consequently, preserved paralogs may either maintain the same function, a quota of the

¹Department of Zoology, School of Biology, Faculty of Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece. ²Institute for Marine Biology, Biotechnology and Aquaculture (IMBBC), Hellenic Centre for Marine Research (HCMR), Heraklion, Greece. ³Centre for Environment Fisheries and Aquaculture Science, (Cefas), Weymouth, UK. ⁴Institut National de la Recherche Agronomique (INRA), Génétique et Amélioration des Fruits et Légumes (GALF), Montfavet Cedex, France. Correspondence and requests for materials should be addressed to E.S. (email: sarris@hcmr.gr)

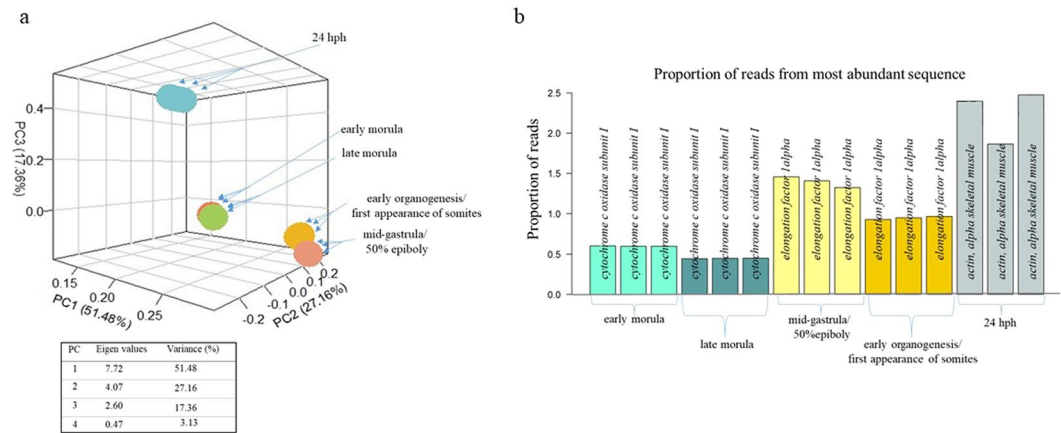


Figure 1. (a) Principal component analysis of developmental stages and (b) the most abundant sequence of each sample, as a proportion of the total reads. (a) PCA grouped together the 3 biological replicates for each developmental stage. Eigen values, as well as the percentage of variance covered by the first four components, are also shown. Three principal components with cumulative percentage of variance explained: >95%, were chosen. (b) Replicates of each stage, as well as neighbouring stages, displayed identical most abundant transcripts. These transcripts were: *cytochrome oxidase subunit I* for both early and late morula stage, for mid-gastrula/50% epiboly and early organogenesis/first appearance of somites and actin alpha, skeletal muscle for the last studied stage: 24 hph.

original function (sub-functionalization), or acquire a completely new role (neo-functionalization)¹⁵. These processes underline the importance of investigating paralogous genes in a broad range of biological processes, and especially teleost-specific paralogs are discussed in functional studies which investigated their expression in, for instance, sex determination¹⁶, hormonal system¹⁷, osmoregulation¹⁸ as well as during development^{15,19}.

In the present study, we examined the transcriptomic profiles of five key developmental stages in the three-spined stickleback, ranging from early morula to 24 hours post-hatching (hph). Therefore, high throughput sequencing was carried out followed by transcript annotation and differential gene expression analysis. The subsequent enrichment analysis aimed to reveal the presence of known and unknown genes playing a pivotal role during development. The generated transcriptome profiles were further explored for known and unknown paralogous genes differentially expressed during three-spined stickleback development. Emphasis was given to two groups of genes: *hox* genes, a group of duplicated genes with a fundamental role in the development of all bilateral animals, as well as to a group showing high expression plasticity, i.e., genes involved in body pigmentation processes.

Results

Data description. Sequencing of 15 libraries corresponding to three biological replicates of the five developmental stages, produced ~275 million reads. After quality trimming, ~230 million reads (~85%) were used for downstream analysis (Supplementary Table S1). The majority of the trimmed reads were successfully mapped to the genome of the three-spined stickleback with an average percentage among stages of ~77%. Successfully mapped reads were used to generate a transcriptome assembly consisting of 101,792 transcripts. The generated transcriptome assembly (Supplementary Table S2) was used as a reference transcriptome in the present study. Out of the 101,792 mapped-to-genome transcripts, 22,892 (~22.5%) were assigned to three-spined stickleback genes already identified and characterized. Blastx search against the nr database of NCBI of the generated reference transcriptome resulted in 37,413 (~36.7%) annotated transcripts.

Evaluation of obtained data matrix (read counts). Evaluation of the read counts was performed by two different approaches: (i) PCA analysis (Fig. 1a) and (ii) bar plots illustrating arithmetic information relevant to transcripts (most abundant transcripts, null counts, and sum of all counts) (Figs 1b and S1). PCA analysis revealed that the three replicates of early morula and the three replicates of late morula had almost identical principal component coordinates (Fig. 1a). On the contrary, replicates of the other three stages were well separated.

Concerning transcript abundance analysis, *cytochrome oxidase subunit I* was identified as the most abundant transcript in early and late morula. Similarly, *elongation factor 1-alpha* ($ef\alpha$) was found to be most abundant in mid-gastrula/50% epiboly and early organogenesis/first appearance of somites, while in the 24 hph stage, *actin-alpha skeletal muscle* was found to be the most abundant transcript (Fig. 1b). Additionally, the proportion of null counts, as well as total read counts, were similar among the replicates of each stage (Supplementary Fig. S1).

Differential expression analysis. Pairwise comparison of each stage with the other four stages (loop-design) resulted in 10 datasets (one for each comparison). The number of transcripts after each comparison for different padj and log2 fold change $\geq |2|$ is shown in Table 1. In the present study the most stringent parameters (transcripts with padj < 0.005 and with log2 fold change $\geq |2|$) were considered as differentially expressed. Hierarchical clustering of all differentially expressed transcripts grouped early developmental stages under one sub-branch,

padj log2 fold change	late morula				mid-gastrula/50% epiboly				early organogenesis/first appearance of somites				24 hph			
	<0.05	<0.01	<0.005	<0.005 ≥ 2	<0.05	<0.01	<0.005	<0.005 ≥ 2	<0.05	<0.01	<0.005	<0.005 ≥ 2	<0.05	<0.01	<0.005	<0.005 ≥ 2
early morula	6,662	4,607	4,114	328	35,447	30,538	29,015	20,339	42,820	37,133	33,876	24,513	49,028	43,315	39,925	33,455
late morula					35,255	30,414	28,867	20,390	41,136	35,661	33,909	24,744	47,602	41,835	39,984	33,685
mid-gastrula/50% epiboly									24,220	19,565	18,234	8,087	38,611	32,709	30,850	23,622
early organogenesis/first appearance of somites													32,227	26,420	24,685	16,036

Table 1. Number of transcripts differentially expressed between stages for different padj threshold as well as for log2 fold change $\geq|2|$.

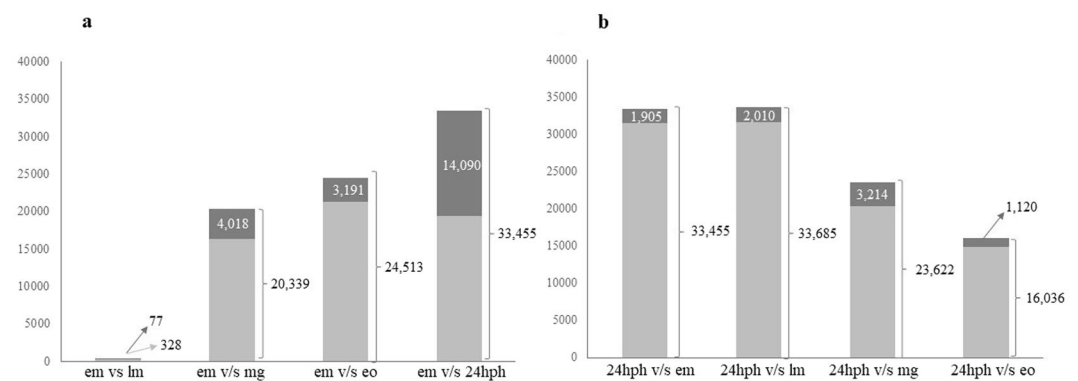


Figure 2. Number of differentially expressed transcripts between two extreme and the rest of the four stages studied. **(a)** Differentially expressed transcripts between the first studied stage (early morula) or **(b)** the last studied stage (24 hph) and the remaining four stages. The number of unique transcripts differentially expressed are shown in dark grey.

the next two under a second sub-branch and the last studied stage (24 hph) was placed on its own in a separate branch (Supplementary Fig. S2). Differentially expressed transcript abundance between the four stages and the two extreme stages studied is shown in Fig. 2 i.e., between the four stages and the early morula (Fig. 2a) and between the four stages and 24 hph (Fig. 2b). Early and late morula comparison produced the lowest number of differentially expressed transcripts (328) while comparing early morula with 24 hph resulted in the highest number of differentially expressed transcripts (33,455). The number of differentially expressed transcripts between 24 hph and early and late morula were similar, while numbers between 24 hph and early organogenesis/first appearance of somites were the lowest among all the comparisons made with 24 hph as a reference (Fig. 2b). Having early morula instead of 24 hph as the reference stage produced higher numbers of unique differentially expressed transcripts (presented in dark grey in Fig. 2a) characterizing each pairwise comparison.

Differentially expressed transcripts were grouped into 11 modules comprising 49,776 transcripts (99.76% of differentially expressed transcripts). About 92% of all transcripts were represented in the first four modules (Fig. 3). Moreover, the average of the expression of transcripts from each module through the developmental stages is shown. Bold lines correspond to four different modules consisting of transcripts that were mainly upregulated either in early and late morula stages (module 1), or mid-gastrula/50% epiboly stage (module 2), or early organogenesis/first appearance of somites stage (module 3) or 24 hph stage (module 4). The expression patterns of each transcript of the four modules are illustrated in the form of heat maps in Fig. 4.

Meta-analysis of differentially expressed transcripts. Modules 1, 2, 3 and 4 were further studied by enrichment analysis. Therefore each of the modules served as a test set and all of the annotated transcripts as a reference set. Enrichment analysis resulted in distinct GO terms, for each of the modules (Supplementary Table S3). Shared and unique GO terms which are present in the four modules are visualized in the form of a Venn diagram (Fig. 5). Module 4 (transcripts mainly upregulated at 24 hph) included the highest number of GOs, followed by module 2 (transcripts upregulated at mid-gastrula/50% epiboly), module 3 (transcripts upregulated at early organogenesis/first appearance of somites) and module 1 (transcripts upregulated at early and late morula). Unique GO terms found in each of the modules 1, 2, 3 and 4 are listed in Supplementary Table S4.

Paralog identification in the transcriptome of three-spined stickleback embryos. In total 2,455 candidate paralogs were identified among differentially expressed transcripts (Supplementary Table S5). Paralogous genes, known to have a decisive role in embryogenesis and later development from studies in other

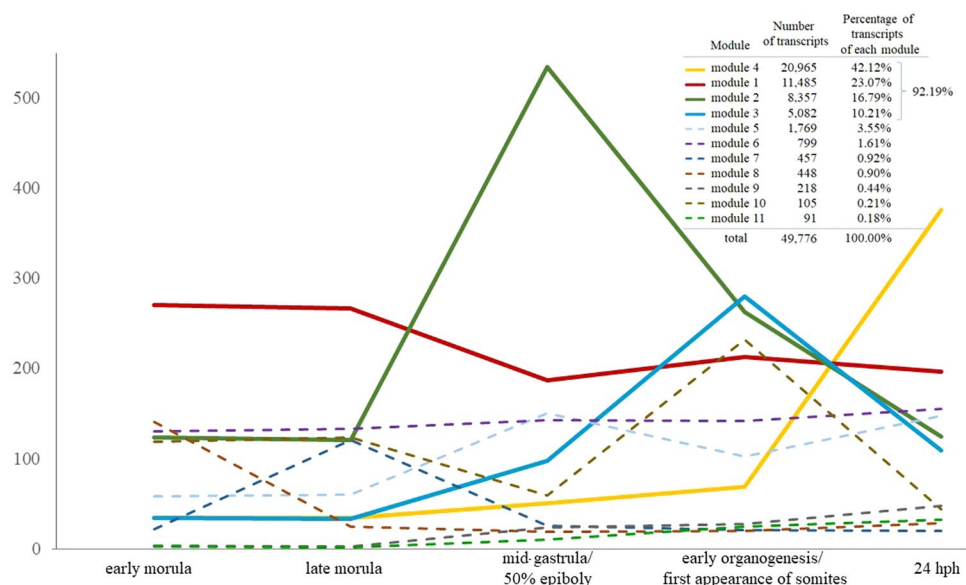


Figure 3. The expression pattern of differentially expressed transcripts of each of the 11 modules. Module 1, 2, 3, and 4 are shown in bold lines and were further analysed with enrichment analysis. Expression pattern corresponding to the rest of the modules are shown as discontinuous lines. Y-axis shows the average normalized counts number of all transcripts of each module. The total number of transcripts included in each module as well as their percentage against all clustered transcripts are also shown.

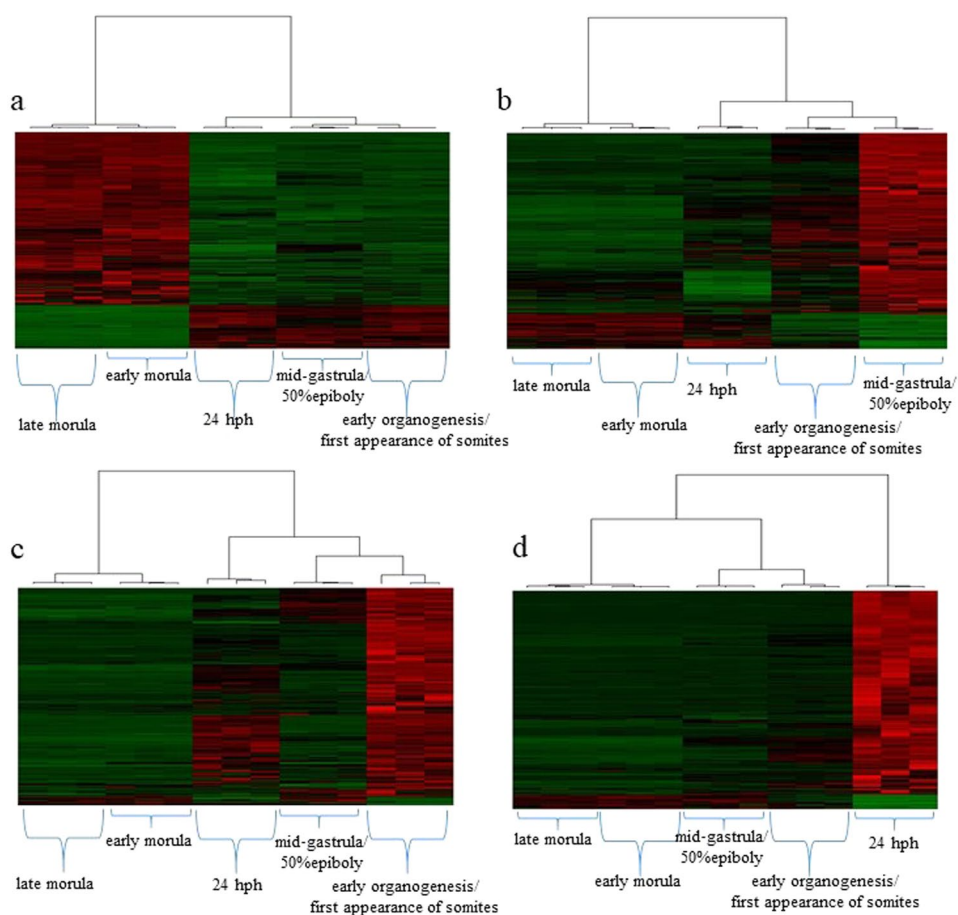


Figure 4. Heat maps illustrating the expression pattern of differentially expressed transcripts of modules (a) 1, (b) 2, (c) 3 and (d) 4. Each row corresponds to a transcript and each column to a biological sample. Upregulation is shown in red colour while downregulation is shown in green.

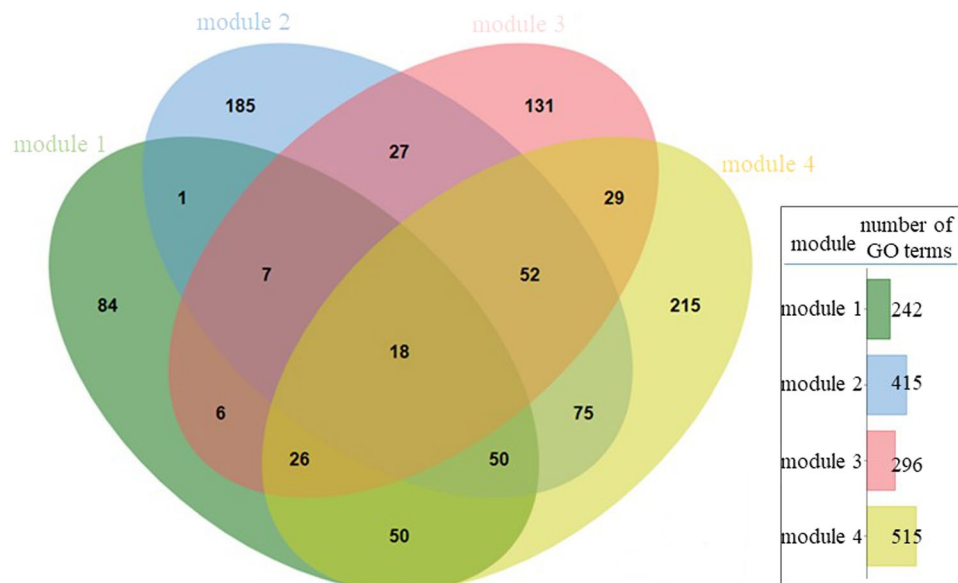


Figure 5. Venn diagram of GO terms under biological process, which are included in modules 1, 2, 3, and 4. The total number of GO terms included in each of the above modules is shown.

teleost species, were searched in our transcriptome and are shown in Tables 2, 3 and Supplementary Table S6. Table 2 contains transcripts with a role in development for which both paralogs were found. On the other hand, Table 3 contains transcripts that were found only in a single copy in the generated transcriptome of the present study but do have a paralogous pair, if not in the stickleback, then in another fish species. Transcripts expressed differentially between at least two stages are noted with an asterisk.

First, we focused on the presence of *hox* paralogous genes, which belong to a well-established group of duplicated genes with evolutionary and developmental interest²⁰. *Hox* genes were searched among all annotated transcripts. In the present study, all genes of both *hoxa* clusters of three-spined stickleback were present. On the contrary, genes of *hox* clusters *b*, *c* and *d* (*hoxb1a*, *-2a*, *-3b*, *-6b*, *-7a*, *-8a*, *hoxc3*, *hoxd11a* and *-11b*) were not found (Supplementary Table S7). The expression patterns of duplicated *hox* genes are illustrated in Fig. 6. The majority of the studied *hox* genes were more highly expressed at early organogenesis/first appearance of somites stage and at 24 hph stage. During the earlier developmental stages (early and late morula) only *hoxa13a* were expressed, while the other members of the family were roughly present. Similarly, expression levels of another group of genes, those involved in body pigmentation, and of their paralogs when present, were also searched in the annotated dataset. The list of genes that participate in body pigmentation (Supplementary Table S8), as well as their categorization depending on the pigmentation pathway (melanin-, pteridine- and iridophore- related genes), was based on the study of Braasch *et al.*²¹. Figure 7 illustrates the expression pattern of pigmentation genes found to be duplicated in the present study. Eight of the genes studied had similar expression patterns with their paralog (*hdacl*, *en1*, *sox9*, *erbb3*, *silv*, *mcoln3*, *mitf*, and *csf1r*). Notably, decreased expression of one paralog concurred with increased levels of the other paralog in the case of *tyr*, *rab38*, *gja5* and *itgb1* of the melanin-pigment related genes and *spr* of the pteridin-pigment related genes.

Discussion

The present study investigated the expression patterns of five key developmental stages of the three-spined stickleback with a focus on paralogous genes. To primarily establish defined stage-specific “expression fingerprints”, the transcriptome patterns during early embryogenesis were explored. The generated transcriptome profiles were used to evaluate the sampling and sequencing procedures. First of all, the three biological replicates of each stage as illustrated in Fig. 1a were grouped together and separated from the other developmental stages validating effective staging and distinct transcriptomic profiles for each stage. Furthermore, a similar percentage of null counts and numbers of reads for each sample indicated comparable sequencing data between replicates of each stage and between stages (Supplementary Fig. S1).

To characterize the functionality of a tissue, it has been shown that a short list of the 10 most expressed transcripts is sufficient²². Abundance analysis in the present study revealed that the most abundant transcript in the first two stages studied (early and late morula) was the *cytochrome oxidase I*, most likely serving the metabolic needs for energy requirements through a period of intense divisions. The increased metabolic demands of the early and late morula were in agreement with the observation that, after fertilization, fish embryo oxygen consumption increases rapidly²³. The increase in the number of cells during early and late morula, as the first embryonic task immediately after fertilization, is also shown by the over-representation of genes involved in molecular mechanisms related to the cell cycle, mitosis (e.g. regulation of G2/M transition of mitotic cell cycle) and chromatin separation (e.g. centromere complex assembly, attachment of spindle microtubules to kinetochore) (Supplementary Table S4). Similar results were shown in a zebrafish transcriptome profile study, during 1–16 and

Gene ID	Chr	References	Gene ID	Chr	References
<i>barhl1</i> (-/*)	scf. 1455	51	<i>bmp1</i> (*/*)	XIII	52
	XIV			XIV	
<i>bmp7</i> (*/*)	XVII	53	<i>casq1</i> (*/*)	III scf. 713	54
	scf. 68				
<i>casq2</i> (-/*)	XVI	54	<i>dmbx1</i> (*/*)	III	55
	I			VIII	
<i>ext1</i> (*/*)	X	56	<i>foxc1</i> (*/*)	III	57
	XX			VIII	
<i>gpi</i> (*/*)	II	58	<i>foxb1</i> (*/*)	XIX	57
	XIX			II	
<i>hey1</i> (*/*)	XX	59	<i>hce1</i> (*/*/*)	XVIII	60
				scf. 48	
	X			scf. 692	
<i>jamb</i> (-/*)	I	61	<i>janc</i> (*/*)	VII	61
	XVI			I	
<i>nanos1</i> (*/*)	IV	62	<i>nrf1</i> (*/-)	V	63
	XV			XI	
<i>pax2</i> (-/*)	VI	64	<i>ppar alpha</i> (*/*)	IV	65
	V			XIX	
<i>tle3 (groucho3)</i> (*/-)	XIX	66	<i>snai1</i> (*/-)	XVII	67
	II			scf. 180	
<i>wnt4</i> (*/*)	XVII	68	<i>pou2f2</i> (*/-)	XX	69
	XX			X	
<i>pou3f2</i> (*/*)	XX	69	<i>stat1</i> (*/*)	I	70
	XVIII			XVI	
<i>tbx5</i> (*/*)	XIV	71	<i>lhx1</i> (*/*)	IX	72
	XIII			VI	
<i>wt1</i> (*/*)	XIX	73	<i>adam17</i> (*/-)	scf. 269	21
	II			XV	
<i>hdac1</i> (*/*)	X	21	<i>en1</i> (*/*)	XVI	21
	XV			I	
<i>gja5</i> (*/-)	VI	21	<i>sox9</i> (*/*)	V	21
	XVI			XI	
<i>tyr</i> (*/*)	I	21	<i>mitf</i> (*/-)	XVII	21
	VII			XII	
<i>rab38</i> (*/*)	I	21	<i>erbb3</i> (*/*)	scf. 27	21
	VII			XII	
<i>silv (pmel)</i> (*/*)	scf. 27	21	<i>itgb1</i> (*/*)	XXI	21
	XII			scf. 169	
<i>mcoln3</i> (-/*)	VIII	21	<i>spr</i> (*/*)	XIII	21
	III			XIV	
<i>csflr</i> (*/*)	IV	21	<i>ghr</i> (*/*)	XIII	21
	VII			XIV	
<i>hoxa2a</i> *	X	31	<i>hoxa9a</i> *	X	31
<i>hoxa2b</i> *	XX		<i>hoxa9b</i> *	XX	
<i>hoxa10a</i> *	X	31	<i>hoxa11a</i> *	X	31
<i>hoxa10b</i> *	XX		<i>hoxa11b</i> *	XX	
<i>hoxa13a</i> *	X	31	<i>hoxb5a</i> *	XI	31
<i>hoxa13b</i> *	XX		<i>hoxb5b</i> *	V	
<i>hoxd4a</i> *	XI	31	<i>hoxd9a</i> *	XI	31
<i>hoxd4b</i> *	VI		<i>hoxd9b</i> *	VI	

Table 2. Transcripts annotated as paralogous genes with a known role in teleost development, where both paralogs were detected in the present study. Transcripts that were differentially expressed between at least two stages are noted with an asterisk.

Two paralogs found in three-spined stickleback in Ensemble and/or Fish-it			One paralog found in three-spined stickleback in Ensemble and/or Fish-it		
Gene ID ^b	Chr.	References	Gene ID ^b	Chr.	References
<i>st8sia3*</i>	XIV (XIII)	74	<i>ppar beta*</i>	XII	65
<i>ahr1*</i>	I (XVI)	75	<i>uts2a*</i>	XII	76
<i>dag1*</i>	XVII (XII)	77	<i>gli2*</i>	XVI	28
<i>marcks*</i>	XVIII (XV)	78	<i>mustn1b*</i>	XVII	79
<i>ncam1*</i>	VII (I)	80	<i>bmp2b*</i>	XVIII	81
<i>stard13</i>	VII (I)	82	<i>cart1*</i>	XV	83
<i>otx2*</i>	XV (XVIII)	84	<i>zic2a*</i>	XVI	21
<i>otx5*</i>	I (VII)	84	<i>mgrn1b*</i>	XI	21
<i>tshr</i>	XVIII (XIV)	85	<i>hoxc6α*</i>	XII	31
<i>nme2a*</i>	scf48 (Un)	86	<i>hoxc11α*</i>	XII	31
<i>irf4</i>	VIII (III)	87	<i>hoxc12α*</i>	XII	31
<i>prox1*</i>	XVIII (VII)	88	<i>hoxc13α*</i>	XII	31
<i>rab32*</i>	XVIII (XV)	21			
<i>pabpc1*</i>	X (XX)	21			
<i>ednrb1*</i>	III (XVI)	21			
<i>atp6v1e1*</i>	IV (XIX)	21			
<i>kit</i>	VIII (IX)	21			
<i>frem2*</i>	I (VII)	21			
<i>myo7a</i>	I (VII)	21			
<i>trpm1*</i>	XIX (II)	21			
<i>egfr*</i>	III (scf 182)	21			
<i>hoxb1b*</i>	V (XI)	31			
<i>hoxb3a*</i>	XI (V)	31			
<i>hoxb6a</i>	XI (V)	31			

Table 3. Transcripts annotated as paralogous genes with a known role in teleost development, where only one paralog was detected in the present study. Transcripts that were differentially expressed between at least two stages are noted with an asterisk.

512 cell stage²⁴. In the third stage studied (the mid-gastrula/50% epiboly stage), the most expressed transcript was *efα*, which is related to the translational machinery²⁵ (Fig. 1b). Similarly in the zebrafish at 50% of epiboly stage, *efα* was also among the 10 most expressed transcripts, serving the translational needs of the embryo²⁴. In addition, GO terms unique in module 2 i.e., transcripts mainly upregulated during the mid-gastrula/50% epiboly, also signalled the next developmental leap: the initiation of organ development (Supplementary Table S4). Although biological processes that set the basis for organogenesis were already present in mid-gastrula/50% epiboly, from what is known, the major part of organogenetic processes is accomplished during the next developmental stage, the organogenesis²⁶. Consistent with this, the transcripts mainly upregulated in early organogenesis/first appearance of somites stage (module 3, Fig. 4c) in the three-spined stickleback were mostly associated with GO terms relevant to organ development and morphogenesis (e.g. cardiac muscle, appendage, kidney vasculature, pectoral fin, neuron projection, blood vessel, rhombomere 4) (Supplementary Table S4). The last studied stage in the present work was revealed to be transcriptionally the most diverse one, as shown by the high degree of differentiation of this stage in the PCA analysis (Fig. 1a) as well as in the high total number of differentially expressed genes compared to all the other stages (Fig. 2). A major developmental step acquired at this stage is that movement is not a reflex but a conscious process. The detection of *actin* as the most abundant transcript in the three-spined stickleback's 24 hph stage (Fig. 1b), appears to be related to the shift from the immobile embryo to the free-moving larvae since actin is a major component of skeletal muscle.

Having established the molecular bases in the form of differentially expressed transcripts among distinct developmental stages, paralogous genes either with different, similar or identical expression patterns among them were identified. Paralogs found in the present study and known to be involved in developmental processes are listed in Tables 2 and 3. Table 2 comprises the cases where both paralogs were found to be present and differentially expressed, while in Table 3 only those paralogs are listed where only one of the paralogs was detected. The occurrence of a single paralog may indicate that only one of the paralogs has a functional role during early development. On the other hand, twelve transcripts were found as a single copy in the present dataset as well as in any of the publicly available databases of the three-spined stickleback. As both paralogs were found in other fish species (Table 3), such as the zebrafish, it may be hypothesized that they have lost their counterpart in the three-spined stickleback. Zebrafish belongs to the Ostariophysi while the three-spined stickleback belongs to the Acanthopterygii. The two superorders were dated to have split approximately 217 ± 4 Ma ago²⁷. In the group of Acanthopterygii eleven of the transcripts were found only in one copy (*mustn1b*; *uts2a*; *bmp2b*, *ppar db/ppar bb*, *cart1*, *mgrn1b*, *zic2a*, *hoxc6a*, *hoxc11a*, *hoxc12a* and *hoxc13a*), while in the Ostariophysi its paralogs (*mustn1a*; *uts2b*; *bmp2a*; *ppar da/ppar ba*, *cart1 second copy*, *mgrn1a*, *zic2b*, *hoxc6b*, *hoxc11b*, *hoxc12b* and *hoxc13b*) were detected. It has recently been shown that

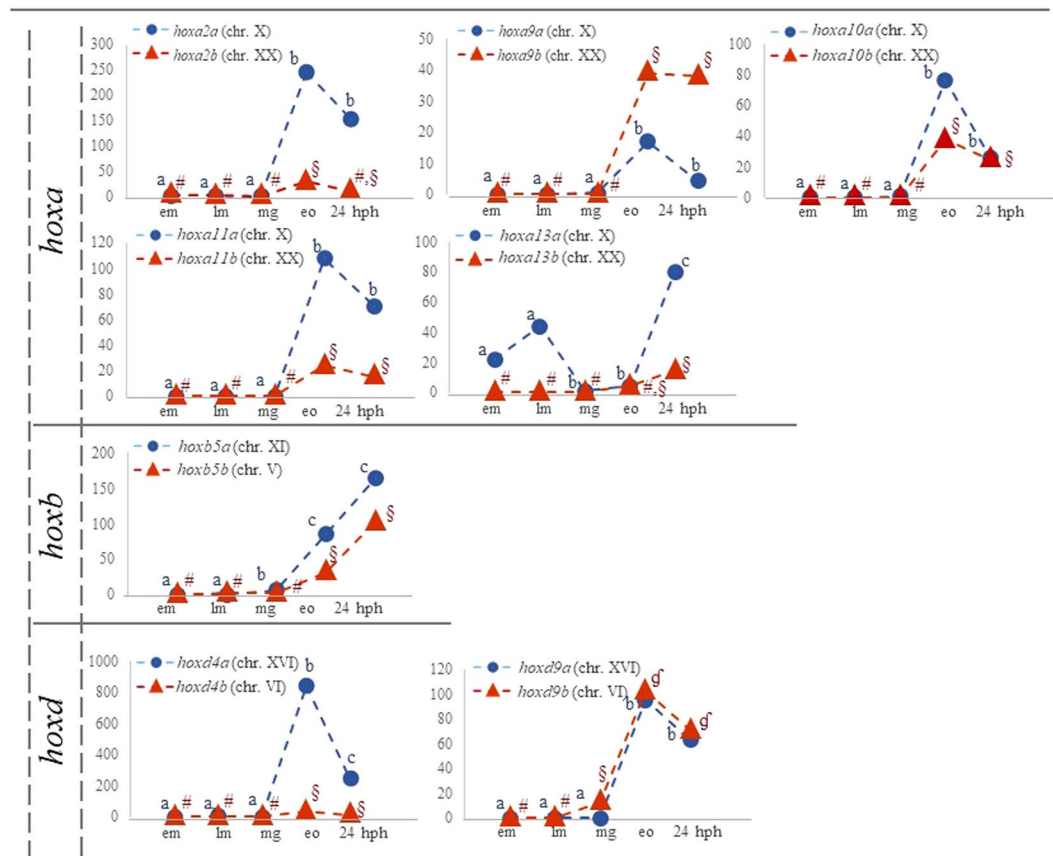


Figure 6. Expression pattern of *hoxa*, *hoxb* and *hoxd* paralogs during three-spined stickleback development. The y-axis shows normalized count values and the x-axis shows the five developmental stages studied. The expression pattern of the alpha paralog is illustrated with blue and of the beta paralog with orange line. The chromosome where each paralog was mapped is shown in parenthesis. Significant different expression values are marked with (a–c) in case of alpha paralog, and #, \$, ? in case of beta paralogs.

the Ostariophysi and the Acanthopterygii have a different paralogous genes retention rate, and seven of the twelve genes i.e., *bmp2*, *mgrn1*, *ppard/pparb*, *hoxc6*, *hoxc11*, *hoxc12* and *hoxc13* belong to the lineage-specific paralogs (LSP)⁷. Thus, this may also explain the absence of *uts2b*, *cart1 second copy*, *zic2b* and *mustn1a* in the three-spined stickleback. With regard to *gli2*, its paralog has been identified in other teleosts belonging to the Acanthopterygii but to the best of our knowledge, not in the three-spined stickleback (Supplementary Fig. S3). *Gli2* is known to be a major mediator of the hedgehog signalling pathway in early development²⁸. Studies in zebrafish have hypothesized that the paralog *gli2a* is superfluous in teleost fish²⁹. However, with only one paralog retained i.e., *gli2a*, clearly the alpha paralog is indispensable in the three-spined stickleback.

Early development is a key life period of any organism including teleosts and the *hox* genes are among the well-studied duplicated genes crucial for proper development³⁰. So far, in the three-spined stickleback, 48 *hox* genes have been identified³¹, located in seven clusters (*aa*, *ab*, *ba*, *bb*, *ca*, *da* and *db*)³². In the present study 39 *hox* genes were found, out of which 16 were paralogs with specific expression patterns as illustrated in Fig. 6. In general, both paralogs showed similar expression pattern with higher expression of the *hoxa* alpha paralogs (chromosome X) in four cases (*hoxa2a*, *hoxa10a*, *hoxa11a*, *hoxa13a*), while in one case the beta paralog (*hoxa9b*; chromosome XX) was only slightly more highly expressed. With the exception of *hoxa13a* (located on chromosome X), low or no expression of the *hox* genes under study was detected during the early stages. The majority of *hox* genes reached a peak at early organogenesis/first appearance of somites stage when the shaping of the body into distinct parts is initiated. This may be justified by the fact that the functional role of *hox* genes lies in the determination and specification of the body segments³³.

Further investigations in the present study were focused on the differentially expressed genes involved in pigmentation, where for eight genes, (*mitf* (both paralogs), *csf1r* (both paralogs), *spr* (paralog chromosome XIV), *tyr* (paralog chromosome VII), *rab 38* (paralog chromosome VII) and *ghr* (paralog chromosome XIII)) one distinct expression peak is seen at the early organogenesis/first appearance of somites (Fig. 7). Among them, three genes (*ghr*²¹, *csf1r*³⁴, and *spr*³²) are involved in the pteridin pigment synthesis, one of the two major synthetic pathways of pigments in the teleost. For each of the three paralog pairs, one paralog was found to be predominantly expressed. For example, the *ghr* paralog in chromosome XIII was already elevated at mid-gastrula/50% epiboly stage, when the neural crest is formed. In medaka, it has been shown that *ghra* (according to²¹ homologue to *ghr*

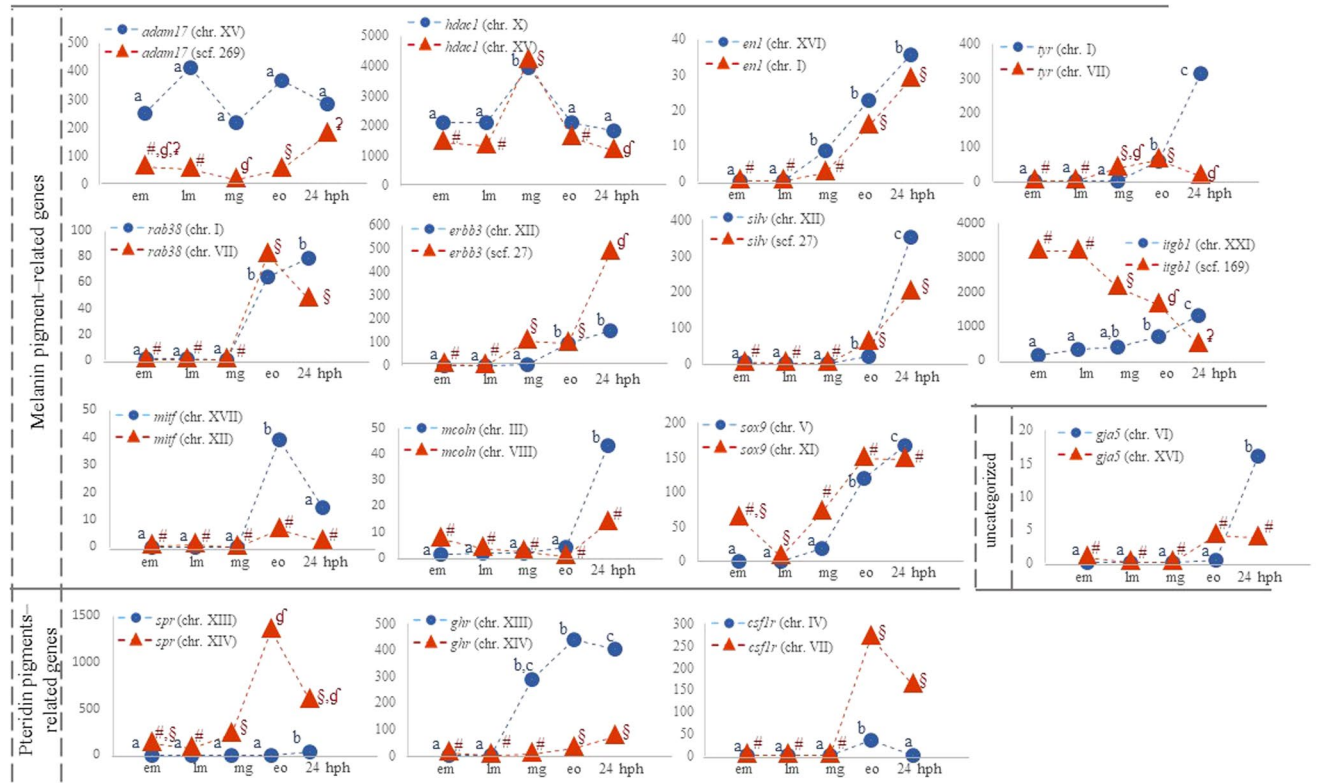


Figure 7. Expression pattern of genes implicated in animal pigmentation, that were found during three-spined stickleback's development in more than one copy. The y-axis shows normalized count values and the x-axis shows the five developmental stages studied. The chromosome where each paralog was mapped is shown in parenthesis. Significant different expression values are marked with (a–c) in case of alpha paralog, and #, \$, \$, \$ in case of beta paralogs.

paralog in chromosome XIII) binds more effectively somatolactin- α , a pituitary secreted hormone that is implicated in chromatophore development³⁵.

Concerning the genes involved in the melanin-pigment synthesis, the second major synthetic pathway of pigments in the teleost, predominant expression of only one paralog has been detected for *miif*³⁶ (paralog on chromosome XVII) and *tyr* (paralog on chromosome I) with an expression peak at early organogenesis/first appearance of somites and at stage 24hph respectively (Fig. 7). This may be explained by the need to serve pigmentation processes as e.g. *tyr* is expressed in developing retina early in zebrafish embryogenesis and preceding melanin accumulation by a few hours³⁷. Zebrafish, however, possesses only a single *tyr* gene located in chromosome 15 (the homologous group to the three-spined stickleback I)²¹.

Opposite expression patterns were found in two of the paralog pairs i.e., *gja5*, and *itgb1*. Concerning the complementary expression of *gja5* at the 24hph stage, it has been shown that the beta paralog is involved in an adult mutant zebrafish leading to a spotted (instead of striped) pattern. On the other hand, knockdown of the alpha paralog had no effect on the zebrafish skin pattern³⁸. In the case of *itgb1*, also known as *cd29*, a particular pattern was observed. The two paralogs displayed a complementary expression throughout the developmental stages studied. *Itgb1* encodes for a cell surface receptor and is involved in the melanophore development²¹. Even though *itgb1* has been connected to melanocyte distribution and normal skin pigmentation in humans³⁹, expression studies in teleost concerning their role in skin colour are absent. In mammals, it has been reported that *itgb1* plays a role in the junctions between germ and Sertoli cells during spermatogenesis⁴⁰. The present finding that the one *itgb1* paralog is highly expressed at the early stages but decreases later on, while the other paralog increases, could provide novel significant insights worthy of further investigation and might pinpoint to a possible paralog sub-functionalization event.

Conclusion

In this paper, we demonstrated that distinct transcriptome profiles amongst developmental stages exist in the three-spined stickleback. The adjacent stages exhibited more similar expression patterns whereas the more distant stages revealed completely different profiles. We further identified paralogs differentially expressed during ontogenesis and demonstrated that paralogous genes, which are known to be involved in teleost development, also have varying expression patterns, with one of the paralogs being dominant and, in some cases, even the only one present. Untangling the specific paralog functions might not only significantly contribute to the understanding of teleost ontogenesis but might also shed light on paralogous genes' evolution.

Materials and Methods

The entire workflow followed in the present study is graphically shown in Supplementary Fig. S4.

Ethical statement. Sampling was carried out at Cefas according to UK legislation. Up until the time point of the samples in our study, sticklebacks have yolk sacs and as such are not considered to be free feeding. Under UK legislation, fish embryos become protected under the Animals (Scientific Procedures) Act from the free feeding stage onwards.

Sampling. Three-spined stickleback fertilised eggs and embryos were collected in triplicate at different time periods from the well-established laboratory colony maintained at the Centre for Environment, Fisheries and Aquaculture Science, at Weymouth, UK. Eggs were collected from a single gravid female via mild abdominal pressure and were fertilised by the sperm of a single male within 1-minute post collection. The fertilised eggs were examined for quality under a microscope and placed in 1 L glass flasks containing dechlorinated freshwater, under aeration (air stone) at 18 °C, 30 minutes post fertilisation. In total, five developmental stages were collected comprising of the: i) early morula, ii) late morula, iii) mid-gastrula/50% epiboly, iv) early organogenesis/first appearance of somites and v) 24 hours post-hatching. The staging observation was performed according to Swarup⁴¹. After designation of the exact developmental stage, all samples were immediately frozen in liquid nitrogen and stored at −80 °C until they were dispatched in dry ice to the Institute of Marine Biology, Biotechnology and Aquaculture at the Hellenic Centre for Marine Research in Crete, Greece.

Total RNA extraction. Total RNA of all samples was extracted using Nucleospin miRNA kits (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. Briefly, pools of eggs or embryos were disrupted with a mortar and pestle in liquid nitrogen and homogenized in lysis buffer by passing the lysate through a 23-gauge (0.64 mm) needle five times. The quantity of extracted RNA was estimated using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). The RNA quality was further evaluated by agarose (1%) gel electrophoresis, as well as by capillary electrophoresis using the RNA Pico Bioanalysis chip (Agilent 2100 Bioanalyzer, Agilent Technologies, Santa Clara, CA 95051, USA). Samples with an RNA integrity number value between 8.9 and 9.9 were used for library construction.

Library preparation and sequencing. Fifteen mRNA libraries (three libraries per developmental stage) were prepared using Truseq stranded mRNA library preparation kit (Illumina, 5200 Illumina Way, San Diego, CA 92122 USA). The magnetic-bead-assisted mRNA purification was followed by mRNA fragmentation. First and second strand synthesis produced cDNA, which was ligated with different adapters, one for each library. Libraries were then amplified by PCR and validated by capillary electrophoresis using a High Sensitivity DNA chip (Bioanalyzer, Agilent Technologies). The quantification of each library was performed by qPCR, using the Kappa Library Quantification kit (Kappa Biosystems, Wilmington, MA 01887, USA). Libraries were paired-end sequenced (150 bp) over two lanes using Illumina HiSeq vs. 2500 at the Norwegian Sequencing Centre, Oslo, Norway.

Read quality, mapping to genome and transcriptome assembly. Sequencing reads generated for each library were initially checked for their quality using FastQC (v0.11.5)⁴². For adaptor sequences as well as low-quality nucleotide reads removal, Trimmomatic (v. 0.32)⁴³ was used. Cleaned sequencing reads were mapped to the three-spined stickleback genome (*Gasterosteus aculeatus*.BROADS1.dna.toplevel.fa, http://ftp.ensembl.org/pub/release-76/fasta/gasterosteus_aculeatus/dna/) applying CRAC (v. 2.5.0)⁴⁴, a software for mapping RNA sequencing reads to the genome with high precision in splice junctions. Genome-guided assembly was performed using Stringtie (v. 1.2.2) assembler⁴⁵ in two steps: the first step assembled the mapped reads of each RNA-Seq sample separately. The resulting assemblies were then merged, at the second step, into one transcriptome containing unique sequences.

Annotation. Annotation of the assembled transcriptome was performed by i) mapping reads to the genome of the three-spined stickleback with default parameters, ii) submission to blastx search against the nr database of NCBI with E-value < 1 10^{−8} and iii) using Blast2GO platform (v 4.1.9) with default parameters⁴⁶. Blast2GO platform was also used for gene ontology (GO) terms assignment.

Differential expression. Trimmed reads of each stage were mapped to the constructed transcriptome with the RSEM estimation method (with align and estimate abundance.pl script of Trinity, r20140717); (parameters used: "RF" library type, "fq" sequence type, bowtie as aligning method)⁴⁷. The generated count matrix served as an input file for differential expression analysis, following the DESeq2 pipeline⁴⁸ integrated into SARTools⁴⁹. After pair-wise comparison, transcripts with padj < 0.005 and log2fold change ≥ |2| between two stages were considered as differentially expressed.

Data evaluation. Principal component analysis (PCA) was used to evaluate the quality of the whole data matrix using plot3D function in R. Furthermore, sample clustering was performed with WGCNA (Weighted Correlation Network Analysis) to detect outliers and group all differentially expressed transcripts according to their expression pattern, using the "block-wise network construction and module detection" R script which is appropriate for large datasets⁵⁰. Heat maps were constructed with heatmap.2 function in R, using normalized count data of differentially expressed transcripts.

Enrichment analysis. Two-tailed enrichment analysis with default parameters (filter value: 0.05 filter mode: FDR, two-tailed) was performed applying Blast2GO (v 4.1.9) with padj ≤ 0.05 filter⁴⁶. All the annotated

transcripts were used as a reference dataset and modules 1, 2, 3 and 4 that counted the majority of the transcripts (Fig. 3) were used as test set.

Paralog identification. To detect paralogous genes, extra gene copies were examined in all differentially expressed transcripts between different developmental stages of the three-spined stickleback's developmental stages. In the present study, two genes were considered as putative paralogs if they combined the following: i) they shared identical or similar (-like) annotation terms and ii) they mapped in different chromosomes or linkage groups. In addition, candidate paralogs were searched among three-spined stickleback paralogs in the Ensembl Compara database.

Data Availability

All reads resulting from Illumina sequencing were submitted to the public database of SRA (Sequence Read Archive) of NCBI under the PRJNA395155 Bioproject, comprising 15 biosamples (one biosample for each library) coded from SAMN07374938 (early morula, replicate 1) to SAMN7374952 (24 hph, replicate 3). This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GHCM00000000. The version described in this paper is the first version, GHCM01000000.

References

- Barber, I. & Arnott, S. A. Split-clutch ivf: a technique to examine indirect fitness consequences of mate preferences in sticklebacks. *Behaviour* (2000).
- Mommer, B. C. & Bell, A. M. Maternal experience with predation risk influences genome-wide embryonic gene expression in threespined sticklebacks (*Gasterosteus aculeatus*). *PLoS One* **9**, e98564 (2014).
- Ahn, D. G. & Gibson, G. Expression patterns of threespine stickleback Hox genes and insights into the evolution of the vertebrate body axis. *Dev. Genes Evol.* (1999).
- Jovelin, R. *et al.* Duplication and Divergence of fgf8 Functions in Teleost Development and Evolution. *J. Exp. Zool. Part B Mol. Dev. Evol.* **308B**, 730–743 (2007).
- Kimmel, C. B. Genetics and early development of zebrafish. *Trends in Genetics* **5**, 283–288 (1989).
- White, R. J. *et al.* A high-resolution mRNA expression time course of embryonic development in zebrafish. *Elife* (2017).
- Garcia de la Serrana, D., Mareco, E. A. & Johnston, I. A. Systematic variation in the pattern of gene paralog retention between the teleost superorders Ostariophysi and Acanthopterygii. *Genome Biol. Evol.* **6**, 981–987 (2014).
- Schartl, M. Beyond the zebrafish: diverse fish species for modeling human disease. *Dis. Model. Mech.* **7**, 181–192 (2014).
- Betancur-R, R. *et al.* The Tree of Life and a New Classification of Bony Fishes. *PLoS Curr. Tree Life* **0732988**, 1–45 (2013).
- Christoffels, A. *et al.* Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* (2004).
- Vandepoele, K., De Vos, W., Taylor, J. S., Meyer, A. & Van de Peer, Y. Major events in the genome evolution of vertebrates: Paralog age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl. Acad. Sci.* (2004).
- Braasch, I. & Postlethwait, J. H. In *Polyploidy and Genome Evolution* (2012).
- Holland, P. W., Garcia-Fernández, J., Williams, N. A. & Sidow, A. Gene duplications and the origins of vertebrate development. *Dev. Suppl.* (1994).
- Ohno, S. *Evolution by gene duplication*. (New York: Springer-Verlag, 1970).
- Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
- Mank, J. E. & Avise, J. C. Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sex. Dev.* **3**, 60–67 (2009).
- Roch, G. J., Wu, S. & Sherwood, N. M. Hormones and receptors in fish: Do duplicates matter? *Gen. Comp. Endocrinol.* **161**, 3–12 (2009).
- Cerdà, J. & Finn, R. N. Piscine aquaporins: An overview of recent advances. *J. Exp. Zool. Part A Ecol. Genet. Physiol.* **313 A**, 623–650 (2010).
- Kaitetzidou, E., Xiang, J., Antonopoulou, E., Tsigenopoulos, C. S. & Sarropoulou, E. Dynamics of gene expression patterns during early development of the European seabass (*Dicentrarchus labrax*). *Physiol. Genomics* (2015).
- Hrycaj, M. S. & Wellik, M. D. Hox genes and chordate evolution. *F1000Research* **5** (2016).
- Braasch, I., Brunet, F., Volf, J. N. & Schartl, M. Pigmentation Pathway Evolution after Whole-Genome Duplication in Fish. *Genome Biol. Evol.* (2010).
- Nishida, Y., Mayumi, Y. & St-Amand, J. The top 10 most abundant transcript are sufficient to characterize the organs functional specificity: evidences from the cortex, hypothalamus and pituitary gland. *Gene* **344**, 133–141 (2005).
- Boulekbache, H. Energy Metabolism in Fish Development I. *Am. Zool.* **21**, 377–389 (1981).
- Vesterlund, L., Jiao, H., Unneberg, P., Hovatta, O. & Kere, J. The zebrafish transcriptome during early development. *BMC Dev. Biol.* **11**, 30 (2011).
- Condeelis, J. Elongation factor 1 alpha, translation and the cytoskeleton. *Trends Biochem. Sci.* **20**, 169–170 (1995).
- Gilbert, S. F. *Developmental Biology*. (Sinauer Associates Inc., U.S.; 6th Revised edition edition, 2000).
- Steinke, D., Salzburger, W. & Meyer, A. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *J. Mol. Evol.* (2006).
- Ke, Z., Emelyanov, A., Lim, S. E. S., Korzh, V. & Gong, Z. Expression of a novel zebrafish zinc finger gene, gli2b, is affected in Hedgehog and Notch signaling related mutants during embryonic development. *Dev. Dyn.* **232**, 479–486 (2005).
- Wang, X. *et al.* Targeted inactivation and identification of targets of the Gli2a transcription factor in the zebrafish. *Biol. Open* (2013).
- Hoegg, S. & Meyer, A. Hox clusters as models for vertebrate genome evolution. *Trends in Genetics* (2005).
- Hoegg, S., Boore, L. J., Kuehl, V. J. & Meyer, A. Comparative phylogenomic analyses of teleost fish Hox gene clusters: Lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* **8**, 317 (2007).
- Kuraku, S. & Meyer, A. The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. *Int. J. Dev. Biol.* **53**, 765–773 (2009).
- Kmita, M. & Duboule, D. Organizing axes in time and space; 25 years of collinear thinking. *Science (80-)*. **301**, 331–333 (2003).
- Ziegler, I., McDonald, T., Hesslinger, C., Pelletier, I. & Boyle, P. Development of the pteridine pathway in the zebrafish, *Danio rerio*. *J. Biol. Chem.* (2000).
- Komine, R. *et al.* Transgenic medaka that overexpress growth hormone have a skin color that does not indicate the activation or inhibition of somatolactin- α signal. *Gene* (2016).
- Bejar, J. Mitf expression is sufficient to direct differentiation of medaka blastula derived stem cells to melanocytes. *Development* (2003).
- Camp, E. & Lardelli, M. Tyrosinase gene expression in zebrafish embryos. *Dev. Genes Evol.* **211**, 150–153 (2001).
- Watanabe, M. *et al.* Spot pattern of leopard Danio is caused by mutation in the zebrafish connexin41.8 gene. *EMBO Rep.* **7**, 893–897 (2006).

39. Swope, V. B., Supp, A. P., Schwemberger, S., Babcock, G. & Boyce, S. Increased expression of integrins and decreased apoptosis correlate with increased melanocyte retention in cultured skin substitutes. *Pigment Cell Res.* **19**, 424–433 (2006).
40. Siu, M. K. Y., Mruk, D. D., Lee, W. M. & Cheng, C. Y. Adhering junction dynamics in the testis are regulated by an interplay of β 1-integrin and focal adhesion complex-associated proteins. *Endocrinology* (2003).
41. Anatomy, C. Stages in the Development of the Stickleback *Gasterosteus aculeatus* (L.). *J. Embryology Exp. Morphol.* **6**, 373–383 (1958).
42. Andrews, S. FastQC: A quality control tool for high throughput sequence data, [Http://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/](http://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/) (2010).
43. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
44. Philippe, N., Salson, M., Combes, T. & Rivals, E. CRAC: an integrated approach to the analysis of RNA-seq reads. *Genome Biol.* **14**, R30 (2013).
45. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
46. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
47. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biol.* **15**, 1–21 (2014).
49. Varet, H., Brilllet-Guéguen, L., Coppée, J.-Y. & Dillies, M.-A. SARTools: A DESeq. 2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS One* **11**, e0157022 (2016).
50. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
51. Schuhmacher, L. N., Albadri, S., Ramialison, M. & Poggi, L. Evolutionary relationships and diversification of barhl genes within retinal cell lineages. *BMC Evol. Biol.* **11**, 340 (2011).
52. Huxley-Jones, J. *et al.* The evolution of the vertebrate metzincins: Insights from *Ciona intestinalis* and *Danio rerio*. *BMC Evol. Biol.* **7**, 1–20 (2007).
53. Shawi, M. & Serluca, F. C. Identification of a BMP7 homolog in zebrafish expressed in developing organ systems. *Gene Expr. Patterns* **8**, 369–375 (2008).
54. Infante, C., Ponce, M. & Machado, M. Duplication of casequestrin genes in teleosts: Molecular characterization in the Senegalese sole (*Solea senegalensis*). *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **158**, 304–314 (2011).
55. Wong, L., Weadick, C. J., Kuo, C., Chang, B. S. W. & Tropepe, V. Duplicate dmbx1 genes regulate progenitor cell cycle and differentiation during zebrafish midbrain and retinal development. *BMC Dev. Biol.* **10**, 100 (2010).
56. Siekmann, A. F. & Brand, M. Distinct tissue-specificity of three zebrafish ext1 genes encoding proteoglycan modifying enzymes and their relationship to semitic Sonic Hedgehog signaling. *Dev. Dyn.* **232**, 498–505 (2005).
57. Zhao, X. F., Suh, C. S., Prat, C. R., Ellingsen, S. & Fjose, A. Distinct expression of two foxg1 paralogues in zebrafish. *Gene Expr. Patterns* **9**, 266–272 (2009).
58. Lin, W. W., Chen, L. H., Chen, M. C. & Kao, H. W. Differential expression of zebrafish gpia and gpib during development. *Gene Expr. Patterns* **9**, 238–245 (2009).
59. Zhou, M. *et al.* Comparative and evolutionary analysis of the HES/HEY gene family reveal exon/intron loss and teleost specific duplication events. *PLoS One* **7** (2012).
60. Kawaguchi, M. *et al.* Evolution of hatching enzyme genes in teleost. *Zool. Sci.* **22**, 1437–1438 (2005).
61. Powell, G. T. & Wright, G. J. Jamb and jamc are essential for vertebrate myocyte fusion. *PLoS Biol.* (2011).
62. Aoki, Y., Nakamura, S., Ishikawa, Y. & Tanaka, M. Expression and Syntenic Analyses of Four *nanos* Genes in Medaka. *Zoolog. Sci.* **26**, 112–118 (2009).
63. Sant, K. E. *et al.* The role of Nrf1 and Nrf2 in the regulation of glutathione and redox dynamics in the developing zebrafish embryo. *Redox Biol.* **13**, 207–218 (2017).
64. Bassham, S., Cañestro, C. & Postlethwait, J. H. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. *BMC Biol.* **6**, 1–17 (2008).
65. Den Broeder, M. J., Kopylova, V. A., Kamminga, L. M. & Legler, J. Zebrafish as a Model to Study the Role of Peroxisome Proliferating-Activated Receptors in Adipogenesis and Obesity. *PPAR Res.* **2015** (2015).
66. Aghaallaei, N., Bajoghli, B., Walter, I. & Czerny, T. Duplicated members of the Groucho/Tle gene family in fish. *Dev. Dyn.* **234**, 143–150 (2005).
67. Liedtke, D., Erhard, I. & Schartl, M. Snail gene expression in the medaka, *Oryzias latipes*. *Gene Expr. Patterns* **11**, 181–189 (2011).
68. Nicol, B., Guerin, A., Fostier, A. & Guiguen, Y. Ovary-predominant *wnt4* expression during gonadal differentiation is not conserved in the rainbow trout (*Oncorhynchus mykiss*). *Mol. Reprod. Dev.* **79**, 51–63 (2012).
69. Brombin, A. *et al.* Genome-wide analysis of the POU genes in medaka, focusing on expression in the optic tectum. *Dev. Dyn.* **240**, 2354–2363 (2011).
70. Song, H., Yan, Y. lin, Titus, T., He, X. & Postlethwait, J. H. The role of *stat1b* in zebrafish hematopoiesis. *Mech. Dev.* (2011).
71. Parrie, L. E. *et al.* Zebrafish *tbx5* paralogs demonstrate independent essential requirements in cardiac and pectoral fin development. *Dev. Dyn.* **242**, 485–502 (2013).
72. Wotton, K. R., Weierud, F. K., Dietrich, S. & Lewis, K. E. Comparative genomics of Lbx loci reveals conservation of identical Lbx ohnologs in bony vertebrates. *BMC Evol. Biol.* **8**, 1–15 (2008).
73. Bollig, F. *et al.* Identification and comparative expression analysis of a second *wnt1* gene in zebrafish. *Dev. Dyn.* **235**, 554–561 (2006).
74. Bentrop, J., Marx, M., Schattschneider, S., Rivera-Milla, E. & Bastmeyer, M. Molecular evolution and expression of zebrafish *St8SialII*, an alpha-2,8-sialyltransferase involved in myotome development. *Dev. Dyn.* **237**, 808–818 (2008).
75. Jenny, M. J. *et al.* Distinct roles of two zebrafish AHR repressors (AHRRA and AHRRB) in embryonic development and regulating the response to 2,3,7,8-Tetrachlorodibenzo-p-dioxin. *Toxicol. Sci.* **110**, 426–441 (2009).
76. Parmentier, C. *et al.* Occurrence of two distinct urotensin II-related peptides in zebrafish provides new insight into the evolutionary history of the urotensin II gene family. *Endocrinology* **152**, 2330–2341 (2011).
77. Pavoni, E. *et al.* Duplication of the dystroglycan gene in most branches of teleost fish. *BMC Mol. Biol.* **8**, 34 (2007).
78. Ott, L. E. *et al.* Two myristoylated alanine-rich C-kinase substrate (MARCKS) paralogs are required for normal development in zebrafish. *Anat. Rec.* (2011).
79. Østbye, T. K. K. *et al.* Myostatin (MSTN) gene duplications in Atlantic salmon (*Salmo salar*): Evidence for different selective pressure on teleost MSTN-1 and -2. *Gene* **403**, 159–169 (2007).
80. Langhauser, M. *et al.* *Ncam1a* and *Ncam1b*: Two carriers of polysialic acid with different functions in the developing zebrafish nervous system. *Glycobiology* **22**, 196–209 (2012).
81. Wise, S. B. & Stock, D. W. Conservation and divergence of *Bmp2a*, *Bmp2b*, and *Bmp4* expression patterns within and between dentitions of teleost fishes. *Evol. Dev.* **8**, 511–523 (2006).
82. Teng, H. *et al.* Genome-wide identification and divergent transcriptional expression of STAR-related lipid transfer (START) genes in teleosts. *Gene* **519**, 18–25 (2013).

83. Bonacic, K., Martínez, A., Martín-Robles, Á. J., Muñoz-Cueto, J. A. & Morais, S. Characterization of seven cocaine- and amphetamine-regulated transcripts (CARTs) differentially expressed in the brain and peripheral tissues of *Solea senegalensis* (Kaup). *Gen. Comp. Endocrinol.* (2015).
84. Suda, Y. *et al.* Evolution of Otx paralogue usages in early patterning of the vertebrate head. *Dev. Biol.* **325**, 282–95 (2009).
85. Ponce, M., Infante, C. & Machado, M. Molecular characterization and gene expression of thyrotropin receptor (TSHR) and a truncated TSHR-like in Senegalese sole. *Gen. Comp. Endocrinol.* **168**, 431–439 (2010).
86. Desvignes, T., Pontarotti, P., Fauvel, C. & Bobe, J. Nme protein family evolutionary history, a vertebrate perspective. *BMC Evol. Biol.* (2009).
87. Ai, K. *et al.* Expression pattern analysis of IRF4 and its related genes revealed the functional differentiation of IRF4 paralogues in teleost. *Fish Shellfish Immunol.* **60**, 59–64 (2017).
88. Deguchi, T., Fujimori, K. E., Kawasaki, T., Ohgushi, H. & Yuba, S. Molecular cloning and gene expression of the prox1a and prox1b genes in the medaka. *Oryzias latipes*. *Gene Expr. Patterns* **9**, 341–347 (2009).

Acknowledgements

Financial support for this study has been provided by the Ministry of Education and Religious Affairs, under the Call “ARISTEIA I” of the National Strategic Reference Framework 2007–2013 (ANnOTATE), co-funded by the EU and the Hellenic Republic through the European Social Fund. We would further like to thank the Informatics group of IMBBC for computational support.

Author Contributions

E.K. carried out laboratory experiments, sequence analysis, meta-analysis, prepared the figures and drafted the manuscript, I.K. provided developmental stages of three-spined stickleback, J.L. supported bio-informatic analysis, E.A. contributed in manuscript writing, E.S. conceived the study, were in charge of overall direction and planning as well as manuscript writing.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-40127-2>.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019