Developmental constraints on genome evolution in four bilaterian model species

Jialin Jiu[1,2], Marc Robinson-Rechavi[1,2*]

1 Department of Ecology and Evolution, University of Lausanne, Switzerland

2 Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland

* corresponding author: marc.robinson-rechavi@unil.ch

Abstract

Developmental constraints on genome evolution have been suggested to follow either an early conservation model or an "hourglass" model. Both models agree that late development (after the morphological 'phylotypic' period: both late and post-embryonic development) diverges between species, but debate on which developmental period is the most conserved. Here, based on a modified "Transcriptome Age Index" approach, we analyzed the constraints acting on three evolutionary traits of protein coding genes (strength of purifying selection on protein sequences, phyletic age, and duplicability) in four species: *C. elegans*, *D. melanogaster*, *D. rerio* and *M. musculus*. In general, we found that both models can be supported from different genomic properties. The evolution of phyletic age and of duplicability follow an early conservation model in all species, but sequence evolution follows different models in different species: an hourglass model in both *D. rerio* and *M. musculus,* and an early conservation model in *D. melanogaster*. Further analyses indicate that stronger purifying selection on sequences in the early development (before the morphological 'phylotypic' period) of *D. melanogaster* and in the middle development (the morphological 'phylotypic' period) of *D. rerio* are driven by temporal pleiotropy of these genes. In addition, inspired by the "new genes out of the testis" hypothesis, we report evidence that expression in late development is enriched with retrogenes. This implies that expression in late development could facilitate transcription, and eventually acquisition of function, of new genes. Thus, it provides a model for why both young genes and high duplicability genes trend to be expressed in late development. Finally, we suggest that dosage imbalance could also be one of the factors that cause depleted expression of young genes and of high duplicability genes in early development, at least in *C. elegans*.

Introduction

Evolutionary changes in the genome can cause changes in development, which are subject to natural selection. This leads developmental processes to constrain genome evolution. More precisely, selection on the output of development affects the genomic elements active in development. Currently, based on morphological similarities during development, two popular models have been proposed to bridge developmental and evolutionary biology. The early conservation model, modified from the “third laws” of Von Baer (1828) (as cited in Kalinka and Tomancak 2012), suggested that the highest morphological similarities among species from the same phylum occurred in early development, followed by a progressive evolutionary divergence over ontogeny. It should be noted that Von Baer in fact based his observations on post-gastrulation embryos (Kalinka and Tomancak 2012; Abzhanov 2013). The "developmental burden" concept was proposed to explain this model. It suggested that the development of later stages is dependent on earlier stages, so that higher conservation should be found in the earlier stages of development (Garstang 1922; Riedl 1978) (as discussed in Irie and Kuratani 2014). Based on renewed observations in modern times, however, Duboule (1994) and Raff (1996) proposed the developmental "hourglass model". This model suggested that a "phylotypic period" (Richardson 1995) in middle development has higher morphological similarities than early or late development. Several mechanisms have been proposed to explain this observation. Duboule (1994) proposed that it may be due to co-linear Hox cluster gene expression in time and space. Raff (1996) suggested a high inter-dependence in signaling among developmental modules in middle development. Galis and Metz (2001) also highlighted the high number of interactions at this period, although Comte et al. (2010) did not find any molecular evidence for these interactions. It is worth noting that, moreover, the hourglass model was not supported by a comprehensive study of vertebrate embryonic morphology variation (Bininda-Emonds et al. 2003). A number of alternatives have been proposed, for example the “adaptive penetrance model” (Richardson et al. 1997) and the “ontogenetic adjacency model” (Poe and Wake 2004).

Both models have been supported by recent genomic level studies based on different properties (such as expression divergence, sequence divergence, duplication, or

phyletic age), different species, and different analysis methods. Concerning expression divergence, interestingly, all studies are consistent across different species and research groups (Kalinka et al. 2010; Irie and Kuratani 2011; Yanai et al. 2011; Levin et al. 2012; Wang et al. 2013; Gerstein et al. 2014; Ninova et al. 2014; Zalts et al. 2017). All of them suggested that middle development has the highest transcriptome conservation. On the other hand, when animals are compared between different phyla, middle development appears to have the highest divergence (Levin et al. 2016) (but see Dunn et al. 2017). From other properties, however, the results are inconclusive based on different methods (Castillo-Davis and Hartl 2002; Cutter and Ward 2005; Davis et al. 2005; Hazkani-Covo et al. 2005; Hanada et al. 2007; Irie and Sehara-Fujisawa 2007; Cruickshank and Wade 2008; Roux and Robinson-Rechavi 2008; Artieri et al. 2009; Domazet-Loso and Tautz 2010; Quint et al. 2012; Piasecka et al. 2013; Cheng et al. 2015; Drost et al. 2015).

Generally, the methods used to measure developmental constraints at the genomics level can be divided into three categories: proportion based analysis, module analysis, and transcriptome index analysis.

Proportion based analysis consists in testing the proportion of genes with a given property within all expressed genes (Roux and Robinson-Rechavi 2008). The method is less used following the emergence of accurate transcriptome-scale data, since it does not take into account the contributions of expression abundance.

Module analysis consists in studying evolutionary properties of distinct sets of genes (modules) which are specifically expressed in groups of developmental stages (Piasecka et al. 2013). This method can avoid problems caused by genes expressed over all or a large part of development. For example, trends might be diluted by highly expressed housekeeping genes, which contribute to the average expression at all developmental stages. However, this approach can only measure the developmental constraints for a specific subset of genes, instead of considering the composition of the whole transcriptome.

Transcriptome index analysis is a weighted mean: the mean value of an evolutionary parameter is weighted by each gene's expression level (Domazet-Loso and Tautz 2010). This method has the benefit of detecting evolutionary constraints on the whole transcriptome, but patterns can be driven by a subset of very highly expressed genes, or even by a few outliers, because the difference between highly and lowly expressed genes can span several orders of magnitude. For instance, Domazet-Loso and Tautz (2010) reported that transcriptomes of middle development stages of *D. rerio* have a higher proportion of old genes than transcriptomes of early and late development stages, using the transcriptome age index. However, Piasecka et al. (2013) re-analyzed the same data and reported that the highest proportion of old genes was in transcriptomes of early development stages, once a standard log-transformation of microarray signal intensities was done, a result confirmed by module analysis and proportion based analysis.

In addition, several statistical methods have been proposed to distinguish the hourglass model from the early conservation model.

The parabolic test is based on fitting both first degree and second degree polynomial models (Roux and Robinson-Rechavi 2008). The hourglass model is supported if the parabolic function provides a significantly better fit and its minimum corresponds to middle development. This method has been criticized for being too specific and insensitive to other non-parabolic hourglass patterns (Drost et al. 2015).

The flat line test simply tests whether variance of transcriptome indexes across development is significantly higher than variance from random samples (Domazet-Loso and Tautz 2010; Quint et al. 2012). But a significant difference does not necessarily imply the existence of an hourglass pattern (Drost et al. 2015).

Since these two methods are either too strict or without power to distinguish hourglass model, Drost et al. (2015) proposed a "reductive hourglass test" which focuses on testing the presence of an hourglass pattern: high-low-high. For this, the development can be divided into three modules (early, phylotypic, and late), based on the known phylotypic period from morphological studies. Then, a permutation method is used to

test whether the mean value in the phylotypic module is significantly lower than in early and late modules.

Overall, the transcriptome index analysis should be the best method to measure developmental constraints on the whole transcriptome, if care is taken to properly transform the expression values, as well as evolutionary parameters if necessary. Moreover, the reductive hourglass test should be used to objectively test the hourglass model, alone or in combination with other methods.

Because previous studies used different methodologies, and few studies adopted log-transformed transcriptome index analysis, their conclusions cannot be compared consistently, making a biological conclusion concerning developmental constraints across species and features difficult. What's more, while many studies focus on distinguishing between early conservation model and hourglass conservation model, we still know very little of the factors driving these patterns.

To measure developmental constraints on genome evolution, we calculated transcriptome indexes over the development of four species (*C. elegans, D. melanogaster, D. rerio* and *M. musculus*), for three evolutionary parameters: strength of purifying selection on coding sequences (omega0: purifying selection dN/dS from the branch-site model; omega: global dN/dS; see Methods), phyletic age, and duplicability (paralog number). All expression levels were log-transformed before use, as were two evolutionary parameters, strength of purifying selection and paralog number. For *C. elegans*, dN/dS (both omega0 and omega) was not reliably estimated, with no data in the Selectome database (Moretti et al. 2014) and very high values of estimated synonymous distances (dS) from Ensembl Metazoa (Kersey et al. 2016) (Figure S1); thus we did not include this parameter in the study of *C. elegans*. Our analysis of the developmental constraints on sequence evolution and phyletic age of *D. melanogaster* differ from that of Drost et al. (2015), although they also used a transcriptome index. Firstly, they did not apply log-transformation to expression data, and thus the patterns they observed might be driven by the subset of genes with very high expression (Piasecka et al. 2013). Secondly, their analysis of sequence evolution was based on sequence divergence strata. They assigned genes into ten discrete

deciles according to their omega values, thus losing information from this continuous variable.

In general, we found results consistent with early conservation for phyletic age and paralog number in the four species, but different models for sequence evolution in different phyla.

Results and discussion

Variation of evolutionary transcriptome indexes across development

In order to objectively distinguish the hourglass model from the early conservation model, we used a permutation test method similar to that of Drost et al. (2015) (see Methods). For all parameters considered the highest divergence is observed in late development. Thus a significant *p*-value for lower divergence in middle *vs*. early development supports the hourglass model, whereas a lack of significance supports the early conservation model. We consider early conservation to cover both stronger conservation in early than middle development, and similar strong conservation over early and middle development.

For the transcriptome index of omega0 (Transcriptome Divergence Index: TDI), we observed different patterns in different species (Figure 1). In *D. melanogaster,* there is an early conservation pattern of TDI: similar low TDI in early and middle development, high TDI in late development. In *D. rerio,* however, there is an hourglass pattern of TDI: medium TDI in early development, low TDI in middle development, and high TDI in late development. In *M. musculus,* the pattern resembles an hourglass like pattern, but the *p*-value is not significant. In addition, for *D. melanogaster* and *M. musculus,* we also calculated TDI based on omega (Figure S2). In *D. melanogaster*, the results are consistent with those based on omega0. In *M. musculus,* the TDI based on omega has a significant *p*-value for the hourglass.

For the transcriptome indexes of phyletic age (Transcriptome Age Index: TAI) and of paralog number (Transcriptome Paralog Index: TPI), in all four species, we observed that genes with higher duplicability and younger phyletic age trend to be expressed at later developmental stages, which corresponds to the early conservation pattern

(Figure 1). Because testis expression is enriched with new genes (Kaessmann 2010), the higher proportion of young genes in late development might be driven by testis. In order to test this, we excluded testis genes from TAI analysis for both *D. melanogaster* and *M. musculus*, where the information of testis gene expression was available. Results were essentially unchanged (Figure S3), indicating that the early conservation pattern of TAI is not caused by late expression of testis genes.

In *D. melanogaster*, we did not confirm the results of Drost et al. (2015) for either phyletic age or sequence evolution. For TAI, after log-transformation of expression data, we found an early conservation pattern instead of the hourglass pattern which they reported (Figure S4B). It appears that the hourglass pattern of phyletic age in their study is driven by a few highly expressed genes, consistently with our previous observations in *D. rerio* (Piasecka et al. 2013). This is verified by excluding the top 10% most expressed genes and analyzing without log-transformation, as in Drost et al. (2015) (Figure S4C). For TDI, the different patterns could be in part due to distinct measurements of sequence evolution: Drost et al. (2015) analyzed discrete sequence divergence strata, whereas we used continuous values. In *D. rerio*, we also found different patterns of phyletic age relative to Drost et al. (2015) and to Domazet-Loso and Tautz (2010). Again, this appears due to the log-transformation or not of expression data.

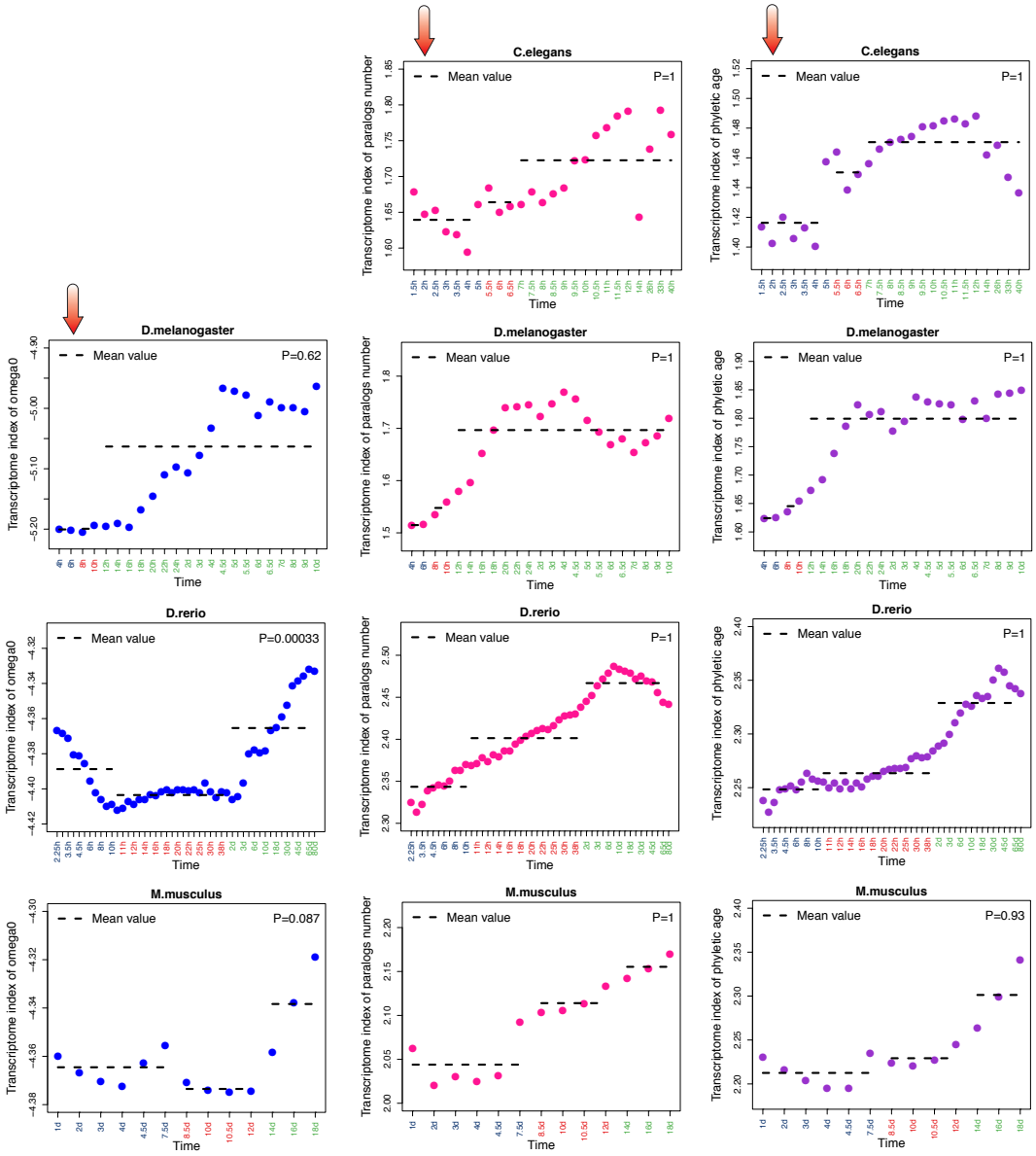

Figure 1: Evolutionary transcriptome indexes (TEI) across development in different species.

Blue, red and green marked time points in the x-axis represent early developmental stages, middle developmental stages and late developmental stages respectively. The direction of the large arrow indicates the strength of developmental constraints. The *p*-values for supporting the hourglass model (permutation test, early *vs*. middle development) are indicated in the top-right corner of each graph.

Overall, these results suggest that genes under strong purifying selection on their protein sequence trend to be expressed in early development in *D. melanogaster*, but in middle development for two vertebrates; it remains to be seen how much these

observations extend to more arthropods or chordates. They also extend our previous observations that genes expressed in earlier stage have a lower duplicability and older age (Roux and Robinson-Rechavi 2008; Piasecka et al. 2013)

Pleiotropic effect

Several models have been proposed to explain why some developmental stages are more conserved than others. Garstang and Riedl proposed that later stages of embryogenesis rely on primary elements established during earlier stages, and thus earlier stages are more conserved than later stages (Garstang 1922; Riedl 1978) (as discussed in Irie and Kuratani 2014). Duboule (1994) and Raff (1996) proposed that changes in the network of middle development could have detrimental effects because of high inter-dependence, leading to evolutionary conservation. In all models, a common point is that high conservation is caused by selection against deleterious pleiotropic effects of mutations. This implies that higher sequence conservation in early or middle developmental stages is caused by higher pleiotropy in these stages, because pleiotropy is one of the major factors that constrain sequence evolution (Fraser et al. 2002).

In order to test this hypothesis, we used one type of development related pleiotropic effect: temporal pleiotropy (Artieri et al. 2009) (expression breadth across development). This is similar to spatial pleiotropy (Larracuente et al. 2008; Kryuchkova-Mostacci and Robinson-Rechavi 2015) (expression breadth across tissues) or connective pleiotropy (Fraser et al. 2002) (protein-protein connectivity). The more stages a gene is expressed in, the more traits it could affect, so it could be under stronger evolutionary constraints (Wagner and Zhang 2011). For *D. melanogaster* and *C. elegans*, we defined FPKM>1 as expressed. For *D. rerio,* we set genes with microarray signal rank in top 70% as expressed. Since the time series transcriptome data sets of *M. musculus* come from two different research groups, with signs of batch effects (Figure S5), we did not integrate it into this analysis.

Firstly, we calculated the proportion of potentially pleiotropic genes, as expressed in more than 50% of development stages. We found pleiotropic genes enriched in the middle development of *D. rerio*, but in both early and middle development of *D.*

melanogaster and *C. elegans* (Figure 2). We also found similar patterns when we define pleiotropic genes as expressed in more than 70% of development stages (Figure S6). For *D. rerio*, in addition, we observed consistent results based on setting expressed genes as microarray signal rank in the top 90% or 50% (Figure S7). Because RNA-seq is more efficient to detect specifically expressed genes than microarrays (Kryuchkova-Mostacci and Robinson-Rechavi 2016a), for both *D. melanogaster* and *C. elegans* with RNA-seq data, we also calculated a stage specificity index (Tau) of gene expression, based on the tissue specificity index (Yanai et al. 2005; Kryuchkova-Mostacci and Robinson-Rechavi 2016a). Genes with lower Tau are expressed across more developmental stages with little variation in level of expression. With a transcriptome index of Tau (Transcriptome Tau Index: TTI), we observed very similar results as well (Figure S8). Similar observations of higher temporal pleiotropy for genes in middle development in vertebrates were recently reported by Hu et al (in press; pers. comm.).

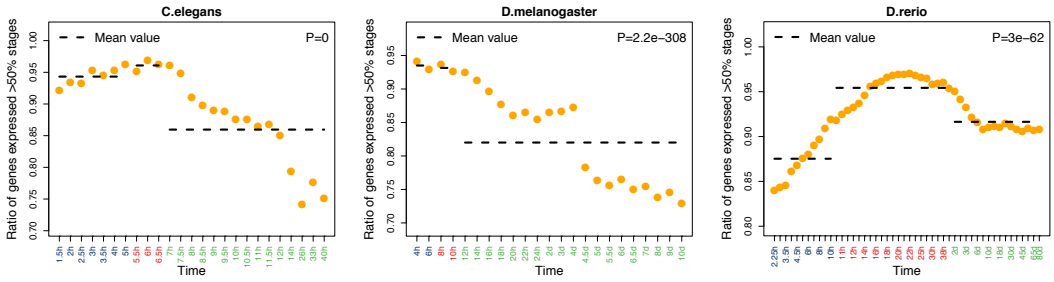

Figure 2: Proportion of temporal pleiotropic genes across development.
Blue, red and green marked time points in the x-axis represent early developmental stages, middle developmental stages, and late developmental stages respectively. The *p*-values from chi-square goodness of fit test are indicated in the top-right corner of each graph. Pleiotropic genes are defined as expressed in more than 50% of stages sampled.

Based on these observations, we further checked whether higher temporal pleiotropic constraint could explain stronger purifying selection on sequence evolution. As expected, we found that pleiotropic genes have lower omega0 than non-pleiotropic genes (Figure S9). Finally, we re-calculated TDI separately for pleiotropic genes and non-pleiotropic genes. In both *D. melanogaster* and *D. rerio*, the pattern observed

over all genes appears to be driven by the pleiotropic genes (Figure 3): early conservation for *D. melanogaster*, and hourglass for *D. rerio*.

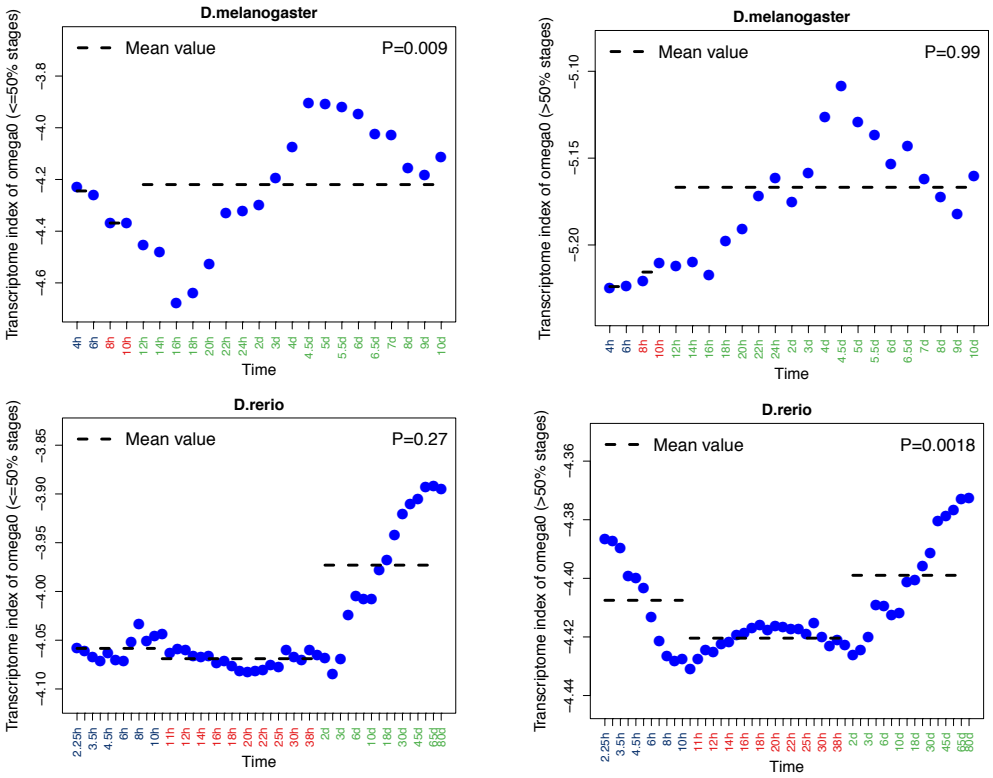

Figure 3: Transcriptome divergence index of omega0 (TDI) across development according to temporal pleiotropy.

Blue, red and green marked time points in the x-axis represent early developmental stages, middle developmental stages and late developmental stages respectively. The *p*-values for supporting the hourglass model are indicated in the top-right corner of each graph.

In summary, it seems that development stages with a higher proportion of broadly expressed genes are under stronger pleiotropic constraint on sequence evolution. And thus, that different sequence constraint models in different species are driven by different distributions of pleiotropic genes.

Higher expression of retrogenes in later development stages

Why do young genes trend to be expressed in late development stages? Inspired by the "new genes out of the testis hypothesis" (Kaessmann et al. 2009; Kaessmann 2010;

Soumillon et al. 2013), we suggest that expression in late development might, like in testis, promote the fixation and functional evolution of new genes. Like testis constitutes the most rapidly evolving organ transcriptome, late development represents the most rapidly evolving stage transcriptome owing to both relaxed purifying selection (Artieri et al. 2009) and increased positive selection (Liu and Robinson-Rechavi 2017). Thus, late development could provide a better environment for the fixation of new genes.

In order to test this, we analyzed the expression of retrogenes across development. Since retrogenes are usually expected to lack regulatory elements, most of them fail to acquire transcription and achieve function (Kaessmann et al. 2009). So, if late development can facilitate the transcription and promote the fixation of retrogenes genes, we should observe higher expression of retrogenes in later developmental stages. Thus we computed the ratio of mean expression of retrogenes to mean expression of non-retrogenes. To display the variation of this ratio across development, we then fitted polynomial models of the first degree and the second-degree. We kept the second-degree polynomial model (parabola) only if it provided a significantly better fit (tested with ANOVA, $p<0.05$). Because retrogenes have higher expression in testis, and testis is already differentiated after middle development, we also excluded testis genes in our analyses for *D. melanogaster* and *M. musculus*. We found that the mean expression of retrogenes is at its maximum in late development (Figure 4), except in *C. elegans*. However, we hesitate to over interpret this inconsistent pattern of *C. elegans*, because we found different patterns by using different data sources for *C. elegans*, whereas results across data sources were consistent for the three other species (Figure S10).

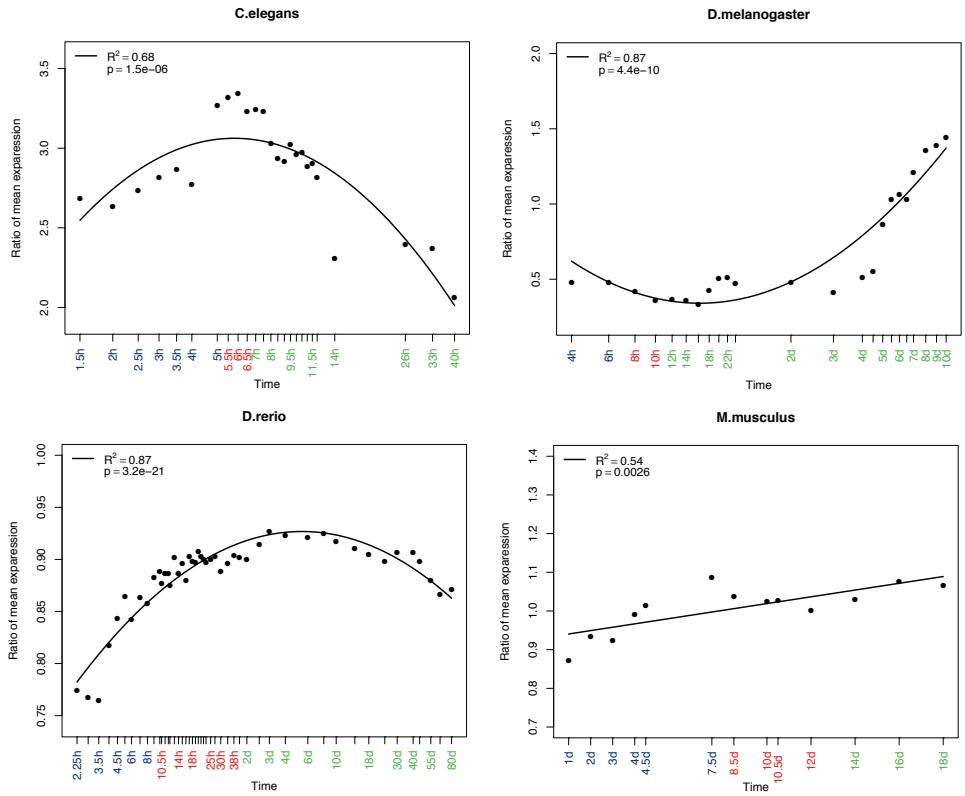

Figure 4: Expression of retrogenes in development.

Each solid circle represents the ratio of the mean expression for retrogenes to the mean expression for non-retrogenes. The ratio is fitted by regression (the first degree of polynomial for *M. musculus* and the second degree of polynomial for other species), whose R^2 and p-value are indicated in the top-left corner of each graph. Blue, red and green marked time points in the x-axis represent early developmental stages, middle developmental stages and late developmental stages respectively. The x-axis for *C. elegans*, *D. rerio* and *D. melanogaster* is in logarithmic scale, while the x-axis for *M. musculus* is in natural scale.

These results confirm that late development could allow more transcription of newly originated gene copies, which usually lack efficient regulatory elements and transcriptional activities. Since the first step to functionality is acquiring transcription, we suggest that the functional acquisition and survival at the beginning of life history for new genes could be promoted by expression in late development. When beneficial mutations come, a subset of these new gene candidates could subsequently obtain beneficial functions in late development and evolve efficient regulatory elements and finally be retained long term in the genome. Thus, the higher proportion of young

genes expressed in later development stages can be in part explained because these stages increased the probability of fixation of new genes. Similarly, based on this transcriptional facilitation model, we suggest that late development could also boost the fixation of duplicates. So, higher duplicability genes trend to be expressed in later development stages. Of course, this is not exclusive with an adaptive scenario that early stages lack opportunities for neo- or sub-functionalization, because of simpler anatomical structures, which could also diminish fixation of duplicates in early development.

Connectivity and dosage imbalance

It has previously been found that, in both *S. cerevisiae* and *C. elegans*, gene duplicability is negatively correlated with protein connectivity (Hughes and Friedman 2005; Prachumwat and Li 2006) which might be explained by dosage balance (Veitia 2002; Papp et al. 2003). Firstly, we checked the relationship of connectivity and duplicability in our datasets. We found, indeed, a negative relationship in *C. elegans* (Figure S11). In contrast, we observed a positive relationship in *M. musculus*, which is consistent with previous mouse results (Liang and Li 2007), although the relation is weak. In *D. melanogaster* and in *D. rerio*, interestingly, there is a strong non-linear pattern of duplicability with connectivity: increasing first, and then decreasing. Secondly, we calculated a transcriptome index of connectivity (Transcriptome Connectivity Index: TCI). In *C. elegans*, earlier developmental stages have higher TCI, which means these stages trend to have higher expression of more connected genes (Figure 5). In *D. melanogaster*, we observed a similar pattern even though there is a trend of increased TCI in Pupae stages. In *D. rerio*, however, there is an hourglass like pattern, although not significant. Interestingly, in *M. musculus*, we detected the inverse of an early conservation pattern, i.e. less connectivity of early expressed genes.

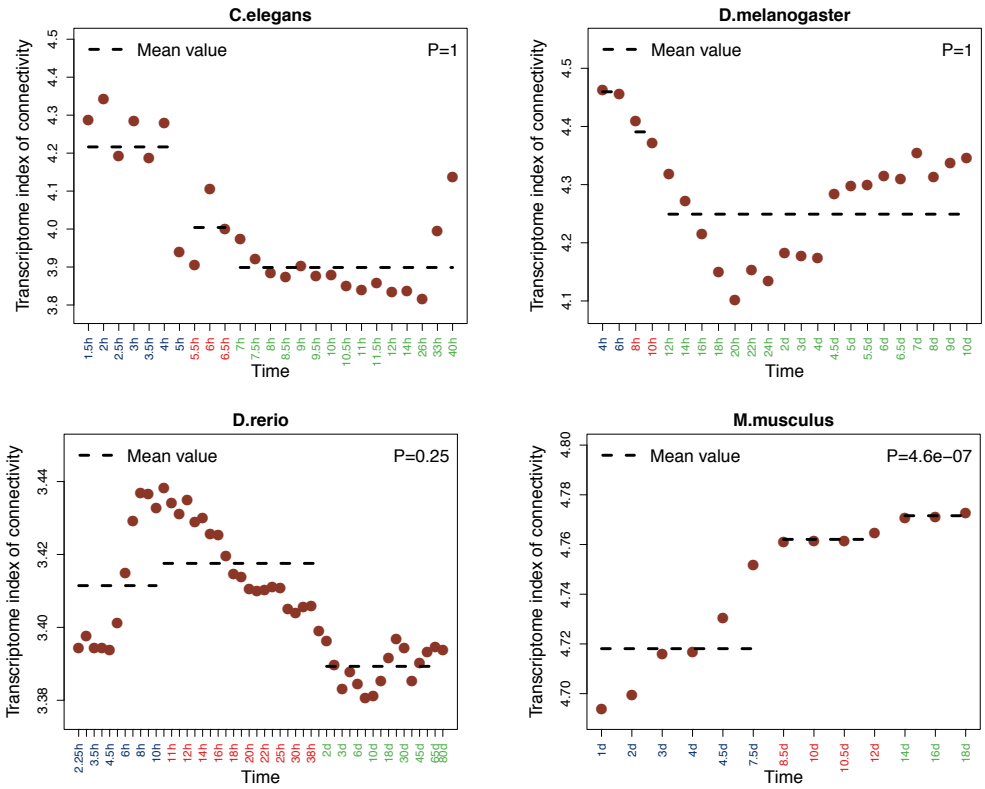

Figure 5: Transcriptome index of connectivity (TCI) across development.
Blue, red and green marked time points in the x-axis represent early developmental stages, middle developmental stages and late developmental stages respectively. Of note, the *p*-value of the *M. musculus* sub-figure doesn't mean it supports the hourglass model, because while middle development has significant higher TCI than early development, late development has even higher TCI than middle development.

These results indicate that, in *C. elegans*, earlier stages trend to express higher connectivity genes, which are less duplicable because more sensitive to dosage imbalance, but that this cannot be generalized to other animals. Because new genes mainly arise from duplication (Kaessmann 2010; Chen et al. 2013), lower duplicability for earlier stages genes could also explain why earlier stages have lower proportion of young genes in *C. elegans*. In addition, the TCI results contradict the hypothesis of a high number of interactions at middle development (Raff 1996; Galis and Metz 2001), consistent with Comte et al. (2010).

Conclusion

Our results concern both patterns and processes of evolution over development. For patterns, we tested the early conservation and hourglass models by using three evolutionary properties: strength of purifying selection, phyletic age and duplicability. Both duplicability and phyletic age support the early conservation model. Less duplicated genes and phyletically older genes are more expressed at earlier stages. The strength of purifying selection on protein sequence supports the early conservation model in *D. melanogaster* but the hourglass model in the vertebrates *D. rerio* and *M. musculus*.

For processes, we investigated the potential causes of the observed patterns. The enrichment of high duplicability genes and young phyletic age genes in late development might be related to a testis-like role of late development that facilitates the expression of retrogenes. The different models of sequence evolution in different species appear to be driven by temporal pleiotropy of gene expression, since temporal pleiotropic genes are enriched in the early development of *D. melanogaster* but in the middle development of *D. rerio*. Finally, in *C. elegans*, connectivity appears to be the main force explaining the observed pattern.

Materials and Methods

Expression data sets

For *D. rerio*, we retrieved the log-transformed and normalized microarray data from our previous study (Piasecka et al. 2013). This data originally comes from (Domazet-Loso and Tautz 2010), which includes 60 stages from egg to adult.

For *M. musculus*, two processed (log-transformed and normalized) microarray data sets were retrieved from Bgee (release 13.1, July 2015; Bastian et al., 2008), a database for gene expression evolution. The first data set contains six stages from zygote to Theiler 6 (Wang et al. 2004), and the second one includes eight stages from Theiler 11 to Theiler 26 (Irie and Kuratani 2011).

For *D. melanogaster* and *C. elegans*, we obtained normalized RNA-seq data from (Li et al. 2014), which originally comes from (Gerstein et al. 2010; Graveley et al. 2011). The *D. melanogaster* data set covers embryo, larva, pupae and adult, including 27

stages. The *C. elegans* data set covers 30 stages from embryo to larval and to adult. All the genes with expression <1 RPKM were set as not expressed (Kryuchkova-Mostacci and Robinson-Rechavi 2016b), replaced by 1.0001 (this value is smaller than the smallest value of expressed genes), and log2 transformed.

For all data sets, we removed stages which precede the maternal to zygote transition (MZT) because these data sets are dominated by maternal transcripts (Tadros and Lipshitz 2009). In addition, we also excluded all adult stages, because we are focusing on developmental processes.

Omega0

The omega0 values were downloaded from Selectome (Moretti et al. 2014), a database of positive selection based on branch-site model (Zhang et al. 2005). Selectome excludes ambiguously aligned regions before model fitting. Omega0 is the dN/dS ratio (dN is the rate of non-synonymous substitutions, dS is the rate of synonymous substitutions) of the subset of codons which have evolved under purifying selection according to the branch-site model. We used omega0 from the Clupeocephala branch, the Murinae branch, and the Melanogaster group branch for *D. rerio*, *M. musculus*, and *D. melanogaster*, respectively. One gene could have two omega0 values in the focal branch because of duplication events. In this case, we keep the value of the branch following the duplication and exclude the value of the branch preceding the duplication.

Omega

The rate of non-synonymous substitutions dN and the rate of synonymous substitutions dS were retrieved from either Ensembl release 84 (for *M. musculus*) (Yates et al. 2016) or Ensembl Metazoa release 34 (for *D. melanogaster* and *C. elegans*) (Kersey et al. 2016), using BioMart (Kinsella et al. 2011). For *D. melanogaster*, the dN and dS values were calculated pairwise using one-to-one orthologs in *D. simulans*. For *M. musculus*, the dN and dS values were calculated pairwise using one-to-one orthologs in *R. norvegicus*. For *C. elegans*, the dN and dS values were calculated pairwise using one-to-one orthologs in *C. briggsae, C. brenneri*, or *C. remanei*.

phyletic age data

Phyletic ages were retrieved from Ensembl version 84 (Yates et al. 2016) using the Perl API. For each gene, we browsed its gene tree from the root and dated it by the first appearance. We assigned the oldest genes with phyletic age value of 1 and the youngest genes with the highest phyletic age value. So, genes can be split into discrete "phylostrata" by phyletic age. We classified 3 phylostrata, 4 phylostrata, 9 phylostrata and 18 phylostrata respectively for *C. elegans*, *D. melanogaster*, *D. rerio* and *M. musculus*.

The number of paralogs

We retrieved the number of paralogs from Ensembl release 84 (Yates et al. 2016) using BioMart (Kinsella et al. 2011).

Retrogene data

For each species, we downloaded retrogenes from two different resources in order to maximize the robustness of results. For *C. elegans,* we retrieved 83 retrogenes from (Zhou et al. 2015) and 51 retrogenes from (Mahmood 2010). For *D. melanogaster,* we retrieved 72 retrogenes from retrogeneDB (Kabza et al. 2014) and 50 retrogenes from (Bai et al. 2007). For *D. rerio* we retrieved 156 retrogenes from (Fu et al. 2010) and 16 genes from retrogeneDB (Kabza et al. 2014). For *M. musculus* we retrieved 268 retrogenes from retrogeneDB (Kabza et al. 2014) and 134 retrogenes from (Potrzebowski et al. 2008). The main figure for retrogene expression across development comes from the data sets with the higher number of retrogenes for each species. The supplementary figure of retrogene expression generated by the data sets with a lower number of retrogenes.

Connectivity data

We retrieved connectivity (protein-protein interactions) data from the OGEE database (Chen et al. 2012).

Transcriptome index analysis for different evolutionary parameters

The TEI (transcriptome evolutionary index) is calculated as:

$$\mathrm{TEI}_s = \frac{\sum_{i=1}^{n} E_i e_{is}}{\sum_{i=1}^{n} e_{is}},$$

where s is the developmental stage, E_i is the relevant evolutionary parameter (omega0, omega, paralog number, phyletic age, stage specificity, or protein connectivity) of gene i, n is the total number of genes, and e_{is} is the log-transformed expression level of gene i in developmental stage s. We also log-transformed all parameters except phyletic age, since the difference of phyletic age is within only two orders of magnitude. Because values of 0 are not manageable with log-transformation, we added the smallest non-zero value to each element before log-transformation.

Permutation test

We first assigned all development stages into three broad development periods (early, middle, and late) based on previous morphological and genomic studies. Next, we calculated the difference of mean transcriptome indexes between the early module and the middle module ($\Delta e\text{-}m$). Then, we randomly sampled the values of the relevant parameter (omega0, omega, paralog number, phyletic age, stage specificity or protein connectivity) from the original data set 10,000 times without replacement. Finally, we approximated a normal distribution for $\Delta e\text{-}m$ based on 10,000 $\Delta e\text{-}m$ values computed from the permutated samples. The p-value of the hourglass model *vs*. the early conservation model for each parameter is the probability of a randomly sampled $\Delta e\text{-}m$ exceeding the observed $\Delta e\text{-}m$. For protein connectivity, the p-value of the hourglass model is the probability that a randomly sampled $\Delta e\text{-}m$ lower than the observed $\Delta e\text{-}m$.

Stage specificity index (Tau)

We calculated stage specificity index as:

$$\mathrm{Tau} = \frac{\sum_{i=1}^{n}(1-\hat{x}_i)}{n-1}; \quad \hat{x}_i = \frac{x_i}{\max_{1\le i\le n}(x_i)},$$

where n is number of stages, x_i is the expression of the gene in stage i. This measure is a modified estimation of tissue specificity index of expression (Yanai et al. 2005; Kryuchkova-Mostacci and Robinson-Rechavi 2016a) and has already been applied to calculate stage specificity (Tian et al. 2013). This index ranges from zero (broadly expressed genes) to one (genes specific to one stage). All genes that were not expressed in at least one stage were removed from the analysis.

Testis specific genes

Similar to the measure of stage specificity, we calculated tissue specificity for *M. musculus* and *D. melanogaster*. We retrieved processed RNA-seq data of 22 *M. musculus* tissues and 6 *D. melanogaster* tissues from (Kryuchkova-Mostacci and Robinson-Rechavi 2016b). We defined genes with highest expression in testis and with tissue specificity value ≥ 0.8 as testis specific genes.

Acknowledgements

We thank Marie Zufferey and Nadezda Kryuchkova for help with preliminary work. We thank the Bgee team, Barbara Piasecka and Nadezda Kryuchkova for help with data retrieval, Julien Roux for help with the Ensembl API, Kamil Jaron and Andrea Komljenovic for help with programing, members of the Robinson-Rechavi lab for helpful discussions. Part of the computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics. JL is supported by Swiss National Science Foundation grant 31003A_153341 / 1.

Author contributions

JL and MRR designed the work. JL performed the data gathering and analysis. JL and MRR interpreted the results. JL wrote the first draft of the paper. JL and MRR finalized the paper.

References

Abzhanov A. 2013. von Baer's law for the ages: lost and found principles of developmental evolution. Trends Genet. 29:712–722.

Artieri CG, Haerty W, Singh RS. 2009. Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of Drosophila. BMC Biol. 7:42.

Bai Y, Casola C, Feschotte C, Betrán E. 2007. Comparative genomics reveals a

constant rate of origination and convergent acquisition of functional retrogenes in Drosophila. Genome Biol. 8:R11.
Bastian F, Parmentier G, Roux J, Moretti S, Laudet V, Robinson-Rechavi M. 2008. Bgee: Integrating and comparing heterogeneous transcriptome data among species. In. DILS: Data Integration in the Life Sciences. p. 124-131.
Bininda-Emonds ORP, Jeffery JE, Richardson MK. 2003. Inverting the hourglass: quantitative evidence against the phylotypic stage in vertebrate development. Proc. Biol. Sci. 270:341–346.
Castillo-Davis CI, Hartl DL. 2002. Genome evolution and developmental constraint in Caenorhabditis elegans. Mol. Biol. Evol. 19:728-735.
Chen S, Krinsky BH, Long M. 2013. New genes as drivers of phenotypic evolution. Nat. Rev. Genet. 14:645–660.
Chen W-H, Minguez P, Lercher MJ, Bork P. 2012. OGEE: an online gene essentiality database. Nucleic Acids Res. 40:D901-6.
Cheng X, Hui JHL, Lee YY, Wan Law PT, Kwan HS. 2015. A "Developmental Hourglass" in Fungi. Mol. Biol. Evol. 32:1556–1566.
Comte A, Roux J, Robinson-Rechavi M. 2010. Molecular signaling in zebrafish development and the vertebrate phylotypic period. Evol. Dev. 12:144–156.
Cruickshank T, Wade MJ. 2008. Microevolutionary support for a developmental hourglass: gene expression patterns shape sequence variation and divergence in *Drosophila*. Evol. Dev. 10:583–590.
Cutter AD, Ward S. 2005. Sexual and Temporal Dynamics of Molecular Evolution in C. elegans Development. Mol. Biol. Evol. 22:178–188.
Davis JC, Brandman O, Petrov DA. 2005. Protein evolution in the context of Drosophila development. J. Mol. Evol. 60:774–785.
Domazet-Loso T, Tautz D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature 468:815–818.
Drost H-G, Gabel A, Grosse I, Quint M. 2015. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. Mol. Biol. Evol. 32:1221–1231.
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. Development 1994:135–142.
Dunn CW, Zapata F, Munro C, Siebert S, Hejnol A. 2017. Pairwise comparisons are problematic when analyzing functional genomic data across species. bioRxiv. doi:10.1101/107177.
Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary Rate in the Protein Interaction Network. Science 296:750-752.
Fu B, Chen M, Zou M, Long M, He S. 2010. The rapid generation of chimerical genes expanding protein diversity in zebrafish. BMC Genomics 11:657.
Galis F, Metz JA. 2001. Testing the vulnerability of the phylotypic stage: on modularity and evolutionary conservation. J. Exp. Zool. 291:195–204.
Garstang W. 1922. The Theory of Recapitulation: A Critical Re-statement of the Biogenetic Law. J. Linn. Soc. London, Zool. 35:81–101.
Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al. 2010. Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. Science 330:1775–1787.
Gerstein MB, Rozowsky J, Yan K-K, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, et al. 2014. Comparative analysis of the transcriptome

across distant species. Nature 512:445–448.
Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. 2011. The developmental transcriptome of Drosophila melanogaster. Nature 471:473–479.
Hanada K, Shiu S-H, Li W-H. 2007. The nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. Mol. Biol. Evol. 24:2235–2241.
Hazkani-Covo E, Wool D, Graur D. 2005. In search of the vertebrate phylotypic stage: a molecular examination of the developmental hourglass model and von Baer's third law. J. Exp. Zool. B. Mol. Dev. Evol. 304:150–158.
Hughes AL, Friedman R. 2005. Gene Duplication and the Properties of Biological Networks. J. Mol. Evol. 61:758–764.
Irie N, Kuratani S. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. Nat. Commun. 2:248.
Irie N, Kuratani S. 2014. The developmental hourglass model: a predictor of the basic body plan? Development 141:4649-4655.
Irie N, Sehara-Fujisawa A. 2007. The vertebrate phylotypic stage and an early bilaterian-related stage in mouse embryogenesis defined by genomic information. BMC Biol. 5:1.
Kabza M, Ciomborowska J, Maka owska I. 2014. RetrogeneDB--A Database of Animal Retrogenes. Mol. Biol. Evol. 31:1646–1648.
Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res. 20:1313–1326.
Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. Nat. Rev. Genet. 10:19–31.
Kalinka AT, Tomancak P. 2012. The evolution of early animal embryos: conservation or divergence? Trends Ecol. Evol. 27:385–393.
Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P. 2010. Gene expression divergence recapitulates the developmental hourglass model. Nature 468:811–814.
Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, et al. 2016. Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res. 44:D574–D580.
Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database 2011:bar030.
Kryuchkova-Mostacci N, Robinson-Rechavi M. 2015. Tissue-Specific Evolution of Protein Coding Genes in Human and Mouse. PLoS One 10:1–15.
Kryuchkova-Mostacci N, Robinson-Rechavi M. 2016a. A benchmark of gene expression tissue-specificity metrics. Brief. Bioinform bbw008:1-10.
Kryuchkova-Mostacci N, Robinson-Rechavi M. 2016b. Tissue-Specificity of Gene Expression Diverges Slowly between Orthologs, and Rapidly between Paralogs. PLOS Comput. Biol. 12:e1005274.
Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG, Zuckerkandl E, et al. 2008. Evolution of protein-coding genes in Drosophila. Trends Genet. 24:114–123.
Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, et al. 2016. The mid-developmental transition and the evolution of animal body plans. Nature 531:637–641.
Levin M, Hashimshony T, Wagner F, Yanai I, Arbeitman MN, Furlong EE, Imam F,

Johnson E, Null BH, Baker BS, et al. 2012. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev. Cell 22:1101–1108.
Li JJ, Huang H, Bickel PJ, Brenner SE. 2014. Comparison of D. melanogaster and C. elegans developmental stages, tissues, and cells by modENCODE RNA-seq data. Genome Res. 24:1086–1101.
Liang H, Li W-H. 2007. Gene essentiality, gene duplicability and protein connectivity in human and mouse. Trends Genet. 23:375–378.
Liu J, Robinson-Rechavi M. 2017. Adaptive evolution of proteins expressed in late and post-embryonic development in animals. bioRxiv. doi:10.1101/161711.
Mahmood S. 2010. A Survey of Functional Retroposed Genes: H. sapiens, M. musculus, D. melanogaster, and C. elegans. (Doctoral dissertation: https://tspace.library.utoronto.ca/handle/1807/24603)
Moretti S, Laurenczy B, Gharib WH, Castella B, Kuzniar A, Schabauer H, Studer RA, Valle M, Salamin N, Stockinger H, et al. 2014. Selectome update: Quality control and computational improvements to a database of positive selection. Nucleic Acids Res. 42:917–921.
Ninova M, Ronshaugen M, Griffiths-Jones S. 2014. Conserved temporal patterns of microRNA expression in Drosophila support a developmental hourglass model. Genome Biol. Evol. 6:2459–2467.
Papp B, Pál C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194–197.
Piasecka B, Lichocki P, Moretti S, Bergmann S, Robinson-Rechavi M. 2013. The Hourglass and the Early Conservation Models—Co-Existing Patterns of Developmental Constraints in Vertebrates. PLoS Genet. 9:e1003476.
Poe S, Wake MH. 2004. Quantitative tests of general models for the evolution of development. Am. Nat. 164:415–422.
Potrzebowski L, Vinckenbosch N, Marques AC, Chalmel F, Jégou B, Kaessmann H. 2008. Chromosomal Gene Movements Reflect the Recent Origin and Biology of Therian Sex Chromosomes. PLoS Biol. 6:e80.
Prachumwat A, Li W-H. 2006. Protein function, connectivity, and duplicability in yeast. Mol. Biol. Evol. 23:30–39.
Quint M, Drost H-G, Gabel A, Ullrich KK, Bönn M, Grosse I. 2012. A transcriptomic hourglass in plant embryogenesis. Nature 490:98–101.
Raff RA. 1996. The shape of life : genes, development, and the evolution of animal form. Chicago (IL): University of Chicago Press.
Richardson MK. 1995. Heterochrony and the phylotypic period. Dev. Biol. 172:412–421.
Richardson MK, Hanken J, Gooneratne ML, Pieau C, Raynaud A, Selwood L, Wright GM. 1997. There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. Anat. Embryol. (Berl). 196:91–106.
Riedl R. 1978. Order in living organisms. West Sussex, UK: Wiley-interscience.
Roux J, Robinson-Rechavi M. 2008. Developmental constraints on vertebrate genome evolution. PLoS Genet. 4:e1000311.
Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthès P, Kokkinaki M, Nef S, Gnirke A, et al. 2013. Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. Cell Rep. 3:2179–2190.
Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. Development 136:3033–3042.

Tian X, Strassmann JE, Queller DC. 2013. Dictyostelium Development Shows a Novel Pattern of Evolutionary Conservation. Mol. Biol. Evol. 30:977–984.
Veitia RA. 2002. Exploring the etiology of haploinsufficiency. BioEssays 24:175–184.
Von-Baer KE. 1828. Über Entwickelungsgeschichte der Tiere: Beobachtung und Reflexion. Königsberg: Gebrüder Bornträger.
Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. Nat. Rev. Genet. 12:204–213.
Wang QT, Piotrowska K, Ciemerych MA, Milenkovic L, Scott MP, Davis RW, Zernicka-Goetz M. 2004. A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. Dev. Cell 6:133–144.
Wang Z, Pascual-Anaya J, Zadissa A, Li W, Niimura Y, Huang Z, Li C, White S, Xiong Z, Fang D, et al. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. Nat. Genet. 45:701–706.
Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21:650–659.
Yanai I, Peshkin L, Jorgensen P, Kirschner MW. 2011. Mapping gene expression in two Xenopus species: evolutionary constraints and developmental flexibility. Dev. Cell 20:483–496.
Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, et al. 2016. Ensembl 2016. Nucleic Acids Res. 44:D710–D716.
Zalts H, Yanai I, Huber W, Salzberg SL, Hunter CP. 2017. Developmental constraints shape the evolution of the nematode mid-developmental transition. Nat. Ecol. Evol. 1:0113.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22:2472–2479.
Zhou K, Huang B, Zou M, Lu D, He S, Wang G. 2015. Genome-wide identification of lineage-specific genes within Caenorhabditis elegans. Genomics 106:242–248.