

OPEN

Developmental transcriptomes of the sea star, *Patiria miniata*, illuminate how gene expression changes with evolutionary distance

Tsvia Gildor¹, Gregory A. Cary², Maya Lalzar³, Veronica F. Hinman² & Smadar Ben-Tabou de-Leon^{1*}

Understanding how changes in developmental gene expression alter morphogenesis is a fundamental problem in development and evolution. A promising approach to address this problem is to compare the developmental transcriptomes between related species. The echinoderm phylum consists of several model species that have significantly contributed to the understanding of gene regulation and evolution. Particularly, the regulatory networks of the sea star, *Patiria miniata* (*P. miniata*), have been extensively studied, however developmental transcriptomes for this species were lacking. Here we generated developmental transcriptomes of *P. miniata* and compared these with those of two sea urchins species. We demonstrate that the conservation of gene expression depends on gene function, cell type and evolutionary distance. With increasing evolutionary distance the interspecies correlations in gene expression decreases. The reduction is more severe in the correlations between morphologically equivalent stages (diagonal elements) than in the correlation between morphologically distinct stages (off-diagonal elements). This could reflect a decrease in the morphological constraints compared to other constraints that shape gene expression at large evolutionary divergence. Within this trend, the interspecies correlations of developmental control genes maintain their diagonality at large evolutionary distance, and peak at the onset of gastrulation, supporting the hourglass model of phylotypic stage conservation.

Embryo development is controlled by regulatory programs encoded in the genome and executed during embryogenesis¹. Genetic changes in these programs that occur in evolutionary time scales lead to alterations in body plans and ultimately, to biodiversity¹. Comparing developmental gene expression between diverse species can illuminate evolutionary conservation and changes that underlie morphological similarity and divergence. To understand how developmental programs change with increasing evolutionary distances it is necessary to compare closely related and further diverged species within the same phylum.

The echinoderm phylum provides an excellent system for comparative studies of developmental gene expression dynamics. Echinoderms have two types of feeding larvae: the pluteus-like larvae of sea urchins and brittle stars, and the auricularia-like larvae of sea cucumbers and sea stars². Of these, the sea urchin and the sea star had been extensively studied both for their embryogenesis and their gene regulatory networks^{3–13}. Sea urchins and sea stars diverged from their common ancestor about 500 million years ago, yet their endoderm lineage and the gut morphology show high similarity between their embryos⁶. On the other hand, the mesodermal lineage diverged to generate novel cell types in the sea urchin embryo (Fig. 1A)^{8,13,14}. Specifically, the skeletogenic mesoderm lineage, which generates the larval skeletal rods that underlie the sea urchin pluteus morphology¹³, and the mesodermal pigment cells that give the sea urchin larva its red pigmentation (arrow and arrowheads in Fig. 1A). Thus, there are both morphological similarities and differences between sea urchin and sea star larval body plans that make these classes very interesting for comparative genetic studies.

The models of the gene regulatory networks that control the development of the sea urchin and those that control cell fate specification in the sea star are the state of the art in the field^{3–11}. The endodermal and ectodermal

¹Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa, 31905, Israel.

²Departments of Biological Sciences and Computational Biology, Carnegie Mellon University, Pittsburgh, PA, 15213, USA. ³Bionformatics Core Unit, University of Haifa, Haifa, 31905, Israel. *email: sben-tab@univ.haifa.ac.il

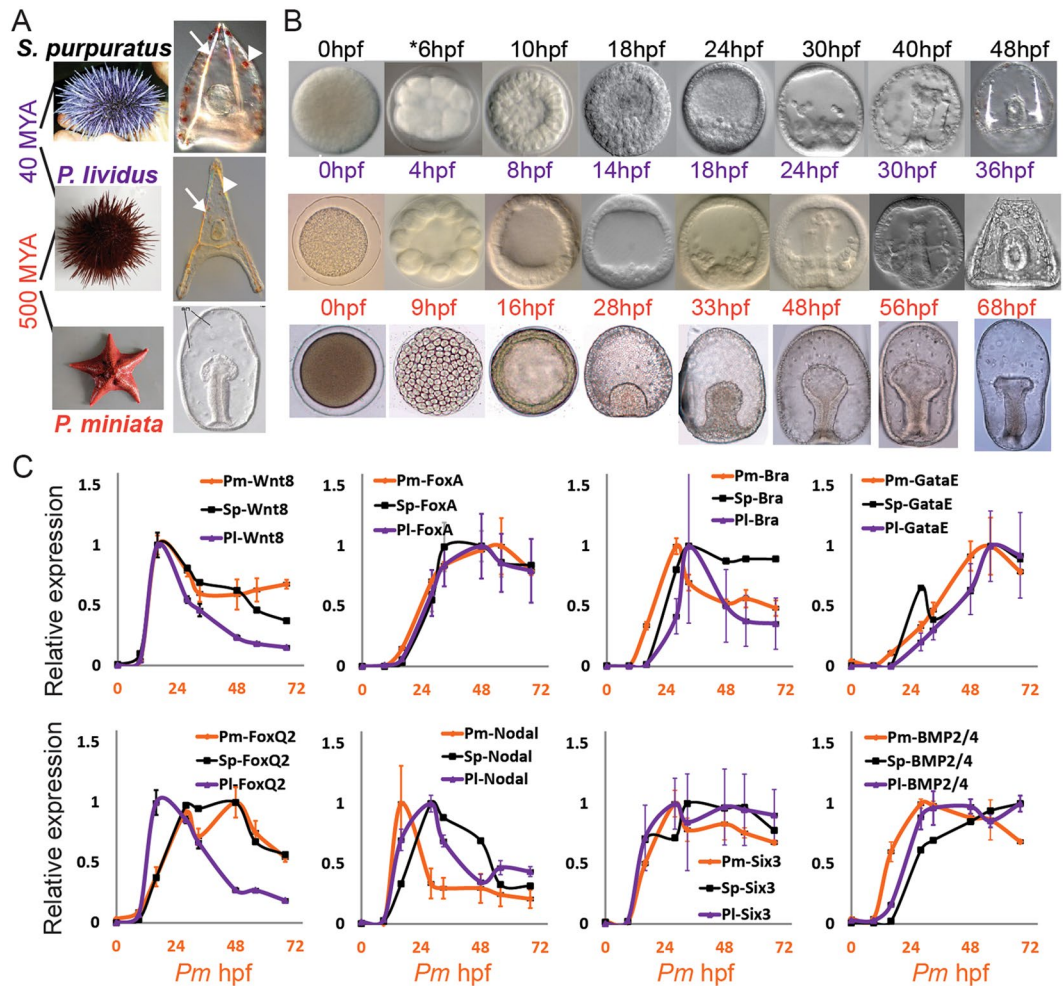


Figure 1. Developmental time points studied and examples for gene expression profiles on the three species. (A) Images of adult and larval stage of *P. miniata*, *P. lividus* and *S. purpuratus*. Arrows point to the sea urchin skeletogenic rods and arrow heads point to the sea urchin pigments. (B) Images of *P. miniata*, *P. lividus* and *S. purpuratus* embryos at the developmental stages that were studied in this work. Time point 6 hpf in *S. purpuratus* does not have RNA-seq data. (C) Relative gene expression in the three species measured in the current paper by RNA-seq for *P. miniata* (orange curves), by RNA-seq for *P. lividus*¹⁹ (purple curves) and by nanostring for *S. purpuratus*²⁸ (black curves). Error bars indicate standard deviation. To obtain relative expression levels for each species we divide the level at each time point in the maximal mRNA level measured for this species in this time interval; so 1 is the maximal expression in this time interval.

gene regulatory networks show high levels of conservation between the sea urchin and the sea star in agreement with the overall conserved morphology of these two germ layers^{6,9,11}. Surprisingly, most of the transcription factors active in the skeletogenic and the pigment mesodermal lineages are also expressed in the sea star mesoderm^{7,8}. The expression dynamics of key endodermal, mesodermal and ectodermal regulatory genes were compared between the Mediterranean sea urchin, *Paracentrotus lividus* (*P. lividus*) and the sea star, *Patiria miniata* (*P. miniata*)¹⁵. Despite the evolutionary divergence of these two species, an impressive level of conservation of regulatory gene expression in all the embryonic territories was observed. This could suggest that novel mesodermal lineages diverged from an ancestral mesoderm through only a few regulatory changes that drove major changes in downstream gene expression and embryonic morphology^{7,8}.

Insight on the genome-wide changes in developmental gene expression can be gained from comparative transcriptome studies of related species^{16–21}. We previously investigated different aspects of interspecies conservation of gene expression between *P. lividus* and the pacific sea urchin, *Strongylocentrotus purpuratus* (*S. purpuratus*)^{18,19}. These species diverged from their common ancestor about 40 million years ago and have a highly similar embryonic morphology (Fig. 1A). We observed high conservation of gene expression dynamics of both genes that regulate developmental processes and of housekeeping genes (e.g., ribosomal and mitochondrial genes)¹⁹. Yet, the interspecies correlations of the expression levels of these two sets of genes show distinct patterns.

Before we describe the observed patterns, we would like to describe the structure and biological meaning of the two dimensional correlation matrix between different developmental stages in two species^{19,22,23}. In such a matrix, the columns correspond to the developmental stages in one species and the rows to the developmental

stages in the other species. The diagonal elements of the matrix, have the same column and row indices, ii , and show the correlations in gene expression between morphologically equivalent developmental stages in the two species. The off-diagonal elements of the matrix, have distinct column and row indices, $ij, i \neq j$, and present the correlations in gene expression between morphologically distinct developmental stages. With increasing evolutionary distance, it becomes harder to properly identify morphologically equivalent stages. However, usually it is still possible for animals in the same phylum and straightforward for the comparison of the mentioned two sea urchin species¹⁹.

Thus, for *P. lividus* and *S. purpuratus*, the interspecies correlations of the expression levels of developmental genes are high between morphologically similar stages and decrease sharply between diverse developmental stages, resulting with highly diagonal correlation matrices¹⁹. The correlations peak at mid-development, at the onset of gastrulation, in agreement with the hourglass model of developmental conservation^{19–21,24,25}. Conversely, the interspecies correlations of housekeeping gene expression increase with developmental time and are high between all post-hatching time points, resulting with distinct off-diagonal elements of the correlation matrix¹⁹. This indicates that the expression levels of housekeeping genes are correlated throughout the developmental of the two species, irrespective of specific morphological stages; which is reasonable, as these genes are expressed in all the cells throughout development. This could suggest that when the off-diagonal elements of the interspecies correlation matrix are of the same scale as the diagonal, the correlation in gene expression might reflect cellular and not development constraints.

Apparent differences in the conservation patterns between developmental and housekeeping genes were observed in other studies of closely related species²⁴ and were a reason to exclude housekeeping genes from comparative studies of developing embryos²⁵. Yet, embryogenesis progression depends on the dynamic expression of housekeeping genes and therefore we believe that these genes and their contribution to the overall correlation patterns should be considered¹⁹. Here we aim to decipher how the interspecies correlations in gene-expression change with increasing evolutionary distance. To this end, we generated and analyzed *de-novo* developmental quantitative transcriptomes of the sea star, *P. miniata*, and compared them with the published developmental transcriptomes of *P. lividus*¹⁹ and *S. purpuratus*²⁶ at equivalent developmental stages (Fig. 1B). Our studies illuminate how the correlations of gene expression levels change with increasing evolutionary distance in both correlation strength and correlation matrix diagonality for different functional classes and embryonic territories.

Results

Developmental transcriptomes of the sea star, *P. miniata*. To study the transcriptional profiles of the sea star species, *P. miniata*, from fertilized egg to late gastrula stage and compare them to those in the sea urchin species, *S. purpuratus* and *P. lividus* we collected embryos at eight developmental stages matching to those studied the sea urchin species^{19,26,27} (Fig. 1B). Details on reference transcriptome assembly, quantification and annotations are provided in the Materials and Methods section. Quantification and annotations of all identified *P. miniata* transcripts are provided in Dataset S2. The temporal expression profiles of selected regulatory genes show variable levels of similarity between *P. miniata* (RNA-seq, current study) and the sea urchin species, *S. purpuratus* (nanosttring²⁸) and *P. lividus* (RNA-seq¹⁹, Fig. 1C). The temporal expression of some genes seem to be highly conserved (e.g., *foxa* and *six3*) while others show some divergence (e.g., *nodal* and *bra*) in agreement with our previous study (Fig. 1C)¹⁴.

Identification of 1:1:1 homologous gene set and quantitative data website. To compare developmental gene expression between the sea star, *P. miniata*, and the two sea urchins, *P. lividus* and *S. purpuratus* we identified 8735 1:1:1 putative homologous genes, as described in the Materials and Methods section. Quantification and annotations of all these homologous genes in *P. miniata*, *P. lividus* and in *S. purpuratus* are provided in Dataset S3 (based on¹⁹ for *P. lividus* and on^{26,27} for *S. purpuratus*). Within this set, 6593 genes were expressed in the three species, 1093 genes are expressed only in the two sea urchin species, 187 genes were expressed only in the sea star, *P. miniata*, and the sea urchin, *S. purpuratus*, and 430 genes are expressed only in *P. miniata* and *P. lividus* (Fig. 2A). We looked for enrichment of gene ontology (GO) terms within these different gene sets but did not identify enrichment of specific developmental processes (GOseq²⁹ with *S. purpuratus* annotations, Fig. S1 and Dataset S4).

We uploaded the data of the 1:1:1 homologous genes expressed in the three species to Echinobase where they are available through gene search at www.echinobase.org/shiny/quantdevPm³⁰. In this Shiny web application³¹, genes can be searched either by their name or by their *P. miniata*, *S. purpuratus* or *P. lividus* transcript identification number. The application returns our quantitative measurements of gene expression for each stage of *P. miniata*, plots of transcript expression level, mRNA sequences, a link to the corresponding records in Echinobase³² and links to the loci in the *S. purpuratus* and *P. miniata* genome browsers. We hope that this resource will be useful for the community. Further analyses of the 6593 1:1:1 homologous genes expressed in all three species are described below.

Main variations in gene expression profiles are due to developmental progression and evolutionary distance.

To identify the general trends of gene expression over developmental time in the three species, we performed a non-metric multidimensional scaling analysis (NMDS, Fig. 2B). *S. purpuratus* RNA-seq data does not include the early development time point equivalent to the *P. miniata* 9 hours post fertilization (hpf) and *P. lividus* 4 hpf²⁶. The NMDS maps the developmental trajectories of the two sea urchin species closely to each other (purple and black tracks), while the sea star samples are relatively distinct (orange track). This is in agreement with the phylogenetic relationships between the three species (Fig. 1A). The developmental progression in all three species is along a similar trajectory in the NMDS two-dimensional space, possibly reflecting the resemblance in the overall morphology between these three echinoderm species. Overall, the NMDS indicates

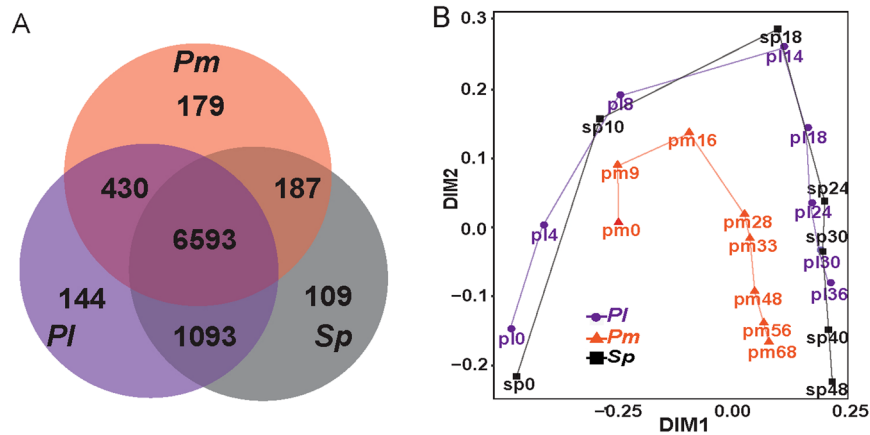


Figure 2. Venn diagram and NMDS analysis of 1:1:1 homologous genes. (A) Venn diagram showing the number of 1:1:1 homologous genes expressed in all three species, in two of the species or only in one species. (B) First two principal components of expression variation (NMDS) between different developmental time points in *P. miniata* (orange), *P. lividus* (purple) and *S. purpuratus* (black).

that major sources of variation in these data sets are developmental progression and evolutionary distance, in agreement with previous comparative analysis of the transcriptomes of three other echinoderm species¹⁷.

Interspecies correlations decrease and become less diagonal with evolutionary distance. We wanted to investigate how the pattern of the interspecies correlations of gene expression changes with evolutionary distance for different classes of genes. To be able to compare the interspecies correlations between the three species we included only time points that had data for all species. Explicitly, we excluded *P. lividus* 4 hpf and *P. miniata* 9 hpf that do not have an equivalent time point in the *S. purpuratus* data. We calculated the Pearson correlations of gene expression levels between *P. lividus* and *S. purpuratus*, between *P. lividus* and *P. miniata* (Fig. 3) and between *S. purpuratus* and *P. miniata* (Fig. S2). We did that for subsets of homologous genes with specific GO terms that describe developmental, housekeeping, response or metabolic function. As expected, the strength of the interspecies correlation decrease at increasing evolutionary distance, for all classes of genes (Figs 3 and S2). Explicitly, the correlations are stronger between the two sea urchin than between the sea urchins and the sea star (Figs 3 and S2). This is in agreement with our previous observation that the expression kinetics and initiation times of key developmental genes show higher conservation between the two sea urchins than between the sea urchins and the sea star¹⁴. Additionally, the correlations between morphologically equivalent stages become more similar to the correlations between distinct stages, that is, the *Pm-Pl* and *Pm-Sp* matrices are less diagonal compared to the *Pl-Sp* matrices (Figs 3 and S2).

For a better assessment of the correlation patterns we sought to quantify these two distinct properties: the correlation strength between equivalent developmental stages and the diagonality of the matrix. As a measure of the correlation strength between morphologically similar time points, we defined the parameter AC = Average Correlation, which is the average of the diagonal elements of the correlation matrix (the matrix elements that have the same column and row indices and show the correlations in gene expression between morphologically equivalent stages in the two species). That is, AC = 1 corresponds to perfect correlation throughout all equivalent developmental stages and AC = 0 corresponds to no correlation. Evidently, AC is not a measure of the diagonality of the matrix as it doesn't consider the off-diagonal elements (the matrix elements that have distinct column and row indices and show the correlations in gene expression between morphologically distinct stages in the two species). To quantify the diagonality of the correlation matrices we used a statistical test we developed before to assess whether the interspecies correlation pattern is significantly close to a diagonal matrix (where all the diagonal elements are 1 and the off-diagonal are 0)¹⁹. Briefly, the parameter, matrix diagonality (MD), indicates how many times within 100 subsamples of the tested set of genes, the interspecies correlation matrix was significantly close to a diagonal matrix compared to a random matrix. Hence, MD = 100 is the highest diagonality score and 0 is the lowest (see Materials and Methods section and¹⁹ for explanation, and Dataset S5 for results. MD is equivalent to the parameter 'count significant', or CS, in¹⁹). Importantly, AC will be high for any matrix that has high diagonal elements, but MD will be high only when the off-diagonal matrix elements are much lower than the diagonal.

We measured these parameters for the three comparisons, *Pm-Pl*, *Pl-Sp* and *Pm-Sp* and observed similar trends for both sea urchin–sea star comparisons (*Pm-Pl* and *Pm-Sp*, see Figs 3 and S2). Therefore, for simplicity of the discussion, in the rest of the paper we focus on the quantification results of the *Pm-Pl* and *Pl-Sp* matrices. The reason for preferring *Pm-Pl* over *Pm-Sp* is that the RNA-sequencing of *P. miniata* and *P. lividus* were conducted by us in the same facility and included three biological replicates for each time point (see Materials and Methods for *P. miniata* and¹⁹ for *P. lividus*).

In Fig. 3, we ordered the correlation matrices according to their MD value in *Sp-Pl*, from the most diagonal (Cell differentiation, MD = 99) to the least (RNA processing, MD = 17). This ordering clearly demonstrates the high diagonality of the developmental genes (Fig. 3A–C) vs. the block patterns of the housekeeping genes (Fig. 3G–J) and how the diagonality of the correlation pattern of response to stress genes, metabolic genes and all

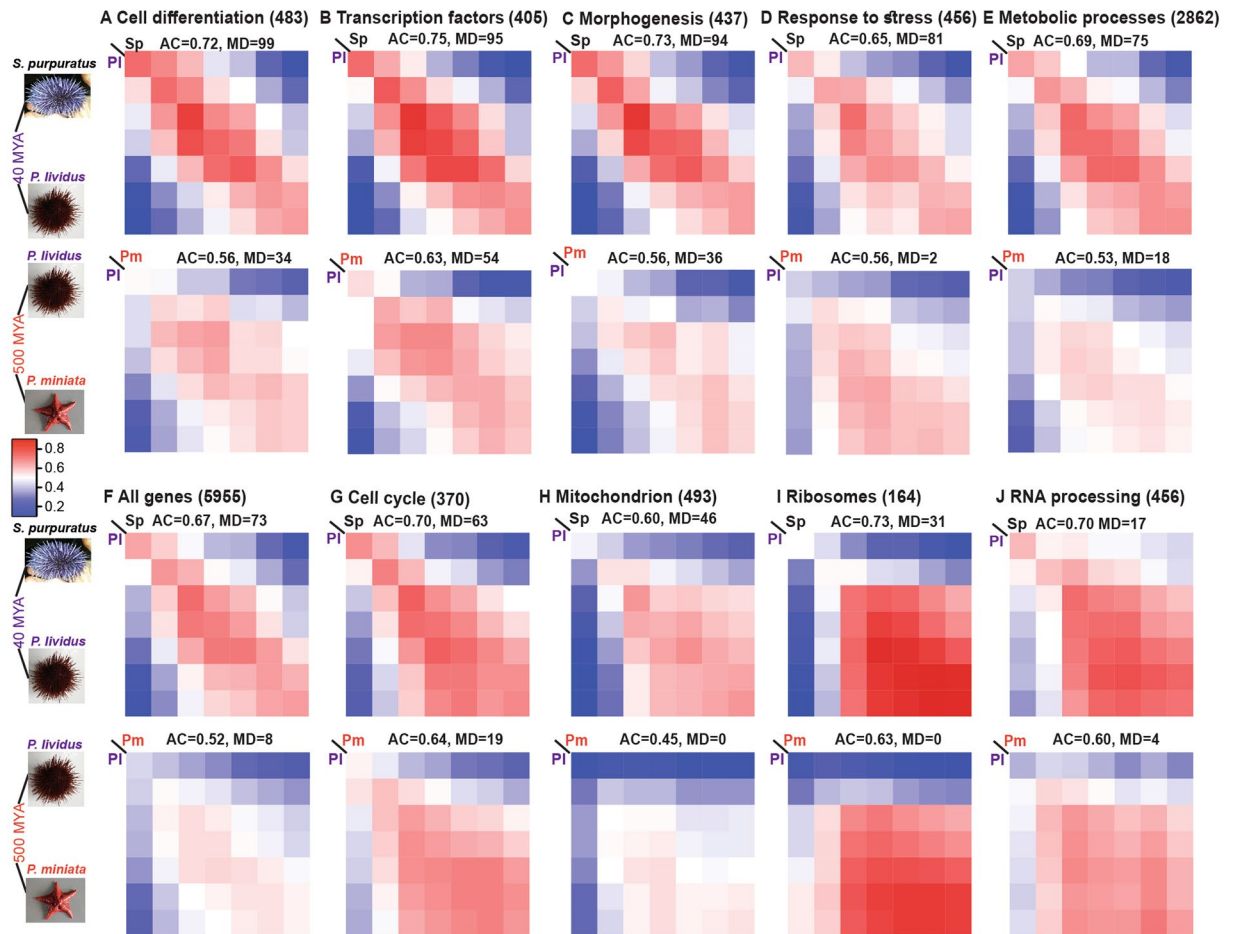


Figure 3. Interspecies Pearson correlations for different GO terms, ordered by the level of matrix diagonality (MD) of *Pl-Sp* matrices. In each panel, from (A–J) we present the Pearson correlation of the expression levels of genes with specific GO term between different developmental stages in two species. Upper matrix in each panel shows the Pearson correlation between the two sea urchins (*P. lividus* and *S. purpuratus*) and the bottom matrix is the Pearson correlation between the sea star, *P. miniata* and the sea urchin *P. lividus*. These matrices include the seven developmental points that have RNA-seq data in all species (Fig. 1B, excluding 9 hpf in *P. miniata* and 4 hpf in *P. lividus*). In each panel we indicate the GO term tested, the number of genes in each set, the average correlation strength in the diagonal (AC) and the matrix diagonality (MD), see text for explanation. Linear color scale of Pearson correlations 0–1 is identical for all graphs and given at the middle of the figure. (F) Shows the Pearson interspecies correlation for all 1:1:1 genes.

genes combined are in between (Fig. 3D–F). Both the average correlation and the matrix diagonality are lower between the sea urchin and the sea star than between the two sea urchins, that is, both parameters decrease with evolutionary distance (compare bottom to upper panels in Fig. 3). The reduction in AC indicates that the correlation strength between morphologically equivalent stages decreases, which could be due to drift, environmental adaptation, or morphological divergence at increasing evolutionary distance. The reduction in matrix diagonality means that the correlation in gene expression between morphologically similar developmental stages becomes of the same order as the correlation between distinct developmental stages. Thus, for an increasing number of gene sets at large evolutionary distance the correlation of gene expression levels between different stages become similar, regardless of morphological similarity. This could suggest that the morphological constraints are less dominant in controlling the conservation of gene expression compared to other cellular or metabolic constraints, with increasing evolutionary distance.

The interspecies correlations vary between different embryonic territories. Our analysis is based on RNA-seq on whole embryos, yet we wanted to see if gene sets enriched in specific cell populations show a difference in their interspecies correlation pattern. In a previous work conducted in the sea urchin *S. purpuratus* embryos, the cells of six distinct embryonic territories were isolated based on cell-specific GFP reporter expression. Gene expression levels in the isolated cells were studied and compared to gene expression levels in the rest of the embryo by RNA-seq³³. This analysis identified genes whose expression is enriched in specific cell populations at the developmental time when the isolation was done³³. We calculated the Pearson interspecies correlations for

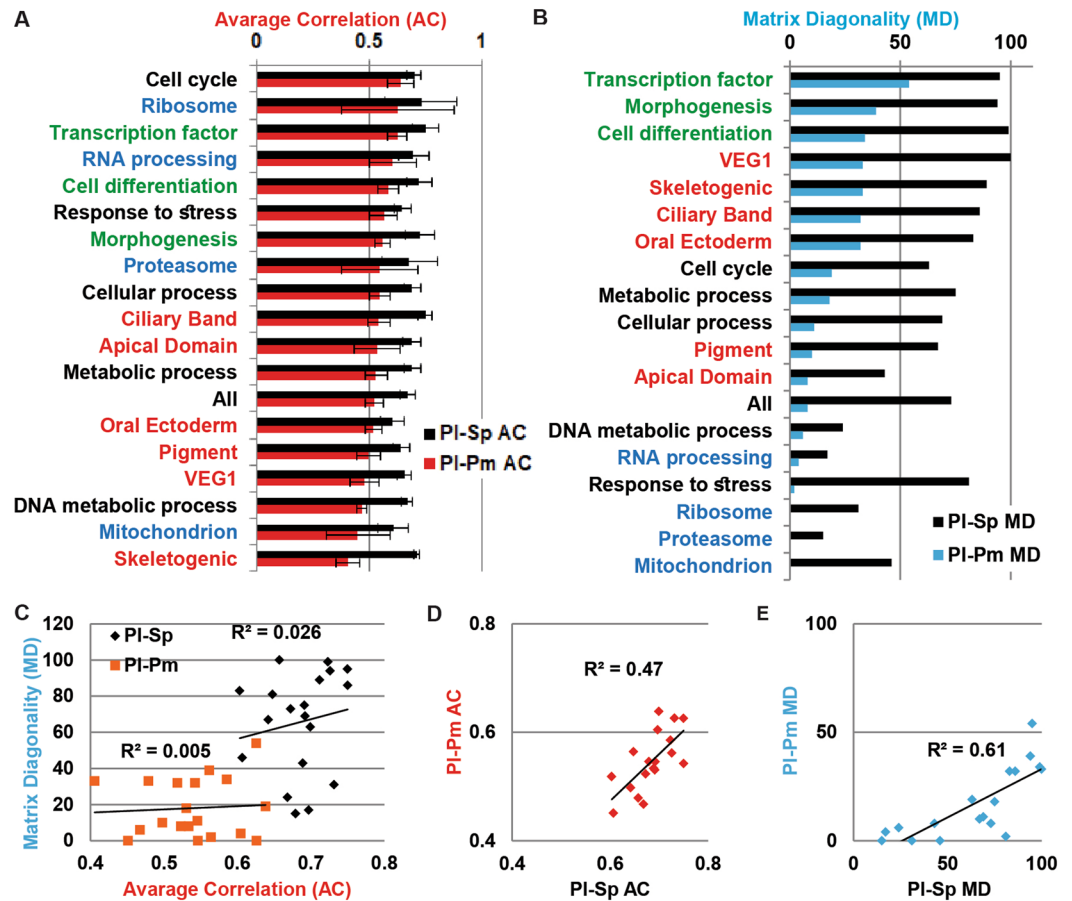


Figure 4. The average correlation strength and the matrix diagonality are independent parameters that reflect different properties of expression conservation. **(A)** The average correlation strength (AC) between *Pl-Sp* (black bars) and *Pl-Pm* (red bars) in receding order of *Pl-Pm* correlation strength. Error bars indicate standard deviation of the correlation strength along the matrix diagonal. **(B)** Matrix diagonality (MD) of the interspecies correlations between *Pl-Sp* (black bars) and *Pl-Pm* (cyan bars) in receding order of *Pl-Pm* matrix diagonality. In **(A,B)** the different gene sets are colored by the following key: developmental GO terms in green, housekeeping GO terms in blue, lineage specific genes in red and all the other sets in black. **(C)** The matrix diagonality changes independently of the average correlation for both *Pl-Sp* (black dots) and *Pl-Pm* (orange dots). **(D)** The interspecies average correlation between *S. purpuratus* and *P. lividus* corresponds to the interspecies average correlation between *P. lividus* and *P. miniata* (R pearson = 0.68). **(E)** The matrix diagonality of the interspecies correlations between *S. purpuratus* and *P. lividus* relates to the matrix diagonality between *P. lividus* and *P. miniata* (R Pearson = 0.78). In **(D,E)** we didn't include the skeletogenic lineage data point as it is an outlier.

the subsets of genes enriched in each of the six sea urchin embryonic territories, between the two sea urchins and between the sea urchins and the sea star (Figs S2 and S3).

Both the matrix diagonality and average correlation strength vary between the different embryonic territories and decrease with evolutionary distance. Interestingly, the genes enriched in sea urchin pigment cells, a lineage that is lacking in the sea star, show similar *Pl-Pm* average correlation strength compared to the matrices of the embryonic territories that are common to the sea urchins and the sea star (*Pl-Sp* AC = 0.64 and *Pl-Pm* AC = 0.5, Fig. S3B). As mentioned above, the regulatory state in the mesoderm of the sea star and the sea urchin are quite similar⁷. Possibly, there is also similarity in the downstream genes active in sea star blastocoelar cells and sea urchin pigment cells, as both lineages function as immune cells in the sea urchin³⁴. On the other hand, the sharp decrease in *Pl-Pm* average correlation strength compared to *Pl-Sp* is for the genes enriched in the sea urchin skeletogenic cells, another lineage that is absent in the sea star (*Pl-Sp* AC = 0.71 while *Pl-Pm* AC = 0.4, Fig. S3C). Interestingly, the *Pl-Pm* matrix diagonality of these genes is comparable to the *Pl-Pm* matrix diagonality of the common lineages (*Pl-Pm* MD = 33, Fig. 4C). It is important to note that key sea urchin skeletogenic matrix proteins were not found in the sea star skeleton³⁵ and the genes encoding them were not found in the sea star genome. Therefore, these key skeletogenic genes are missing from our 1:1:1 homologous genes that include only genes that are common to the three species. Thus, we are probably underestimating the differences in skeletogenic and mesodermal gene expression between the sea urchin and the sea star, which might explain the relatively high diagonality of the skeletogenic gene correlation matrix. Overall, genes enriched in different cells populations

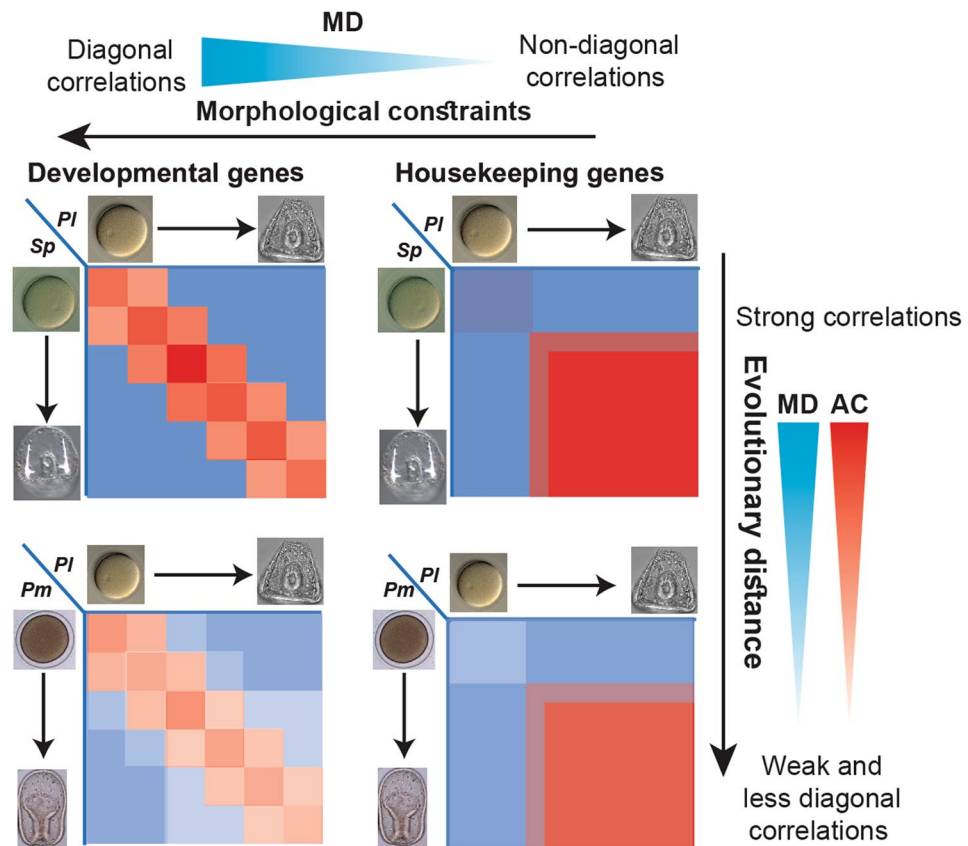


Figure 5. The correlation matrix diagonality, MD, reflects how the dominance between cellular and developmental constraints changes with evolutionary distance. Illustration of typical interspecies correlation matrices of developmental control genes and housekeeping genes between closely related and further diverged species (upper and lower panels, respectively). With increasing evolutionary distance, that is, between the sea urchin and the sea star, the average correlation and the diagonality decrease for all gene sets but the diagonality of developmental control genes is least affected and they still maintain the hourglass pattern (left panels). On the other hand, the interspecies correlation matrices of housekeeping genes are strong and non-diagonal even between the two sea urchins and remain non-diagonal between the sea urchin and the sea star (right panels).

differ in their correlation patterns even between the closely related sea urchins and show distinct differences in the correlation pattern with evolutionary distance.

Matrix diagonality and correlation strength describe different properties of gene expression conservation. The average correlation strength and the matrix diagonality seem to change independently from each other for different gene functions and cell populations (Figs 3, S2 and S3). To further investigate that we plot, separately, the average correlation and the matrix diagonality for different GO terms and cell populations, in decreasing strength in *Pl-Pm* and observed two distinct orders (Fig. 4A,B). Ordering by matrix diagonality separates between genes with developmental GO terms or specific cell populations (high diagonality) and genes with housekeeping GO terms (low diagonality, Fig. 4B). Ordering by the average correlation does not show this separation (Fig. 4A). Moreover, the matrix diagonality and the average correlation strength change independently of each other in both *Pl-Sp* and *Pl-Pm* correlation matrices (Fig. 4C). On the other hand, the average correlations in *Pl-Sp* matrices seem to correspond to the average correlation in *Pl-Pm* (Fig. 4D), and the same is true for the matrix diagonality (Fig. 4E).

These different behaviors of the average correlation and the matrix diagonality could suggest that these two parameters describe different properties of the conservation in gene expression, as we propose in Fig. 5. The average correlation strength seems to reflect the conservation in gene expression levels: the higher it is, the more the gene-set is constrained against expression change. The matrix diagonality on the other hand, seems to reflect the link between the expression of a gene-set and morphological constraints on this set: the more diagonal is the correlation pattern of a gene set, the more dominant is the morphological constraint on the expression conservation within the set (Fig. 5). For example, the correlation matrices of transcription factors and cell differentiation genes show strong correlations and high diagonality even between the sea urchin and the sea star (Relatively high AC and MD, Figs 3A,B and 5). Conversely, ribosomal gene expression is highly conserved (high AC) but this conservation is not related to morphological conservation (Low MD, Figs 3I and 5). Overall, the average correlation strength and matrix diagonality seem to carry complementary information about the relationship between the conservation of gene expression and morphological conservation at varying evolutionary distances (Fig. 5).

Discussion

In this paper we generated the developmental transcriptomes of the Pacific sea star, *P. miniata*, and studied them in comparison with the published developmental transcriptomes of two sea urchin species, *P. lividus*¹⁹ and *S. purpuratus*²⁶. We generated a web application where the *P. miniata* time courses and sequences can be publicly viewed to facilitate the common use of this data³⁰. We studied the interspecies correlation patterns of different gene sets including, housekeeping, developmental, response and metabolic genes (Figs 3 and S2), as well as genes that are enriched in specific cell populations in the sea urchin embryo (Figs S2 and S3). We defined two parameters that describe different properties of the conservation strength: the average correlation strength in the diagonal, AC, and the matrix diagonality, MD. We noticed that these parameters vary independently between different functional groups and cell populations and decrease with evolutionary distance, possibly reflecting different constraints on gene expression (Fig. 4). The correlation strength seems to indicate the evolutionary constraint on gene expression level while the matrix diagonality seems to reflect the dominance of morphological constraints on the conservation of gene expression within a gene set (Fig. 5). As we suggested previously, parallel embryonic transcriptional programs might be responsible for different aspects of embryo development and evolve under distinct constraint¹⁹, as can be inferred from analyzing these two parameters.

Previous studies have shown that housekeeping genes and tissue specific genes have different chromatin structures³⁶ and distinct core promoters³⁷. Furthermore, Enhancers of developmental genes were shown to be de-methylated during the vertebrates' phylotypic period, suggesting another unique epigenetic regulation of this set of genes³⁸. These epigenetic and *cis*-regulatory differences could underlie the separation of the regulation of developmental and housekeeping gene expression, leading to dissimilar evolutionary conservation patterns of these two functionally distinct gene sets.

We observed a clear reduction of the matrix diagonality with increasing evolutionary distance for all sets of genes (Figs 3–5, S2 and S3). This is in agreement with the morphological and evolutionary divergence between the sea urchin and the sea star. However, it is important to note that in the comparisons conducted in this work the phylogenetic distance and morphological divergence are interconnected: the two sea urchins are phylogenetically close and morphologically similar, while the sea urchins and the sea star are phylogenetically distant and morphologically divergent. Thus, in these comparisons it is impossible to disentangle the effects of phylogenetic distance and morphological divergence. It would be interesting to study the correlation strength and matrix diagonality between phylogenetically close but morphologically diverged species. Possible example could be sea urchin species in the genus *Heliocidaris* that rapidly transitioned from feeding to nonfeeding larval development and therefore show significant morphological divergence within a very short phylogenetic distance¹⁷. We hypothesize that in this kind of comparison, the average correlation strength will associate with the phylogenetic distance and will be strong within the closely related and morphologically diverged *Heliocidaris* species. On the other hand, the matrix diagonality will correlate with the morphological similarity and will show higher diagonality between phylogenetically distance species that have similar morphology. Thus, we expect these two parameters to show opposite trends when the phylogenetic distance and morphological similarities are distinct.

A recent study has shown that the correlation matrices for all homologous genes between 10 different phyla are strictly not-diagonal²². At this large evolutionary distance and extreme morphological divergence, the dominant constraint is probably the cellular requirements for differential transcript abundance, which are unrelated to morphological similarity. Therefore the opposite hourglass pattern found for the diagonal elements of the interspecies correlation matrices between different phyla might not be indicative for morphological divergence at the phylotypic stage²². Overall, to better estimate the relationship between the conservation of gene expression levels to morphological conservation, both correlation strength and correlation matrix diagonality should be assessed and the focus should be on developmental control genes.

Materials and Methods

Sea star embryo cultures and RNA extraction. Adult *P. miniata* sea stars were obtained in Long Beach, California, from Peter Halmay. Embryos were cultured at 15 °C in artificial sea water. Total RNA was extracted using Qiagen mini RNeasy kit from thousands of embryos. RNA samples were collected from eight embryonic stages, from fertilized egg to late gastrula stage (Fig. 1B). For each embryonic stage, three biological replicates from three different sets of parents were processed, except for the last time point for which only two biological replicates were sampled (23 samples in all). To match *P. miniata* time points to those of the published *S. purpuratus* and *P. lividus* transcriptomes^{19,27} we used the linear ratio between the developmental rates of these species found in^{14,39}. The corresponding time points in both species are presented in Fig. 1B. The time points, 9hpf in *P. miniata* and 4hpf in *P. lividus* do not have a comparable time point in *S. purpuratus* data. RNA quantity was measured by nanodrop and quality was verified using bioanalyzer.

Transcriptome assembly, annotations and quantification. *RNA-Seq preparation.* The 23 *P. miniata* samples that we analyze here were processed together with 24 *P. lividus* samples that we described in¹⁹. Library preparation was done at the Israel National Center for Personalized Medicine (INCPM) as following. polyA fraction (mRNA) was purified from 500 ng of total RNA per developmental time point followed by fragmentation and generation of double stranded cDNA. Then, end repair, A base addition, adapter ligation and PCR amplification steps were performed. Libraries were evaluated by Qubit and TapeStation. Sequencing libraries were constructed with barcodes to allow multiplexing of 47 samples. On average, 20 million single-end 60-bp reads were sequenced per sample on Illumina HiSeq 2500 V4 using four lanes. Exact number of reads for each sample is provided in Dataset S1.

RNA-Seq datasets used. Three datasets were used for transcripts quantification: (1) newly sequenced *P. miniata* Single End (SE) reads of 23 transcriptome samples generated as explained above; (2) publicly available SE

reads of *P. lividus* transcriptome samples from eight developmental stages (NCBI project PRJNA375820)¹⁹; (3) publicly available Paired End (PE) reads of seven *S. purpuratus* transcriptome samples (NCBI project PRJNA81157)²⁷. For generating reference transcriptome for *P. miniata* we used PE reads of *P. miniata* from different developmental stages, testis and ovaries, accessions: SRR6054712, SRR5986254, SRR2454338, SRR1138705, SRR573705-SRR573710 and SRR573675.

RNA-seq quality filtering. RNA-Seq reads from the above datasets were adapter-trimmed using cutadapt 1.15 (<https://cutadapt.readthedocs.io>), then low-quality regions were removed with Trimmomatic 0.3⁴⁰, and further visually inspected in fastq-screen (www.bioinformatics.babraham.ac.uk).

***P. miniata* transcriptome assembly.** While *P. miniata* genome based gene predictions are available for genome assembly v.2 (echinobase.org/Echinobase), many gene sequences are fragmented or duplicated within the genome, and therefore we decided to generate a reference transcriptome *de-novo*. Accordingly, *P. miniata* RNA-Seq reads, based on the available PE and SE data, were assembled using Trinity 2.4^{41,42} with the same Trinity parameters as we used for *P. lividus* in¹⁹. Trinity produced 1,610,829 contigs (the Trinity equivalent of transcripts), within 679,326 Trinity gene groups. Transcriptome completeness was tested using BUSCO⁴³, and by comparing the contigs to genome-based protein annotations *P. miniata* v.2.0 and *S. purpuratus* v.3.1 (WHL22), (echinobase.org/Echinobase).

***P. miniata* data availability.** Illumina short read sequences generated in this study were submitted to the NCBI Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>), under bio-project PRJNA522463 (<http://www.ncbi.nlm.nih.gov/bioproject/522463>). Fastq read accessions: SRR8580044–SRR8580066, assembled *P. miniata* transcriptome accession: SAMN10967027.

Transcriptome homology. We searched for homologous genes in the *P. miniata* transcriptome, *P. lividus* transcriptome, *S. purpuratus* genome. Since the *S. purpuratus* genome-based gene predictions dataset currently includes the most non-redundant and complete data among the three datasets, it was used as a reference dataset. Accordingly, the largest isoforms of new *P. miniata* transcriptome (see below), and the publicly available *P. lividus* transcriptome¹⁹, were both compared to the *S. purpuratus* genome-based predicted protein annotations (Echinobase v.3.1), using CRB-BLAST⁴⁴. CRB-BLAST reports relationships of 1:1 (Reciprocal hits), and 2:1 matches (when Blastx and tBlastn results are not reciprocal), where only matches with e-values below a conditional threshold are reported. From the 2:1 cases, we selected the query-target pair with the lowest e-value. Using CRB-BLAST, 11,291 and 12,720 *P. miniata* and *P. lividus* Trinity genes, were identified as homologous to *S. purpuratus* proteins, respectively. We considered *P. miniata* and *P. lividus* query genes that share the same *S. purpuratus* target gene, as homologous. As the gene expression analysis shows (see next sections), 8,735 homologous genes are expressed in at least one of the three tested species during development, and 6,593 are expressed in all the three.

Gene-level transcripts abundance. For *P. miniata* and *P. lividus* transcriptomes, transcripts abundance was estimated using kallisto-0.44.0⁴⁵, and a further quantification at the gene-level, and read-count level, was done using tximport⁴⁶ on R3.4.2. Expression analysis at read-count level was conducted at gene-level in Deseq2⁴⁷. *S. purpuratus* PE reads, of 7 developmental samples were mapped to the *S. Purpuratus* 3.1 genome assembly, using STAR v2.4.2a⁴⁸, quantitated in Htseq-count v2.7⁴⁹, and analyzed in Deseq2 at read-count level. For *P. miniata* data, only contigs mapped to the Echinobase v.2 *P. miniata* proteins were considered. Prior to Deseq2 read count standard normalization and expression analysis, genes with <1 CPM (Count Per Million) were removed. Overall, most input reads were mapped and quantified, as further detailed in Dataset S1. Samples were clustered using Non-metric multi-dimensional scaling (NMDS) ordination in Vegan (<https://cran.r-project.org/web/packages/vegan/index.html>), based on log10 transformed FPKM values, and Bray-Curtis distances between samples. Since the NMDS results indicate that all *P. miniata* and *P. lividus* samples are affected by the ‘batch’ factor (see Dataset S1), we removed the estimated effect of this factor on FPKMs (Fragment per Kilobase Million) values using “removeBatchEffect” function in EdgeR⁵⁰, in order to obtain corrected FPKM values. Quantification and annotations of 34,307 identified *P. miniata* transcripts with FPKM >3 in at least one time point, are provided in Dataset S2. NMDS analysis of the biological replicates of all time points in *P. miniata* show that similar time points at different biological replicate map together indicating high reproducibility of our gene expression analysis (Fig. S4). Comparison between our RNA-seq quantification of gene expression and previous QPCR quantification¹⁴ for a subset of genes show high agreement between the two measurements (Fig. S5). Quantification and annotations of the 8,735 identified *P. miniata*, *P. lividus* and *S. purpuratus* 1:1:1 homologous genes are provided in Dataset S3 and are publicly available through gene search in Echinobase at www.echinobase.org/shiny/quantdevPm.

Gene Ontology functional enrichment analysis. Functional enrichment analysis was conducted using TopGo in R3.4.2 (bioconductor.org). A custom Goseq GO database was built using the publicly available Blast2Go *S. purpuratus* v.3.1 (WHL22) version (<http://www.echinobase.org/Echinobase/rnaseq/download/blast2go-whl.annot.txt.gz>).

Pearson correlations for subsets of genes. The interspecies Pearson correlations for different sets of genes presented in Figs 3, S2 and S3 were calculated using R3.4.2. For the analysis in Fig. S3, we selected genes that have 1:1:1 homologs in *P. miniata* and *P. lividus* to the *S. purpuratus* genes that are enriched in a specific *S. purpuratus* cell population with p-value < 0.05, based on³³.

Cross species analysis of matrix diagonality (MD). We used a statistical test described in detail in¹⁹. Shorty, the main goal of this procedure is to test the probability that a set of homologous genes from two species, S_1 and S_2 , show the most similar expression patterns in equivalent developmental times, namely: is the interspecies correlation pattern significantly close to a diagonal matrix? Here, S_1 and S_2 represent *P. miniata* vs. *P. lividus*, or *P. lividus* vs. *S. purpuratus*. This test was conducted using all homologous genes, and for specific subsets of genes belonging to specific GO categories as well as for genes enriched in specific cell population³³. Here, only samples from the $n = 5$ late embryonic stages were used, and the mean of FPKM values of all samples belonging to the same stages were taken. First, n_{s1} by n_{s2} matrix of Pearson correlations, C , was produced, where n_{s1} by n_{s2} are the number of time point measurements in each species (here $n_{s1} = n_{s2} = 5$). Each of the C matrix positions, C_{ij} , represents a Pearson's correlation value calculated based on n_g FPKM values, between stage i in one species and stage j in the other, where n_g is the count of homologous gene pairs tested. Next, the C matrix was compared to an "ideal" time-dependent correlation matrix I , in which a correlation of 1 exists along the diagonal line, and 0 in other positions, to obtain a diagonality measure, d . Overall, $n_g = 50$ genes were resampled $n_{resamp} = 100$ times, and for each resampling-iteration, a permutation test based on value d was applied using $n_{perm} = 1000$ permutations. Accordingly, each of the above $n_{resamp} = 100$ permutation tests indicate the probability that S_1 and S_2 show non-random diagonality, on a subset of $n_g = 50$ genes. Then, the count of permutation tests indicating a significant similarity to the ideal diagonal matrix, out of the $n_{resamp} = 100$ subsamples is used for estimating diagonality of the correlation matrix (matrix diagonality = MD). Our results are presented in Dataset S5 and within Figs 3, 4, S2 and S3.

Received: 19 March 2019; Accepted: 11 October 2019;

Published online: 07 November 2019

References

- Peter, I. S. & Davidson, E. H. *Genomic Control Process: Development and Evolution*. (Academic Press, 2015).
- Raff, R. A. & Byrne, M. The active evolutionary lives of echinoderm larvae. *Heredity* **97**, 244–252, <https://doi.org/10.1038/sj.hdy.6800866> (2006).
- Oliveri, P., Tu, Q. & Davidson, E. H. Global regulatory logic for specification of an embryonic cell lineage. *Proc Natl Acad Sci USA* **105**, 5955–5962, <https://doi.org/10.1073/pnas.0711220105> (2008).
- Ben-Tabou de-Leon, S., Su, Y. H., Lin, K. T., Li, E. & Davidson, E. H. Gene regulatory control in the sea urchin aboral ectoderm: Spatial initiation, signaling inputs, and cell fate lockdown. *Dev Biol*, **374**, 245–254, <https://doi.org/10.1016/j.ydbio.2012.11.013> (2012).
- Li, E., Materna, S. C. & Davidson, E. H. Direct and indirect control of oral ectoderm regulatory gene expression by Nodal signaling in the sea urchin embryo. *Dev Biol* **369**, 377–385, <https://doi.org/10.1016/j.ydbio.2012.06.022> (2012).
- Hinman, V. F., Nguyen, A. T., Cameron, R. A. & Davidson, E. H. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc Natl Acad Sci USA* **100**, 13356–13361, <https://doi.org/10.1073/pnas.2235868100> (2003).
- McCauley, B. S., Weideman, E. P. & Hinman, V. F. A conserved gene regulatory network subcircuit drives different developmental fates in the vegetal pole of highly divergent echinoderm embryos. *Dev Biol* **340**, 200–208, <https://doi.org/10.1016/j.ydbio.2009.11.020> (2010).
- McCauley, B. S., Wright, E. P., Exner, C., Kitazawa, C. & Hinman, V. F. Development of an embryonic skeletogenic mesenchyme lineage in a sea cucumber reveals the trajectory of change for the evolution of novel structures in echinoderms. *Evodevo* **3**, 17, <https://doi.org/10.1186/2041-9139-3-17> (2012).
- Yankura, K. A., Martik, M. L., Jennings, C. K. & Hinman, V. F. Uncoupling of complex regulatory patterning during evolution of larval development in echinoderms. *BMC Biol* **8**, 143, <https://doi.org/10.1186/1741-7007-8-143> (2010).
- McCauley, B. S., Akyar, E., Filliger, L. & Hinman, V. F. Expression of wnt and frizzled genes during early sea star development. *Gene Expr Patterns* **13**, 437–444, <https://doi.org/10.1016/j.gep.2013.07.007> (2013).
- Yankura, K. A., Koechlein, C. S., Cryan, A. F., Cheatle, A. & Hinman, V. F. Gene regulatory network for neurogenesis in a sea star embryo connects broad neural specification and localized patterning. *Proc Natl Acad Sci USA* **110**, 8591–8596, <https://doi.org/10.1073/pnas.1220903110> (2013).
- Yaguchi, S., Yaguchi, J., Angerer, R. C. & Angerer, L. M. A Wnt-FoxQ2-nodal pathway links primary and secondary axis specification in sea urchin embryos. *Dev Cell* **14**, 97–107, <https://doi.org/10.1016/j.devcel.2007.10.012> (2008).
- Morgulis, M. *et al.* Possible cooption of a VEGF-driven tubulogenesis program for biomineralization in echinoderms. *Proc Natl Acad Sci USA*, **116**, 12353–12362, <https://doi.org/10.1073/pnas.1902126116> (2019).
- Gildor, T., Hinman, V. & Ben-Tabou-De-Leon, S. Regulatory heterochronies and loose temporal scaling between sea star and sea urchin regulatory circuits. *Int J Dev Biol* **61**, 347–356, <https://doi.org/10.1387/ijdb.160331sb> (2017).
- Cary, G. A. & Hinman, V. F. Echinoderm development and evolution in the post-genomic era. *Dev Biol* **427**, 203–211, <https://doi.org/10.1016/j.ydbio.2017.02.003> (2017).
- Dylus, D. V., Czarkwiani, A., Blowes, L. M., Elphick, M. R. & Oliveri, P. Developmental transcriptomics of the brittle star *Amphiura filiformis* reveals gene regulatory network rewiring in echinoderm larval skeleton evolution. *Genome Biol* **19**, 26, <https://doi.org/10.1186/s13059-018-1402-8> (2018).
- Israel, J. W. *et al.* Comparative developmental transcriptomics reveals rewiring of a highly conserved gene regulatory network during a major life history switch in the sea urchin genus *Heliocidaris*. *PLoS Biol* **14**, e1002391, <https://doi.org/10.1371/journal.pbio.1002391> (2016).
- Gildor, T. & Ben-Tabou-De-Leon, S. Comparative studies of gene expression kinetics: methodologies and insights on development and evolution. *Frontiers in genetics* **9**, 339, <https://doi.org/10.3389/fgene.2018.00339> (2018).
- Malik, A., Gildor, T., Sher, N., Layous, M. & Ben-Tabou de-Leon, S. Parallel embryonic transcriptional programs evolve under distinct constraints and may enable morphological conservation amidst adaptation. *Dev Biol*, **430**, 202–213, <https://doi.org/10.1016/j.ydbio.2017.07.019> (2017).
- Irie, N. & Kuratani, S. The developmental hourglass model: a predictor of the basic body plan? *Development* **141**, 4649–4655, <https://doi.org/10.1242/dev.107318> (2014).
- Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun* **2**, 248, <https://doi.org/10.1038/ncomms1248> (2011).
- Levin, M. *et al.* The mid-developmental transition and the evolution of animal body plans. *Nature* **531**, 637–641, <https://doi.org/10.1038/nature16994> (2016).
- Yanai, I., Peshkin, L., Jorgensen, P. & Kirschner, M. W. Mapping gene expression in two *Xenopus* species: evolutionary constraints and developmental flexibility. *Dev Cell* **20**, 483–496, <https://doi.org/10.1016/j.devcel.2011.03.015> (2011).

24. Kalinka, A. T. *et al.* Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814, <https://doi.org/10.1038/nature09634> (2010).
25. Piasecka, B., Lichocki, P., Moretti, S., Bergmann, S. & Robinson-Rechavi, M. The hourglass and the early conservation models—co-existing patterns of developmental constraints in vertebrates. *PLoS Genet* **9**, e1003476, <https://doi.org/10.1371/journal.pgen.1003476> (2013).
26. Tu, Q., Cameron, R. A. & Davidson, E. H. Quantitative developmental transcriptomes of the sea urchin *Strongylocentrotus purpuratus*. *Dev Biol* **385**, 160–167, <https://doi.org/10.1016/j.ydbio.2013.11.019> (2014).
27. Tu, Q., Cameron, R. A., Worley, K. C., Gibbs, R. A. & Davidson, E. H. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res* **22**, 2079–2087, <https://doi.org/10.1101/gr.139170.112> (2012).
28. Materna, S. C., Nam, J. & Davidson, E. H. High accuracy, high-resolution prevalence measurement for the majority of locally expressed regulatory genes in early sea urchin development. *Gene Expr Patterns* **10**, 177–184, <https://doi.org/10.1016/j.gep.2010.04.002> (2010).
29. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**, R14, <https://doi.org/10.1186/gb-2010-11-2-r14> (2010).
30. Gildor, T., Cary, G., Lalar, M., Hinman, V. & Ben-Tabou de-Leon, S. Echinobase viewer of *P. miniata* quantitative developmental transcriptomes, www.echinobase.org/shiny/quantdevPm.
31. Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. Framework for R. R package version 1.2.0, <https://CRAN.R-project.org/package=shiny> (2018).
32. Cameron, R. A., Samanta, M., Yuan, A., He, D. & Davidson, E. SpBase: the sea urchin genome database and web site. *Nucleic Acids Res* **37**, D750–754, <https://doi.org/10.1093/nar/gkn887> (2009).
33. Barsi, J. C., Tu, Q., Calestani, C. & Davidson, E. H. Genome-wide assessment of differential effector gene use in embryogenesis. *Development* **142**, 3892–3901, <https://doi.org/10.1242/dev.127746> (2015).
34. Solek, C. M. *et al.* An ancient role for Gata-1/2/3 and Scl transcription factor homologs in the development of immunocytes. *Dev Biol* **382**, 280–292, <https://doi.org/10.1016/j.ydbio.2013.06.019> (2013).
35. Flores, R. L. & Livingston, B. T. The skeletal proteome of the sea star *Patiria miniata* and evolution of biomineralization in echinoderms. *BMC Evol Biol* **17**, 125, <https://doi.org/10.1186/s12862-017-0978-z> (2017).
36. She, X. *et al.* Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC Genomics* **10**, 269, <https://doi.org/10.1186/1471-2164-10-269> (2009).
37. Zabidi, M. A. *et al.* Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559, <https://doi.org/10.1038/nature13994> (2015).
38. Bogdanovic, O. *et al.* Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nat Genet* **48**, 417–426, <https://doi.org/10.1038/ng.3522> (2016).
39. Gildor, T. & Ben-Tabou de-Leon, S. Comparative study of regulatory circuits in two sea urchin species reveals tight control of timing and high conservation of expression dynamics. *PLoS Genet* **11**, e1005435, <https://doi.org/10.1371/journal.pgen.1005435> (2015).
40. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
41. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512, <https://doi.org/10.1038/nprot.2013.084> (2013).
42. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644–652, <https://doi.org/10.1038/nbt.1883> (2011).
43. Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*, **35**, 543–548, <https://doi.org/10.1093/molbev/msx319> (2017).
44. Aubry, S., Kelly, S., Kumpers, B. M., Smith-Unna, R. D. & Hibberd, J. M. Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C4 photosynthesis. *PLoS Genet* **10**, e1004365, <https://doi.org/10.1371/journal.pgen.1004365> (2014).
45. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527, <https://doi.org/10.1038/nbt.3519> (2016).
46. Sonesson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521, <https://doi.org/10.12688/f1000research.7563.2> (2015).
47. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, <https://doi.org/10.1186/s13059-014-0550-8> (2014).
48. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, <https://doi.org/10.1093/bioinformatics/bts635> (2013).
49. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, <https://doi.org/10.1093/bioinformatics/btu638> (2015).
50. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, <https://doi.org/10.1093/bioinformatics/btp616> (2010).

Acknowledgements

We thank Assaf Malik from the bioinformatics service unit at the University of Haifa for the analyses of the transcriptome data. We thank the Genomics unit at the Grand Israel National Center for Personalized Medicine (G-INCPM) for their assistance with library preparation and sequencing. We thank Paola Oliveri for the idea to study specific cell populations. This work was supported by the Binational Science Foundation Grant Number 2015031 to S.B.D. and V.F.H. and the Israel Science Foundation Grants Number 2304/15 (ISF-INCPM) and 41/14 (I.S.F.) to S.B.D., the National Science Foundation grants IOS 1557431 and MCB 1715721 to VFH and the National Institute of Health grant P41HD071837 to V.F.H.

Author contributions

T.G., S.B.D. and V.F.H. designed the project. Embryo cultures and mRNA extractions were done by T.G. with the help of V.F.H. S.B.D., G.A.C. and M.L. analyzed the RNA-seq data and all authors helped in the interpretation of the data. G.A.C. generated the web viewer. S.B.D. wrote the paper with a significant help from T.G. and G.A.C. All authors reviewed the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-52577-9>.

Correspondence and requests for materials should be addressed to S.B.-T.d.-L.

Reprints and permissions information is available at www.nature.com/reprints.

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