Developmental transcriptomes of the sea star, *Patiria miniata*, illuminate the relationship between conservation of gene expression and morphological conservation

Tsvia Gildor[1], Gregory Cary[3] , Maya Lalzar[2], Veronica Hinman[3] and Smadar Ben-Tabou de-Leon[1,*]

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

[2]Bionformatics Core Unit, University of Haifa, Haifa 31905, Israel.

[3]Departments of Biological Sciences and Computational Biology, Carnegie Mellon University

Pittsburgh, PA 15213, USA

*Corresponding author sben-tab@univ.haifa.ac.il

Abstract

Evolutionary changes in developmental gene expression lead to alteration in the embryonic body plan and biodiversity. A promising approach for linking changes in developmental gene expression to altered morphogenesis is the comparison of developmental transcriptomes of closely related and further diverged species within the same phylum. Here we generated quantitative transcriptomes of the sea star, *Patiria miniata* (*P. miniata*) of the echinoderm phylum, at eight embryonic stages. We then compared developmental gene expression between *P. miniata* and the sea urchin, *Paracentrotus lividus* (~500 million year divergence) and between *Paracentrotus lividus* and the sea urchin, *Strongylocentrotus purpuratus* (~40 million year divergence). We discovered that the interspecies correlations of gene expression level between morphologically equivalent stages decreases with increasing evolutionary distance, and becomes more similar to the correlations between morphologically distinct stages. This trend is observed for different sub-sets of genes, from various functional classes and embryonic territories, but is least severe for developmental genes sub-sets. The interspecies correlation matrices of developmental genes show a consistent peak at the onset of gastrulation, supporting the hourglass model of phylotypic stage conservation. We propose that at large evolutionary distance the conservation of relative expression levels for most sets of genes is more related to the required quantities of transcripts in a cell than to conserved morphogenesis processes. In these distances, the information about morphological similarity is reflected mostly in the interspecies correlations between the expressions of developmental control genes.

Author summary

Understanding the relationship between the interspecies conservation of gene expression and morphological similarity is a major challenge in modern evolutionary and developmental biology. The Interspecies correlations of gene expression levels have been used extensively to illuminate these relationships and reveal the developmental stages that show the highest conservation of gene expression, focusing on the diagonal elements of the correlation matrices. Here we generated the developmental transcriptomes of the sea star, *Patiria miniata,* and used them to study the interspecies correlations between closely related and further diverged species within the echinoderm phylum. Our study reveals that the diagonal elements of the correlation matrices contain only partial information. The off-diagonal elements, that compare gene expression between distinct developmental stages, indicate whether the conservation of gene expression is indeed related to similar morphology or instead, to general cellular constraints that linger throughout development. With increasing evolutionary distances the diagonal elements decrease and become similar to the off-diagonal elements, reflecting the shift from morphological to general cellular constraints. Within this trend, the interspecies correlations of developmental control genes maintain their diagonality even at large evolutionary distance, and peak at the onset of gastrulation, supporting the hourglass model of phylotypic stage conservation.

Introduction

Embryo development is controlled by regulatory programs encoded in the genome that are executed during embryogenesis, which usually last days to months [1]. Genetic changes in these programs that occur in evolutionary time scales - millions to hundreds of million years - lead to alterations in body plans and ultimately, biodiversity [1]. Comparing developmental gene expression between diverse species can illuminate evolutionary conservation and modification in these programs that underlie morphological conservation and divergence. To understand how these programs change with increasing evolutionary distances it is important to study closely related and further divergent 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 stars [2]. Of these, the sea urchin and the sea star had been extensively studied both for their embryogenesis and their gene regulatory networks [3-11]. Sea urchins and sea stars diverged from their common ancestor about 500 million years ago, yet their endoderm lineage and the gut

65　morphology show high similarity between their embryos [6]. On the other hand, the mesodermal

66　lineage diverged to generate novel cell types in the sea urchin embryo (Fig. 1A, [8, 12]).

67　Specifically, the skeletogenic mesoderm lineage, which generates the larval skeletal rods that

68　underlie the sea urchin pluteus morphology, and the mesodermal pigment cells that give the sea

69　urchin larva its red pigmentation (arrow and arrowheads in Fig. 1A). Thus, there are both

70　morphological similarities and differences between sea urchin and sea star larval body plans that

71　make these classes very interesting for comparative genetic studies.

72　The models of the gene regulatory networks that control the development of the sea urchin and

73　those that control cell fate specification in the sea star are the state of the art in the field [3-11].

74　The endodermal and ectodermal gene regulatory networks show high levels of conservation

75　between the sea urchin and the sea star in agreement with the overall conserved morphology of

76　these two germ layers [6, 9, 11]. Surprisingly, most of the transcription factors active in the

77　skeletogenic and the pigment mesodermal lineages are also expressed in the sea star mesoderm

78　[7, 8]. Furthermore, the expression dynamics of key endodermal, mesodermal and ectodermal

79　regulatory genes were compared between the Mediterranean sea urchin, *Paracentrotus lividus* (*P.*

80　*lividus*) and the sea star, *Patiria miniata* (*P. miniata*) [13]. Despite the evolutionary divergence of

81　these two species, an impressive level of conservation of regulatory gene expression in all the

82　embryonic territories was observed. This could suggest that novel mesodermal lineages diverged

83　from an ancestral mesoderm through only a few regulatory changes that drove major changes in

84　downstream gene expression and embryonic morphology [7, 8].

85　Insight on the genome-wide changes in developmental gene expression can be gained from

86　comparative transcriptome studies of related species at equivalent developmental stages [14-17].

87　We previously investigated different aspects of interspecies conservation of gene expression

88　between *P. lividus* and the pacific sea urchin, *Strongylocentrotus purpuratus* (*S. purpuratus*) [16,

89　17]. These species diverged from their common ancestor about 40 million years ago and have a

90　highly similar embryonic morphology (Fig. 1A). We observed high conservation of gene

91　expression dynamics of both developmental genes (expressed differentially and regulate specific

92　embryonic lineages) and housekeeping genes (expressed in all the cells throughout developmental

93　time) [17]. Yet, the interspecies correlations of the expression levels of these two sets of genes

94　show distinct behaviors.

95　The interspecies correlations of the expression levels of developmental genes are high between

96　morphologically similar stages in the two sea urchins, and decrease sharply between diverse

97　developmental stages, resulting with highly diagonal correlation matrices [17]. The correlations

3

98   peak at mid-development, at the onset of gastrulation, in agreement with the hourglass model of

99   developmental conservation [17-19]. Conversely, the interspecies correlations of housekeeping

100   gene expression increase with developmental time and are high between all post-hatching time

101   points, resulting with distinct off-diagonal elements of the correlation matrix [17]. This indicates

102   that the ratio between the expression levels of sets of housekeeping genes is conserved and

103   maintained throughout embryogenesis, irrespective of specific developmental stages.

104   Interestingly, another situation where the interspecies correlation matrix is not diagonal was

105   observed for highly diverged species that belong to different phyla [20]. In these large

106   evolutionary distances, the interspecies correlation has an opposite hourglass behavior with high

107   correlations between all early time points and between all late time points [20]. This suggests that

108   the off-diagonal elements of the interspecies correlation matrix might contain information about

109   the relationship between the conservation of gene expression and the conservation of

110   morphology.

111   Apparent differences in the conservation patterns between developmental and housekeeping

112   genes were observed in other studies of closely related species [18] and were a reason to exclude

113   housekeeping genes from comparative studies of developing embryos [19]. Yet, embryogenesis

114   progression depends on the dynamic expression of housekeeping genes that is highly conserved

115   between closely related species [17]. Here we aim to decipher how the patterns of gene-

116   expression conservation change between closely to distantly related species using the sea star and

117   the sea urchin as our model systems. To this end, we generated and analyzed *de-novo*

118   developmental quantitative transcriptomes of the sea star, *P. miniata,* and compared them with

119   the published developmental transcriptomes of *P. lividus* [17] and *S. purpuratus* [21] at

120   equivalent developmental stages (Fig. 1B). Our studies illuminate how the interspecies

121   conservation of gene expression levels for different functional classes and cell lineages change

122   with evolutionary distance.

Results

Developmental transcriptomes of the sea star, *P. miniata*

125   To study the transcriptional profiles of the sea star species, *P. miniata*, and compare them to those

126   of the sea urchin species, *S. purpuratus* and *P. lividus* we collected embryos at eight

127   developmental stages matching to those studied the sea urchin species [17, 21, 22] , from

128   fertilized egg to late gastrula stage (Fig. 1B). Details on reference transcriptome assembly,

129   quantification and annotations are provided in the Materials and Methods section. Quantification

130 and annotations of all identified *P. miniata* transcripts are provided in Table S2. The expression

131 profiles of selected regulatory genes show mild heterochronies and overall similarity between *P.*

132 *miniata* (RNA-seq, current study) and the sea urchin species, *S. purpuratus* (nanostring, [23]) and

133 *P. lividus* (RNA-seq, [17]), in agreement with our previous study (Fig. 1C, [12]).

134 **Identification of 1:1:1 orthologues gene set and quantitative data access**

135 To compare developmental gene expression between the sea star, *P. miniata,* and the two sea

136 urchins, *P. lividus* and *S. purpuratus* we identified 8735 1:1:1 putative homologous genes, as

137 described in the Materials and Methods section. Quantification and annotations of all these

138 homologous genes in *P. miniata*, *P. lividus* and in *S. purpuratus* are provided in Table S3 (based

139 on [17] for *P. lividus* and on [21, 22] *for S. purpuratus*). Within this set, 6593 genes were

140 expressed in the three species, 1093 genes are expressed only in the two sea urchin species, 187

141 genes were expressed only in the sea star, *P. miniata,* and the sea urchin, *S. purpuratus,* and 430

142 genes are expressed only in *P. miniata* and *P. lividus* (Fig. 2A). We looked for enrichment of

143 gene ontology (GO) terms within these different gene sets but did not identify enrichment of

144 specific developmental processes (GOseq [24] with *S. purpuratus* annotations, Fig. S1 and Table

145 S4).

146 We uploaded the data of the 1:1:1 homologous genes expressed in the three species to Echinobase

147 where they are available through gene search at www.echinobase.org/shiny/quantdevPm [25]. In

148 this Shiny web application [26], genes can be searched either by their name or by their *P. miniata*,

149 *S. purpuratus* or *P. lividus* transcript identification number. The application returns our

150 quantitative measurements of gene expression for each stage of *P. miniata*, plots of transcript

151 expression level, mRNA sequences, a link to the corresponding records in Echinobase [27] and

152 links to the loci in the *S. purpuratus* and *P. miniata* genome browsers. This will hopefully be a

153 useful resource for the community. Further analyses of the 6593 1:1:1 homologous genes

154 expressed in all three species are described below.

155 **Similarity in gene expression profiles depends on developmental stage and**
156 **evolutionary distance**
157
158 To identify the most similar stages in terms of gene expression within and between species, we

159 performed a Non-metric multidimensional scaling analysis (NMDS, Fig. 2B). *S. purpuratus*

160 RNA-seq data does not include the early development time point equivalent to the *P. miniata*

161 9hpf and *P. lividus* 4hpf [21]. The NMDS maps the equivalent developmental time points of the

162 two sea urchin species closely to each other (purple and black tracks), while the sea star samples

5

163   are relatively distinct (orange track). This is in agreement with the relative evolutionary distances

164   between the three species (Fig. 1A). However, the developmental progression in all three species

165   is along a similar trajectory in the NMDS two dimensional space, which probably reflects the

166   resemblance in the overall morphology between these three echinoderm species. Overall, the

167   NMDS indicates that major sources of variation in these data sets are developmental progression

168   and evolutionary distance.

Interspecies correlations decrease and the pattern becomes less diagonal with evolutionary distance

171   We wanted to study how the pattern of the interspecies correlations of gene expression changes

172   with evolutionary distance for different classes of genes. To be able to compare the interspecies

173   correlations between the three species we included only time points that had data for all species.

174   Explicitly, we excluded *P. lividus* 4hpf and *P. miniata* 9hpf that do not have an equivalent time

175   point in the *S. purpuratus* data. We calculated the Pearson correlations of gene expression levels

176   between *P. lividus* and *S. purpuratus* and between *P. lividus* and *P. miniata* for subsets of

177   homologous genes with specific developmental, housekeeping or metabolic function, according

178   to their GO terms (Fig. 3). As expected, the interspecies correlation patterns differ in strength and

179   diagonality between different functional classes for both *Pl-Sp* and *Pl-Pm* comparison.

180   For a better assessment of the correlation pattern we sought to quantify these two distinct

181   properties: the correlation strength between equivalent developmental stages and the diagonality

182   of the matrix. As a measure of the correlation strength between morphologically similar time

183   points, we defined the parameter **AC = Average Correlation, which is the average of the**

184   **diagonal elements of the correlation matrix**. That is, AC=1 corresponds to perfect correlation

185   throughout all equivalent developmental stages and AC=0 corresponds to no correlation.

186   Evidently, AC is not a measure of the diagonality of the matrix as it doesn't consider the off-

187   diagonal terms. To quantify the diagonality of the correlation matrices we used a statistical test

188   we developed before to assess whether the interspecies correlation pattern is significantly close to

189   a diagonal matrix [17]. Briefly, the parameter, **matrix diagonality (MD), indicates how many**

190   **times within 100 subsamples of the tested set of genes, the interspecies correlation matrix**

191   **was significantly close to a diagonal matrix compared to a random matrix**. Hence, MD=100

192   is the highest diagonality score and 0 is the lowest (see materials and methods and [17] for

193   explanation, and Table S5 for results. MD is equivalent to 'count significant', or CS, in [17]).

194   Importantly, AC will be high for any matrix that has high diagonal elements, but MD will be high

195   only when the off-diagonal matrix elements are much lower than the diagonal, and low otherwise.

196   We ordered the correlation matrices according to their MD value in *Sp-Pl,* from the most diagonal

197   (Cell differentiation, MD=99) to the least (RNA processing, MD=17). This ordering clearly

198   demonstrates the high diagonality of the developmental genes (Fig. 3A-C) vs. the block patterns

199   of the housekeeping genes (Fig. 3G-J) and how the diagonality of the correlation pattern of all

200   genes combined is in between (Fig. 3F). Both the average correlation and the matrix diagonality

201   are lower between the sea urchin and the sea star than between the two sea urchins, that is, both

202   parameters decrease with evolutionary distance (compare bottom to upper panels in Fig. 3A-I).

203   Low correlation matrix diagonality indicates that the off-diagonal elements have a similar

204   strength to the diagonal elements. This means that the relationship between the expression levels

205   of a set of genes is maintained not only between morphologically similar developmental stages,

206   but throughout development. In the case of housekeeping genes that form complex structures, like

207   ribosomal proteins, this could be explained by stoichiometric ratio requirements between different

208   proteins that form the structure. With increasing evolutionary distance and morphological

209   divergence there is a reduction in matrix diagonality for all sets of genes (Fig. 3). Possibly, the

210   dominant constraints in large evolutionary distances are the cellular requirement for different

211   levels of transcripts that have different functions. These constraints are cellular and not related to

212   specific morphogenetic processes, which might explain the even correlation strength between

213   different developmental stages and hence, the lower diagonality at large evolutionary distances.

214   **The interspecies correlation vary between different cell lineages**

215   Our analysis is based on RNA-seq on whole embryos, yet we wanted to see if gene sets enriched

216   in a specific cell lineage show a difference in their interspecies correlation pattern. In a previous

217   work conducted in the sea urchin *S. purpuratus* embryos, the cells of six distinct embryonic

218   lineages were isolated based on cell-specific GFP reporter expression. Gene expression levels in

219   the isolated cells were studied and compared to gene expression levels in the rest of the embryo

220   by RNA-seq [28]. This analysis identified genes that their expression is enriched in the specific

221   lineages at the time when the isolation was done [28]. We calculated the Pearson interspecies

222   correlations for the subsets of genes enriched in each of the six sea urchin lineages, between the

223   two sea urchins and between the Mediterranean sea urchin and the sea star (Fig. 4). The sea

224   urchin cell lineages at the time of cell isolation are illustrated above the correlation patterns and

225   the developmental stage is indicated as a black square within the *Pl-Sp* correlation matrices.

226   Both the matrix diagonality and average correlation strength vary between the different cell types

227   and decrease with evolutionary distance. Interestingly, the genes enriched in sea urchin pigment

228   cells, a lineage that is lacking in the sea star, show similar *Pl-Pm* average correlation strength

229   compared to the matrices of the conserved lineages between the sea urchin and the sea star (*Pl-Sp*

230   AC=0.64 and *Pl-Pm* AC=0.5, Fig. 4B). As mentioned above, the regulatory state in the

231   mesoderm of the sea star and the sea urchin are quite similar [7]. Possibly, there is also similarity

232   in the downstream genes active in sea star blastocoelar cells and sea urchin pigment cells, as both

233   lineages have hematopoietic function [29]. On the other hand, the sharpest decrease in *Pl-Pm*

234   average correlation strength compared to *Pl-Sp* is for the genes enriched in the sea urchin

235   skeletogenic cells, another lineage that is absent in the sea star (*Pl-Sp* AC=0.71 while *Pl-Pm*

236   AC=0.4, Fig. 4C). Interestingly, the *Pl-Pm* matrix diagonality of these genes is comparable to the

237   *Pl-Pm* matrix diagonality of the conserved lineages (*Pl-Pm* MD=0.33, Fig. 4C). It is important to

238   note that key sea urchin skeletogenic matrix proteins were not found in the sea star skeleton [30]

239   and the genes encoding them were not found in the sea star genome. Therefore these key

240   skeletogenic genes are missing from our 1:1:1 homologous genes that include only genes that are

241   common to the three species. Thus, we are probably underestimating the differences in

242   skeletogenic and mesodermal gene expression between the sea urchin and the sea star, which

243   might explain the relatively high diagonality for the skeletogenic gene correlation matrix. Overall,

244   genes enriched in different cells lineages differ in their correlation pattern even between the

245   closely related sea urchins and show distinct differences in the correlation pattern with

246   evolutionary distance.

Matrix diagonality and correlation strength describe different properties of gene expression conservation

249   Both the average correlation strength and the matrix diagonality depend on gene function, cell

250   lineage and decrease with evolutionary distance (Figs. 3, 4). When we plot, separately, the

251   average correlation and the matrix diagonality for different GO terms and cell lineages, in

252   decreasing strength in *Pl-Pm* we see two distinct orders (Fig. 5A and B). While the matrix

253   diagonality separates clearly the developmental genes with high diagonality from the

254   housekeeping genes with low diagonality, the average correlation does not show this distinction

255   (Fig. 5A, B). Moreover, the matrix diagonality and the average correlation strength change

256   independently of each other in both *Pl-Sp* and *Pl-Pm* correlation matrices (Fig. 5C). On the other

257   hand, the average correlations in *Pl-Sp* matrices seem to correspond to the average correlation in

258   *Pl-Pm* (Fig. 5D), and the same is true for the matrix diagonality (Fig. 5E).

259   These analyses imply that these two parameters describe different properties of the conservation

260   in gene expression, as we illustrate in Fig. 6. Apparently, the average correlation strength is a

261   measure of the conservation in gene expression levels: the higher it is, the more the gene-set is

262 constrained against expression change. The matrix diagonality on the other hand, seems to reflect

263 the link between the expression of a gene-set and morphological conservation: the more diagonal

264 is the correlation pattern of a gene set, the more related is the gene expression conservation within

265 the set to morphological similarity (Fig. 6). For example, the correlation matrices of transcription

266 factors and cell differentiation genes show strong correlations and high diagonality even between

267 the sea urchin and the sea star (Relatively high AC and MD, Fig. 3A,B, 6). Conversely, ribosomal

268 gene expression is highly conserved (high AC) but this conservation is not related to

269 morphological conservation (Low MD, Fig. 3I, 5A,B, 6). Overall, the average correlation strength

270 and matrix diagonality seem to carry complementary information about the relationship between

271 the conservation of gene expression and morphological conservation at varying evolutionary

272 distances (Fig. 6).

273 **Discussion**

274 In this paper we generated the developmental transcriptomes of the pacific sea star, *P. miniata,*

275 and studied them in comparison with the published developmental transcriptomes of two sea

276 urchin species, *P. lividus* [17] and *S. purpuratus* [21]. We generated a web application where the

277 *P. miniata* time courses and sequences can be publically viewed to facilitate the common use of

278 this data [25]. We studied the interspecies correlation patterns of different gene sets including,

279 housekeeping, developmental and metabolic genes (Fig. 3), as well as genes that are enriched in a

280 specific cell lineage in the sea urchin embryo (Fig. 4). We defined two parameters that describe

281 different properties of the conservation strength: the average correlation strength in the diagonal,

282 AC, and the matrix diagonality, MD. We noticed that these parameters vary independently

283 between different functional groups and cell types and decrease with evolutionary distance,

284 possibly reflecting different constraints on gene expression (Fig. 5). The correlation strength is an

285 indication of the evolutionary constraint on gene expression level while the matrix diagonality

286 seems to reflect the relationship between the gene set and morphological similarity. As we

287 suggested previously, parallel embryonic transcriptional programs might be responsible for

288 different aspects of embryo development and evolve under distinct constraint [17], as can be

289 inferred from analyzing these two parameters.

290 Previous studies have shown that housekeeping genes and tissue specific genes have different

291 chromatin structures [31] and distinct core promoters [32]. Furthermore, Enhancers of

292 developmental genes were shown to be de-methylated during the vertebrates' phylotypic period,

293 suggesting another unique epigenetic regulation of this set of genes [33]. These epigenetic and

294 *cis*-regulatory differences could underlie the separation of the regulation of developmental and

9

295   housekeeping gene expression, leading to dissimilar evolutionary conservation patterns of these
296   two functionally distinct gene sets.

297   We observed a clear reduction of the matrix diagonality with increasing evolutionary distance for
298   all sets of genes, including developmental genes and transcription factors (Fig. 3-6). This is in
299   agreement with the morphological divergence between the sea urchin and the sea star. As
300   mentioned above, a recent study had shown that the correlation matrices for all homologous
301   genes between 10 different phyla are strictly not-diagonal [20]. At this large evolutionary distance
302   and extreme morphological divergence the only constraint that remains seem to be the cellular
303   requirements for differential transcript abundance, which are unrelated to morphological
304   similarity. Therefore the opposite hourglass pattern found for the diagonal elements of the
305   interspecies correlation matrices might not be indicative for morphological divergence at the
306   phylotypic stage. Overall, to better estimate the relationship between the conservation of gene
307   expression levels to morphological conservation, both correlation strength and correlation matrix
308   diagonality should be assessed and the focus should be on developmental control genes.

Materials and Methods

Sea star embryo cultures and RNA extraction

311   Adult *P. miniata* sea stars were obtained in Long Beach, California, from Peter Halmay. Embryos
312   were cultured at 15°C in artificial sea water. Total RNA was extracted using Qiagen mini RNeasy
313   kit from thousands of embryos. RNA samples were collected from eight embryonic stages, from
314   fertilized egg to late gastrula stage (Fig. 1B). For each embryonic stage, three biological
315   replicates from three different sets of parents were processed, except for the last time point for
316   which only two biological replicates were samples (23 samples in all). To match *P. miniata* time
317   points to those of the published *S. purpuratus* and *P. lividus* transcriptomes [17, 22] we used the
318   linear ratio between the developmental rates of these species found in [12, 34]. The corresponding
319   time points in both species are presented in Fig. 1B. The time points, 9 hour post fertilization
320   (hpf) in *P. miniata* and 4hpf in *P. lividus* do not have a comparable time point in *S. purpuratus*
321   data. RNA quantity was measured by nanodrop and quality was verified using bioanalyzer.

Transcriptome assembly, annotations and quantification

Transcriptome assembly, annotations and quantification

324   **RNA-Seq preparation:** Library preparation was done at the Israel National Center for
325   Personalized Medicine (INCPM) using their RNA-seq protocol. Briefly, polyA fraction (mRNA)

10

326 was purified from 500 ng of total RNA per developmental time point following by fragmentation

327 and generation of double stranded cDNA. Then, end repair, A base addition, adapter ligation and

328 PCR amplification steps were performed. Libraries were evaluated by Qubit and TapeStation.

329 Sequencing libraries were constructed with barcodes to allow multiplexing of 47 samples in one

330 lane. On average, 20 million single-end 60-bp reads were sequenced per sample on Illumina

331 HiSeq 2500 V4 using four lanes.

332 **RNA-Seq datasets used:** Four datasets were used: (1) newly sequenced *P. miniata* Single End

333 (SE) reads of 23 transcriptome samples generated as explained above; (2) publically available

334 Paired End (PE) reads of *P. miniata* from different developmental stages, testis and ovaries,

335 accessions: SRR6054712, SRR5986254, SRR2454338, SRR1138705, SRR573705-SRR573710

336 and SRR573675; (3) publically available SE reads of *P. lividus* transcriptome samples from eight

337 developmental stages (NCBI project PRJNA375820, [17]) ; (4) publically available PE reads of

338 seven *S. purpuratus* transcriptome samples (NCBI project PRJNA81157, [22]).

339 **RNA-seq quality filtering:** RNA-Seq reads from the above datasets were adapter-trimmed using

340 cutadapt 1.15 (https://cutadapt.readthedocs.io), then low-quality regions were removed with

341 Trimmomatic 0.3 [35], and further visually inspected in fastq-screen

342 (www.bioinformatics.babraham.ac.uk).

343 *P. miniata* **transcriptome assembly** While *P. miniata* genome based gene predictions are

344 available for genome assembly v.2 (echinobase.org/Echinobase), many gene sequences are

345 fragmented or duplicated within the genome, and therefore we decided to generate a reference

346 transcriptome *de-novo*. Accordingly, *P. miniata* RNA-Seq reads, based on the available PE and

347 SE data, were assembled using Trinity 2.4 [36, 37] with the same Trinity parameters as before we

348 used for *P. lividus* in [17]. Trinity produced 1,610,829 contigs (the Trinity equivalent of

349 transcripts), within 679,326 Trinity gene groups. Transcriptome completeness was tested using

350 BUSCO [38], and by comparing the contigs to genome-based protein annotations *P. miniata* v.2.0

351 and *S. purpuratus* v.3.1 (WHL22), (echinobase.org/Echinobase).

352 *P. miniata* **data availability:** Illumina short read sequences generated in this study was submitted

353 to the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra), under bio-project

354 PRJNA522463 (http://www.ncbi.nlm.nih.gov/bioproject/522463). Fastq read accessions:

355 SRR8580044 - SRR8580066, assembled *P. miniara* transcriptome accession: SAMN10967027.

356 **Transcriptome homology:** We searched for homologous genes in the *P. miniata* transcriptome,

357 *P. lividus* transcriptome*, S. purpuratus* genome. Since the *S. purpuratus* genome-based gene

358 predictions dataset currently includes the most non-redundant and complete data among the three

359 datasets, it was used as a reference dataset. Accordingly, the largest isoforms of new *P. miniata*

360 transcriptome (see below), and the publicly available *P. lividus* transcriptome [17], were both

361 compare to the *S.purpuratus* genome-based predicted protein annotations (Echinobase v.3.1),

362 using CRB-BLAST [39]. CRB-BLAST reports relationships of 1:1 (Reciprocal hits), and 2:1

363 matches (when Blastx and tBlastn results are not reciprocal), where only matches with e-values

364 below a conditional threshold are reported. From the 2:1 cases, we selected the query-target pair

365 with the lowest e-value. Using CRB-BLAST, 11,291 and 12,720 *P. miniata* and *P. lividus* Trinity

366 genes, were identified as homologous to *S. purpuratus* proteins, respectively. We considered *P.*

367 *miniata* and *P. lividus* query genes that share the same *S. purpuratus* target gene, as homologous.

368 As the gene expression analysis shows (see next sections), 8,735 homologous genes are expressed

369 in at least one of the three tested species during development, and 6,593 are expressed in all the

370 three.

371 **Gene-level transcripts abundance**: For *P. miniata* and *P. lividus* transcriptomes, transcripts

372 abundance was estimated using kallisto-0.44.0 [40], and a further quantification at the gene-level,

373 and read-count level, was done using tximport [41] on R3.4.2. Expression analysis at read-count

374 level was conducted at gene-level in Deseq2 [42]. *S. purpuratus* PE reads, of 7 developmental

375 samples were mapped to the *S. Purpuratus 3.1* genome assembly, using STAR v2.4.2a [43],

376 quantitated in Htseq-count v2.7 [44], and analyzed in Deseq2 at read-count level. For *P. miniata*

377 data, only contigs mapped to the Echinobase v.2 *P. miniata* proteins were considered. Prior to

378 DEseq2 read count standard normalization and expression analysis, genes with <1 CPM (Count

379 Per Million) were removed. Overall, most input reads were mapped and quantified, as further

380 detailed in Table S1. Samples were clustered using Non-metric multi-dimentional scaling

381 (NMDS) ordination in Vegan (https://cran.r-project.org/web/packages/vegan/index.html), based

382 on log10 transformed FPKM values, and Bray-Curtis distances between samples. Since the

383 NMDS results indicate that all *P. miniata* and *P. lividus* samples are affected by the 'batch' factor

384 (see Table S1), we removed the estimated effect of this factor on FPKMs (Fragment per Kilobase

385 Million) values using "removeBatchEffect" function in EdgeR [45], in order to obtain corrected

386 FPKM values. Quantification and annotations of 34,307 identified *P. miniata* transcripts with

387 FPKM>3 in at least one time point, are provided in Table S2. NMDS analysis of the biological

388 replicates of all time points in *P. miniata* show that similar time points at different biological

389 replicate map together indicating high reproducibility of our gene expression analysis (Fig. S2).

390 Comparison between our RNA-seq quantification of gene expression and previous QPCR

391 quantification [12] for a sub-set of genes show high agreement between the two measurements

392   (Fig. S3). Quantification and annotations of the 8,735 identified *P. miniata, P. lividus and S.*

393   *purpuratus* 1:1:1 homologous genes are provided in Table S3 and are publicly available through

394   gene search in Ehinobase at www.echinobase.org/shiny/quantdevPm.

395   **Gene Ontology functional enrichment analysis:** Functional enrichment analysis was conducted

396   using TopGo in R3.4.2 (bioconductor.org). A custom GOseq GO database was built using the

397   publically   available   Blast2Go   *S.   purpuratus   v3.1*   (WHL22)   version

398   (http://www.echinobase.org/Echinobase/rnaseq/download/blast2go-whl.annot.txt.gz).

399   **Pearson correlations for subsets of genes**: The interspecies Pearson correlations for different

400   sets of genes presented in Figs. 3 and 4 were calculated using R3.4.2. For the analysis in Fig. 4,

401   we selected genes that have 1:1:1 homologs in *P. miniata* and *P. lividus* to the *S. purpuratus*

402   genes that are enriched in a specific *S. purpuratus* cell lineage with p-value<0.05, based on [28].

403   **Cross species analysis of matrix diagonality (MD):** We used a statistical test described in

404   detail in [17]. Shorty, the main goal of this procedure is to test the probability that a set of

405   homologous genes from two species, S_1 and S_2, show the most similar expression patterns in

406   equivalent developmental times, namely: is the interspecies correlation pattern significantly close

407   to a diagonal matrix? Here, S_1 and S_2 represent *P. miniata* vs. *P. lividus*, or *P. lividus* vs. *S.*

408   *purpuratus*. This test was conducted using all homologous genes, and for specific subsets of

409   genes belonging to specific GO categories as well as for genes enriched in specific cell lineage

410   [28]. Here, only samples from the $n=5$ late embryonic stages were used, and the mean of FPKM

411   values of all samples belonging to the same stages were taken. First, n_{s1} by n_{s2} matrix of Pearson

412   correlations, C, was produced, where n_{s1} by n_{s2} are the number of time point measurements in

413   each species (here $n_{s1}=n_{s2}=5$). Each of the C matrix positions, C_{ij}, represents a Pearson's

414   correlation value calculated based on n_g FPKM values, between stage i in one species and stage j

415   in the other, where n_g is the count of homologous gene pairs tested. Next, the C matrix was

416   compared to an "ideal" time-dependent correlation matrix I, in which a correlation of 1 exists

417   along the diagonal line, and 0 in other positions, to obtain a diagonality measure, d. Overall,

418   $n_g=50$ genes were resampled $n_{resamp}=100$ times, and for each resampling-iteration, a permutation

419   test based on value d was applied using $n_{perm}=1000$ permutations. Accordingly, each of the above

420   $n_{resamp}=100$ permutation tests indicate the probability that S_1 and S_2 show non-random diagonality,

421   on a subset of $n_g=50$ genes. Then, the count of permutation tests indicating a significant similarity

422   to the ideal diagonal matrix, out of the $n_{resamp}=100$ subsamples is used for estimating diagonality

13

423 of the correlation matrix (matrix diagonality = MD). Our results are presented in Table S5 and
424 within Figs. 3-5.

Acknowledgements

References

435 1.   Peter IS, Davidson EH. Genomic Control Process: Development and Evolution. New York:
436 Academic Press; 2015.
437 2.   Raff RA, Byrne M. The active evolutionary lives of echinoderm larvae. Heredity.
438 2006;97(3):244-52. Epub 2006/07/20. doi: 10.1038/sj.hdy.6800866. PubMed PMID: 16850040.
439 3.   Oliveri P, Tu Q, Davidson EH. Global regulatory logic for specification of an embryonic
440 cell lineage. Proc Natl Acad Sci U S A. 2008;105(16):5955-62. Epub 2008/04/17. doi:
441 10.1073/pnas.0711220105. PubMed PMID: 18413610; PubMed Central PMCID: PMC2329687.
442 4.   Ben-Tabou de-Leon S, Su YH, Lin KT, Li E, Davidson EH. Gene regulatory control in the
443 sea urchin aboral ectoderm: Spatial initiation, signaling inputs, and cell fate lockdown. Dev Biol.
444 2012. Epub 2012/12/06. doi: 10.1016/j.ydbio.2012.11.013. PubMed PMID: 23211652.
445 5.   Li E, Materna SC, Davidson EH. Direct and indirect control of oral ectoderm regulatory
446 gene expression by Nodal signaling in the sea urchin embryo. Dev Biol. 2012;369(2):377-85.
447 Epub 2012/07/10. doi: 10.1016/j.ydbio.2012.06.022. PubMed PMID: 22771578; PubMed Central
448 PMCID: PMC3423475.
449 6.   Hinman VF, Nguyen AT, Cameron RA, Davidson EH. Developmental gene regulatory
450 network architecture across 500 million years of echinoderm evolution. Proc Natl Acad Sci U S A.
451 2003;100(23):13356-61. Epub 2003/11/05. doi: 10.1073/pnas.2235868100

452 2235868100 [pii]. PubMed PMID: 14595011; PubMed Central PMCID: PMC263818.
453 7.   McCauley BS, Weideman EP, Hinman VF. A conserved gene regulatory network
454 subcircuit drives different developmental fates in the vegetal pole of highly divergent
455 echinoderm embryos. Dev Biol. 2010;340(2):200-8. Epub 2009/11/28. doi: S0012-
456 1606(09)01375-X [pii]

457 10.1016/j.ydbio.2009.11.020. PubMed PMID: 19941847.
458 8.   McCauley BS, Wright EP, Exner C, Kitazawa C, Hinman VF. Development of an embryonic
459 skeletogenic mesenchyme lineage in a sea cucumber reveals the trajectory of change for the
460 evolution of novel structures in echinoderms. Evodevo. 2012;3(1):17. Epub 2012/08/11. doi:
461 10.1186/2041-9139-3-17. PubMed PMID: 22877149.

462 9.   Yankura KA, Martik ML, Jennings CK, Hinman VF. Uncoupling of complex regulatory
463 patterning during evolution of larval development in echinoderms. BMC Biol. 2010;8:143. Epub
464 2010/12/02. doi: 1741-7007-8-143 [pii]

465 10.1186/1741-7007-8-143. PubMed PMID: 21118544; PubMed Central PMCID: PMC3002323.
466 10.   McCauley BS, Akyar E, Filliger L, Hinman VF. Expression of wnt and frizzled genes during
467 early sea star development. Gene Expr Patterns. 2013;13(8):437-44. Epub 2013/08/01. doi:
468 10.1016/j.gep.2013.07.007. PubMed PMID: 23899422.
469 11.   Yankura KA, Koechlein CS, Cryan AF, Cheatle A, Hinman VF. Gene regulatory network for
470 neurogenesis in a sea star embryo connects broad neural specification and localized patterning.
471 Proc Natl Acad Sci U S A. 2013;110(21):8591-6. Epub 2013/05/08. doi:
472 10.1073/pnas.1220903110. PubMed PMID: 23650356; PubMed Central PMCID: PMC3666678.
473 12.   Gildor T, Hinman V, Ben-Tabou-De-Leon S. Regulatory heterochronies and loose
474 temporal scaling between sea star and sea urchin regulatory circuits. Int J Dev Biol. 2017;61(3-4-
475 5):347-56. doi: 10.1387/ijdb.160331sb. PubMed PMID: 28621432.
476 13.   Cary GA, Hinman VF. Echinoderm development and evolution in the post-genomic era.
477 Dev Biol. 2017;427(2):203-11. doi: 10.1016/j.ydbio.2017.02.003. PubMed PMID: 28185788.
478 14.   Dylus DV, Czarkwiani A, Blowes LM, Elphick MR, Oliveri P. Developmental
479 transcriptomics of the brittle star Amphiura filiformis reveals gene regulatory network rewiring
480 in echinoderm larval skeleton evolution. Genome Biol. 2018;19(1):26. doi: 10.1186/s13059-018-
481 1402-8. PubMed PMID: 29490679; PubMed Central PMCID: PMCPMC5831733.
482 15.   Israel JW, Martik ML, Byrne M, Raff EC, Raff RA, McClay DR, et al. Comparative
483 Developmental Transcriptomics Reveals Rewiring of a Highly Conserved Gene Regulatory
484 Network during a Major Life History Switch in the Sea Urchin Genus Heliocidaris. PLoS Biol.
485 2016;14(3):e1002391. doi: 10.1371/journal.pbio.1002391. PubMed PMID: 26943850; PubMed
486 Central PMCID: PMCPMC4778923.
487 16.   Gildor T, Smadar BD. Comparative Studies of Gene Expression Kinetics: Methodologies
488 and Insights on Development and Evolution. Frontiers in genetics. 2018;9:339. doi:
489 10.3389/fgene.2018.00339. PubMed PMID: 30186312; PubMed Central PMCID:
490 PMCPMC6113378.
491 17.   Malik A, Gildor T, Sher N, Layous M, Ben-Tabou de-Leon S. Parallel embryonic
492 transcriptional programs evolve under distinct constraints and may enable morphological
493 conservation amidst adaptation. Dev Biol. 2017. doi: 10.1016/j.ydbio.2017.07.019. PubMed
494 PMID: 28780048.
495 18.   Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, et al. Gene
496 expression divergence recapitulates the developmental hourglass model. Nature.
497 2010;468(7325):811-4. doi: 10.1038/nature09634. PubMed PMID: 21150996.
498 19.   Piasecka B, Lichocki P, Moretti S, Bergmann S, Robinson-Rechavi M. The hourglass and
499 the early conservation models--co-existing patterns of developmental constraints in vertebrates.
500 PLoS Genet. 2013;9(4):e1003476. doi: 10.1371/journal.pgen.1003476. PubMed PMID:
501 23637639; PubMed Central PMCID: PMC3636041.
502 20.   Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, et al. The mid-developmental
503 transition and the evolution of animal body plans. Nature. 2016;531(7596):637-41. doi:
504 10.1038/nature16994. PubMed PMID: 26886793; PubMed Central PMCID: PMCPMC4817236.
505 21.   Tu Q, Cameron RA, Davidson EH. Quantitative developmental transcriptomes of the sea
506 urchin Strongylocentrotus purpuratus. Dev Biol. 2014;385(2):160-7. Epub 2013/12/03. doi:
507 10.1016/j.ydbio.2013.11.019. PubMed PMID: 24291147; PubMed Central PMCID: PMC3898891.

508 22.   Tu Q, Cameron RA, Worley KC, Gibbs RA, Davidson EH. Gene structure in the sea urchin
509 Strongylocentrotus purpuratus based on transcriptome analysis. Genome Res.
510 2012;22(10):2079-87. doi: 10.1101/gr.139170.112. PubMed PMID: 22709795; PubMed Central
511 PMCID: PMCPMC3460201.
512 23.   Materna SC, Nam J, Davidson EH. High accuracy, high-resolution prevalence
513 measurement for the majority of locally expressed regulatory genes in early sea urchin
514 development. Gene Expr Patterns. 2010;10(4-5):177-84. Epub 2010/04/20. doi:
515 10.1016/j.gep.2010.04.002. PubMed PMID: 20398801; PubMed Central PMCID: PMC2902461.
516 24.   Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq:
517 accounting for selection bias. Genome Biol. 2010;11(2):R14. Epub 2010/02/06. doi: 10.1186/gb-
518 2010-11-2-r14. PubMed PMID: 20132535; PubMed Central PMCID: PMC2872874.
519 25.   Tsvia Gildor GC, Maya Lalzar, Veronica Hinman and Smadar Ben-Tabou de-Leon.
520 Echinobase viewer of *P. miniata* quantitative developmental transcriptomes
521 www.echinobase.org/shiny/quantdevPm.
522 26.   Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Framework for R. R package version
523 1.2.0. https://CRANR-projectorg/package=shiny 2018.
524 27.   Cameron RA, Samanta M, Yuan A, He D, Davidson E. SpBase: the sea urchin genome
525 database and web site. Nucleic Acids Res. 2009;37(Database issue):D750-4. Epub 2008/11/18.
526 doi: gkn887 [pii]

527 10.1093/nar/gkn887. PubMed PMID: 19010966; PubMed Central PMCID: PMC2686435.
528 28.   Barsi JC, Tu Q, Calestani C, Davidson EH. Genome-wide assessment of differential
529 effector gene use in embryogenesis. Development. 2015;142(22):3892-901. doi:
530 10.1242/dev.127746. PubMed PMID: 26417044; PubMed Central PMCID: PMCPMC4712884.
531 29.   Solek CM, Oliveri P, Loza-Coll M, Schrankel CS, Ho EC, Wang G, et al. An ancient role for
532 Gata-1/2/3 and Scl transcription factor homologs in the development of immunocytes. Dev Biol.
533 2013;382(1):280-92. Epub 2013/06/25. doi: 10.1016/j.ydbio.2013.06.019. PubMed PMID:
534 23792116.
535 30.   Flores RL, Livingston BT. The skeletal proteome of the sea star Patiria miniata and
536 evolution of biomineralization in echinoderms. BMC Evol Biol. 2017;17(1):125. doi:
537 10.1186/s12862-017-0978-z. PubMed PMID: 28583083; PubMed Central PMCID:
538 PMCPMC5460417.
539 31.   She X, Rohl CA, Castle JC, Kulkarni AV, Johnson JM, Chen R. Definition, conservation and
540 epigenetics of housekeeping and tissue-enriched genes. BMC Genomics. 2009;10:269. doi:
541 10.1186/1471-2164-10-269. PubMed PMID: 19534766; PubMed Central PMCID:
542 PMCPMC2706266.
543 32.   Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, et al. Enhancer-core-
544 promoter specificity separates developmental and housekeeping gene regulation. Nature.
545 2015;518(7540):556-9. doi: 10.1038/nature13994. PubMed PMID: 25517091.
546 33.   Bogdanovic O, Smits AH, de la Calle Mustienes E, Tena JJ, Ford E, Williams R, et al. Active
547 DNA demethylation at enhancers during the vertebrate phylotypic period. Nat Genet.
548 2016;48(4):417-26. doi: 10.1038/ng.3522. PubMed PMID: 26928226; PubMed Central PMCID:
549 PMCPMC5912259.
550 34.   Gildor T, Ben-Tabou de-Leon S. Comparative Study of Regulatory Circuits in Two Sea
551 Urchin Species Reveals Tight Control of Timing and High Conservation of Expression Dynamics.
552 PLoS Genet. 2015;11(7):e1005435. doi: 10.1371/journal.pgen.1005435. PubMed PMID:
553 26230518.

554 35.   Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
555 data. Bioinformatics. 2014;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. PubMed PMID:
556 24695404; PubMed Central PMCID: PMCPMC4103590.
557 36.   Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo
558 transcript sequence reconstruction from RNA-seq using the Trinity platform for reference
559 generation and analysis. Nat Protoc. 2013;8(8):1494-512. Epub 2013/07/13. doi:
560 10.1038/nprot.2013.084. PubMed PMID: 23845962.
561 37.   Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length
562 transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol.
563 2011;29(7):644-52. Epub 2011/05/17. doi: 10.1038/nbt.1883. PubMed PMID: 21572440;
564 PubMed Central PMCID: PMC3571712.
565 38.   Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, et al.
566 BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol
567 Evol. 2017. doi: 10.1093/molbev/msx319. PubMed PMID: 29220515; PubMed Central PMCID:
568 PMCPMC5850278.
569 39.   Aubry S, Kelly S, Kumpers BM, Smith-Unna RD, Hibberd JM. Deep evolutionary
570 comparison of gene expression identifies parallel recruitment of trans-factors in two
571 independent origins of C4 photosynthesis. PLoS Genet. 2014;10(6):e1004365. doi:
572 10.1371/journal.pgen.1004365. PubMed PMID: 24901697; PubMed Central PMCID:
573 PMCPMC4046924.
574 40.   Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
575 quantification. Nat Biotechnol. 2016;34(5):525-7. doi: 10.1038/nbt.3519. PubMed PMID:
576 27043002.
577 41.   Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level
578 estimates improve gene-level inferences. F1000Res. 2015;4:1521. doi:
579 10.12688/f1000research.7563.2. PubMed PMID: 26925227; PubMed Central PMCID:
580 PMCPMC4712774.
581 42.   Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
582 RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8.
583 PubMed PMID: 25516281; PubMed Central PMCID: PMCPMC4302049.
584 43.   Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
585 universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. doi:
586 10.1093/bioinformatics/bts635. PubMed PMID: 23104886; PubMed Central PMCID:
587 PMCPMC3530905.
588 44.   Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput
589 sequencing data. Bioinformatics. 2015;31(2):166-9. doi: 10.1093/bioinformatics/btu638.
590 PubMed PMID: 25260700; PubMed Central PMCID: PMCPMC4287950.
591 45.   Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
592 expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40. Epub
593 2009/11/17. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308; PubMed Central
594 PMCID: PMC2796818.

595

596

597   **Figures**

598
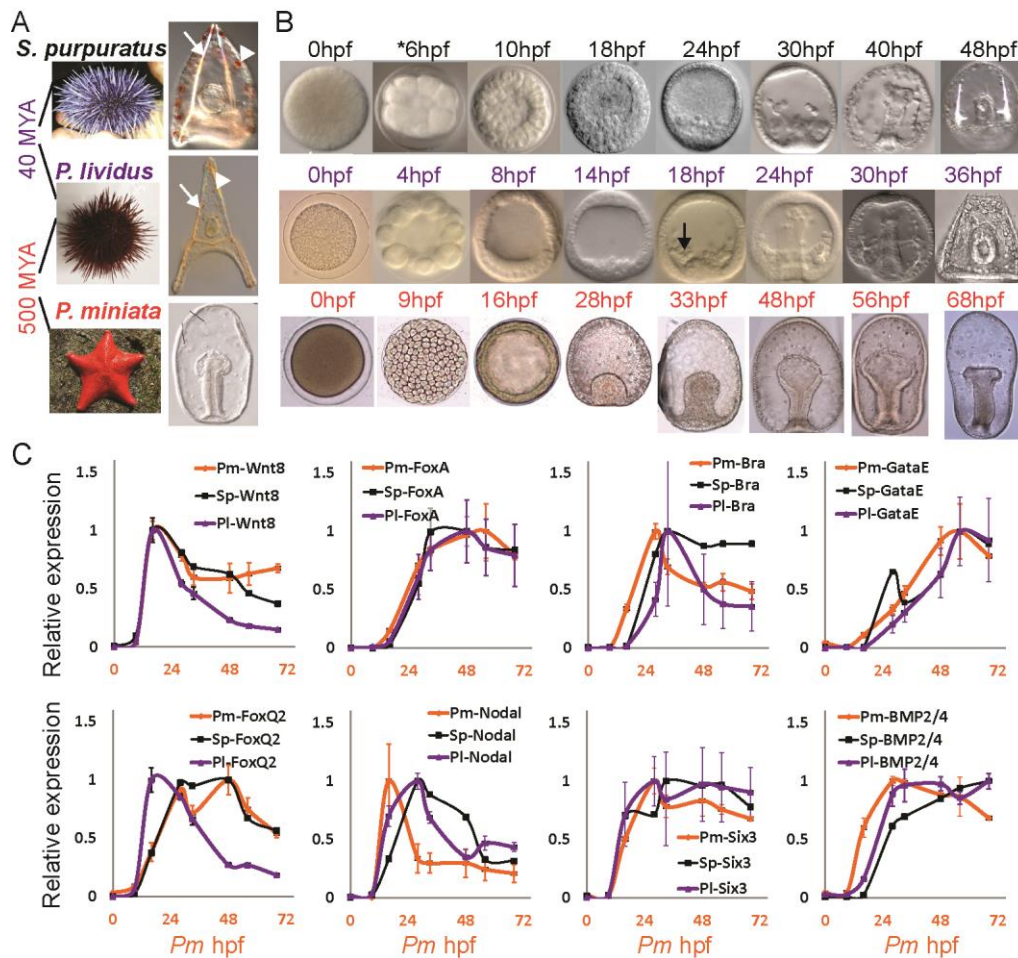
599   **Figure 1 Developmental time points studied and examples for gene expression profiles on**
600   **the three species.** (A) Images of adult and larval stage of *P. miniata*, *P. lividus* and *S. purpuratus*
601   Arrows point to the sea urchin skeletogenic rods and arrow heads point to the sea urchin
602   pigments. (B) Images of *P. miniata*, *P. lividus* and *S. purpuratus* embryos at the developmental
603   stages that were studied in this work. Time point 6hpf in *S. purpuratus* does not have RNA-seq
604   data. (C) Relative gene expression in the three species measured in the current paper by RNA-seq
605   for *P. miniata* (orange curves), in [17] by RNA-seq for *P. lividus* (purple curves) and in [23] by
606   nanostring for *S. purpuratus* (black curves). Error bars indicate standard deviation. To obtain
607   relative expression levels for each species we divide the level at each time point in the maximal
608   mRNA level measured for this species in this time interval; so 1 is the maximal expression in this
609   time interval.

610
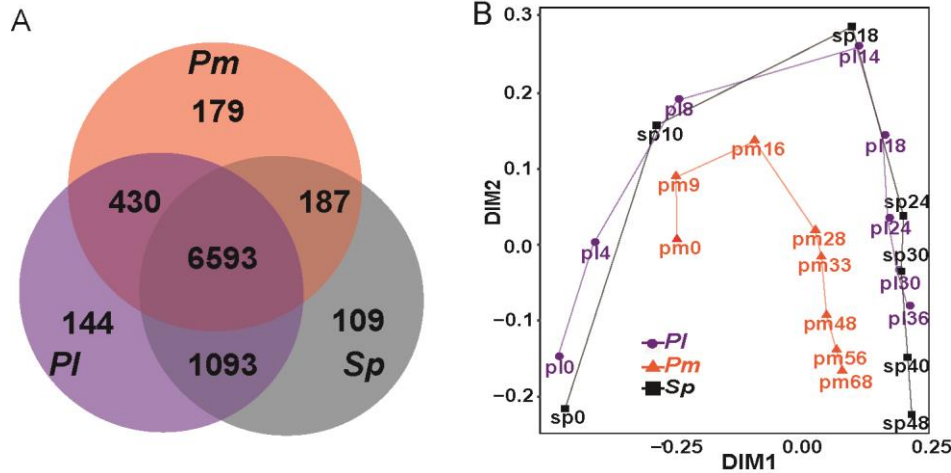
Figure 2 Venn diagram and NMDS analysis of 1:1:1 orthologues genes. (A) Venn diagram showing the number of 1:1:1 orthologues genes expressed in all three species, in two of the species or only in one species. (B) First two principal components of expression variation (NMDS) between different developmental time points in *P. miniata* (orange), *P. lividus* (purple) and *S. purpuratus* (black).

19
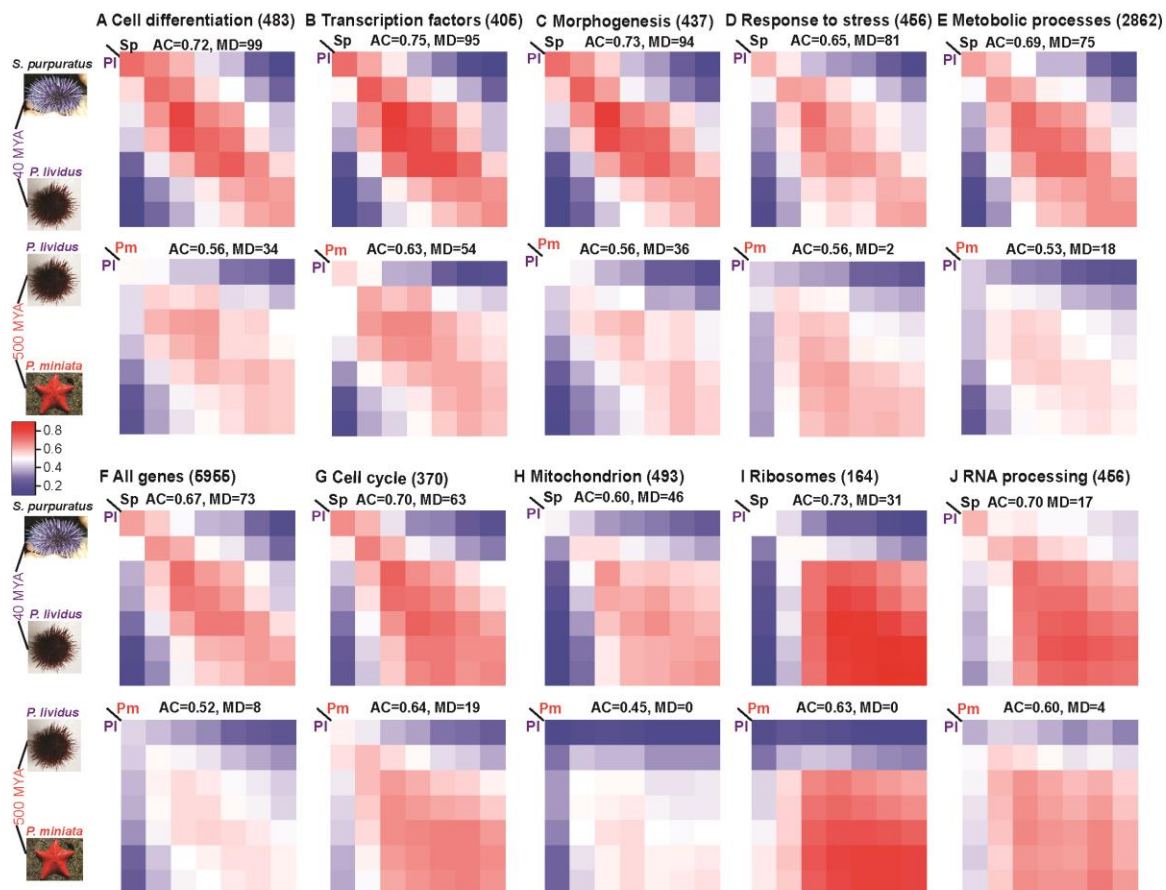
Figure 3 Interspecies Pearson correlations for different GO terms, ordered by the level of matrix diagonality (MD) of *Pl-Sp* matrices. In each panel, from A to J, we present the Pearson correlation of the expression levels of genes with specific GO term between different developmental stages in the three 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 9hpf in *P. miniata* and 4hpf 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. Color scale is similar for all graphs and given at the middle of the figure. F, shows the Pearson interspecies correlation for all 1:1:1 genes.

20
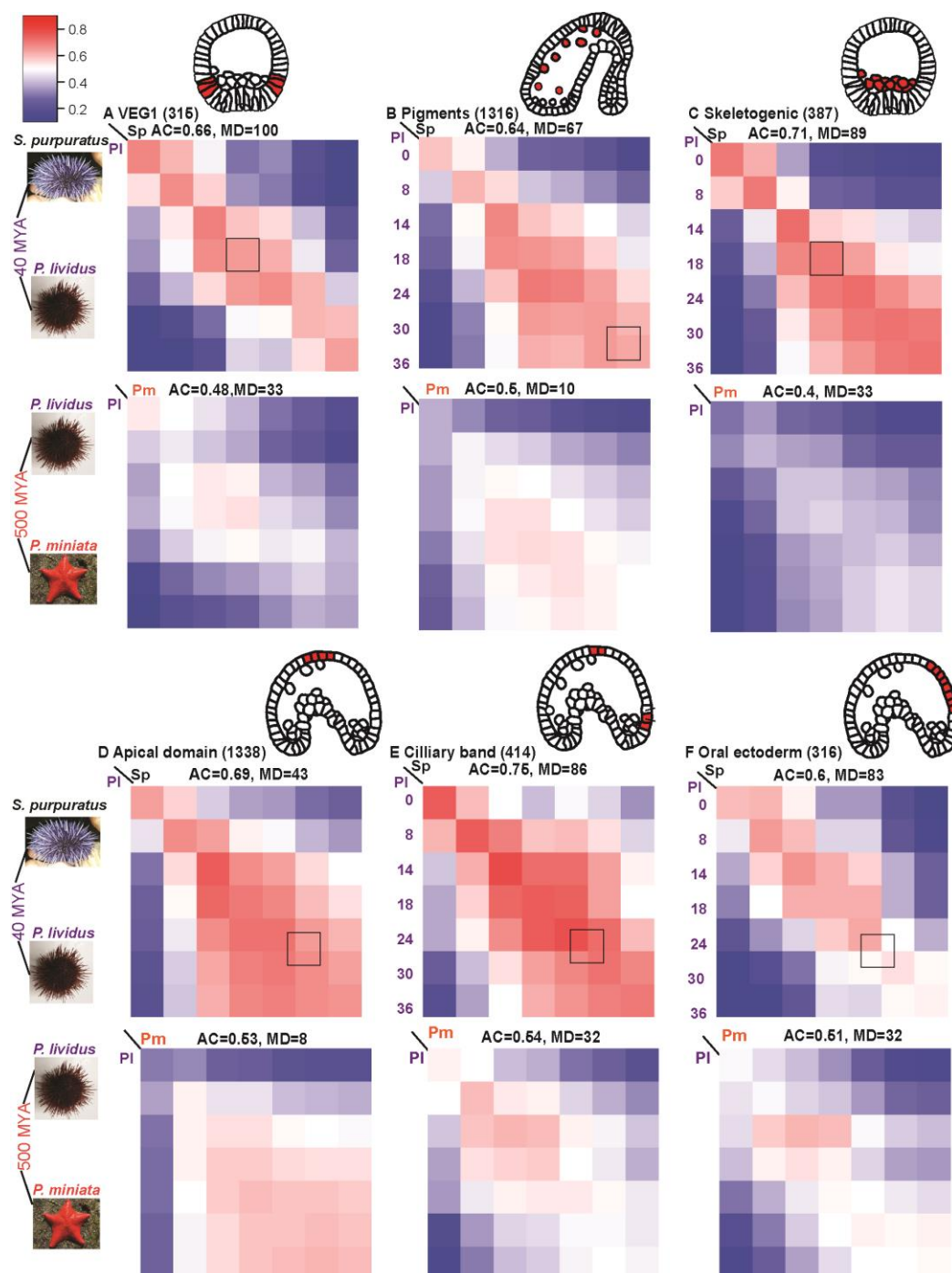
Figure 4 Interspecies Pearson correlations in gene expression for genes enriched in specific sea urchin lineages. A-F, each panel shows the interspecies Pearson correlation between the developmental stages in the three species for genes that their expression enriched at specific time point in a particular cell lineage in the sea urchin, *S. purpuratus* according to [28]. The cell lineages at the time where the enrichment was observed are illustrated by the embryo diagrams above the relevant correlation pattern[28], the time point is also marked in a black square in each *Pl-Sp* panel. In each panel we indicate lineage were these genes are enriched, the number of genes in each set, the average correlation strength in the diagonal (AC) and the matrix diagonality (MD), see text for explanation. Color scale is similar for all graphs and given at the top.
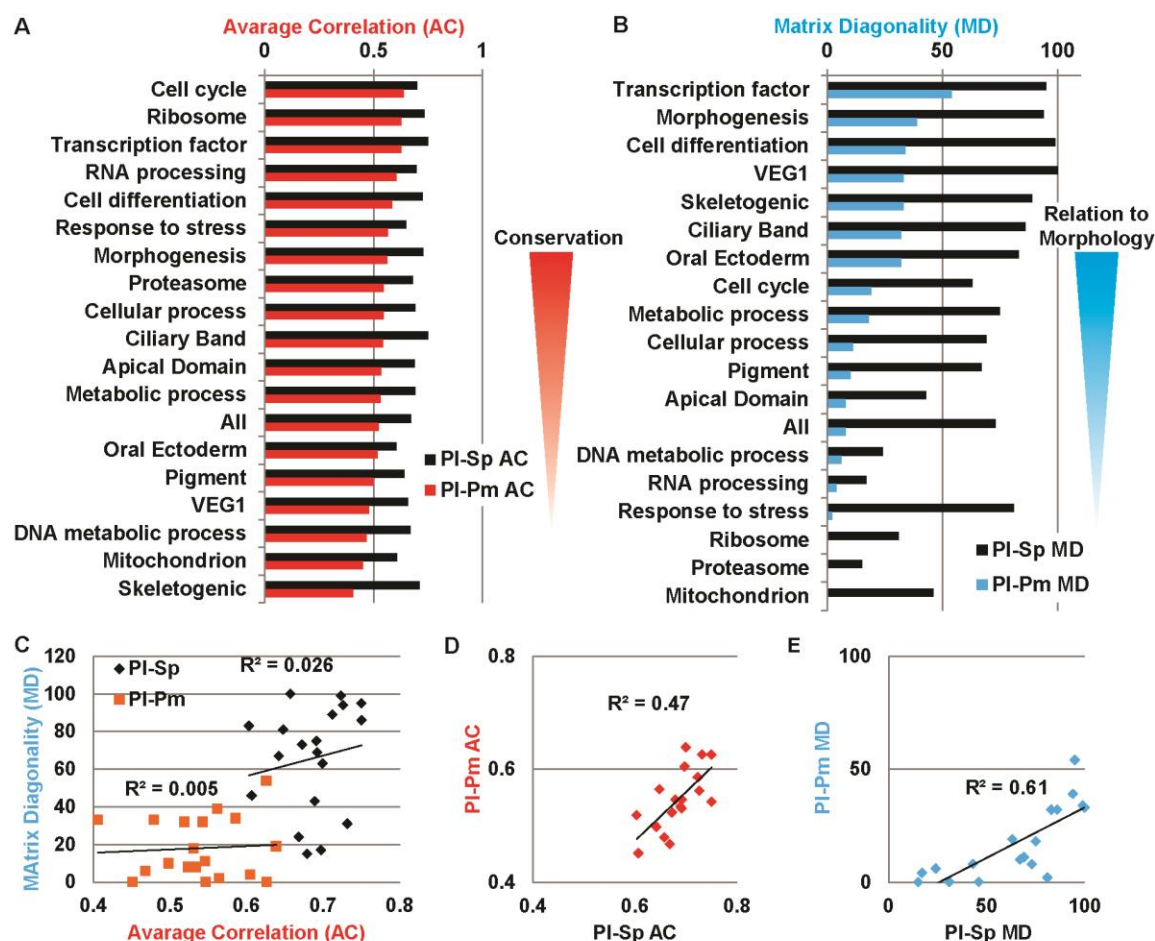
Figure 5 The average correlation strength and the matrix diagonality are independent parameters that reflect different properties of expression conservation. A, the average correlation strength in the diagonal (AC) between *Pl-Sp* (black bars) and *Pl-Pm* (red bars) in receding order of *Pl-Pm* correlation strength. 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. 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* (excluding skeletogenic genes, 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* (excluding skeletogenic genes, R Pearson = 0.78).

22

652
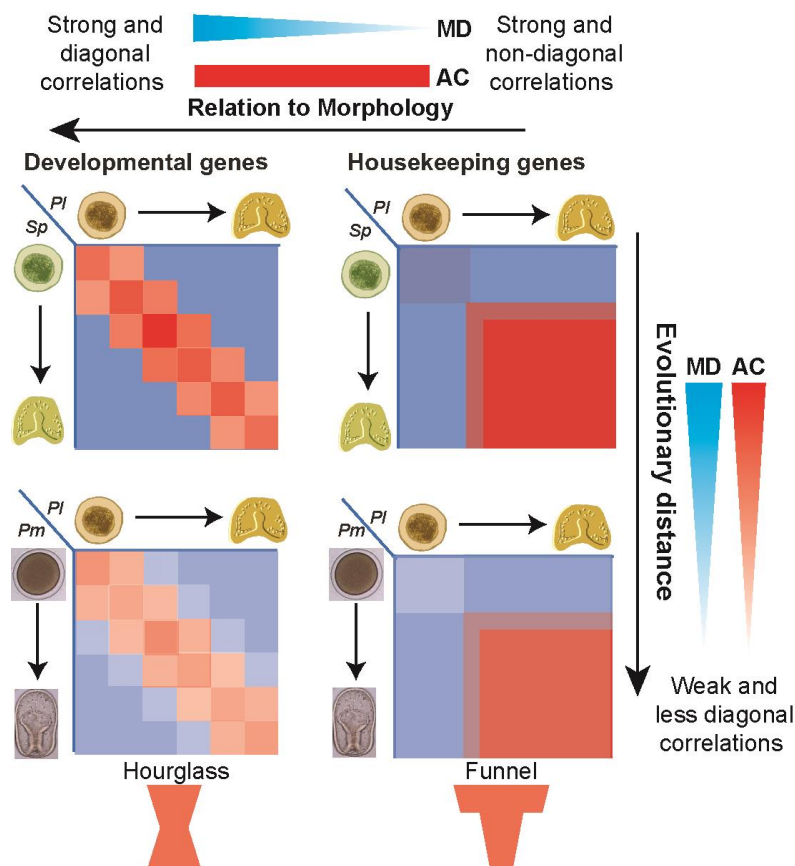
Figure 6 The correlation matrix diagonality, MD, reflects how the dominance between cellular and developmental constraints changes with evolutionary distance for different functional classes. Illustration of typical interspecies correlation matrices of developmental control genes and housekeeping genes between closely related and further diverged (lower 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 that still maintain the hourglass pattern (lower left panel). 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).

663