Gene expression patterns in chordate embryonic development suggest partial applicability of Haeckel's postulates

Song Guo[1,†], Haiyang Hu[2,†], Chuan Xu[2,†], Naoki Irie[3,4], Philipp Khaitovich[1,*]

[1] Skolkovo Institute of Science and Technology, 121205, Moscow, Russia

[2] CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute for Nutrition and Health, CAS, 320 Yue Yang Road, 200031, Shanghai, China.

[3] Department of Biological Sciences, University of Tokyo, Tokyo, Japan.

[4] Universal Biology Institute, The University of Tokyo, Tokyo, Japan.

[†] Equally contributed to this work.

[*] Correspondence should be addressed to: P.Khaitovich@skoltech.ru

Abstract

The relationship between embryonic development and evolution historically investigated based on embryo morphology, could now be reassessed using mRNA expression endophenotype. Here, we investigated the applicability of von Baer's and Haeckel's arguments at mRNA expression level by comparing the developmental changes among nine evolutionarily distinct species: from oyster to mouse. In agreement with models based on von Baer's postulates, up to 36% of mRNA expression indicated nearly linear conservation of species' developmental programs. By contrast, 5-15% of developmental expression profiles, enriched in neural genes, displayed an alignment pattern compatible with the terminal edition paradigm proposed by Haeckel. Thus, the development-evolution relationship based on mRNA expression agrees with early concepts based on embryo morphology and demonstrates that the corresponding patterns coexist in chordate development.

Keywords: evolution, evo-devo, Haeckel, von Baer, development

Introduction

The general relationship between ontogeny and phylogeny, or development and evolution, has long been discussed. Among cornerstones of the evolutionary developmental theory are the arguments formulated in von Baer's laws of embryology and Ernst Haeckel's Biogenetic law. Although von Baer did not accept the concept of

evolution, his idea that the earlier stages of embryogenesis reflect shared traits of a broader taxonomic group(Von Baer, 1828) laid the foundation for the current evolutionary views on developmental programs' conservation. Ernst Haeckel, who advocated evolution, proposed in his Biogenetic law that the embryogenesis replays the species' evolutionary past. Thus, according to this concept, the new developmental stages will be added to the ancestral embryonic program to produce more recently evolved species(Haeckel, 1866). After a long debate, the general concept of early embryogenesis conservation, reflecting a more ancestral stage, continued in a form of a monotonic developmental conservation model known as the funnel model. More recently, a developmental conservation model called the developmental hourglass model was proposed, which placed the most significant stage conservation at a mid-embryonic part of embryogenesis, defined as the phylotypic period(Duboule, 1994). The applicability of the hourglass model to the conservation of developmental stages at mRNA expression level has been supported both in animals(Domazet-Loso & Tautz, 2010; Hu et al., 2017; Irie & Kuratani, 2011; Irie & Sehara-Fujisawa, 2007; Kalinka et al., 2010) and plants(Cheng, Hui, Lee, Wan Law, & Kwan, 2015; Cridge, Dearden, & Brownfield, 2016; Quint et al., 2012). Recent studies, however, demonstrated the applicability of the funnel model or the coexistence and validity of both funnel and hourglass models at different levels of trait evolution in vertebrates(Artieri, Haerty, & Singh, 2009; Bininda-Emonds, Jeffery, & Richardson, 2003; Comte, Roux, & Robinson-Rechavi, 2010; Hu et al., 2017; Levin et al., 2016; Piasecka, Lichocki, Moretti, Bergmann, & Robinson-Rechavi, 2013; Roux &

Robinson-Rechavi, 2008; Uesaka, Kuratani, Takeda, & Irie, 2019). Even the broadly criticized Biogenetic law, has been pointed out to have potential applicability of its principles(Richardson & Keuck, 2002).

One of the major barriers in testing the relationship between development and evolution, including the applicability of concepts proposed in Biogenetic law or the von Baer's law of embryology, is the difficulty in identifying evolutionarily-homologous developmental stages among distant species. While pioneering studies focused on the investigation of embryonic morphology(Jeffery, Bininda-Emonds, Coates, & Richardson, 2002), recent works relied on developmental changes in mRNA expression profiles as an informative endophenotype(Bozinovic, Sit, Hinton, & Oleksiak, 2011; J. J. Li, Huang, Bickel, & Brenner, 2014; Piasecka et al., 2013; Yanai, Peshkin, Jorgensen, & Kirschner, 2011). The use of gene expression facilitates inter-species comparisons, as orthologous genes can be matched among species, and their expression profiles could be traced at all stages of development. Several studies examined developmental gene expression patterns in mammals(Wagner, Tabibiazar, Liao, & Quertermous, 2005), vertebrates(Piasecka et al., 2013), chordates(Levin et al., 2016; Yanai et al., 2011), and fruit fly species(Tomancak et al., 2007). These studies, however, either focused on gene expression patterns within a specific species or compared developmental stage conservation within the concepts of the hourglass model paradigm.

Here, we directly tested the compatibility of developmental expression patterns with predictions of the Biogenetic law and von Baer's law of embryology. To do so, we performed temporal alignment of mRNA developmental expression trajectories among eight chordate species and oyster. Haeckel's Biogenetic law proposing the addition of new parts to the ancestral developmental program (terminal addition) could be described as a "progressive" developmental model concerning the ontogenetic gene expression pattern alignment. By contrast, Baer's law and its later modifications presume the existence of a general developmental program spanning the entire development, while more conserved at early embryonic stages. Such a model could be termed as a "continuous" developmental model concerning the ontogenetic gene expression pattern alignment.

Results

Alignment of developmental gene expression patterns among species

We investigated the relationship between ontogeny and phylogeny of chordate species at the level of molecular phenotype using RNA-sequencing (RNA-seq) data collected over the entire course of embryonic development in eight chordate species of different organizational complexity (amphioxus, ciona, zebrafish, two species of frogs, turtle, chicken, and mouse) and an outgroup species (oyster)(Hu et al., 2017; Zhang et al., 2012). In each species, the data were collected at 11 to 20 developmental stages and

104 measured in duplicates (Figure 1A; Supplementary file 1: Table S1 and Figure
105 1A–figure supplement 1).

106 Within species, differences among developmental stages explained approximately
107 90% of total expression variation (Figure 1B,C and Figure 1C–figure supplement 2).
108 By contrast, other factors, such as biological replicates and sequence coverage,
109 explained 1% and 0.6% of variation each (Figure 1B). Accordingly, on average, 85%
110 of detected genes showed significant expression differences among developmental
111 stages in each species (development-related genes) (polynomial test,
112 Benjamini-Hochberg (BH) corrected $p < 0.05$; Figure 1D and Figure 1D -Source data
113 1).

114 We searched for the best temporal alignment between each pair of species using
115 genes preferentially expressed at a particular developmental stage (stage-associated
116 genes; Methods; Supplementary file 2: Table S2)(J. J. Li et al., 2014). Surprisingly,
117 despite substantial differences in organizational complexity, an approximately linear
118 alignment of developmental stages was a dominant stage-matching pattern in each
119 pair of species (Figure 1E and Figure 1E–figure supplement 3,4).

120 ***Two alternative evolutionary models of species' development***

121 We next sought to test, from the perspective of gene expression, two models, the
122 progressive and continuous one, describing embryogenesis rooted in von Baer's and
123 Haeckel's ideas. Specifically, we based the progressive model (PM) on Haeckel's
124 postulate that embryogenesis replays the evolutionary history of the species. Based on

this view, the developmental gene expression profiles of a species with more ancestral body plan organization will align the best to early embryonic stages of species with more recently evolved organization, "replaying" the evolutionary history of this body plan (Figure 2A; Figure 2A–figure supplement 5). By contrast, evolutionary developmental models rooted in von Baer's ideas, such as the hourglass and the funnel model, presume general preservation of the developmental profiles across species (Figure 2A). Accordingly, we based an alternative continuous model (CM) on this assumption, thus predicting the best alignment between complete, non-truncated, developmental expression profiles of the species.

We found that out of all development-related genes that aligned significantly to the mouse developmental trajectory, 19-36% aligned linearly, thus behaving in agreement with the continuous model predictions (CM genes) (correlation test, $r > 0.7$, $p < 0.05$; Figure 2B,E; Figure 2B–figure supplement 6 and Figure 2E – source data 1). By contrast, 5-15% of development-related genes aligned significantly better to the truncated sets (PM genes) (permutations, BH-corrected $p < 0.05$; Figure 2B,C,E; Figure 2B–figure supplement 6 and Figure 2E – source data 2). Further, in each pairwise alignment, PM genes did not align uniformly to all shortened mouse developmental intervals but tended to peak at a particular fragment. For instance, oyster PM genes aligned best to the mouse developmental fragments ending at stages 3-4, while amphioxus PM genes – to the fragments ending at stages 5-6 (Figure 2B,C and Figure 2C–figure supplement 7). Overall, PM genes in species evolutionarily

proximal to the mouse tended to align to increasingly longer truncated mouse developmental sets (Figure 2D). This phylogenetic ordering of ontogenetic patterns was not caused by the alignment procedure, which was not biased to any developmental stage, and robust to the use of evolutionarily old genes present in each of the nine investigated species (Supplementary file 3: Figure S1). At the same time, this phenomenon matches the phylogeny-ontogeny relationship among multiple species proposed by Haeckel, even though we did not include this aspect in PM model formulation. By contrast, repeating the alignment procedure using species with more ancient body plans, such as amphioxus or ciona, instead of the mouse, did not yield a consistent significant excess of genes following PM predictions (Supplementary file 4: Figure S2).

Properties of CM and PM genes

While we defined CM and PM gene sets independently for each species, each set overlapped significantly between species (Methods, subsampling, Bonferroni-corrected $p < 0.05$; Supplementary file 5: Table S3), indicating the conservation of CM and PM patterning across chordate embryogenesis and presumably convergent functionality specific to each model.

Indeed, CM genes showed enrichment in Gene Ontology (GO) terms corresponding to general cellular functions, such as spliceosome, RNA transport, DNA replication, and several metabolic processes (Hypergeometric test, $p < 0.05$; Figure 2F – source data 1). By contrast, PM genes were primarily enriched in neural

functions, including axon guidance, glutamatergic synapse, dopaminergic synapse, and MAPK signaling pathways (Hypergeometric test, $p < 0.05$; Figure 2F and Figure 2F – source data 1). In line with functional enrichment results, expression of CM genes displayed significantly lower tissue-specificity in mouse(B. Li et al., 2017) compared to PM genes, and all expressed genes (one-sided Fisher's exact test, $p < 0.05$; Supplementary file 6: Figure S3). Nonetheless, PM genes were more conserved at the amino acid sequence level compared to CM genes (Ka/Ks, one-sided Wilcoxon rank-sum test, $p < 0.0005$) (Figure 2G and Figure 2G–figure supplement 8,9).

Visualization of developmental expression patterns

We next investigated developmental expression trajectories of the 244 orthologous genes classified as development-related in all nine species using the nearly linear, species' alignment (Figure 3A). These genes were grouped into six clusters based on their developmental profiles in the unsupervised clustering analysis (Figure 3B). Remarkably, 66% of the genes fell within a single cluster representing a decreasing expression pattern conserved across all nine species (CL9.1) (Figure 3C). By contrast, the expression trajectories in the other clusters differed more among species, with the extent of differences being directly proportional to corresponding phylogenetic distances (Figure 3E and Figure 3E–figure supplement 10). Accordingly, CL9.1 genes substantially overlapped with CM genes (one-sided Fisher's exact test, odds ratio = 3, $p < 0.0001$) and were shown in the same biological processes as CM genes, including spliceosome and RNA processing (Supplementary file 7: Table S4).

188 The same analysis conducted using 2,038 development-related orthologs shared
189 among six vertebrate species revealed a CL9.1-like developmental pattern represented
190 by a single cluster (CL6.1) (Figure 3D). Genes in clusters CL9.1 and CL6.1
191 overlapped significantly and were enriched in similar GO terms (Figure 3F and
192 Supplementary file 8: Table S5).

193 The second-largest cluster found in vertebrates (CL6.2) showed an opposite,
194 ascending expression pattern conserved among species and contained genes enriched
195 in signaling pathways, such as cGMP-PKG signaling pathway, oxytocin signaling
196 pathway, and Renin secretion (Supplementary file 8: Table S5). Notably, this cluster
197 overlapped with the chordate cluster CL9.4, where the ascending pattern was
198 conserved among the seven species, excluding ciona and oyster (Figure 3C).

199

200 **Discussion**

201 Our goal was to explore the general relationship between ontogeny and phylogeny
202 based on the comparison of developmental gene expression trajectories among nine
203 species separated by approximately 800 million years of evolution. Our results
204 indicate that developmental patterns consistent with evolutionary models rooted in
205 von Baer's and Haeckel's ideas match approximately half of all detected mRNA
206 developmental expression profiles. Between the two historical concepts, the
207 continuous developmental model consistent with Baer's arguments was nearly
208 three-fold more prominent compared to the progressive model compatible with the

209 Biogenetic law. Still, a measurable fraction of genes displayed the developmental
210 expression pattern matching Haeckel's predictions, indicating the need for a further
211 critical assessment of this historical concept. These results do not conflict with the
212 hourglass model of development. Our previous work demonstrated the applicability of
213 this model to all nine species' development(Hu et al., 2017). While the hourglass
214 model compares relative conservation of developmental stages, we focused on the
215 alignment of expression differences among stages.

216 Expression profiles conserved along the entire species' development included
217 two main patterns: a gradual decrease conserved among all nine species and gradual
218 increase conserved among six vertebrates. The conserved downward pattern coincides
219 with the reported preferential expression of evolutionary older genes at earlier
220 developmental stages(Gao et al., 2018). Consistently, genes forming this pattern tend
221 to display ubiquitous expression among tissues and are involved in essential functions,
222 such as RNA processing (Supplementary file 9: Figure S4). Further, a study
223 examining embryonic gene expression reported the existence of such a pattern in
224 isolated mouse tissues(Sarropoulos, Marin, Cardoso-Moreira, & Kaessmann, 2019),
225 indicating that the descending expression reflects alterations within embryo tissues
226 rather than changes in embryo composition. Overall, expression patterns conserved
227 between species over the entire development involve 19-36% of orthologous
228 development-related genes detected in our study, constituting the major alignment
229 trend.

The second prominent type of developmental relationship among species, involving 5-15% of orthologous development-related genes, was generally consistent with predictions of Haeckel's Biogenetic law. According to Haeckel's views, the best developmental alignment should match stages reflecting the phylogenetic history of the species, rather than the entire developmental span. When aligning developmental profiles of eight species to the mouse, we indeed observed this phenomenon. First, for this gene group, the developmental expression profiles indeed aligned the best to the truncated mouse developmental series lacking late stages. Second, for each species, these alignments peaked at a particular shortened mouse developmental fragment, with species phylogenetically more distant from the mouse aligning best to the increasingly shorter mouse developmental sets. Intriguingly, while all species evolved their developmental programs after branching from the common ancestor, a mirror procedure, alignment to the fragmented developmental series of amphioxus and ciona representing more ancient body plans, revealed potential PM genes in only two of 16 alignments (Supplementary file 4: Figure S2). Further, there was no relationship between the length of the best-aligning developmental fragment and the phylogenetic distance between species.

Similar to the linearly-aligning genes, genes displaying this "Haeckelian" expression behavior overlapped significantly between species (Supplementary file 5: Table S3), indicating their potential functional conservation. These genes were indeed more conserved at the amino acid sequence level than the linearly-aligning genes and

displayed enrichment in specialized functions, such as synaptic signaling and secretion.

Our work highlights the usefulness of quantitative endophenotypes as an essential complement to morphology-based developmental studies. It further suggests partial applicability of Haeckel's ideas to a fraction of developmental expression differences involving genes associated with neural functions.

Methods

RNA-seq data processing

All data in this analysis were downloaded from Hu et al.(Hu et al., 2017). We adopted the same methods to quantify gene expression. In brief, we got the RNA-seq data of *Crassostrea gigas* from GEO database (accession number SRP014559)(Zhang et al., 2012), and the rest of species from DDBJ (accession number DRA003460) (Supplementary file 1: Table S1). To quantify gene expression, we mapped RNA-seq reads to the corresponding genome (Supplementary file 1: Table S1) using Tophat (v2), allowing up to three mismatches and indels, except for ciona (Ci). In the case of ciona, we mapped reads to the genome with up to five mismatches and indels as the ciona RNA-seq data and the genome data have slightly lower quality compared to the rest.

To define the expressed genes, we required the maximal expression across all development stages exceeding 1 FPKM. For each stage, we calculated expression as the mean of expression of the replicated samples. More than 70% of coding genes annotated for a given species were reliably detected, except for amphioxus (41%) (Supplementary file 10: Table S6).

Identification of development-related genes

We defined developmental alterations of gene expression levels using polynomial regression models following the method described in Somel et al.(Somel et al., 2009). For each gene, we chose the best regression model with the developmental stage (by

rank) as predictor and expression level as a response with Benjamini-Hochberg corrected $p < 0.05$. The genes that fit a significant regression model were termed as development-related genes (Figure 1D -Source data 1).

Identification of stage-associated genes

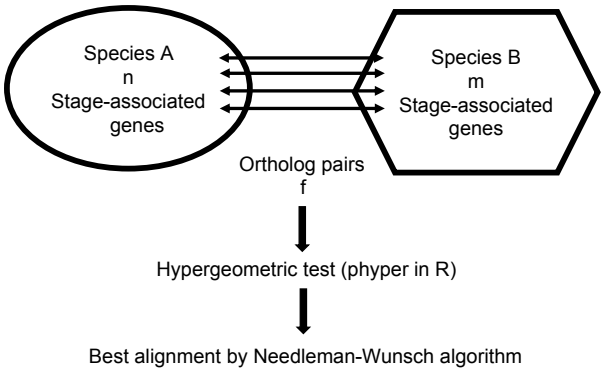
We defined genes preferentially expressed at a particular developmental stage as stage-associated genes in each species following the method described in(Hu et al., 2017; J. J. Li et al., 2014). For each species, we required FPKM (fragments per kilobase of exon model per million reads mapped) > 2 at a particular stage and Z-score (normalized FPKM across samples) representing the difference between this stage and the rest of stages > 1.5. On average, 65%-90% of genes expressed in development were identified as stage-associated genes (Supplementary file 2: Table S2).

Alignment of mouse developmental stage sets to developmental profiles of other species

To assess the transcriptome similarity in the course of developmental stages between mouse (Mm) and other species, we used stage-associated and 1:1 orthologous genes to align the developmental stages between mouse and other species followed by the method reported in(Hu et al., 2017; J. J. Li et al., 2014). The detailed description can be found in the supplementary materials of(Hu et al., 2017; J. J. Li et al., 2014). In brief, we calculated the pairing score using the hypergeometric test to evaluate the ratio of orthologs overlapping within the stage-associated genes pairwise. The pairing score can be written as:

300 $$\text{Paring score} = -\log_{10}(\text{Bonferroni corrected P-values})$$

301 The paring score was used to quantify the significance of the overlap on each pairwise
302 stage comparison between the mouse and another species. From the paring score, we
303 identified the relationship between mouse and other species (Figure 1E). To check the
304 stability of this relationship, we repeated the comparison 500 times with randomly
305 assigned stage-associated genes for each stage of each species using the same
306 procedure.



307

308 According to the above scheme, we further assigned the corresponding stage
309 alignment between mouse and other species using the Needleman-Wunsch algorithm
310 with the gap penalty equal to one. To estimate the stability of alignment based on the
311 mean of the replicates, we randomly chose one individual per stage, aligned two
312 species 500 times, and calculated the frequency for each pair of alignment. The
313 resulting frequency was presented by the thickness of the line in Figure 3A.

314

315 *Identification of CM/PM genes*

316 To identify genes with expression trajectories compatible with the continuous model
317 (CM) or progressive model (PM) predictions, we considered development-related
318 genes as mentioned in "*Identification of development-related genes*" with existing
319 orthologs between mouse and any other species. We then linearly aligned the
320 complete set of developmental stages in a given species (oyster, amphioxus, ciona,
321 zebrafish, frogs, turtle, and chicken) to the increasingly truncated sets of mouse
322 developmental stages, i.e., from first two to all 17 stages (Figure 2A–figure
323 supplement 5). To this end, we interpolated 20 time points uniformly distributed
324 across the whole ontogeny of non-mouse species using cubic smoothing spline (with
325 three degrees of freedom for amphioxus data and six degrees of freedom for other
326 species) or polynomial regression curves (up to the fourth degree). The two methods
327 for interpolating expression values across development were used to ensure the
328 robustness of our CM or PM predictions. Of note, during cubic smoothing spline, the
329 degrees of freedom chosen for amphioxus were fewer than other species, due to the
330 relatively fewer sampled developmental stages of this species; whereas during
331 polynomial regression, we chose up to four degrees to take advantage of the
332 near-fitness of polynomial regression in most time series data.

333 We then aligned the resulting expression trajectories to the gradually increasing
334 sets of mouse developmental stages, where the expression trajectory was interpolated
335 using the same approach (cubic smoothing spline with six degrees of freedom or
336 polynomial regression with up to the fourth degree). The alignments with the Pearson
337 correlation coefficient (PCC) greater than 0.7 and correlation test *p*-value less than

0.05 were considered as valid. We further selected the alignment with the maximum PCC among all valid alignments as the ultimate one. Genes failing to pass the two alignment criteria were considered as non-aligned, regardless of the complete or partial linear alignment paradigms.

Based on the above classification procedure, we summarized for each species, the number of genes aligning to each set of developmental stages of the mouse. Furthermore, we repeated the classification process by shuffling the orthologous relationships between non-mouse species and the mouse (Figure 2B–figure supplement 6). Specifically, we expected that by shuffling, each developmental stage of the mouse would be equally represented in other species, allowing us to calculate the random expectation of alignment occurrences for each stage. Thus, the final number of genes aligning to each set of developmental stages of the mouse was obtained after subtracting these background contaminations. Genes adhering to the continuous model (CM) were defined as those best aligned between mouse and other species over the complete developmental intervals. By contrast, genes following the progressive model (PM) were defined by the following criteria: i) aligned best to a truncated set of mouse developmental stages; ii) displayed a significantly greater number of aligning genes compared to the background distributions, determined using a BH-corrected *p*-value of less than 0.05 as a cutoff; and iii) an increase in the number of genes aligning to particular sets of mouse stages compared to its adjacent sets (Figure 2A, Figure 2E – source data 1 and Figure 2E – source data 2).

To further check the robustness of the alignment results, we restricted our analysis to evolutionarily old genes. To do so, we repeated the alignment procedure using a subset of the development-related genes with inferred Earliest Ortholog Level (EOL) < level 10. This conservation level corresponds to genes that appeared and became fixed within the genomes before the separation of Chordata(Litman & Stein, 2019).

Overlap of CM/PM genes between species

To calculate the significance of the overlap of CM or PM genes between species, we sampled the same number of CM or PM genes in each species from all development-related genes 1,000 times and recalculated the number of overlapping genes to obtain the empirical distribution. The overlap significance *p*-value was calculated as the proportion of the values, which were as larger or greater than the actual overlapping gene number. Given all pairs of species involved, a Bonferroni-corrected *p*-value of less than 0.05 was used as the cutoff of the overlap significance (Supplementary file 5: Table S3).

Amino acid conservation of CM/PM gene

We compared the evolutionary conservation between proteins encoded by CM and PM genes using the Ka/Ks metric(Yang & Nielsen, 2000). Since in this study, we are not studying Ka/Ks variations across the evolutionary branches – we focused on the comparisons of PM versus CM genes within each individual species, we conducted this analysis mainly in a pairwise manner between a given species and mouse. Firstly,

we extracted the coding sequences of each gene based on the corresponding annotation information and then translated the sequences into amino acid sequences using the function "translate" implemented in the Bioconductor package "Biotrings". The longest protein sequence was selected as the gene protein sequence. Next, we performed protein sequence alignments between non-mouse species and the mouse using the function "pairwiseAlignment" in "Biostrings", with the scoring matrix set as BLOSUM50. The resulting protein alignment was translated to the nucleotide level and used as the input for PAML(Yang, 1997) for a Ka/Ks analysis output. This procedure was conducted in a pairwise manner for the mouse and a given non-mouse species to match the corresponding definitions of PM and CM genes carried out in a pairwise manner (Figure 2G–figure supplement 8). To test the robustness of our Ka/Ks calculations in the context of multiple species, we performed multiple sequence alignment using Clustal Omega(Sievers et al., 2011) for all protein sequences mentioned above convert to nucleotide levels. We next utilized the CODEML application in PAML to calculate the Ka/Ks for given cross-species ortholog. The significance of the difference between PM and CM genes in each species was assessed using one-sided Wilcoxon rank-sum test (Figure 2G–figure supplement 9).

Gene expression interpretation to mouse developmental stage

To compare the similarity of the expression profiles across developmental stages of nine species, we used the predicted developmental stage alignment presented in Figure

3A to create a unified alignment of eight species to the mouse developmental curves. To do so, we mapped 33 stages cumulatively interpolated from eight species to the full mouse developmental curve fitted using cubic smoothing spline with ten degrees of freedom. We then compared gene expression curves among nine species based on z-transformed expression of each gene interpolated at these 33 stage points.

Clustering of gene expression profiles in six vertebrate and nine chordate species

To investigate the expression pattern diversity in nine or six species, we used hierarchical clustering (hclust function in R) of z-transformed gene expression trajectories aligned among species with (1 - rho) as the distance measure, where rho is the Spearman correlation coefficient. We chose k equal six, as optimal, based on visual inspection of clusters obtained using different k values.

Functional annotation of developmental expression patterns

To check the functions of genes in each cluster, we applied Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)(Kanehisa & Goto, 2000) enrichment tests. GO annotation of mouse genes was downloaded from Bioconductor package 'org.Mm.eg.db'. Mouse pathway annotation was downloaded from http://rest.kegg.jp/list/pathway/ and http://rest.kegg.jp/link/genes/.

For GO enrichment test, the "elim" algorithm of topGO(Alexa & Rahnenfuhrer, 2016) was chosen to eliminate the hierarchical dependency of the GO terms. Fisher's exact test was applied for each GO term. The background set consisted of all development-related genes orthologous among six vertebrate species (2,038 genes) and

422 nine chordate species (244 genes), respectively. For the test of vertebrate gene clusters,
423 Benjamini-Hochberg correction on "molecular function", "biological process", and
424 "cellular component" was applied independently. GO terms with BH corrected $p <$
425 0.05 were reported. We applied no multiple test correction to chordate' gene cluster
426 enrichment analysis due to low statistical power of the dataset and used a more
427 stringent nominal *p*-value cutoff of $p < 0.001$.

428 For the pathway enrichment test, the reference genes were the same as for
429 GO-based analysis. We used hypergeometric test (phyper in R) to assess the
430 enrichment in each KEGG pathway. Bonferroni correction was applied for genes in
431 vertebrate clusters. Pathways with corrected $p < 0.05$ were reported as significantly
432 enriched. No correction was applied to genes from chordate clusters due to low
433 statistical power, and the nominal pathway enrichment *p*-value was set to $p < 0.05$. This
434 relaxed cutoff was used to assess the potential overlap between enriched functions
435 found in vertebrate and chordate clusters (Supplementary file 8: Table S5).

436 *Species tree construction*

437 The species tree was generated under the web https://phylot.biobyte.de/ with NCBI
438 taxonomy IDs (10090, 9031, 13735, 8364, 7955, 7719, 8355, 7739, 29159).
439 The species separation time to the mouse was obtained from
440 http://www.timetree.org/(Kumar, Stecher, Suleski, & Hedges, 2017).

441

442

443 *Statistical analysis and software*

444 Analyses were conducted in the R environment (http://www.r-project.org/). To
445 minimize the type I error rate, we used multiple test correction for *p*-value calculations,
446 unless specifically indicated otherwise. The statistically significant level used for each
447 was specified in the main text. Additionally, we used Perl, Python, and R packages,
448 including 'topGO', 'reshape', 'RColorBrewer', 'ggplot2',' Biotrings', as well as shell
449 scripts, in the analyses. Pathway visualization was conducted under
450 https://pathview.uncc.edu/ for the spliceosome pathway.

451

452 **Acknowledgments**

453 We thank Michael Lachmann for inspiring this study, Zhisong He, Qian Li and
454 Qianhui Yu for helpful discussions, and Pavel Mazin for critical comments.

455

456 **Funding**

457 This work was supported by the National Natural Science Foundation of China
458 (Grants 91331203 and 31420103920 to P.K.); the National One Thousand Foreign
459 Experts Plan (Grant WQ20123100078 to P.K.); and Grant-in-Aid for Scientific
460 Research on Innovative Areas (17H06387 to N.I.).

461

462

Author contribution

S.G, HY.H and C.X designed and executed the bioinformatics analysis. S.G and PK wrote the manuscript. N.I and P.K revised the manuscript. PK supervised the project.

Conflict of interest

The authors declare no conflict of interest.

References

Alexa, A., & Rahnenfuhrer, J. (2016). topGO: enrichment analysis for Gene Ontology. R Packag. version 2.26.0. *R Package version 2.26.0*.

Artieri, C. G., Haerty, W., & Singh, R. S. (2009). Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of Drosophila. *BMC Biol, 7*, 42. doi:10.1186/1741-7007-7-42

Bininda-Emonds, O. R., Jeffery, J. E., & Richardson, M. K. (2003). Inverting the hourglass: quantitative evidence against the phylotypic stage in vertebrate development. *Proc Biol Sci, 270*(1513), 341-346. doi:10.1098/rspb.2002.2242

Bozinovic, G., Sit, T. L., Hinton, D. E., & Oleksiak, M. F. (2011). Gene expression throughout a vertebrate's embryogenesis. *BMC Genomics, 12*, 132. doi:10.1186/1471-2164-12-132

Cheng, X., Hui, J. H., Lee, Y. Y., Wan Law, P. T., & Kwan, H. S. (2015). A "developmental hourglass" in fungi. *Mol Biol Evol, 32*(6), 1556-1566. doi:10.1093/molbev/msv047

Comte, A., Roux, J., & Robinson-Rechavi, M. (2010). Molecular signaling in zebrafish development and the vertebrate phylotypic period. *Evol Dev, 12*(2), 144-156. doi:10.1111/j.1525-142X.2010.00400.x

Cridge, A. G., Dearden, P. K., & Brownfield, L. R. (2016). Convergent occurrence of the developmental hourglass in plant and animal embryogenesis? *Ann Bot, 117*(5), 833-843. doi:10.1093/aob/mcw024

491 Domazet-Loso, T., & Tautz, D. (2010). A phylogenetically based transcriptome age
492 index mirrors ontogenetic divergence patterns. *Nature, 468*(7325), 815-818.
493 doi:10.1038/nature09632

494 Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for
495 the stability of a vertebrate Bauplan and the evolution of morphologies through
496 heterochrony. *Dev Suppl*, 135-142.

497 Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., . . . Liu, J. (2018). Chromatin
498 Accessibility Landscape in Human Early Embryos and Its Association with
499 Evolution. *Cell, 173*(1), 248-259 e215. doi:10.1016/j.cell.2018.02.028

500 Haeckel, E. (1866). Generelle Morphologie der Organismen. *Berlin: Georg Reimer*.

501 Hu, H., Uesaka, M., Guo, S., Shimai, K., Lu, T. M., Li, F., . . . Consortium, E. (2017).
502 Constrained vertebrate evolution by pleiotropic genes. *Nat Ecol Evol, 1*(11),
503 1722-1730. doi:10.1038/s41559-017-0318-0

504 Irie, N., & Kuratani, S. (2011). Comparative transcriptome analysis reveals vertebrate
505 phylotypic period during organogenesis. *Nat Commun, 2*, 248.
506 doi:10.1038/ncomms1248

507 Irie, N., & Sehara-Fujisawa, A. (2007). The vertebrate phylotypic stage and an early
508 bilaterian-related stage in mouse embryogenesis defined by genomic information.
509 *BMC Biol, 5*, 1. doi:10.1186/1741-7007-5-1

510 Jeffery, J. E., Bininda-Emonds, O. R., Coates, M. I., & Richardson, M. K. (2002).
511 Analyzing evolutionary patterns in amniote embryonic development. *Evol Dev,*
512 *4*(4), 292-302. doi:10.1046/j.1525-142x.2002.02018.x

513 Kalinka, A. T., Varga, K. M., Gerrard, D. T., Preibisch, S., Corcoran, D. L., Jarrells,
514 J., . . . Tomancak, P. (2010). Gene expression divergence recapitulates the
515 developmental hourglass model. *Nature, 468*(7325), 811-814.
516 doi:10.1038/nature09634
517 Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes.
518 *Nucleic Acids Res, 28*(1), 27-30. doi:10.1093/nar/28.1.27
519 Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource
520 for Timelines, Timetrees, and Divergence Times. *Molecular Biology and*
521 *Evolution, 34*(7), 1812-1819. doi:10.1093/molbev/msx116
522 Levin, M., Anavy, L., Cole, A. G., Winter, E., Mostov, N., Khair, S., . . . Yanai, I.
523 (2016). The mid-developmental transition and the evolution of animal body plans.
524 *Nature, 531*(7596), 637-641. doi:10.1038/nature16994
525 Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., . . . Shi, L. (2017). A
526 Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq.
527 *Sci Rep, 7*(1), 4200. doi:10.1038/s41598-017-04520-z
528 Li, J. J., Huang, H., Bickel, P. J., & Brenner, S. E. (2014). Comparison of D.
529 melanogaster and C. elegans developmental stages, tissues, and cells by
530 modENCODE RNA-seq data. *Genome Res, 24*(7), 1086-1101.
531 doi:10.1101/gr.170100.113
532 Litman, T., & Stein, W. D. (2019). Obtaining estimates for the ages of all the
533 protein-coding genes and most of the ontology-identified noncoding genes of the

534 human genome, assigned to 19 phylostrata. *Seminars in Oncology, 46*(1), 3-9.
535 doi:10.1053/j.seminoncol.2018.11.002

536 Piasecka, B., Lichocki, P., Moretti, S., Bergmann, S., & Robinson-Rechavi, M. (2013).
537 The hourglass and the early conservation models--co-existing patterns of
538 developmental constraints in vertebrates. *PLoS Genet, 9*(4), e1003476.
539 doi:10.1371/journal.pgen.1003476

540 Quint, M., Drost, H. G., Gabel, A., Ullrich, K. K., Bonn, M., & Grosse, I. (2012). A
541 transcriptomic hourglass in plant embryogenesis. *Nature, 490*(7418), 98-101.
542 doi:10.1038/nature11394

543 Richardson, M. K., & Keuck, G. (2002). Haeckel's ABC of evolution and
544 development. *Biol Rev Camb Philos Soc, 77*(4), 495-528.
545 doi:10.1017/s1464793102005948

546 Roux, J., & Robinson-Rechavi, M. (2008). Developmental constraints on vertebrate
547 genome evolution. *PLoS Genet, 4*(12), e1000311.
548 doi:10.1371/journal.pgen.1000311

549 Sarropoulos, I., Marin, R., Cardoso-Moreira, M., & Kaessmann, H. (2019).
550 Developmental dynamics of lncRNAs across mammalian organs and species.
551 *Nature, 571*(7766), 510-514. doi:10.1038/s41586-019-1341-x

552 Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D.
553 G. (2011). Fast, scalable generation of high-quality protein multiple sequence
554 alignments using Clustal Omega. *Molecular Systems Biology, 7*, 539.
555 doi:10.1038/msb.2011.75

556 Somel, M., Franz, H., Yan, Z., Lorenc, A., Guo, S., Giger, T., . . . Khaitovich, P.
557 (2009). Transcriptional neoteny in the human brain. *Proc Natl Acad Sci U S A,*
558 *106*(14), 5743-5748. doi:10.1073/pnas.0900544106
559 Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein,
560 V., . . . Rubin, G. M. (2007). Global analysis of patterns of gene expression
561 during Drosophila embryogenesis. *Genome Biol,* *8*(7), R145.
562 doi:10.1186/gb-2007-8-7-r145
563 Uesaka, M., Kuratani, S., Takeda, H., & Irie, N. (2019). Recapitulation-like
564 developmental transitions of chromatin accessibility in vertebrates. *Zoological*
565 *Lett, 5*, 33. doi:10.1186/s40851-019-0148-9
566 Von Baer, K. E. (1828). Uber Entwicklungsgeschichte Der Thiere.
567 Wagner, R. A., Tabibiazar, R., Liao, A., & Quertermous, T. (2005). Genome-wide
568 expression dynamics during mouse embryonic development reveal similarities to
569 Drosophila development. *Dev Biol,* *288*(2), 595-611.
570 doi:10.1016/j.ydbio.2005.09.036
571 Yanai, I., Peshkin, L., Jorgensen, P., & Kirschner, M. W. (2011). Mapping gene
572 expression in two Xenopus species: evolutionary constraints and developmental
573 flexibility. *Dev Cell, 20*(4), 483-496. doi:10.1016/j.devcel.2011.03.015
574 Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum
575 likelihood. *Comput Appl Biosci,* *13*(5), 555-556.
576 doi:10.1093/bioinformatics/13.5.555

577 Yang, Z., & Nielsen, R. (2000). Estimating synonymous and nonsynonymous
578 substitution rates under realistic evolutionary models. *Mol Biol Evol, 17*(1), 32-43.
579 doi:10.1093/oxfordjournals.molbev.a026236
580 Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., . . . Wang, J. (2012). The oyster
581 genome reveals stress adaptation and complexity of shell formation. *Nature,*
582 *490*(7418), 49-54. doi:10.1038/nature11413
583
584

Figures and figure legends

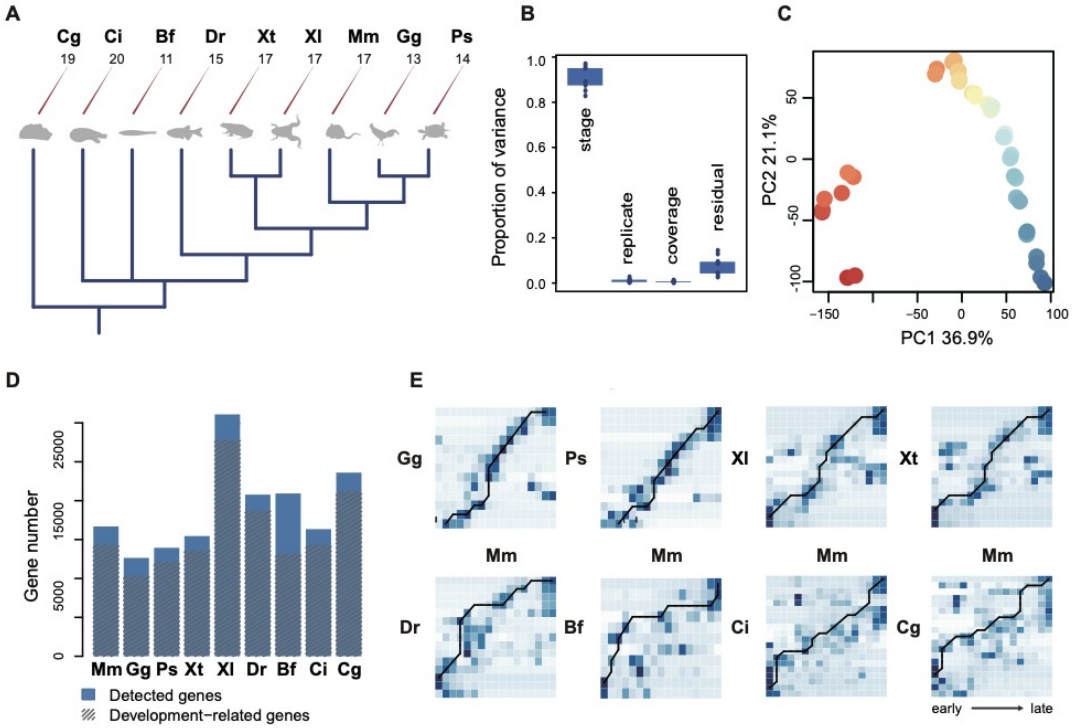
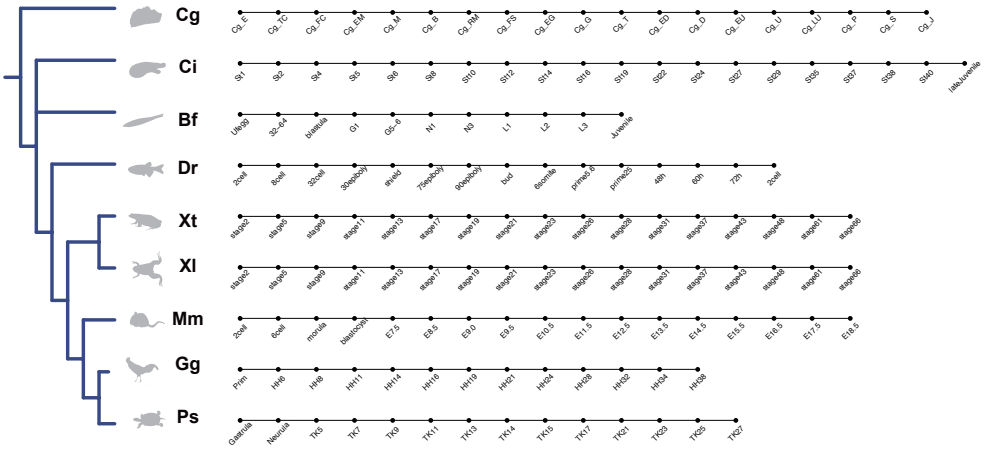


Figure 1. Species phylogeny and alignment of embryonic development stages.

A Experimental scheme showing the phylogenetic relationship among species (dark blue dendrogram) and numbers of sampled developmental stages. The abbreviations here and throughout the figures indicate species: Cg – oyster (*Crassostrea gigas*), Ci – ciona (*Ciona intestinalis*), Bf – amphioxus (*Branchiostoma floridae*), Dr – zebrafish (*Danio rerio*), Xt – Western clawed frog (*Xenopus tropicalis*), Xl – African clawed frog (*Xenopus laevis*), Ps – turtle (*Pelodiscus sinensis*), Gg – chicken (*Gallus gallus*), Mm – mouse (*Mus musculus*). **B** Variance explained by different factors based on ANOVA. Boxes represent the interquartile values of the variance proportion explained by the factors listed on the x-axis among nine species. Dots represent the mean explained variance proportion for each species. **C** PCA plot based on the expression of 16,685 genes in mouse development. Dots represent samples, and color represents developmental stages (red – early, blue – late). **D** The number of detected and development-related genes in each species. **E** Heatmap showing the pairing scores reflecting overlap of stage-associate genes (darker shade of blue representing greater overlap) and the optimal alignment path calculated using the Needleman-Wunsch algorithm (black line).



604

605 **Figure 1A–figure supplement 1. Illustration of sampled developmental stages for**
606 **each species correspondent to Figure 1A.**

607 The dendrogram shows the phylogenetic relationship among species, indicated by
608 silhouette figures and two-letter abbreviated species' classification names. Dots
609 indicate sampled embryonic stages. Detailed sampled stage information is listed in
610 Supplementary file 1: Table S1.

611

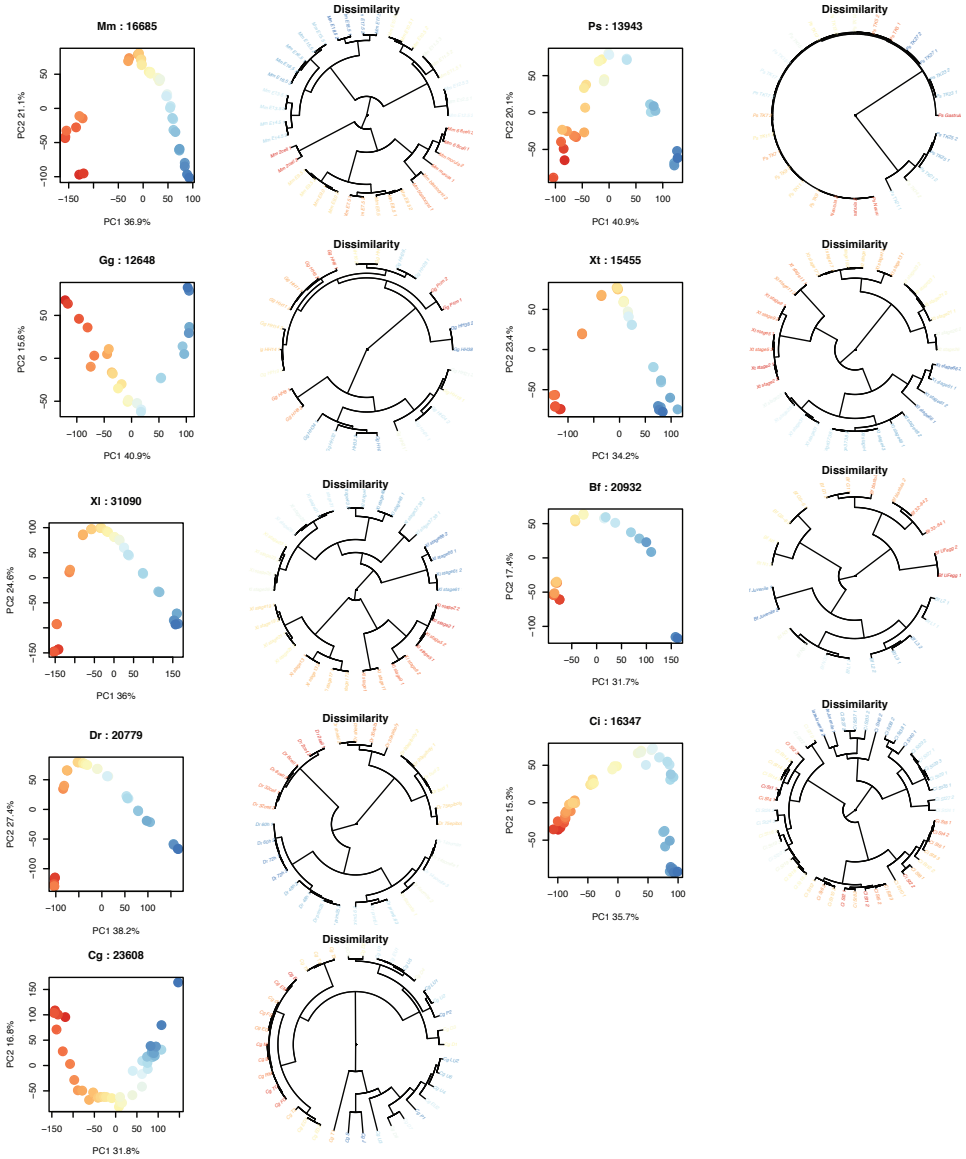
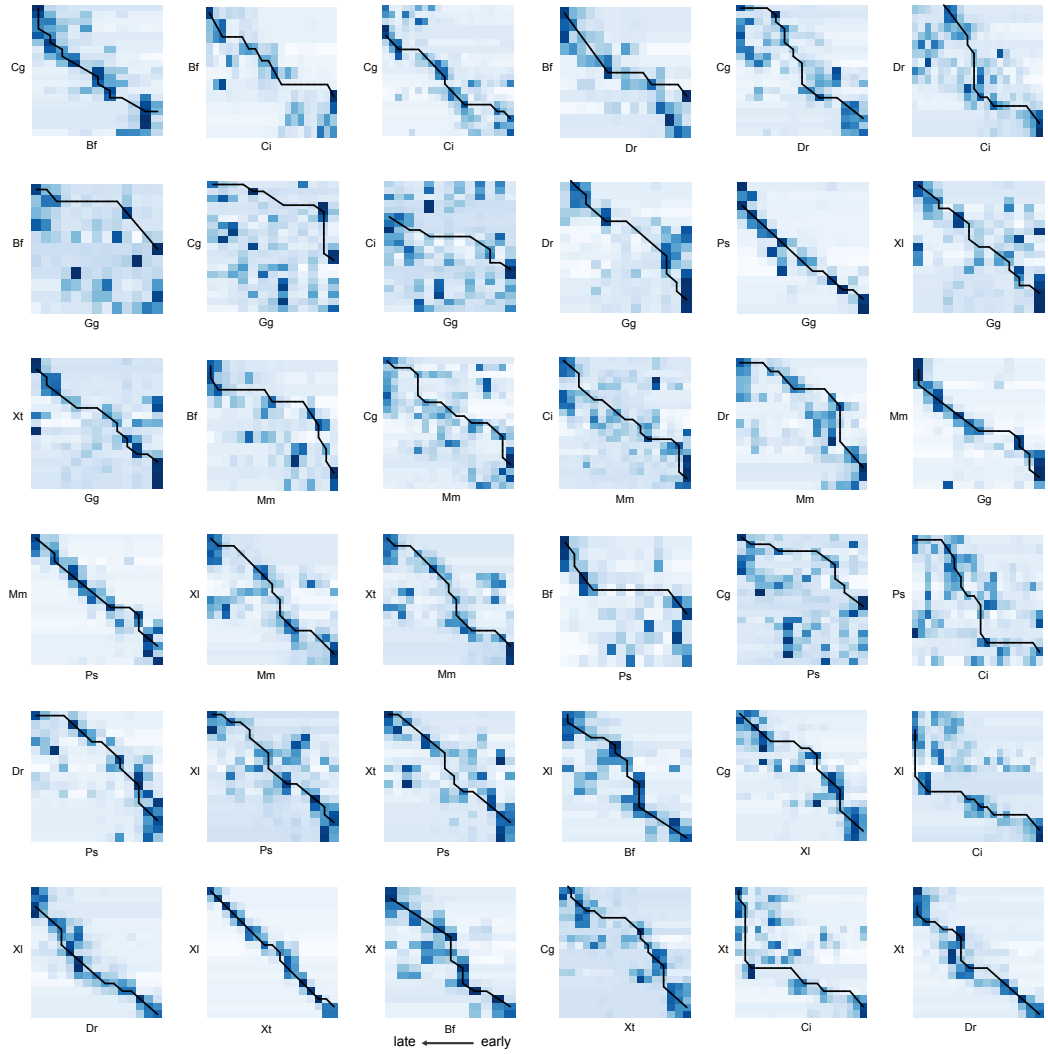


Figure 1C–figure supplement 2. PCA plot and dendrogram based on expression variation within each species.

Dots represent samples, color represents developmental stages (red – early, blue – late). Dendrograms were constructed using 1- Spearman correlation coefficient dissimilarity matrix. Panel titles show abbreviated species' identifiers and the number of expressed genes with FPKM > 1 in at least one developmental stage.

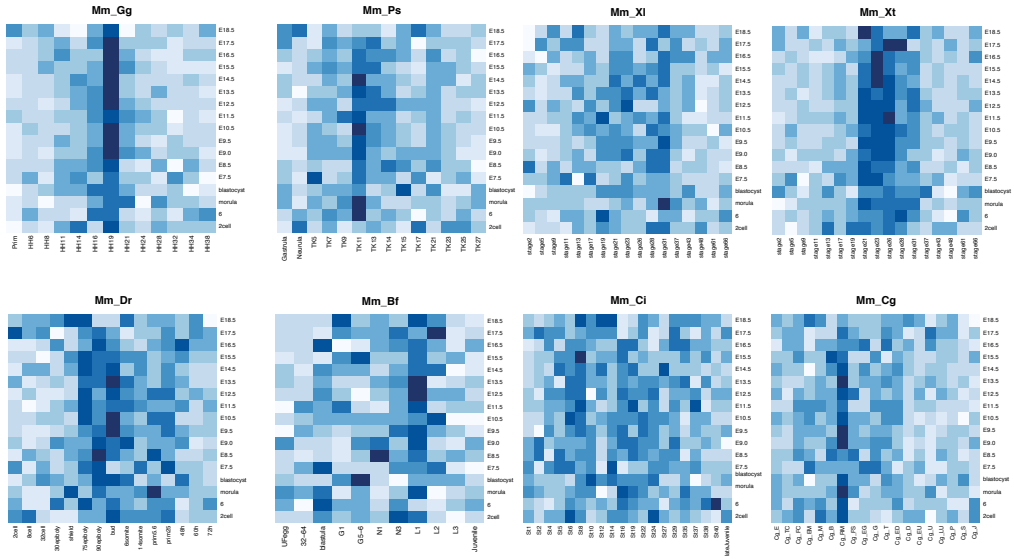


621

622 **Figure 1E–figure supplement 3. Heatmap of the pairing scores reflecting the**
623 **overlap of stage-associate genes correspondent to Figure 1E.**

624 Darker shade of blue represents greater overlap of stage-associate genes. Black line
625 represents the optimal alignment path calculated using the Needleman-Wunsch
626 algorithm.

627

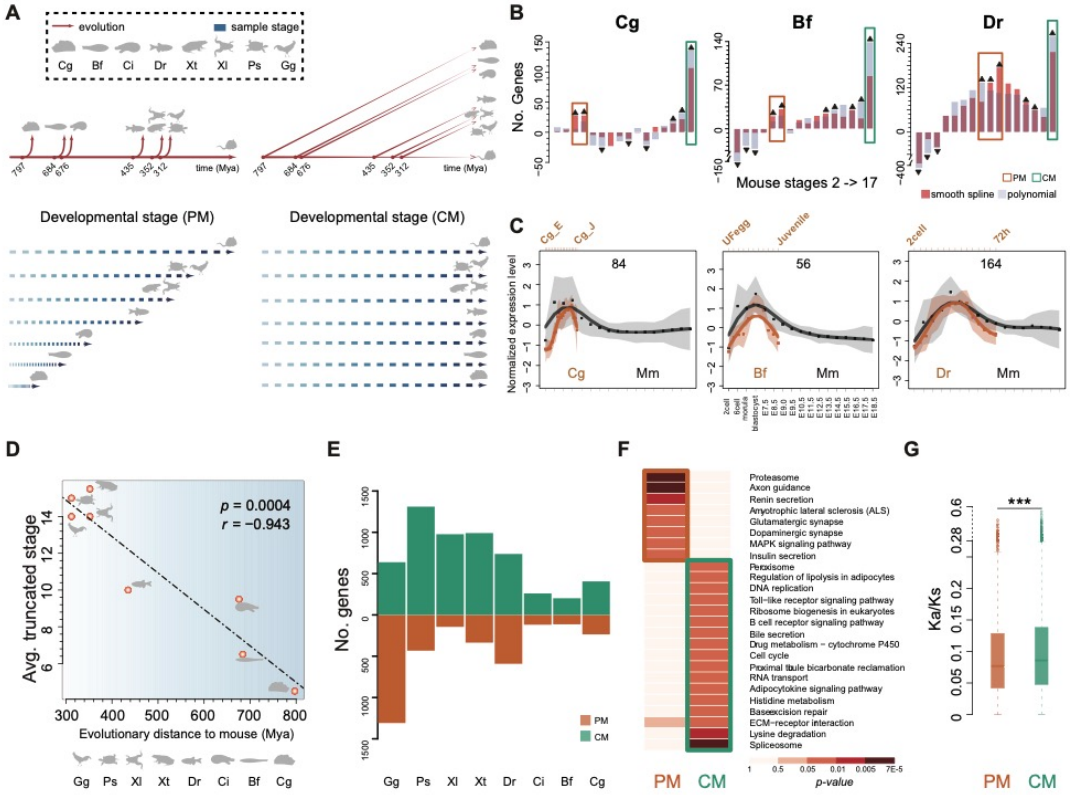


628

629 **Figure 1E–figure supplement 4. Heatmap of the average pairing scores calculated**
630 **randomly assigned stage-associated genes correspondent to Figure 1E**.

631 The heatmap shows pairing scores based on the average of 500 iterations of random
632 assignments of stage-specific gene labels. Panel titles show abbreviated species'
633 identifiers. Darker shades of blue represent greater overlap between stages.

634



635

636 **Figure 2. Relationship of developmental gene expression patterns among species.**
637 **A** Schematic description of the progressive and continuous models and their predicted
638 alignment patterns. The progressive model (PM, left) involves an extension of
639 developmental program in more recently evolved species, thus predicting the best
640 alignment to a shortened mouse developmental course. The continuous model (CM,
641 right) predicts the best alignment to the complete developmental course. **B** Numbers
642 of genes showing the best expression trajectory's alignment between the complete
643 developmental course of the oyster (left), amphioxus (middle), and zebrafish (right)
644 and mouse developmental sets of different lengths: from two to 17 stages. Colors
645 indicate two methods used for developmental expression trajectory calculation: cubic
646 smooth spline (orange) and polynomial regression (gray). Colored rectangles mark
647 stages containing alignments fitting CM (green) or PM (orange) predictions. Black
648 triangles indicate the significantly greater (up) or lesser (down) number of genes than
649 that expected by chance aligning to an indicated mouse developmental interval
650 (permutation test, BH-corrected $p < 0.05$). **C** Examples of PM gene expression
651 patterns. Dots show the cluster-level standardized expression levels at each

652 developmental stage in mouse (black) and the other species (orange). The curves
653 represent average expression profiles, and the shaded regions represent the standard
654 deviation of curve estimates. **D** The relationship between the lengths of truncated sets
655 of mouse stages containing maximal numbers of PM genes for each of the eight
656 species and the phylogenetic distances to the mouse. Dots represent species. The
657 dotted line marks the regression curve. The Pearson correlation coefficient and linear
658 regression p-value are shown at the top right. Mya – millions of years ago. **E**
659 Numbers of CM and PM genes identified in each of the eight species. Note that PM
660 gene number in chicken (Gg) might be inflated due to poor resolution of PM and CM
661 predictions at close phylogenetic distances. **F** KEGG pathways significantly enriched
662 in PM or CM from all eight non-mouse species. The heat map shows the p-values of
663 the enrichment tests. Colored rectangles mark significantly enriched KEGG pathways.
664 **G** Amino acid sequence conservation (Ka/Ks) of PM and CM genes from all eight
665 non-mouse species. Asterisks indicate the significance of the difference (Wilcoxon
666 rank-sum one-sided test, $p < 0.0005$).

667

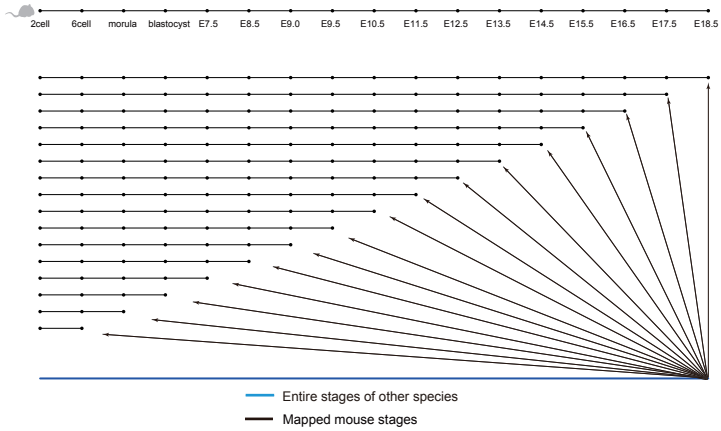
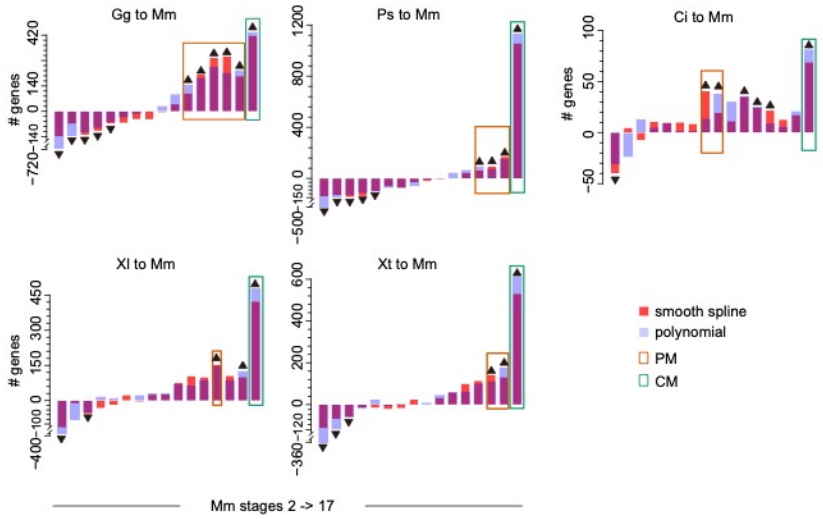


Figure 2A–figure supplement 5. Schematic representation of the alignment procedure used to test the predictions of progressive and continuous models correspondent to Figure 2A.

The scheme depicts an example of all possible alignments between the complete developmental series of a species (any of investigated species except mouse; blue line) and mouse developmental sets (black lines). The mouse developmental sets included the entire developmental series consisting of 17 sampled stages and developmental series progressively shortened by truncation of the last stage, down to the first two developmental stages. For each gene showing development-dependent expression in the two compared species, we then aligned the developmental profile of a species to the entire mouse developmental course and the increasingly truncated sets of mouse developmental stages. The continuous model (CM) predicts the best alignment of the amphioxus development to the complete mouse developmental course. The progressive model (PM) predicts the best alignment to the truncated developmental course. We used all possible truncated mouse developmental series, from 16 to two stages, to identify the best alignment to the mouse developmental series in an unbiased manner.

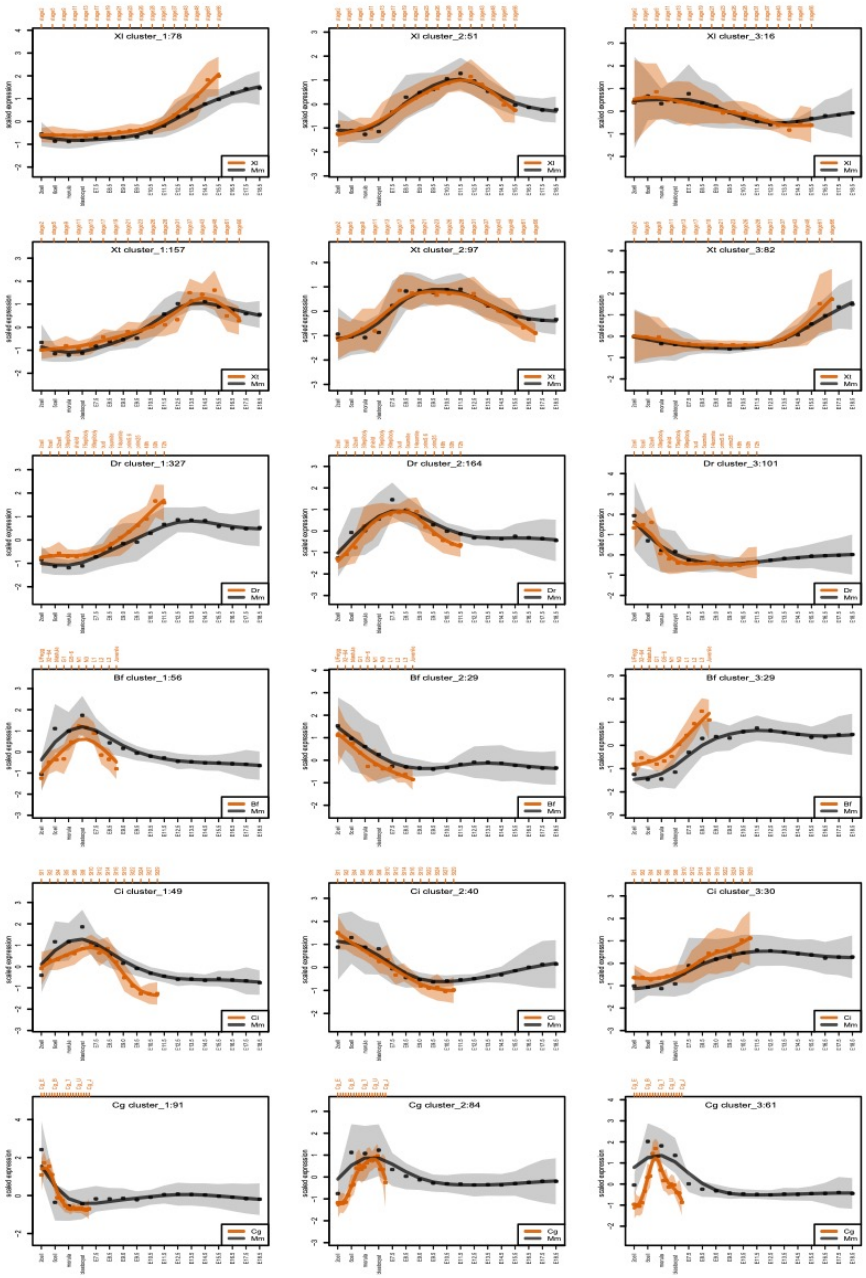


686

687 **Figure 2B–figure supplement 6. Numbers of genes showing the best expression**
688 **trajectory's alignment between the complete developmental course of five species**
689 **and mouse developmental sets of different lengths: from two to 17 stages.**

690 Panel titles show abbreviated species' identifiers. Colors indicate two methods used
691 for developmental expression trajectory calculation: cubic smooth spline (orange) and
692 polynomial regression (gray). Colored rectangles mark stages containing alignments
693 fitting CM (green) or PM (orange) predictions. Black triangles indicate significantly
694 greater (up) or lesser (down) number of genes than that expected by chance aligning
695 to an indicated mouse developmental interval (permutation test, BH-corrected $p <$
696 0.05). The background distribution of gene numbers for subtraction was estimated by
697 shuffling the orthologous relationships of mouse and a given species.

698

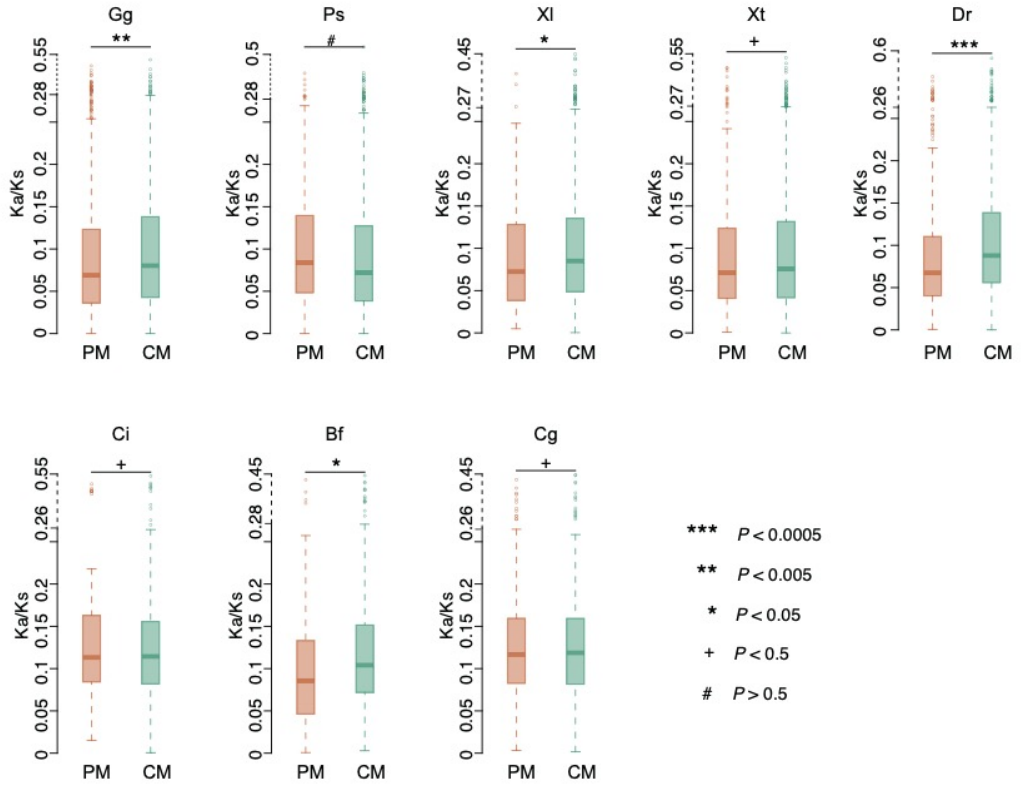


699

700 **Figure 2C – figure supplement 7. Examples of PM gene expression patterns**
701 **correspondent to Figure 2C.**

702 Dots show the cluster-level standardized expression levels at each developmental
703 stage in mouse (black) and the other species (orange). The curves represent average
704 expression profiles and the shaded regions represent the standard deviation of curve
705 estimates.

706



707

708 **Figure 2G – figure supplement 8. The distributions of Ka/Ks values of PM and**
709 **CM genes in each non-mouse species correspondent to Figure 2G**.

710 Significance of the difference between the two distribution was assessed using
711 one-sided Wilcoxon rank-sum test. ***, $p < 0.0005$, **, $p < 0.005$, *, $p < 0.05$, +, $p <$
712 0.5, #, $p > 0.5$.

713

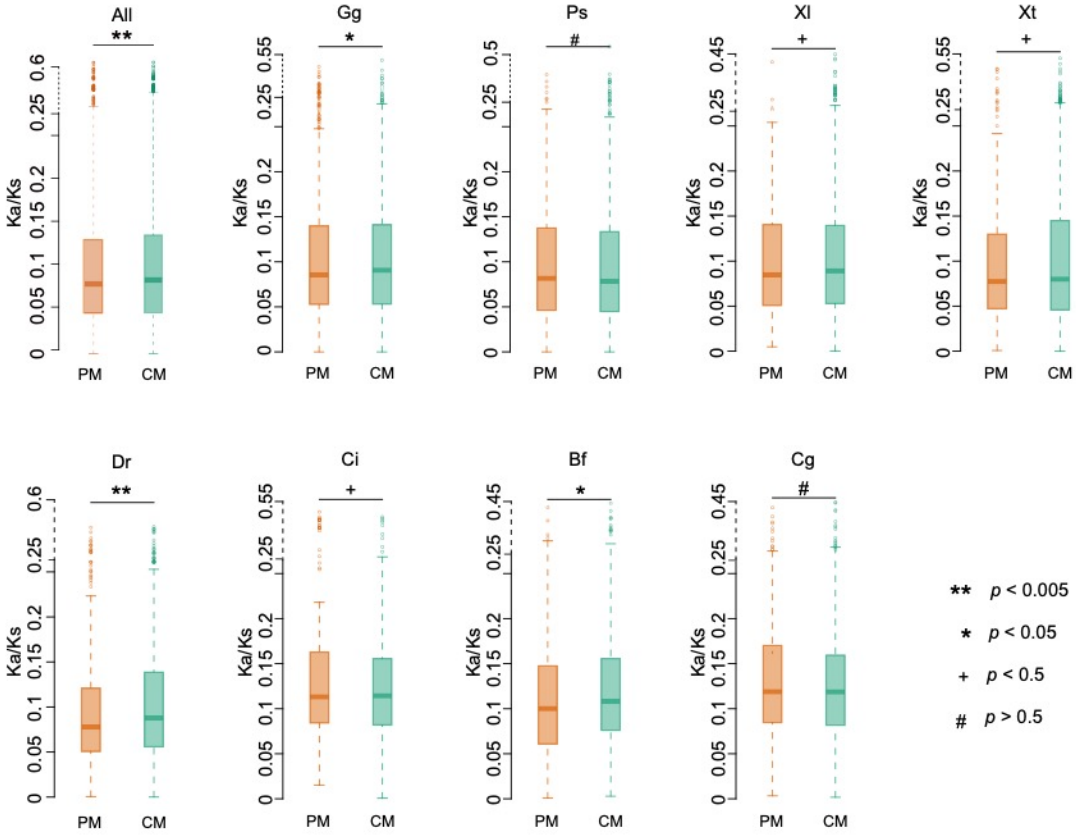


Figure 2G – figure supplement 9. The distributions of Ka/Ks values of PM and CM genes in all species, and in each of the non-mouse species correspondent to Figure 2G.

Significance of the difference between the two distributions was assessed using one-sided Wilcoxon rank-sum test. ***, $p < 0.0005$, **, $p < 0.005$, *, $p < 0.05$, +, $p < 0.5$, #, $p > 0.5$. In contrast to results displayed in Figure 2G and Figure 2G – figure supplement **8**, the Ka/Ks ratios here were estimated using multiple species alignment.

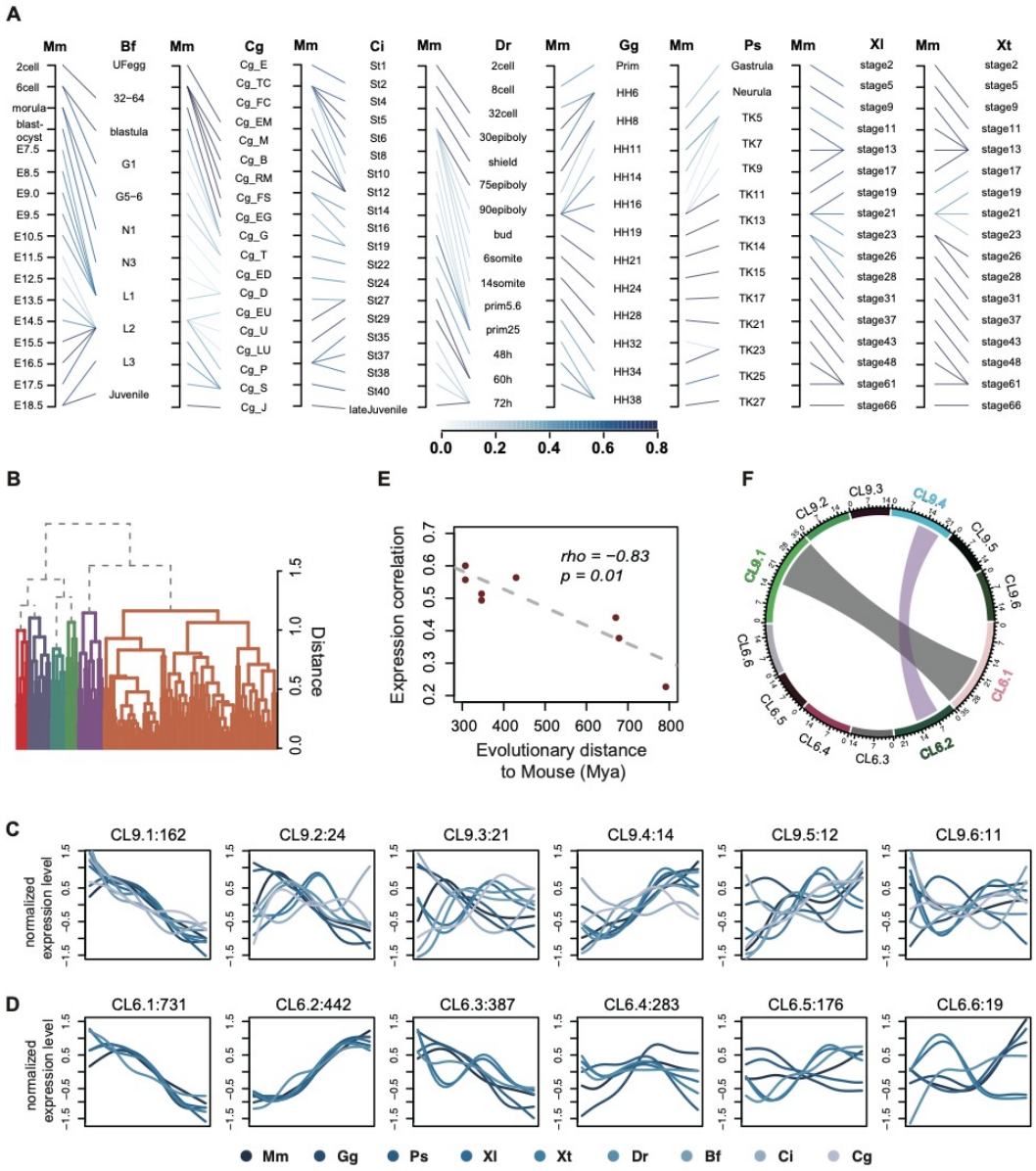
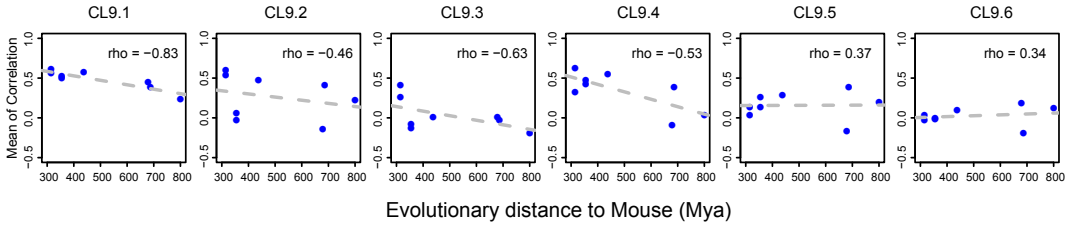


Figure 3. Developmental expression of genes based on the species' alignment.

A Pairwise alignment of developmental stages between mouse and the other species. Thick lines represent alignments supported by technical replicates. Thicker lines represent more stable alignments calculated by random subsampling of samples 500 times. **B** Hierarchical clustering of concatenated developmental gene expression trajectories of 244 gene orthologs shared among nine species. Colors represent clusters. **C** Developmental gene expression patterns in each of six clusters based on nine-species gene orthologs. Colors represent species. Panel titles show cluster identifiers and number of contained genes. **D** Developmental gene expression patterns in each of six clusters based on six vertebrate species gene orthologs. Panel titles

734 show cluster identifiers and number of contained genes. **E** Relationship between the
735 similarity of developmental gene expression patterns and phylogenetic distances. The
736 expression similarity was calculated as the mean of Spearman correlation coefficients
737 between mouse and non-mouse expression trajectories. **F** Chord graph indicating the
738 relationship between clusters obtained using nine-species and six-species ortholog
739 genes. Wider chords represent stronger connection between clusters.

740

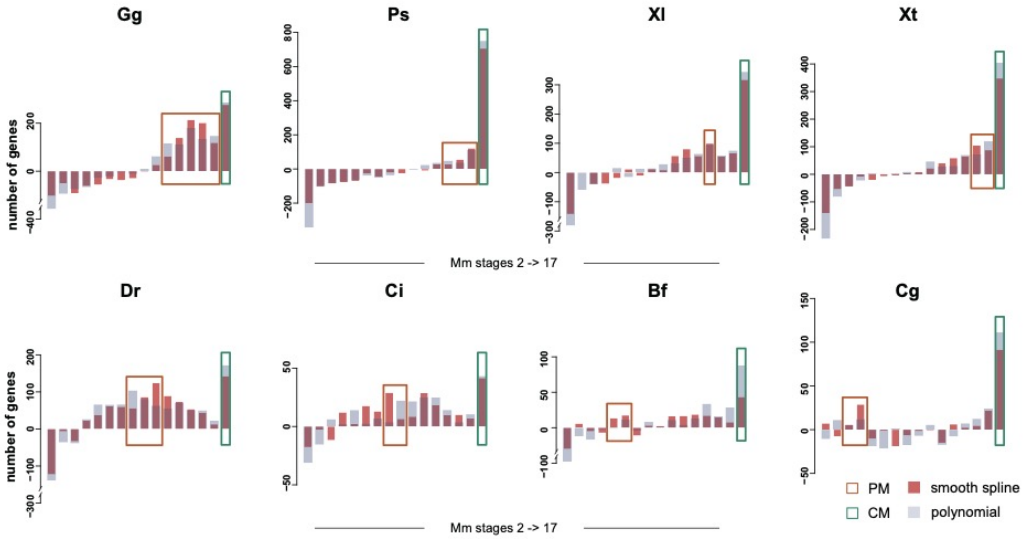


741

742 **Figure 3E – figure supplement 10. Relationship between the similarity of**
743 **developmental gene expression patterns and phylogenetic distances**
744 **correspondent to Figure 3E.**

745 The relationship is shown for each of six clusters of development-related genes
746 orthologous among nine species. The expression similarity was calculated as the
747 mean of Spearman correlation coefficients between mouse and non-mouse expression
748 trajectories. The x-axis shows the phylogenetic distance to the mouse.

749

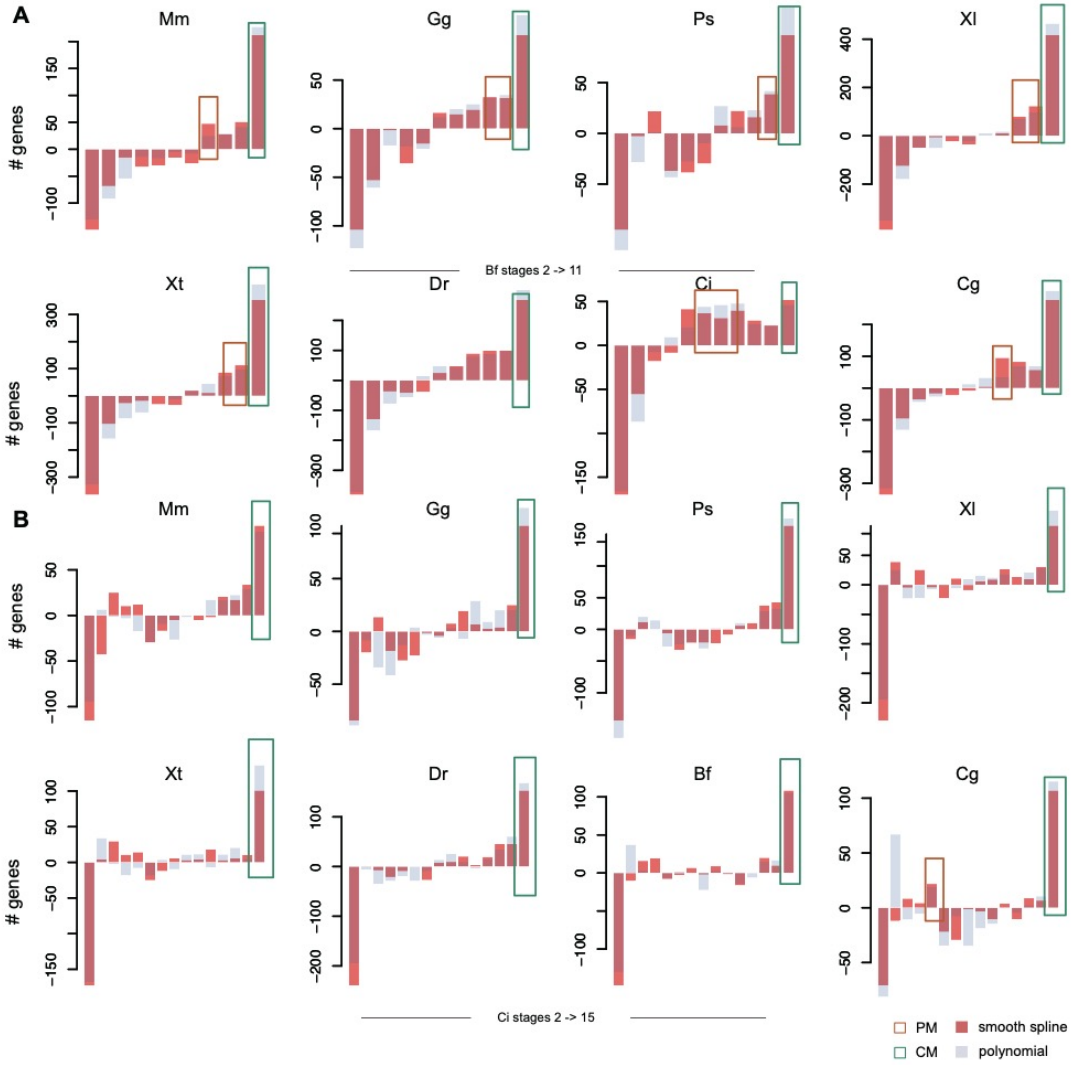


750

751 **Supplementary file 3: Figure S1. Numbers of genes showing the best expression**
752 **trajectory's alignment between the complete developmental course of eight**
753 **species and mouse developmental sets of different lengths: from two to 17 stages.**

754 Panel titles show abbreviated species' identifiers. Bar colors indicate two methods
755 used for developmental expression trajectory calculation: cubic smooth spline (orange)
756 and polynomial regression (purple). Colored rectangles mark stages containing
757 alignments fitting CM (green) or PM (orange) predictions as in Figure 2B. The
758 analysis was restricted to a subset of evolutionarily old genes inferred to exist back in
759 *chordata* ancestor or earlier.

760

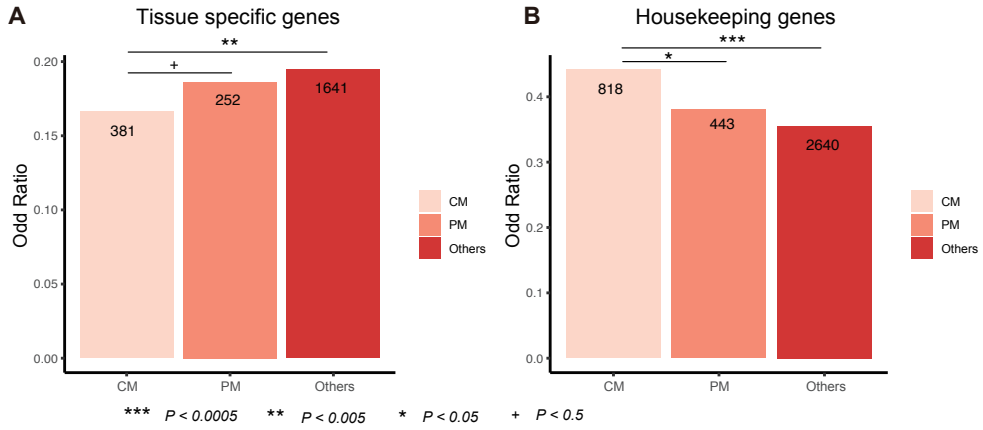


761

762 **Supplementary file 4: Figure S2. Numbers of genes showing the best expression**
763 **trajectory's alignment between the complete developmental course of eight**
764 **species and amphioxus (panel A) and ciona (panel B) developmental sets of**
765 **different lengths.**

766 Panel titles show abbreviated species' identifiers. Bar colors indicate two methods
767 used for developmental expression trajectory calculation: cubic smooth spline (orange)
768 and polynomial regression (purple). Orange and green colored rectangles mark stages
769 corresponding to alignments with significant excess of genes following PM and CM
770 predictions respectively, as in Figure 2B.

771

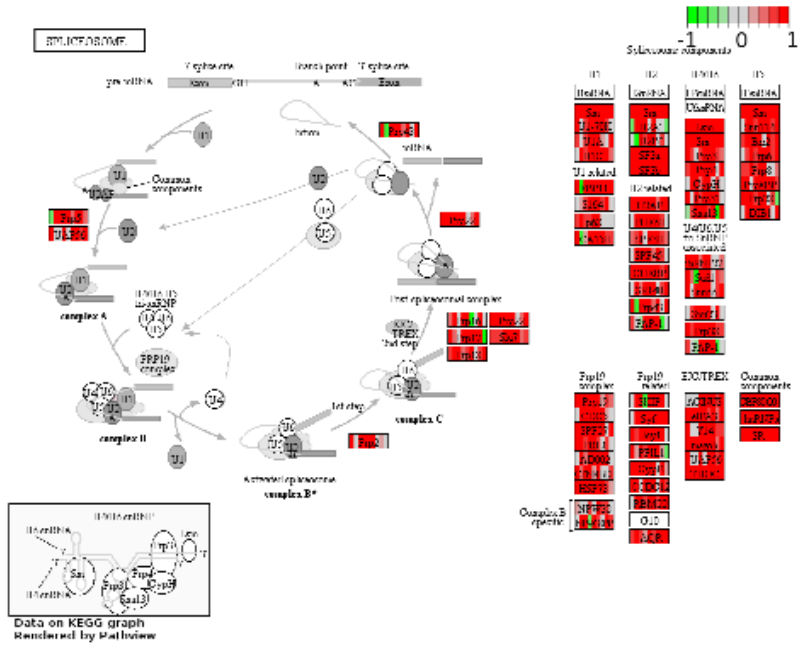


772

773 **Supplementary file 6: Figure S3. Tissue specificity of CM and PM genes.**

774 **A** Bars represent the ratio of tissue-specific and non-tissue specific genes for three
775 gene groups. Numbers within bars show the number of tissue-specific genes in each
776 group. **B** Bars represent the ratio of housekeeping and non-housekeeping genes.
777 Numbers within bars show the number of housekeeping genes in each group. The
778 significance of the differences between gene groups was assessed using one-sided
779 Fisher's exact test. ***, $p < 0.0005$, **, $p < 0.005$, *, $p < 0.05$, +, $p < 0.5$.

780



781

782 **Supplementary file 9: Figure S4. Visualization of spliceosome pathway showing**

783 **developmental expression profile correlation among nine species.**

784 Scheme shows mouse orthologs of spliceosome pathway genes present in nine species.

785 Colors indicate the coefficients of developmental expression profile correlations

786 estimated in all pairwise comparisons between mouse and the other eight species.

787

Additional files

Figure 1D – source data1. Developmental related gene list of each species correspondent to Figure 1D.

Figure 2E – source data1. The list of continuous model genes for each species

Figure 2E – source data2. The list of progressive model genes for each species

Figure 2F – source data 1. Enriched KEGG pathways for progressive model (PM) and continuous model(CM) genes correspondent to Figure 2F.

Supplementary file 1: Table S1. Sample information used in our study with raw file lists and the reads number.

Supplementary file 2: Table S2. The number of stages associated gene for each species

Supplementary file 5: Table S3. Bonferroni-corrected p-values of overlapping CM genes between pairwise species

Supplementary file 7: Table S4. Annotations and GO terms of overlapped genes between CM and CL9.1

Supplementary file 8: Table S5. Enriched GO functional annotation for developmental expression patterns CL1-6

Supplementary file 10: Table S6. Summary of expressed gene number for each species