Comparative transcriptomics reveal distinct patterns of gene expression conservation through vertebrate embryogenesis

Megan E. Chan[1,2], Pranav S. Bhamidipati[1,2], Heather J. Goldsby[3], Arend Hintze[3], Hans A. Hofmann[1,2,4], and Rebecca L. Young[1,2,*]

Affiliations:

[1]Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA

[2]Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, TX, USA

[3]Department of Integrative Biology, Michigan State University, East Lansing, MI, USA

[4]Institute for Cellular and Molecular Biology, Institute for Neuroscience, The University of Texas at Austin, Austin, TX, USA

*Correspondence to Dr. Rebecca L. Young, youngrl@utexas.edu.

ABSTRACT

Despite the diversity of life, studies of variation across animals often remind us of our deep evolutionary past. Abundant genome sequencing over the last ~25 years reveals remarkable conservation of genes and recent analyses of gene regulatory networks illustrate that not only genes but entire pathways are conserved, reused, and elaborated in the evolution of diversity. Predating these discoveries, 19th-century embryologists observed that though morphology at birth varies tremendously, certain stages of embryogenesis appear remarkably similar across vertebrates. Specifically, while early and late stages are variable across species, anatomy of mid-stages embryos (the phylotypic stage) is conserved. This model of vertebrate development and diversification has found mixed support in recent analyses comparing gene expression across species possibly owing to differences across studies in species, embryonic stages, and gene sets compared. Here we perform a meta-analysis of 186 microarray and RNA-seq expression data sets covering embryogenesis in six vertebrate species spanning ~420 million years of evolution. We use an unbiased clustering approach to group stages of embryogenesis by transcriptomic similarity and ask whether gene expression similarity of clustered embryonic stages deviates from the null hypothesis of no relationship between timing and diversification. We use a phylogenetic comparative approach to characterize expression conservation pattern (i.e., early conservation, hourglass, inverse hourglass, late conservation, or no relationship) of each gene at each evolutionary node. We find an enrichment of early conservation and hourglass patterns and a large depletion of genes exhibiting no distinguishable pattern of conservation. Using this approach, we ask whether the proportions of genes following distinct evolutionary conservation patterns change through evolutionary time and whether genes consistently follow the same pattern across nodes of the vertebrate phylogeny. We find that genes exhibiting an hourglass pattern at one node of the phylogeny are more likely to show an hourglass pattern at other nodes with 89 hourglass genes shared in at least three of the four nodes compared to only six early conservation genes. Consistent with the hourglass hypothesis, this finding suggests that genes following an hourglass pattern are more conserved over evolutionary time.

INTRODUCTION

During embryogenesis, metazoan organisms transition from a single-cell zygote to a multicellular, functional embryo. Given the complexity of the process and the astonishing diversity of resultant phenotypes, the similarities in the developmental processes and anatomy of embryogenesis across species are striking and have captivated the imagination of biologists for nearly 200 years (von Baer, 1828). For example, vertebrates establish a highly conserved body plan ('*bauplan*') from which species-specific variation and elaborations develop. Yet, whether there are generalizable rules that direct diversification of embryogenesis across distantly related species remains controversial (Richardson et al., 1997; Bininda-Emonds et al., 2003). One hypothesis, inspired by von Baer's (1828) pioneering observations, suggests that early and late phases of embryogenesis are variable across species (owing to diversity and species specificity of reproductive modes and post-body plan elaboration, respectively), while anatomy mid-embryogenesis is conserved (Fig. 1). According to this developmental hourglass hypothesis, similarity of the mid-embryogenesis 'phylotypic stage' (Sander, 1983; Richardson, 1995) reflects developmental constraints of body plan formation, including global signaling interdependence and interactions (Raff, 1996; Galis and Metz, 2001) and temporal and spatial patterns of Hox expression (Duboule, 1994). Still others hypothesize that early stages of embryogenesis are most similar resulting from lethality and later developmental consequences ('developmental burden') of early developmental anomalies (Riedl, 1978; discussed in Irie and Kuratani, 2014) (Fig. 1).

In recent years a number of studies have leveraged the ever-increasing amount of sequencing data and approaches to test transcriptome-level predictions of the hourglass hypothesis and its underlying mechanistic basis. Support for the hourglass model of development varies across studies depending on the species included, the methods of analysis, and the approach to selecting and comparing stages of embryogenesis. A number of studies comparing gene expression variation through embryogenesis across species have found support for an increase in expression conservation mid-embryogenesis (Domazet-Loso and Tautz, 2010; Irie and Kuratani, 2011, 2014; Yanai et al., 2011; Levin et al., 2012; Wang et al., 2013; Gerstein et al., 2014; Zalts and Yanai, 2017). Still others find that diversification in gene expression is not consistent with the hourglass model (Piasecka et al., 2013; Tian et al., 2013; Wu et al., 2019). These analyses vary in the evolutionary distances among species investigated, sometimes spanning more than one phylum (Tian et al., 2013; Gerstein et al., 2014) and often testing predictions of the hourglass hypothesis in only one or two species (e.g., zebrafish, Domazet-Loso and Tautz, 2010; soft-shell turtle and chicken, Wang et al., 2013; *Caenorhabditis elegans*, Zalts and Yanai, 2017).

Any test of the developmental hourglass hypothesis is hampered by heterochrony of developmental events as well as species-specific nomenclature and sampling procedures, which together complicate the proper

alignment of stages of embryogenesis (Mungall et al., 2012). Irie and Kuratani (2011) circumvent this challenge by directly comparing only a subset of well-defined stages across species, but it is unclear to what extent the discovery of an hourglass pattern might be biased by the stages selected for comparison or what the null expectation for expression conservation is. Finally, selecting the appropriate gene set to compare across species is unclear. Most studies examine conservation of the entire transcriptome or at least the comparable set of orthologous genes. However, others suggest that abundant expression of housekeeping genes may bias discovery of gene expression conservation across species (Piasecka et al., 2013). In fact, it is clear that the genomic and developmental processes underlying even anatomically similar and homologous phenotypes can diverge via developmental drift or selective processes (de Beer, 1971; Wagner, 1989; True and Haag, 2001; Wilkins, 2002; Young and Wagner, 2011). As a result, some argue that studies examining the evolution of organismal phenotypes should focus on a core set of regulatory genes critical to the initiation of the specific developmental program of that character (designated as “kernels” by Davidson and Erwin, 2006, or "Character Identity Networks," by Wagner, 2007). How to identify the relevant gene set (i.e., kernel or ChIN) that is fundamental for shared developmental processes of vertebrate embryogenesis is unclear.

While studies consistently find patterned expression divergence (i.e., early conservation or hourglass patterns) through embryogenesis across species, what pattern is followed and whether that diversification pattern differs over evolutionary time remains unclear. One study comparing expression divergence of animals from different phyla reported an inverse hourglass – where expression differences were highest mid-embryogenesis; however, this finding has been refuted due to the methodological approach that does not account for phylogenetic non-independence in sampling (Dunn et al., 2018). Here we perform a meta-analysis of 186 publicly available microarray and RNA-seq expression data sets covering embryogenesis in six vertebrate species spanning ~420 million years of evolution. We use an unbiased clustering approach to group stages of embryogenesis by transcriptomic similarity and ask whether gene expression similarity of clustered embryonic stages deviates from the null hypothesis of no relationship between timing and diversification. Second, using a phylogenetic comparative approach, we characterize the expression conservation pattern (i.e., early conservation, hourglass, inverse hourglass, late conservation, or no relationship) of each gene at each evolutionary node. Using this approach, we ask whether the proportions of genes following distinct evolutionary conservation patterns changes through evolutionary time and whether genes consistently follow the same pattern across nodes of the vertebrate phylogeny.

METHODS

Obtaining and preprocessing transcriptomic data

Gene expression profiles through embryogenesis were obtained from publicly available repositories for six vertebrate species. In total 112 microarray datasets from five species and 74 RNA-seq datasets from four species were included in the analysis. Developmental time points included for each species and sequencing platform are detailed in Supplementary Tables 1 and 2 (microarray and RNA-seq, respectively). Data sets include: zebrafish, *Danio rerio* (microarray: Roux and Robinson-Rechavi, 2008; RNA-seq: Comte et al., 2010; Yang et al., 2013); chicken, *Gallus gallus* (microarray: Irie and Kuratani, 2011; RNA-seq: Wang et al., 2013), Chinese soft-shell turtle, *Pelodiscus sinensis* (RNA-seq: Wang et al., 2013), mouse, *Mus musculus* (microarray: Irie and Kuratani, 2011; Xue et al., 2013); African clawed frog, *Xenopus laevis* (microarray: Yanai et al., 2011), and Western clawed frogs, *Xenopus tropicalis* (microarray: Yanai et al., 2011; RNA-seq: Tan et al., 2013). *D. rerio* expression data at the zygote developmental stage (0.25 hpf) was excluded because of likely abundance of maternal transcripts, and time points after 4 dpf were excluded due to completion of the developmental program (after Kimmel et al., 1995; Yang et al., 2013).

Preprocessing microarray and RNA-seq data

Affymetrix and Agilent microarray data were imported directly using the R packages *simpleaffy* and *limma*, respectively (Wilson and Miller, 2005; Ritchie et al., 2015). For both microarray platforms, preprocessing consisted of RMA background correction with quantile normalization (Irizarry et al., 2003). This information was automatically attained by *limma* for Agilent data sets. For Affymetrix and Agilent data, probe sets that mapped to multiple genes or no genes at all were excluded from further analysis. All expression values were then transformed to log-scale using the function $log_2(x)$ ("Log2-transformed"). For RNA-seq data, raw reads were checked for quality using FastQC (Andrews, 2010). All datasets were good quality with less than 10% adaptor contamination, thus, no trimming was required. After quality control raw reads were pseudoaligned to species-specific transcriptomes using Kallisto to produce read counts (Bray et al., 2016). Read counts were transformed to transcripts per million mapped reads (TPMs). The package *biomaRt* was used to assign corresponding ENSEMBL gene ID(s) to each Affymetrix probe set and RNA-seq transcript (Durinck et al., 2009). For microarray, the signals of all probe sets mapping to the same gene were averaged. For RNA-seq counts mapped to different transcripts of the same gene were summed. Expression of each gene was averaged across biological replicates for developmental time point. These were the expression values used for downstream analysis.

Clustering of embryonic stages

A major challenge in comparing embryogenesis across species is aligning developmental stages. We used transcriptomic similarities to generate unbiased clusters of embryonic stages for each species (Fig. 2A). We determined the number of clusters using an elbow plot method. For each species and sequencing platform, we performed *k*-means clustering using gene expression for all developmental stages. We varied the number of clusters from k = 1-9 for microarray and k = 1-7 for RNA-seq and computed the sums of squares error (SSE, or variance within cluster) for all k. To determine an appropriate number of clusters, we used a) the "elbow" effect (or determined the k at which additional cluster no longer results in a large reduction in SSE) and b) determined the k where clustering of groups maintained temporal order of embryogenesis (i.e., no late stages cluster with early rather than other late stages). Using these criteria, we selected k = 5 for downstream analyses (Fig. 2). To generate the five clusters of embryonic stages for each species, we used Spearman correlations. Spearman's rank correlations in gene expression for all pairwise stages were used to generate hierarchical clustering of stages for each species and resulting dendrogram was partitioned to generate five clusters. Description of developmental events were obtained from species-specific references including: zebrafish (Kimmel et al., 1995), chicken (Hamburger and Hamilton, 1992), softshell turtle (Tokita and Kuratani, 2001), both *Xenopus* species (Nieuwkoop and Faber, 1994), and mouse (Graham et al., 2015). For clusters containing more than one embryonic stage, an expression mean was used as the representative expression for that cluster.

Ortholog calling

To identify orthologous genes across taxa we used the sequence-based ortholog calling software package using OrthoMCL (Li et al., 2003) for microarray data and FastOrtho (Wattam et al., 2013; an implementation of OrthoMCL) for RNA-seq data. Though FastOrtho is computationally much more efficient, the implementation of FastOrtho was not yet widely distributed when microarray orthologs calling was performed. Protein sequences of the reference genomes were organized into orthologous gene groups based on sequence similarity. For each reference proteome, protein and corresponding gene ids were grouped assigned as paralogs when sequence similarity was higher among genes within species than between species. To facilitate downstream analysis of expression conservation across taxa, we removed any orthologous gene groups containing paralogs or losses/absences in one or more species. The resulting one-to-one orthologs were used for all downstream analyses. To assess similarity in microarray and RNA-seq comparison, we compared the one-to-one ortholog sets the three species (zebrafish, chicken, and the Western clawed frog).

Comparing transcriptomes through embryogenesis across species

For both microarray and RNA-seq data, we assessed transcriptomic similarity at early, middle and late phases of embryogenesis across species by calculating the Spearman's rank correlation for all pairwise comparisons of species for each of the five clusters of embryonic stages. For microarray data, we excluded one frog (*X. laevis*) from the pairwise comparisons to prevent biasing the outcome as a consequence of high correlations in gene expression between the two frog species at each embryonic time cluster. Due to the high correlations between these two species, similar results were recovered when *X. tropicalis* was removed from the analysis instead of *X. laevis*. We used permutation analysis to assess whether correlations are higher or lower than expected by chance. Specifically, for each species, we randomly assigned stages to a cluster maintaining the original number of stages included in the observed cluster and computed the rank correlation for all pairwise species comparison. We repeated the permutation analysis 1000 times and assessed significance by comparing the observed rank correlation to distribution of rank correlations generated by permutation analysis.

Characterizing expression conservation of each gene through embryogenesis across vertebrates

To assess conservation of gene expression for each gene at each cluster of embryonic stages and each node, we calculated a difference in expression rank scaled by the divergence time between the groups (Supplementary Fig. 1). For this analysis, we focus on microarray data since we have four phylogenetic nodes (i.e., frogs, amniotes, tetrapods, and vertebrates) compared to only three for the RNA-seq data (i.e., amniotes, tetrapods, and vertebrates). At each phylogenetic node, we characterized patterns of expression conservation across embryogenesis using the R package Clustering of Time Series Gene Expression Data (*ctsGE*; Sharabi-Schwager and Ophir, 2019). Using *ctsGE*, gene conservation scores were median-scaled and converted into conservation indices that indicated whether gene conservation was above (1), below (-1), or within (0) the cutoff range (+/- 0.7) of the median value at each time step (here cluster of embryonic stages). Each index of expression conservation across embryogenesis was assigned to a conservation pattern based on median transitions across the assigned significance cutoff (+/- 0.7). For early conservation: similarity decreases significantly; hourglass: similarity increases and then decreases: inverse hourglass: similarity decreases and then increases; late conservation: similarity increases; no relationship: similarity does not vary or does not follow other conservation patterns through embryogenesis. Indices assignments are provided in Supplementary Table 5. R package *UpsetR* was used to identify the genes assigned to the same patterns across evolutionary nodes (Conway et al., 2017).

To examine enrichment of genes in each conservation pattern, we first calculated the proportion of indices in each conservation pattern. The null expectation assumes that the proportion of genes for each conservation pattern should be equal to the proportion of patterns. Deviation from the null expectation was assessed at each evolutionary node using a Chi-square Goodness of Fit Test.

RESULTS

Variation in gene sets and alignment of embryonic stage clusters across species and sequencing platforms

We identified 1626 and 1793 one-to-one orthologs for microarray and RNA-seq, respectively, consistent with other comparative gene expression studies spanning vertebrates (Young et al., 2019). One-to-one orthologs from microarray and RNA-seq datasets from the same species were largely non-overlapping. 335, 264, 345 genes overlapped in the zebrafish, chicken, and western clawed frog, respectively. This represented 20.6%, 18.7%, and 21.2% of the zebrafish, chicken, and western clawed frog microarray orthologs and 16.2%, 18.0%, and 19.2% of the zebrafish, chicken, and western clawed frog RNA-seq orthologs. Consistency of this overlap across species suggests that microarray and RNA-seq approaches may target distinct aspects of the transcriptome.

When we clustered the embryonic stages according to their gene expression patterns we found that the temporal order of stages across embryogenesis was maintained. However, inconsistent sampling across studies as well as heterochrony across species resulted in some variation in the major events contained in each cluster of embryonic stages across species and between microarray and RNA-seq platforms (Fig. 3). Stages contained in each cluster and a biological description of embryonic events is provided in Fig. 3 and Supplementary Tables 1 and 2 (microarray and RNA-seq, respectively). The number of stages collapsed into each cluster also varied across species and sequencing platform; however, no systematic bias between cluster timing in embryogenesis and the number of stages included was apparent (Fig. 4).

Conservation of gene expression at the transcriptome level

We then asked how variable the developmental clusters identified in Fig. 3 are across species. Clusters that show less variability would be considered more conserved. We found that the pattern of pairwise rank correlations for all one-to-one orthologs (1626 and 1793 for microarray and RNA-seq, respectively) through embryogenesis differed for microarray and RNA-seq comparisons (Fig 5). In the microarray comparison, median pairwise rank correlations increased mid-embryogenesis, with clusters 2, 3, and 4 having the highest median correlation, and decreased early and late in embryogenesis, with cluster 5 having the lowest median,

suggesting an hourglass. However, when compared to the null expectation, generated by permuting the cluster assignment of each stage, the observed pattern did not differ from the null (cluster 1, p = 0.57; cluster 2, p = 0.25; cluster 3, p = 0.47; cluster 4, p = 0.65; cluster 5, p = 0.512). In the RNA-seq comparison, median pairwise rank correlations increased to its highest median score at cluster 2 and dropped through later stages of embryogenesis 3, with cluster 5 having the lowest median. When compared to the null expectation, median pairwise rank correlation of cluster 2 was significantly higher and correlations of clusters 4 and 5 were significantly lower than expected by chance (cluster 1, p = 0.24; cluster 2, p = 0.005; cluster 3, p = 0.51; cluster 4, p = 0.003; cluster 5, p = 0.008).

Enrichment and overlap of expression conservation patterns across the phylogeny

Next, we used ctsGE time series analysis to characterize patterns of gene expression conservation across embryogenesis at each phylogenetic node. For the microarray data set, gene conservation scores yielded a total of 90, 90, 87, 89 patterns of expression conservation in frogs, amniotes, tetrapods, and vertebrates, respectively. Interestingly, the same three, thirteen, thirteen and three early conservation, hourglass, inverse hourglass, and late conservation patterns were described at all nodes of the phylogeny. Also, 55 'no relationship' patterns were shared at all nodes of the phylogeny with an additional two patterns shared in frogs, amniotes and vertebrates. We calculated the expected number of genes for each pattern as equivalent to the proportion of total patterns. For all patterns, except no relationship, had more genes than expected (Fig. 6). This deviation from the null expectation was significant for all nodes of the phylogeny (frogs: X^2 = 297.23, df = 4, p-value < 0.001; amniotes: X^2 = 418.96, df = 4, p-value < 0.001; tetrapods: X^2 = 261.4, df = 4, p-value < 0.001; vertebrates: X^2 = 285.17, df = 4, p-value < 0.001). The proportion of genes observed for each pattern is provided in Fig. 6B.

An exploration of overlap in gene conservation across nodes of the phylogeny revealed that genes exhibiting an hourglass pattern are the most consistent across the phylogeny. 89 hourglass genes were overlapping across at least three out of four nodes as compared to six early conservation genes, 41 inverse hourglass, and two late conservation genes (Fig. 6; Supplementary Fig. 2).

DISCUSSION

Over the past decade, the debate on whether diversification of embryogenesis follows generalizable rules has been reinvigorated by the ability to test predictions of the hourglass, developmental burden, and other hypotheses (Fig. 1A) at the transcriptome- and genome-levels. Enabled by the increase in 'omics-level data

and next-generation sequencing accessibility, a number of studies have explored patterns of diversification in gene expression though embryogenesis across species. These studies have found mixed support for the hourglass and other models of diversification across species (reviewed in: Irie, 2017; Liu and Robinson-Rechavi, 2018). Differences among studies could reflect differences in species compared as some studies span phyla (Tian et al., 2013; Gerstein et al., 2014) and others are restricted to internal nodes of the phylogeny (e.g., amniotes; Wang et al., 2013). Alternatively, selection of stages or gene sets to compare may lead to different results and interpretations. Finally, studies comparing gene expression similarity across the transcriptome directly test against a null hypothesis. Such a test is critical because the degree of variation expected through developmental stages across species is unknown (Young and Hofmann, 2019). Here, we examine patterns of expression diversification throughout embryogenesis in six vertebrate species using a meta-analysis of 112 microarray and 74 RNA-seq data sets (Fig. 1B). First, using an unbiased approach we cluster stages of embryogenesis within species and compare expression of those clusters across species (Figs. 2-4; Supplementary Tables 1 and 2). This approach allows for the inclusion of all available stages of embryogenesis and removes bias that could result from comparing only selected stages. Second, we use a permutation analysis to generate a null expectation. We test observed conservation estimates against this null expectation to characterize transcriptome-level diversification through embryogenesis across species (Fig. 3). Finally, we characterize the expression conservation of each gene at each node of the phylogeny (Fig. 1B; Supplementary Fig. 1) to examine how expression conservation patterns vary through evolutionary time (Fig. 5).

Aligning stages is complicated by taxon-specific sampling practices as well as heterochrony in developmental events across distantly related species, leading to some researchers to question the validity of anatomical hourglass hypotheses (Bininda-Emonds et al., 2003). We found that while expression varies in similar ways through embryogenesis across species, stages of embryogenesis did not always consistently cluster together within species (Fig.3 and 4; Supplementary Tables 1 and 2). These differences likely reflect both biological variation in the molecular timing of developmental events, technical variation in sampling procedures across species, and a lack of available data sets particularly at early time points. Because clustering has the advantage of being unbiased and no systematic bias in sampling was apparent, we moved forward by comparing gene expression at each embryonic stage cluster across species. Future studies with a systematic sampling of embryogenesis across species could disentangle the source (biological or technical) of variation in stage clustering across species.

Consistent with mixed support for the hourglass and other models of developmental divergence (e.g., early conservation or inverse hourglass) found across studies, our comparisons of expression variation in all one-to-one orthologs present across species in the microarray and RNA-seq data yielded different patterns.

While observed gene expression correlations (as a measure of conservation) differed through embryogenesis for both microarray and RNA-seq, only the RNA-seq data showed a pattern that differed from the null expectation of no relationship between timing and diversification (Fig. 5). For RNA-seq, we found a significant increase in expression correlation over the null expectation in cluster 2 followed by a significant reduction in expression correlation in late clusters 4 and 5. Though clusters 1 and 2 are similar in gene expression correlation suggesting an early conservation pattern, cluster 1 does not differ from the null expectation. Inconsistencies between the microarray and RNA-seq data sets could result from differences in species inclusion (Fig. 1B) and/or variation in aspects of the transcriptome captured by these distinct sequencing platforms. In fact, we found little overlap of one-to-one orthologs between microarray and RNA-seq datasets, with only ~20% of the genes contained in both analyses. This illustrates a major challenge for meta-analyses, like our present study, aiming to capitalize on existing data sets to test biological hypotheses. Importantly, this challenge exists not only for comparisons of microarray and RNA-seq but also other sequencing approaches including bulk versus single-cell RNA-seq, especially as more and more single-cell data become available (e.g., Hicks et al., 2017).

Inconsistency across studies and sequencing platforms also suggests that comparing the whole transcriptome may not appropriate. Specifically, patterns of expression conservation at the whole transcriptome-level may be biased by abundant and constitutively expressed genes (e.g., see Piasecka et al., 2013). Whole-genome approaches have the great advantage of being unbiased which allows for the identification of novel gene associations with phenotypes and/or gene interactions that would be missed using a candidate gene approach. However, we know that variation is not equivalent across levels of biological organization. For example, studies comparing mRNA and protein levels have found only modest overall correlation (Foss et al., 2007; Fu et al., 2009; Ghazalpour et al., 2011; Vogel and Marcotte, 2012). Further, studies of character homology have shown that even anatomically and physiologically similar homologies can differ in underlying developmental and molecular mechanisms (Wagner, 1989; Wilkins, 2002; Young and Wagner, 2011). As a result, in addition to a transcriptome-level comparison through embryogenesis across species, we also used a time series analysis to characterize the conservation pattern each gene at each node of the phylogeny (Fig. 6). We observed a significant enrichment of all patterned conservation models (early conservation, hourglass, inverse hourglass, and late conservation) above the expected number and a large depletion of genes following no distinguishable pattern (‘no relationship’; Fig. 6A). Both the enrichment of genes exhibiting different patterns and the proportion of overall genes with distinct patterns varied in interesting ways. We found that early conservation and hourglass patterns exhibited higher enrichment overall with the exception of the frogs where both inverse hourglass and late conservation patterns had higher enrichment than early conservation (Fig. 6A). Enrichment of early conservation increased steadily with evolutionary time and was highest the vertebrate node (Fig. 6A).

Further, relative to other comparisons, the frogs are closely related (Fig.1B) Combined, these results suggest that divergence in gene expression through embryogenesis may depend on the evolutionary distance being compared. Follow-up analyses including closely and distantly related species are needed to better understand these patterns. Finally, we found that genes exhibiting an hourglass pattern at one node of the phylogeny were more likely to show an hourglass pattern at other nodes. 89 hourglass genes were shared in at least three of the four nodes compared to only six early conservation genes (Fig. 6C and D). Consistent with the hourglass hypothesis, this suggests that hourglass genes are more likely to be conserved over evolutionary time. In addition to testing hypothesized models of gene expression diversification and revealing interesting variation in patterns of expression diversification over evolution, time series analyses that characterize expression divergence of each gene can be used to test hypothesized mechanisms of constraint. For example, do regulatory interactions or temporal and spatial expression patterns of a gene follow a correlated dynamic pattern through embryogenesis?

ACKNOWLEDGEMENTS

We thank D. Wylie for consulting on data analysis and T. Solomon-Lane, C. Friesen, I. Miller-Crews, and M. Rodriguez-Santiago for discussions that improved this manuscript. This work was funded in part by a University of Texas at Austin Undergraduate Research Fellowship to P.S.B. and a grant from the NSF BEACON Center for the Study of Evolution in Action, DBI-0939454 to R.L.Y., H.J.G, A.H., and H.A.H.

LITERATURE CITED

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

von Baer, K.E. 1828. Uber Entwickelungsgeschichte der Thiere (Königsberg: Bornträger).

de Beer, G. 1971. Homology, An Unsolved Problem, Volume 11 (Oxford: Oxford University Press).

Bininda-Emonds, O.R.P., Jeffery, J.E., and Richardson, M.K. 2003. Inverting the hourglass: quantitative evidence against the phylotypic stage in vertebrate development. *Proc. R. Soc. London B Biol. Sci.* 270: 341–346.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34: 525.

Comte, A., Roux, J., and Robinson-Rechavi, M. 2010. Molecular signaling in zebrafish development and the vertebrate phylotypic period. *Evol. Dev.* 12: 144–156.

Conway, J.R., Lex, A., and Gehlenborg, N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33: 2938–2940.

Davidson, E.H., and Erwin, D.H. 2006. Gene Regulatory Networks and the Evolution of Animal Body Plans. *Science (80-.).* 311: 796–800.

Domazet-Loso, T., and Tautz, D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 468: 815–818.

Duboule, D. 1994. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev.* 1994: 135–142.

Dunn, C.W., Zapata, F., Munro, C., Siebert, S., and Hejnol, A. 2018. Pairwise comparisons across species are problematic when analyzing functional genomic data. *Proc. Natl. Acad. Sci.* .

Durinck, S., Spellman, P., Birney, E., and Huber, W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4: 1184–1191.

Foss, E.J., Radulovic, D., Shaffer, S.A., Ruderfer, D.M., Bedalov, A., Goodlett, D.R., and Kruglyak, L. 2007. Genetic basis of proteome variation in yeast. *Nat. Genet.* 39: 1369–1375.

Fu, J., Keurentjes, J.J.B., Bouwmeester, H., America, T., Verstappen, F.W.A., Ward, J.L., Beale, M.H., de Vos, R.C.H., Dijkstra, M., Scheltema, R.A., et al. 2009. System-wide molecular evidence for phenotypic buffering in Arabidopsis. *Nat. Genet.* 41: 166–167.

Galis, F., and Metz, J.A.J. 2001. Testing the vulnerability of the phylotypic stage: On modularity and evolutionary conservation. *J. Exp. Zool.* 291: 195–204.

Gerstein, M.B., Rozowsky, J., Yan, K.-K., Wang, D., Cheng, C., Brown, J.B., Davis, C.A., Hillier, L., Sisu, C., Li, J.J., et al. 2014. Comparative analysis of the transcriptome across distant species. *Nature* 512: 445–448.

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., et al. 2011. Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLOS Genet.* 7: e1001393.

Graham, E., Moss, J., Burton, N., Armit, C., Richardson, L., and Baldock, R. 2015. The atlas of mouse development eHistology resource. *Development* 142: 1909 LP – 1911.

Hamburger, V., and Hamilton, H.L. 1992. A series of normal stages in the development of the chick embryo. *Dev Dyn* 88: 231–272.

Hicks, S.C., Townes, F.W., Teng, M., and Irizarry, R.A. 2017. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* 19: 562–578.

Irie, N. 2017. Remaining questions related to the hourglass model in vertebrate evolution. *Curr. Opin. Genet. Dev.* 45: 103–107.

Irie, N., and Kuratani, S. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* 2: 248.

Irie, N., and Kuratani, S. 2014. The developmental hourglass model: a predictor of the basic body plan? *Development* 141: 4649–4655.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249–264.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. 1995. Stages of embryonic development of the zebrafish. *Dev. Dyn.* 203: 253–310.

Levin, M., Hashimshony, T., Wagner, F., and Yanai, I. 2012. Developmental Milestones Punctuate Gene Expression in the Caenorhabditis Embryo. *Dev. Cell* 22: 1101–1108.

Li, L., Stoeckert, C.J., and Roos, D.S. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13: 2178–2189.

Liu, J., and Robinson-Rechavi, M. 2018. Developmental Constraints on Genome Evolution in Four Bilaterian Model Species. *Genome Biol. Evol.* 10: 2266–2277.

Mungall, C.J., Torniai, C., Gkoutos, G. V, Lewis, S.E., and Haendel, M.A. 2012. Uberon, an integrative multi-species anatomy ontology. *Genome Biol.* 13: R5.

Nieuwkoop, P.D., and Faber, J. 1994. Normal table of Xenopus laevis (Daudin) : a systematical and chronological survey of the development from the fertilized egg till the end of metamorphosis (New York: Garland Pub.).

Piasecka, B., Lichocki, P., Moretti, S., Bergmann, S., and Robinson-Rechavi, M. 2013. The Hourglass and the Early Conservation Models—Co-Existing Patterns of Developmental Constraints in Vertebrates. *PLoS Genet* 9: e1003476.

Raff, R.A. 1996. The shape of life: Genes, development, and the evolution of animal form (Chicago: University of Chicago Press).

Richardson, M.K. 1995. Heterochrony and the Phylotypic Period. *Dev. Biol.* 172: 412–421.

Richardson, M.K., Hanken, J., Gooneratne, M.L., Pieau, C., Raynaud, A., Selwood, L., and Wright, G.M. 1997. There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. *Anat Embryol* 196:.

Riedl, R. 1978. Order in Living Systems: A Systems Analysis of Evolution (New York: Wiley).

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43: e47–e47.

426 Roux, J., and Robinson-Rechavi, M. 2008. Developmental Constraints on Vertebrate Genome Evolution.
427 *PLoS Genet* 4: e1000311.

428 Sander, K. 1983. The evolution of patterning mechanisms: gleanings from insect embryogenesis and
429 spermatogenesis BT - Development and Evolution. In Development and Evolution, B.C. Goodwin, N.
430 Holder, and C.C. Wylie, eds. (Cambridge: Cambridge University Press), pp. 137–159.

431 Sharabi-Schwager, M., and Ophir, R. 2019. ctsGE: Clustering of Time Series Gene Expression data. R
432 package version 1.12.0, https://github.com/michalsharabi/ctsGE.

433 Tan, M.H., Au, K.F., Yablonovitch, A.L., Wills, A.E., Chuang, J., Baker, J.C., Wong, W.H., and Li, J.B.
434 2013. RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome
435 over development. *Genome Res.* 23: 201–216.

436 Tian, X., Strassmann, J.E., and Queller, D.C. 2013. Dictyostelium Development Shows a Novel Pattern of
437 Evolutionary Conservation. *Mol. Biol. Evol.* 30: 977–984.

438 Tokita, M., and Kuratani, S. 2001. Normal Embryonic Stages of the Chinese Softshelled Turtle
439 *Pelodiscus sinensis* (Trionychidae). *Zoolog. Sci.* 18: 705–715.

440 True, J.R., and Haag, E.S. 2001. Developmental system drift and flexibility in evolutionary trajectories.
441 *Evol. Dev.* 3: 109–119.

442 Vogel, C., and Marcotte, E.M. 2012. Insights into the regulation of protein abundance from proteomic and
443 transcriptomic analyses. *Nat. Rev. Genet.* 13: 227–232.

444 Wagner, G.P. 1989. The Biological Homology Concept. *Annu. Rev. Ecol. Syst.* 20: 51–69.

445 Wagner, G.P. 2007. The developmental genetics of homology. *Nat Rev Genet* 8: 473–479.

446 Wang, Z., Pascual-Anaya, J., Zadissa, A., Li, W., Niimura, Y., Huang, Z., Li, C., White, S., Xiong, Z.,
447 Fang, D., et al. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the
448 development and evolution of the turtle-specific body plan. *Nat Genet* 45: 701–706.

449 Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R.,
450 Hix, D., Kenyon, R., et al. 2013. PATRIC, the bacterial bioinformatics database and analysis resource.
451 *Nucleic Acids Res.* 42: D581–D591.

452 Wilkins, A.S. 2002. The Evolution of Developmental Pathways (Sunderland, MA: Sinauer Associates,
453 Inc).

454 Wilson, C.L., and Miller, C.J. 2005. Simpleaffy: a BioConductor package for Affymetrix Quality Control
455 and data analysis. *Bioinformatics* 21: 3683–3685.

456 Wu, L., Ferger, K.E., and Lambert, J.D. 2019. Gene Expression Does Not Support the Developmental
457 Hourglass Model in Three Animals with Spiralian Development. *Mol. Biol. Evol.* 36: 1373–1383.

458 Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., et al.
459 2013. Genetic programs in human and mouse early embryos revealed by single-cell
460 RNA[thinsp]sequencing. *Nature* 500: 593–597.

461 Yanai, I., Peshkin, L., Jorgensen, P., and Kirschner, M.W. 2011. Mapping Gene Expression in Two
462 Xenopus Species: Evolutionary Constraints and Developmental Flexibility. *Dev. Cell* 20: 483–496.

463 Yang, H., Zhou, Y., Gu, J., Xie, S., Xu, Y., Zhu, G., Wang, L., Huang, J., Ma, H., and Yao, J. 2013. Deep
464 mRNA Sequencing Analysis to Capture the Transcriptome Landscape of Zebrafish Embryos and Larvae.
465 *PLoS One* 8: e64058.

466 Young, R.L., Ferkin, M.H., Ockendon-Powell, N.F., Orr, V.N., Phelps, S.M., Pogány, Á., Richards-
467 Zawacki, C.L., Summers, K., Székely, T., Trainor, B.C., et al. 2019. Conserved transcriptomic profiles
468 underpin monogamy across vertebrates. *Proc. Natl. Acad. Sci.* 116: 1331 LP – 1336.

469 Young, R.L., and Hofmann, H.A. 2019. Reply to Jiang and Zhang: Parallel transcriptomic signature of
470 monogamy: What is the null hypothesis anyway? *Proc. Natl. Acad. Sci.* 116: 17629 LP – 17630.

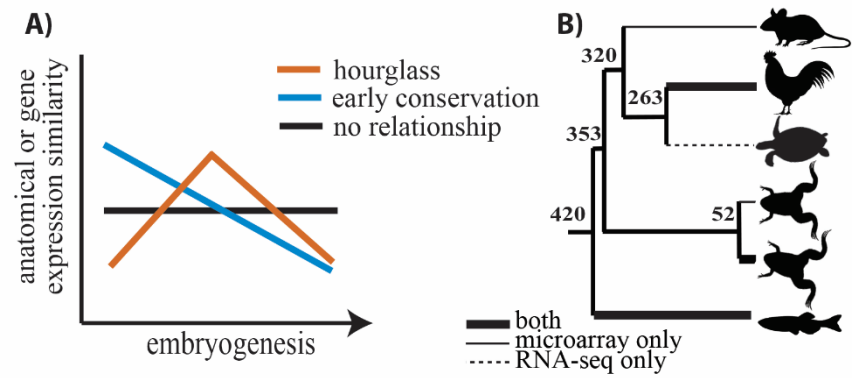
471 Young, R.L., and Wagner, G.P. 2011. Why ontogenetic homology criteria can be misleading: lessons
472 from digit identity transformations. *J. Exp. Zool. B. Mol. Dev. Evol.* 316B: 165–170.

473 Zalts, H., and Yanai, I. 2017. Developmental constraints shape the evolution of the nematode mid-
474 developmental transition. *Nat. Ecol. Evol.* 1: 113.

475

476

477 **FIGURES**



478

479 **Figure 1.** Anatomical and gene expression similarity predicted under different models of conservation
480 through embryogenesis (A). Late conservation where similarity increases through embryogenesis and
481 inverse hourglass where similarity decreases mid-embryogenesis and increases early and late
482 embryogenesis are not shown. Expression conservation was assessed using 186 publicly available
483 microarray and RNA-seq datasets through embryogenesis across a phylogeny of six vertebrate species
484 (B). Diverge times at each node are shown in millions of years.

485

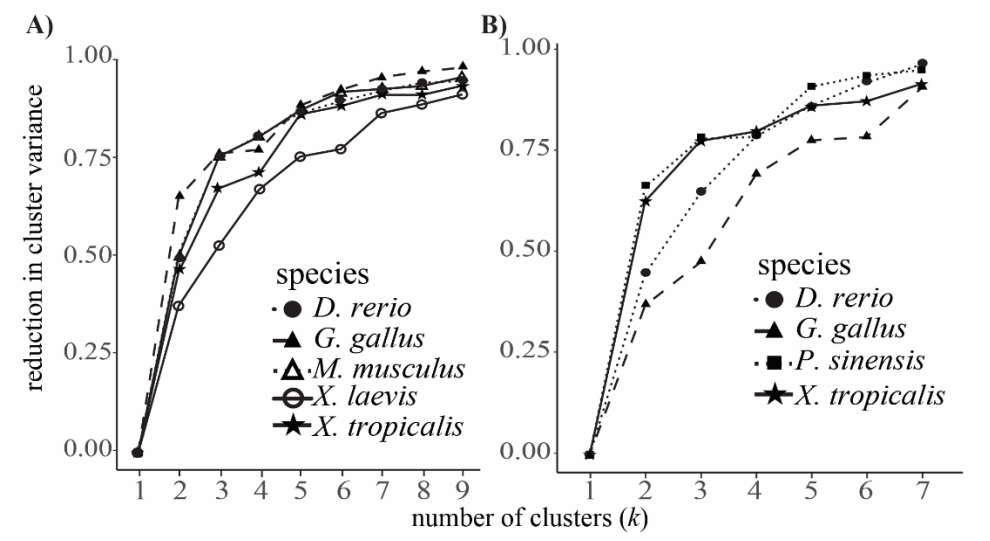
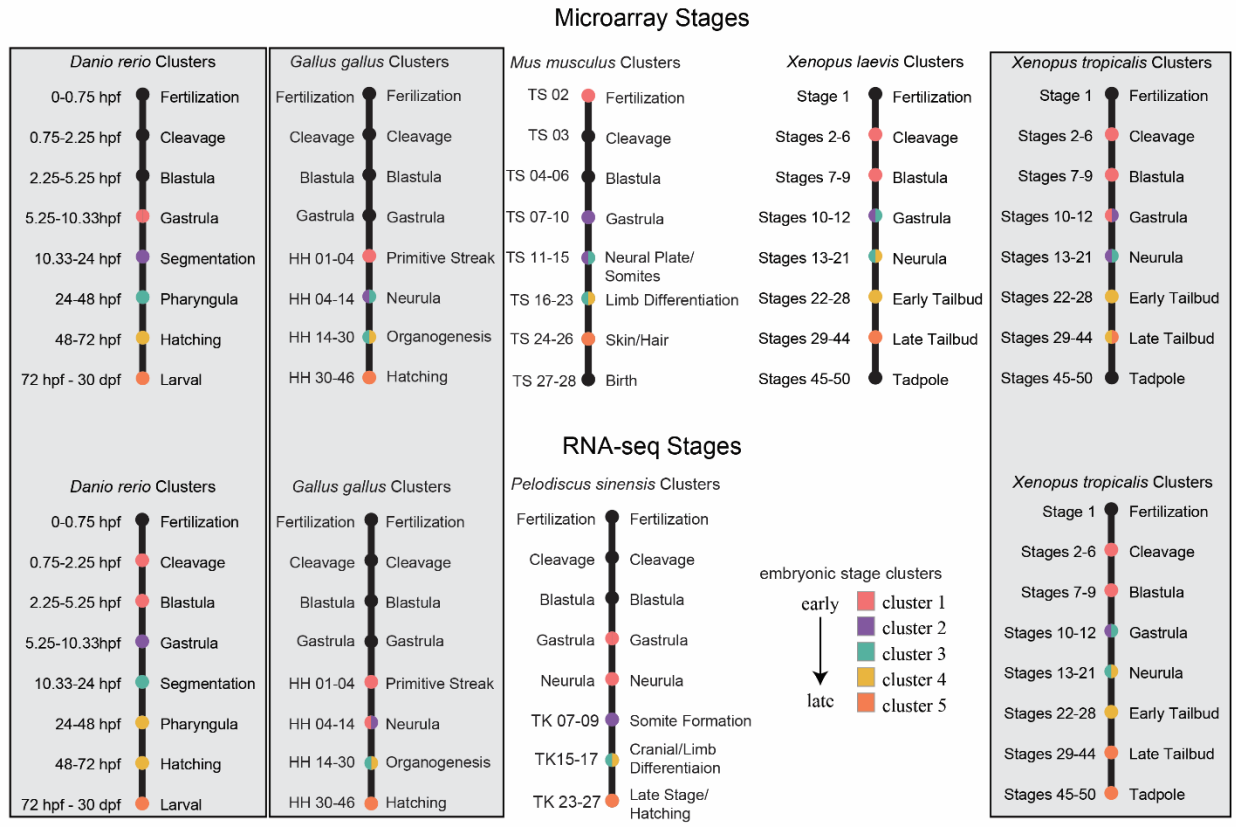


Figure 2. *k*-means clustering of microarray (A) and RNA-seq (B) data sets. Reduction of within cluster variance increases as the number of cluster (*k*) increases. Gains asymptote at approximately $k = 5$.

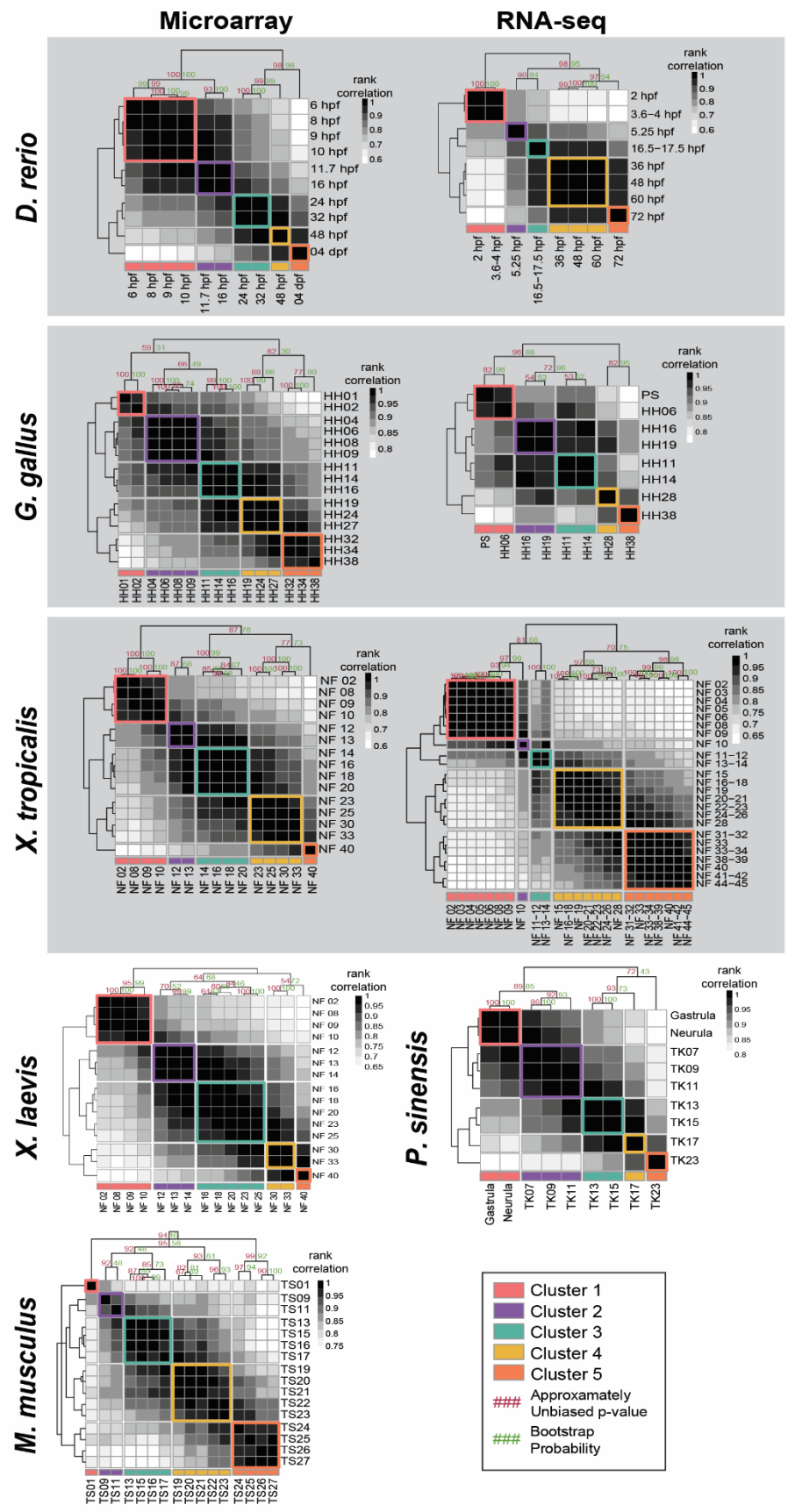
491



492

493 **Figure 3.** Overlap of embryonic stage cluster and major developmental events. Variation in sampling and
494 heterochrony among species result in differences of developmental events captured by the available data
495 across species and sequencing platforms. Gray boxes link species with expression data from both
496 microarray (top) and RNA-seq (bottom) platforms.

497



499 **Figure 4.** Spearman Rank correlations were used to group stages into five clusters. Shown are all pairwise
500 correlations of stages for all species and both sequencing platforms. Grouping of stages is in indicated color.

501

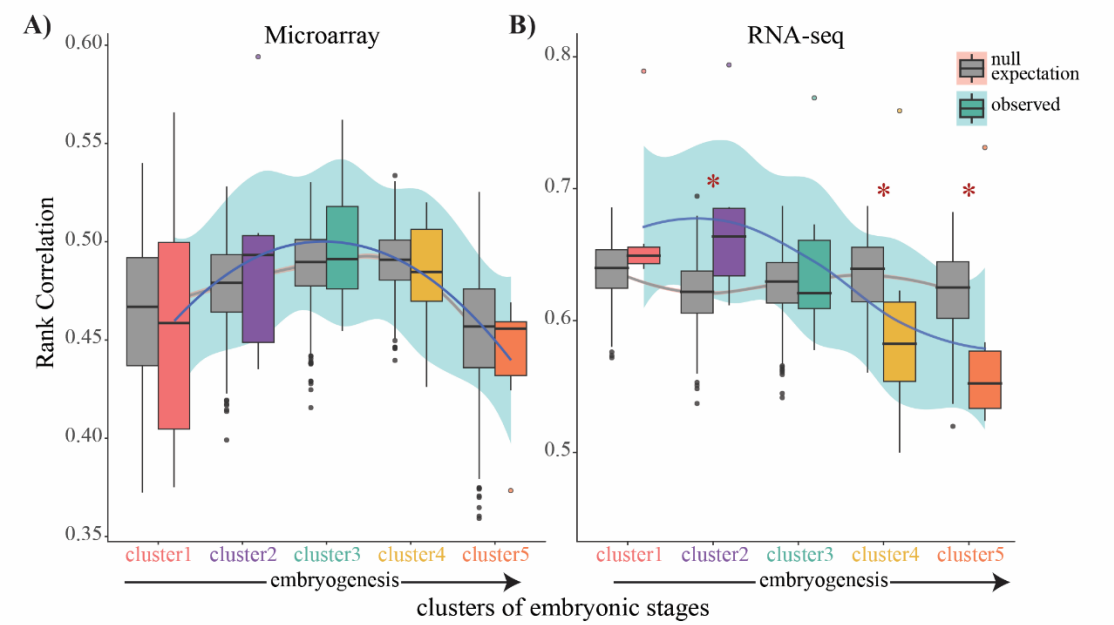


Figure 5. Spearman rank correlations for pairwise comparisons of species at each cluster of embryogenesis for microarray (A) and RNA-seq data (B). Gene expression correlations (as a measure of conservation) differ through embryogenesis for both microarray and RNA-seq and could be perceived as supporting the hourglass, only the RNA-seq data showed a pattern that differed from the null expectation. Colored boxes indicate observed correlations; grey boxes indicate rank correlations of after permutation analysis randomizing stage association with cluster. Asterisks indicate that the observed median correlation differs from the null expectation at $p < 0.001$.

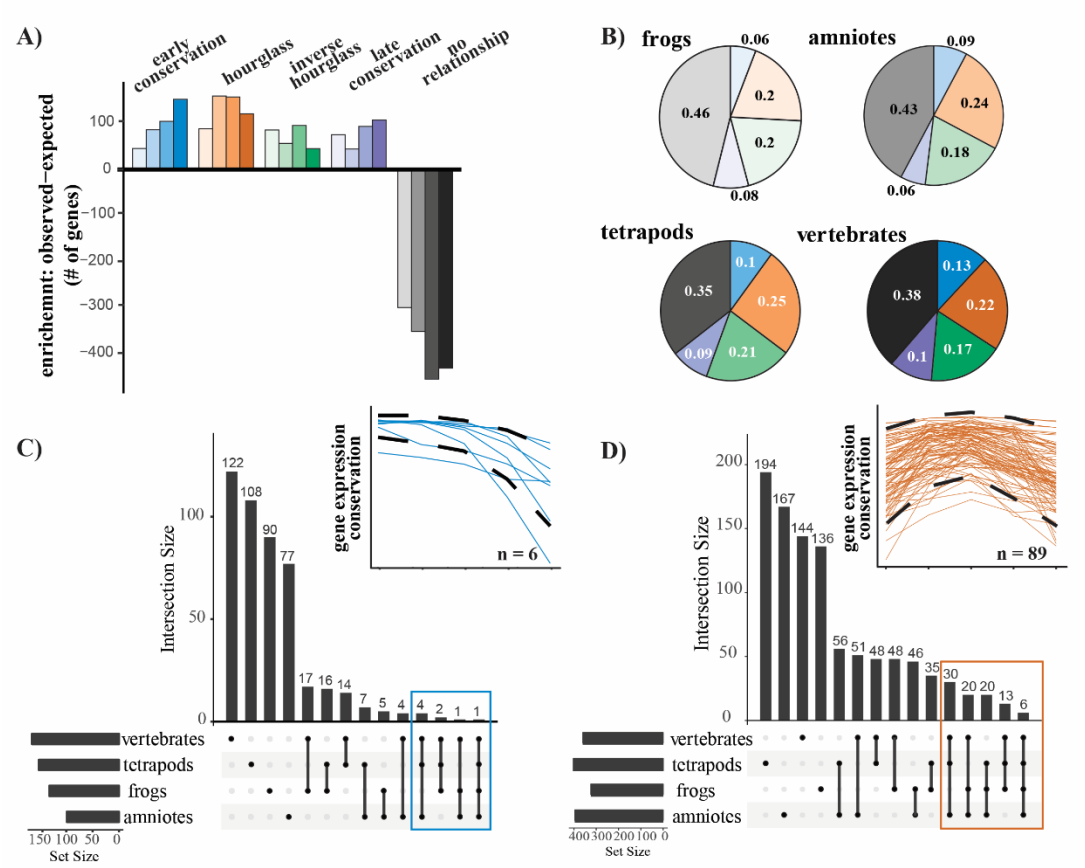
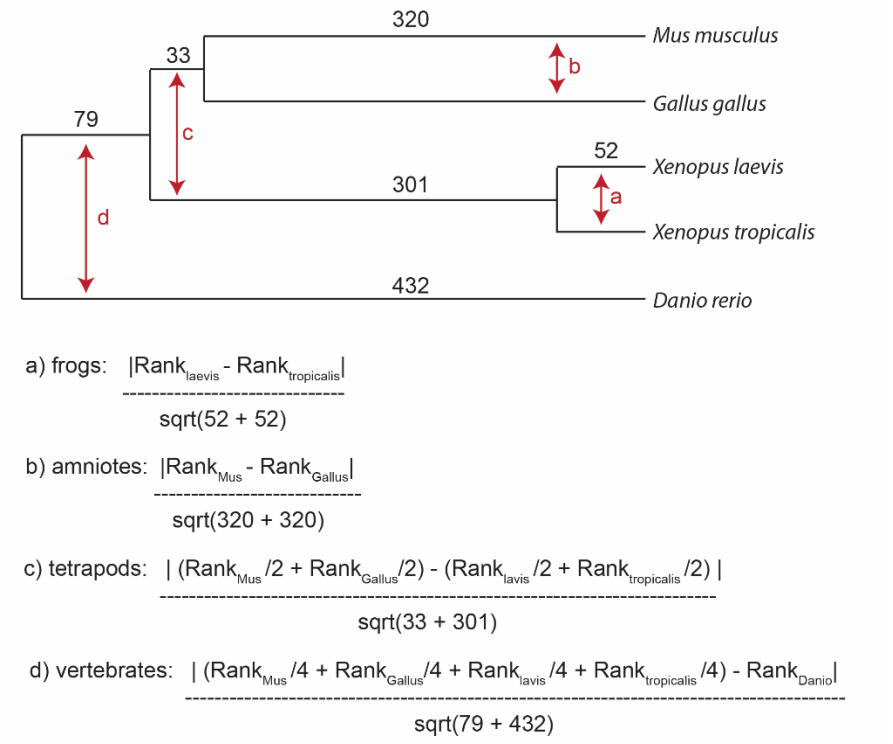


Figure 6. Enrichment and conservation of gene expression patterns across evolutionary nodes. We found an enrichment of genes associated with all gene conservation patterns and a depletion of genes with no relationship between expression conservation across embryogenesis (A). The proportion of total one-to-one genes varies across the nodes (B). The intersection of genes exhibiting early conservation (C) and hourglass patterns (D) across nodes of the phylogeny. Intersections for inverse hourglass, late conservation, and no relationship are provided in Supplementary Fig. 2.

519 **SUPPLEMENTARY FIGURES**

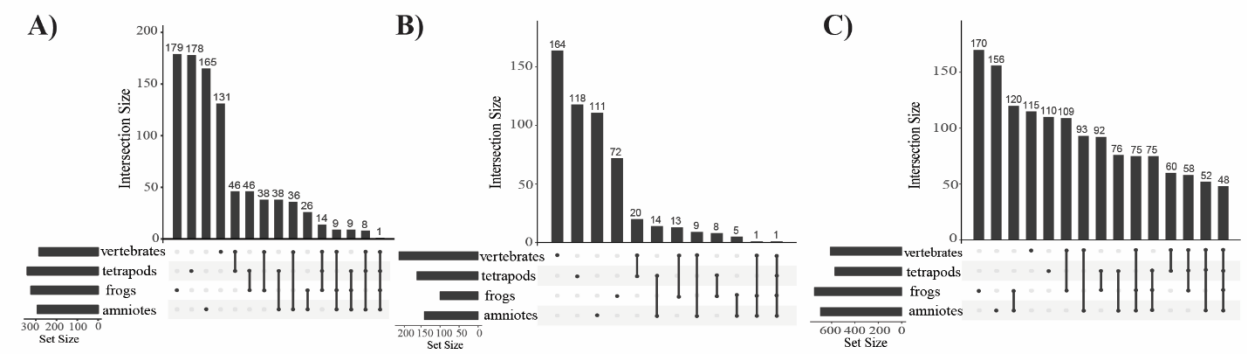
520



521

522 **Supplementary Figure 1.** To characterize expression conservation of each gene at each embryonic stage
523 cluster at each node of the phylogeny we calculated the similarity in rank expression scaled by the
524 divergence times of the included species.

525



Supplementary Figure 2. Intersections of genes exhibiting inverse hourglass (A), late conservation (B), and no relationship (C) across evolutionary nodes.