Adaptive evolution of proteins expressed in late development and adult in animals

Jialin Liu[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

Both the early conservation and hourglass Evo-Devo models suggest that, during development, late development (after the morphological 'phylotypic' period: both late and post-embryonic development) has the highest divergence at the morphological and molecular levels between species. Our aim here is to determine as much as possible the relative contributions of relaxation of purifying selection vs. increase in positive selection, in the observed high sequence divergence of genes from late development. For this, we investigated the evidence of positive selection on protein-coding genes in relation to their expression over development, in *D. rerio*, *M. musculus* and *D. melanogaster*. First, we found that genes specifically and highly expressed in late development or adult have stronger signals of positive selection. Second, over the full transcriptome, genes with evidence for positive selection trend to be expressed in late development and adult. Finally, genes involved in pathways with cumulative evidence of positive selection have higher expression in late development and adult. Overall, there is a strong and consistent signal that positive selection mainly affects genes and pathways expressed in late development and adult. Our results imply that evolution of embryogenesis is mostly conservative, with most adaptive evolution affecting post-embryonic gene expression, and thus post-embryonic phenotypes. This is consistent with the diversity of environmental challenges to which juveniles and adults are exposed.

Introduction

There are two main models to explain the relationship of development and evolutionary divergence. The early conservation model suggests that embryonic morphology between different species within the same phylum progressively diverges across development (Von Baer 1828). In contrast, the hourglass model proposes that middle development (the morphological 'phylotypic' period) has the highest morphological similarity (Duboule 1994; Raff 1996). Based on recent genomic studies, both models have some level of molecular support. Some studies support the early conservation model (Castillo-Davis and Hartl 2002; Roux and Robinson-Rechavi 2008; Artieri et al. 2009; Comte et al. 2010), while most recent ones support the hourglass model (Hazkani-Covo et al. 2005; Irie and Sehara-Fujisawa 2007; Kalinka et al. 2010; Irie and Kuratani 2011; Yanai et al. 2011; Levin et al. 2012; Quint et al. 2012; Cheng et al. 2015; Drost et al. 2015). And in fact the two models may not be mutually exclusive (Piasecka et al. 2013; Liu and Robinson-Rechavi 2017).

Both the early conservation and hourglass models predict that late development has high evolutionary divergence. This high divergence of late development has been interpreted as a consequence of relaxed developmental constraints, i.e. weaker negative selection. For example, Garstang and Riedl suggested that the development of later stages is dependent on earlier stages, so higher divergence should be found in the later stages of development (cited in Irie and Kuratani 2014). Indeed, many studies have found evidence for relaxed purifying selection in late development (Castillo-Davis and Hartl 2002; Roux and Robinson-Rechavi 2008; Artieri et al. 2009; Kalinka et al. 2010; Liu and Robinson-Rechavi 2017). Yet highly divergent late development could also be driven by adaptive evolution (positive selection), at least in part. Weaker negative and stronger positive selection are not mutually exclusive. For example, Cai and Petrov (2010) found the accelerated sequence evolution rate of primate lineage specific genes driven by both relaxed purifying selection and enhanced positive selection. Necsulea and Kaessmann (2014) suggested that the high evolution rate of testis transcriptome could be caused by both sex-related positive selection and reduced constraint on transcription.

As far as we know, few studies have tried to distinguish the roles of adaptation vs. relaxation of constraints in late development (Artieri et al. 2009), and no evidence has shown enhanced adaptive evolution in late development. Yet there is an intuitive case for adaptation to act on phenotypes established in late development, because they will be present in the juvenile and adult and interact with a changing environment.

In the case of detecting individual gene adaptation, one of the best established methods is using the ratio (ω) of non-synonymous (dN) to synonymous substitutions (dS) (Hurst 2002). Because synonymous changes are assumed to be functionally neutral, $\omega>1$ indicates evidence of positive selection. As adaptive changes probably affect only a few codon sites and at a few phylogenetic lineages, branch-site models allow the ω ratio to vary both among codon sites and among lineages (Yang and Nielsen 2002; Zhang et al. 2005). Polymorphism based methods such as frequency spectrum, linkage disequilibrium and population differentiation can also be used to identify changes due to recent positive selection (Vitti et al. 2013).

Since several genes with slight effect mutations can act together to have a strong effect, adaptive evolution can act on the pathway level as well (Daub et al. 2013; Berg et al. 2014). In the case of polygenic adaptation, a gene set enrichment test has successfully been applied to detect gene sets with polygenic adaptive signals (Daub et al. 2013; Daub et al. 2015; Daub et al. 2017). This gene set enrichment analysis allows to detect weak but consistent adaptive signals from whole genome scale, unlike traditional enrichment tests only considering top scoring genes with an artificial significance threshold.

In order to estimate the contribution of positive selection on the highly divergent late development evolution, we have adopted three approaches. First, we used modularity analysis to obtain distinct sets of genes (modules) which are specifically expressed in different meta developmental stages (Piasecka et al. 2013), and compared signal of positive selection across modules. Second, we applied a modified "transcriptome age index" (Domazet-Loso and Tautz 2010) to measure evolutionary adaptation on the whole transcriptome level. Finally, we used a gene set enrichment approach to detect polygenic selection on pathways, and studied the expression of these gene sets over

development. Each approach was applied to developmental transcriptomes and to results of the branch-site test for positive selection in *D. rerio*, *M. musculus* and *D. melanogaster*. All the analyses found a higher rate of adaptation in late development, including adult.

Results

Modularity analysis

For the modularity analysis, we focused on different sets of specifically expressed genes (modules) in each developmental period. Our expectation is that genes in each module have specific involvement during embryonic development (Piasecka et al. 2013), so different adaptation rates of these genes should reflect stage specific adaptability. For *D. rerio*, we obtained seven modules (early modules: cleavage./blastula and gastrula, middle modules: segmentation and pharyngula, late modules: larva, juvenile and adult) from our previous study (Piasecka et al. 2013). For *M. musculus* and *D. melanogaster*, we identified five modules (early modules: cleavage, blastula and neurula, middle module: organogenesis, late module: late embryo) and six modules (early module: early embryo, middle module: middle embryo, late module: late embryo, larva, pupa and adult) respectively (Figure S1, S2, S3).

In order to detect the signal of positive selection, we used the log-likelihood ratio test statistic (ΔlnL) of H_1 to H_0 (models with or without positive selection) from the branch-site model (Zhang et al. 2005) as precomputed in Selectome on filtered alignments (Moretti et al. 2014), and as used in (Roux et al. 2014; Daub et al. 2017). Briefly, ΔlnL represents the evidence for positive selection, thus a branch in a gene tree with a higher value indicates higher evidence for positive selection for this gene over this branch. If late development is under stronger positive selection, we should observe higher ΔlnL in late modules.

Because not all genes have any evidence for positive selection, we first compared the proportion of genes with strong evidence (q-value < 0.05) or weak evidence (no threshold for q-value) of positive selection (ΔlnL > 0) across modules (Figure S4, S5). In *M. musculus* and *D. melanogaster*, the proportion is not significantly different

across modules for either criterion (Figure S4). In contrast, in *D. rerio* we observed a higher proportion in the juvenile and adult modules for the proportion of genes with strong evidence of positive selection (Figure S5).

We then compared the values of ΔlnL for genes with weak evidence of positive selection (Figure 1). In order to improve the normality of non-zero ΔlnL, we transformed ΔlnL with fourth root (Hawkins and Wixley 1986; Roux et al. 2014; Daub et al. 2017). In the two vertebrate species the overall variations of ΔlnL observed across modules are significant (Kruskal-Wallis test).

In *D. rerio*, we detected an hourglass pattern of ΔlnL, highest in late modules, intermediate in early modules and lowest in middle modules. Specifically, in the juvenile module, the mean ΔlnL is significantly higher than the mean ΔlnL for all genes (randomization test, p=0.02; all p are adjusted p-values for false discovery rate, see Methods). We note that the adult module also has higher mean ΔlnL, even though it's not significant (randomization test, p=0.12). In the pharyngula module, the mean ΔlnL is lower than the mean ΔlnL for all genes, and the difference is significant (randomization test, p=0.03). In the other modules, the mean ΔlnL is not significantly different with the mean for all genes.

In *M. musculus*, however, we found an early conservation pattern of ΔlnL, lowest in early modules, intermediate in middle module and highest in late module. Specifically, the late embryo module has a higher mean ΔlnL than all genes (randomization test, p=0.03). Both the cleavage and blastula modules have lower mean ΔlnL than all genes, even though the adjusted p-value is not very low (randomization test, p=0.08 for cleavage module, p=0.06 for blastula module). In addition, for *M. musculus*, we repeated the analysis with ΔlnL values computed on ten older phylogenetic branches (Sciurognathi to Sarcopterygii), and seven of these also show significantly higher mean ΔlnL for the late embryo module (Figure S6).

In *D. melanogaster*, the mean ΔlnL in no module is significantly different from the mean ΔlnL for all genes, but we still can observe a weak early conservation like pattern. In the early embryo module, the mean ΔlnL is slightly lower than the mean

ΔlnL for all genes (randomization test, p=0.29). In the adult module, the mean ΔlnL is slightly higher than the mean ΔlnL for all genes (randomization test, p=0.09).

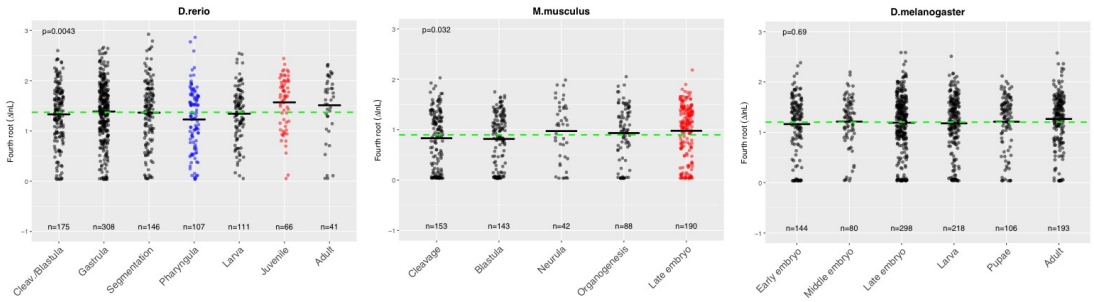

Figure 1: Variation of ΔlnL in different modules.

For each module, dots are values of ΔlnL for individual genes and the black line is the mean of ΔlnL. Red (respectively blue) dots indicate modules for which the mean of ΔlnL is significantly (adjusted p<0.05) higher (respectively lower) than the mean of ΔlnL from all modules. The green dash line denotes the mean value of ΔlnL from all modular genes. The number of genes in each module is indicated below the module. The adjusted p-value of the overall variation across all modules is indicated in the top-left corner of each graph.

Overall, these findings suggest that positive selection is stronger on genes expressed in late development or in adult than in early and middle development. In addition, it also indicates that ΔlnL in different species supports different Evo-Devo models (hourglass vs. early conservation, see also Liu and Robinson-Rechavi (2017)).

Transcriptome index analysis

Modularity analysis measures the strength of positive selection for genes specifically expressed in each meta developmental stage. We next used a modified "Transcriptome Age Index" (Domazet-Loso and Tautz 2010) to calculate the weighted mean of ΔlnL on transcriptome level. Notably, all expression levels were log-transformed before use, unlike in Domazet-Lošo & Tautz (Domazet-Loso and Tautz 2010) (see discussion in (Piasecka et al. 2013)). We named this modified index "Transcriptome Likelihood Index" (TLI). A higher index indicates that the transcriptome has higher expression of transcripts from high ΔlnL genes.

To describe the variation of TLI across development, we fitted both first degree and second degree polynomial models (Roux and Robinson-Rechavi 2008). We keep the second-degree polynomial model (parabola) only if it provides a significantly better fit (tested with ANOVA, $p<0.05$).

In *D. rerio*, the parabola provides a better fit and its minimum corresponds to middle development, suggesting an hourglass-like model of TLI; TLI is highest in late development and adult (Figure 2).

In *M. musculus*, TLI progressively increases with development, supporting an early conservation model. Like for modularity analysis, in *M. musculus,* we verified results with ten older phylogenetic branches. Six of these also supported that late development has highest TLI (Figure S7).

In *D. melanogaster*, although the parabola provides a better fit, its minimum corresponds to early development (between 4-6h), indicating an early conservation model of TLI. Both early development and middle development have similar low TLI, and TLI gradually increases after middle development.

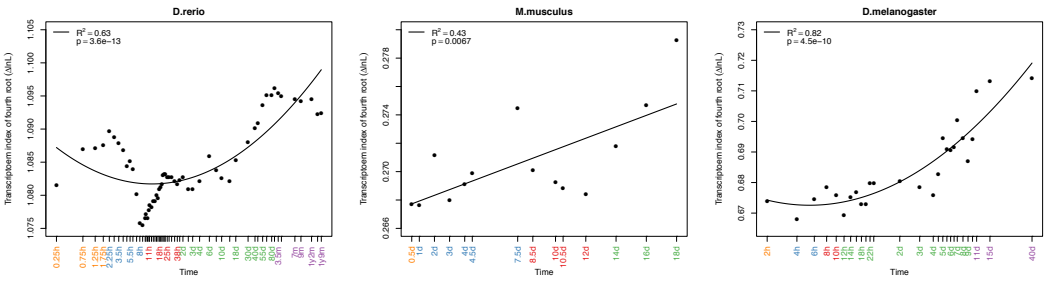

Figure 2: Transcriptome index of ΔlnL (TLI) across development.

The relation between TLI and developmental time is fit by a general linear regression (the second degree of polynomial for *D. rerio* and *D. melanogaster*, the first degree of polynomial for *M. musculus*, based on ANOVA test), whose R^2 and adjusted p-value are indicated in the top-left corner of each graph. Orange, blue, red, green and purple time points represent maternal stages, early developmental stages, middle developmental stages, late developmental stages and adult stages respectively. The x-

axis for *D. rerio* and *D. melanogaster* is in logarithmic scale, while the x-axis for *M. musculus* is in natural scale.

As in the modularity analysis, both *M. musculus* and *D. melanogaster* support the early conservation model, while *D. rerio* follows an hourglass model. Again, from whole transcriptome level, these results indicate that genes with evidence for positive selection are more highly expressed in late development and adult.

Polygenic selection analysis

Positive selection can be detected at the biological pathway level, even when individual genes within the pathway only fix small effect mutations (Daub et al. 2013; Berg et al. 2014; Daub et al. 2017). Briefly, we calculated the sum of ΔlnL (SUMSTAT) for a pathway, and inferred the significance of this SUMSTAT with an empirical null distribution (Tintle et al. 2009; Daub et al. 2013; Daub et al. 2017). In total, we identified 10, 4 and 9 pathways with a significant signal of positive selection, respectively in *D. rerio*, *M. musculus* and *D. melanogaster* (q-value<0.2, Table1). If late development is under stronger positive selection at the pathway level as well, we expect genes involved in pathways with a signal of positive selection to be more highly expressed at this period. Thus we computed the ratio of median expression between positively selected pathway genes and genes included in pathways not positively selected. Generally, consistent with previous results, we found higher ratios in late development and adult. The profile is slightly different among species (Figure 3). In *D. rerio* and *D. melanogaster*, the ratio progressively increases until adult, in *M. musculus* there is an hourglass pattern, but skewed to much higher ratios in late development than in early development.

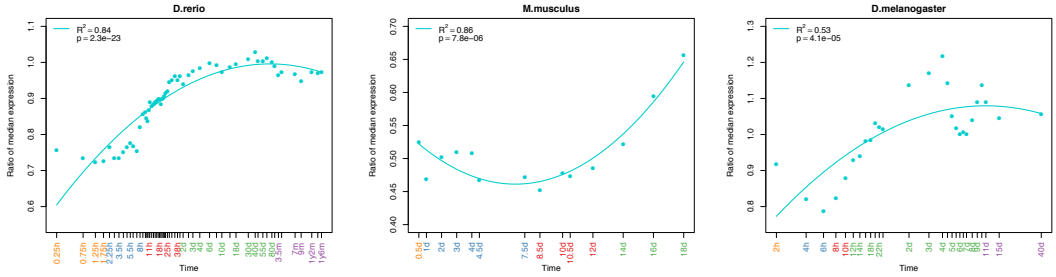

Figure 3: Expression in development for genes involved in pathways enriched with signal of positive selection.

Each solid circle represents the ratio of the median expression for genes involved in pathways enriched with signal of positive selection to the median expression for genes involved in pathways without signal of positive selection. The ratio is fit by a second degree polynomial regression (based on ANOVA test), whose R^2 and adjusted *p*-value are indicated in the top-left corner of each graph. Orange, blue, red, green and purple time points represent maternal stages, early developmental stages, middle developmental stages, late developmental stages and adult stages respectively. The x-axis for *rerio* and *D. melanogaster* is in logarithmic scale, while the x-axis for *M. musculus* is in natural scale.

Since late development and adult tend to express genes involved in pathways enriched for signal of positive selection, this confirms that adaptive evolution at the pathway level mainly affects these stages.

Discussion

Re-unification of structuralist and functionalist comparative biology

There have been two major approaches to comparative biology since the late 18th century: the structuralist approach (which gave rise to Evo-Devo) emphasizes the role of constraints, and focuses on investigating spatial and timing variations of conserved structures in distantly related species. In a modern context, the focus is often on comparing developmental genes' expression between distantly related species. The functionalist or adaptationist approach (which gave rise to the Modern Synthesis and most of evolutionary biology) emphasizes the role of natural selection. In a modern context, the focus is often on investigating adaptive mutations. It has been suggested that these two approaches could not be reconciled (Amundson 2007), since the former

underscores how mutations generate morphological diversity, while the later underscores whether mutations are fixed by positive selection or not. A good example of the differences between structuralist and adaptationist comes from the debate between (Hoekstra and Coyne 2007; Carroll 2008). As a structuralist, Carroll suggested that mutations affecting morphology largely occur in the *cis*-regulatory regions. However, as adaptationists, Hoekstra and Coyne argued that this statement is at best premature. Their main argument was that they didn't find that adaptive evolution more likely occur in *cis*-regulatory elements, but rather in protein coding genes, from both genome-wide surveys and single-locus studies. It is important to note that Carroll's theory is specific to morphological evolution, but not directly related to evolutionary adaptation. Basically, both sides could be correct, and were mostly discussing different things.

Since both adaptation and structure are part of biology, we should be able to explain both in a consistent manner. Here, we try to bridge positive selection and morphological evolution by combining developmental time-series transcriptomes, positive selection inference on protein coding genes, modularity analysis, transcriptome index analysis, and gene set analysis. From both modularity analysis and transcriptome index analysis, we found that genes highly expressed in late development and adult have higher evidence for positive selection. From polygenic analysis, we found that the expression of positively selected pathways is higher in late development and adult. Overall, these results suggest that higher morphological variation in late development could be at least in part driven by adaptive evolution. In addition, coding sequence evolution could also make an significant contribution for the evolution of morphology, as suggested by Hoekstra and Coyne (2007) and Burga et al. (2017). It should be noted that we do not test here whether regulatory sequence evolution plays a similar or greater role, since we do not have equivalent methods to test for positive selection in regulatory regions.

Developmental constraint hypothesis and Darwin's selection opportunity hypothesis

Despite the repeated observation that late development is highly divergent for diverse genomic properties (sequence evolution, duplication, gene age, expression divergence) in all animal species studied, the underlying evolutionary forces driving such a pattern

remain obscure. The developmental constraint hypothesis suggests that this high divergence is due to relaxed purifying selection, whereas Darwin's "selection opportunity" hypothesis proposed stronger positive selection (as discussed in Artieri et al. 2009). The latter could be due to the greater diversity of challenges to which natural selection needs to respond in juvenile and adult life than in early development.

Several studies have found evidence (direct or indirect) to support the importance of developmental constraints (Castillo-Davis and Hartl 2002; Roux and Robinson-Rechavi 2008; Artieri et al. 2009; Kalinka et al. 2010). For example, we (Roux and Robinson-Rechavi 2008) found that genes expressed earlier in development contain a higher proportion of essential genes. Weaker purifying selection in late development would imply that genes expressed in this period have less fitness impact, which is consistent with the paucity of essential genes. Here and in Liu and Robinson-Rechavi (2017) the branch-site codon model allows us to isolate the contribution of purifying selection to coding sequence evolution. We found indeed that genes under weaker purifying selection on the protein sequence trend to be expressed in late development (Liu and Robinson-Rechavi 2017). This provides direct evidence of relaxed purifying selection in late development.

To the best our knowledge, there was no direct systematic evidence so far in support of Darwin's "selection opportunity" hypothesis. One such study was proposed by Artieri et al. (2009). Unfortunately, based on relative poor quantitative expression data (ESTs) and limited time points (embryonic, larval/pupal and adult) they did not find any direct evidence of higher positive selection in late development. Since they noticed that the accelerated sequence evolution of genes expressed at adult stage was confounded by male-biased genes, they argued that the rapidly evolution observed in late development could be due to specific selective pressures such as sexual selection. Our results clearly provide a quantitative test which supports a role of positive selection in the high divergence of late development. We show that this is not due to testis-expressed genes (Figure S12). It should be noted that our results also provide counter evidence to the adaptive penetrance hypothesis, which argues that adaptive evolution mainly occurs in the middle development (Richardson 1999).

Correcting confounding factors

Since some non-adaptive factors (such as gene length, gene tree size (the number of branches in a tree) and branch length) can be correlated with ΔlnL and affect our results (Daub et al. 2017), we investigated the correlation between ΔlnL and these potential confounding factors. Generally, we found a small correlation between ΔlnL and tree size, but a larger correlation between ΔlnL and gene length or branch length (Figure S8). One explanation for this high correlation between ΔlnL and gene length is that long genes could accumulate more mutations than short genes, so we have more power to detect positive selection with higher number of mutations (Fletcher and Yang 2010; Gharib and Robinson-Rechavi 2013). So, we decided to correct the influence of gene length on our results. Because branch length is inferred from the number of mutations, and higher branch length can be driven by higher evolutionary rate due to positive selection, we did not correct the correlation between ΔlnL and branch length.

In order to investigate if gene length might have affected our results, for modularity and TLI analysis, we tested whether patterns purely based on gene length are similar to those based on ΔlnL or not. Surprisingly, we found an opposite pattern between gene length and ΔlnL. For modularity analysis, in *D. rerio* and *M. musculus*, we did not observe a significantly higher mean gene length in the higher ΔlnL modules. In *D. melanogaster*, we even found the larva module to have significantly lower mean gene length than all genes (Figure 9). For TLI analysis, we found a negative correlation pattern between transcriptome index of gene length and developmental stage, suggesting that later stages trend to express shorter genes (Figure 10). These findings imply that the detection of higher positive selection in late development is not driven by gene length.

Different pattern for different species

Developmental patterns of adaptive evolution on protein coding sequences are not consistent in different species. In *D. rerio*, both modularity analysis and transcriptome index analysis support an hourglass pattern, implying that middle development is specifically depleted of adaptive evolution, and under an evolutionary lockdown state. However, in *M. musculus* and *D. melanogaster*, both analyses support the early

conservation model. The different patterns between *D. rerio* and *M. musculus* could be related with different reproductive strategies even they are from same phylum. Because *D. rerio* is oviparous, its embryos are exposed to variable ecological niches. *M. musculus,* however, is viviparous, so the embryos exist in more stable foetal environments. So, higher adaptive evolution in early development than in middle development for *D. rerio* could reflect adaptation to diverse ecological niches (Kalinka and Tomancak 2012). A potential caveat is that the mouse expression data comes from two data sets, so the different patterns between *D. rerio and M. musculus* could also be simply explained by batch effect of mouse data sets.

For polygenic analysis, interestingly, we found a consistent early conservation pattern across the three species. Thus, while there is some signal of early development adaptive evolution on single genes, the later developmental signal is more consistent at the pathway level. Because pathways link genes to phenotypes, this suggests stronger phenotypic adaptation in juveniles and adults, consistent with "Darwin's selection opportunity" hypothesis.

Materials and Methods

Expression data sets

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

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

For *D. melanogaster*, we obtained normalized RNAseq data from (Li et al. 2014), which originally comes from (Graveley et al. 2011). This data has 27 stages from embryo to adult. For the last three stages, since data were available for male and female, we took the mean. All the genes with expression <1 RPKM were set as not expressed (Kryuchkova-Mostacci and Robinson-Rechavi 2016), replaced by 1.0001

(this value is smaller than the smallest RPKM among genes with RPKM>1) and log2 transformed.

Branch-site likelihood test data

The log-likelihood ratio (ΔlnL) values of a test for positive selection were retrieved from Selectome (Moretti et al. 2014), a database of positive selection based on the branch-site likelihood test (Zhang et al. 2005). One major advantage of this test is allowing positive selection to vary both among codon sites and among phylogenetic branches. The branch-site test contrasts two hypotheses: the null hypothesis is that no positive selection occurred (H0) in the phylogenetic branch of interest, and the alternative hypothesis is that at least some codons experienced positive selection (H1). The log likelihood ratio statistic (ΔlnL) is computed as $2 \times (\ln LH1 - \ln LH0)$. Notably, in order to mitigate false positives due to poor sequence alignments, Selectome integrates filtering and realignment steps to exclude ambiguously aligned regions.

We used ΔlnL from the Clupeocephala branch, the Murinae branch and the Melanogaster group branch for *D. rerio*, *M. musculus* and *D. melanogaster* respectively. In addition, for *M. musculus*, we also retrieved ten older branches (Sciurognathi, Rodentia, Glires, Euarchontoglires, Eutheria, Theria, Mammalia, Amniota, Tetrapoda, and Sarcopterygii). One gene could have two ΔlnL 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.

Pathways

We downloaded lists of 1,683 *D. rerio* gene sets, 2,269 *M. musculus* gene sets and 1365 *D. melanogaster* gene sets of type "pathway" from the NCBI Biosystems Database (Geer et al. 2009). This is a repository of gene sets collected from manually curated pathway databases, such as BioCyc (Caspi et al. 2014), KEGG (Kanehisa et al. 2014), Reactome (Croft et al. 2014), The National Cancer Institute Pathway Interaction Database (Schaefer et al. 2009) and Wikipathways (Kelder et al. 2012).

Coding sequence length

We extracted coding sequence (CDS) length from Ensembl version 84 (Yates et al.

2016) using BioMart (Kinsella et al. 2011). For genes with several transcripts, we used the transcript with the maximal CDS length.

Testis specific genes

Testis specific genes for *M. musculus* and *D. melanogaster* were obtained from a parallel study (Liu and Robinson-Rechavi 2017). The testis specific genes were defined as genes with highest expression in testis and with tissue specificity value ≥ 0.8.

Clustering of expression data

For *D. melanogaster*, we applied the Iterative Signature Algorithm (ISA) (Bergmann et al. 2003; Piasecka et al. 2013). The ISA was initialized with six candidate seeds, each consisting of one artificial expression profile corresponding to one of the *D. melanogaster* developmental meta stages (early embryo (0-6h), middle embryo (6-10h), late embryo (10-24h), larva, pupa, and adult. In the early embryo module, since the expression in the first time point (0-2h, median expression 0.00015) is very low compared with the other time points (2-4h, median expression 0.35; 4-6h, median expression 0.72), we exclude the first time point from this analysis. We ran ISA based on two threshold parameters: "thr.row"=1.2 and "thr.col"=1.9. The threshold parameters adjust the size of the modules, higher values resulting in smaller, more correlated modules.

A single experiment covering the entire *M. musculus* developmental transcriptome is not available at present, so we separately applied the Fuzzy C-Means Algorithm (Futschik and Carlisle 2005), which is more sensitive for small datasets, to two datasets (Wang et al. 2004; Irie and Kuratani 2011). For the first data set, we ran the algorithm with parameters c=6 (c is the number of clusters), and m=1.30 (m represents a value which prevents clustering of random data). For the second data set, we ran the algorithm with c=20, and m=1.30. For the first data set, we obtained two useful clusters: cluster 2 and cluster 6 correspond to blastula and cleavage modules respectively (see Figure S2). For the second data set, we obtained a neurula module from cluster 11, an organogenesis module from cluster 1, and a late-embryo module from clusters 3, 12, 13 and 20 (see Figure S3). Since we clustered two data sets

separately, some genes might occur in clusters from both analyses. We excluded these genes for further module-based analyses.

Randomization test of modularity analysis

For each module, we randomly choose the same number of ΔlnL from all modular genes (genes attributed to any module in that species) without replacement and calculated the mean value. We repeated this 10000 times and approximated a normal distribution for the mean value of ΔlnL. The *p*-value that the mean value of interested module is higher (or lower) than the mean value from all modular genes is the probability that the randomly sampled mean value of ΔlnL is higher (or lower) than the original mean value of ΔlnL. In the same way, we also estimated the *p*-value of the median ΔlnL value.

Transcriptome index of log-likelihood ratio (TLI)

The TLI is calculated as:

$$\mathrm{TLI}_s = \frac{\sum_{i=1}^{n} \sqrt[4]{\triangle \mathrm{lnL}_i}\, e_{is}}{\sum_{i=1}^{n} e_{is}},$$

s is the developmental stage, $\Delta\mathrm{lnL}_i$ is the value of log-likelihood ratio for gene i, n the total number of genes and e_{is} is the log-transformed expression level of gene i in developmental stage s. Here, we used all ΔlnL values without applying any cut-off on ΔlnL or the associated *p*-value. For genes with ΔlnL<0, we replaced it with 0. For *M. musculus*, we calculated the TLI from a merged data set, instead of computing it on two data sets separately.

Detection of polygenic selection

We performed a gene set enrichment approach to detect polygenic signals of positive selection on pathways (Ackermann and Strimmer 2009; Daub et al. 2013; Daub et al. 2017). For each pathway, we calculated its SUMSTAT score, which is the sum of ΔlnL of all genes within this pathway. The ΔlnL values were fourth-root transformed. This approach makes the distribution of non-zero ΔlnL approximate normal distribution (Canal 2005; Roux et al. 2014; Daub et al. 2017). So, with fourth-root

transformation, we limit the risk that the significant pathways we found be due to a few outlier genes with extremely high ΔlnL. The SUMSTAT score of a pathway is calculated as:

$$\mathrm{SAMSTAT}_p = \sum_{i \in p} \sqrt[4]{\triangle \mathrm{lnL}_i} ,$$

where p represents a pathway, and $\Delta\mathrm{lnL}_i$ represents the value of log-likelihood ratio for gene i within pathway p. *P*athways less than 10 ΔlnL values were excluded from our analysis. Like in TLI analysis, we used all ΔlnL values and replaced <0 values with 0.

Empirical null distribution of SAMSTAT

We used a randomization test to infer the significance of the SUSMTAT score of a pathway. To correct for the potential bias caused by gene length, we firstly created bins with genes that have similar length (Figure S11). Secondly, we randomly sampled (without replacement) the same number of genes from each bin, to make the total number of genes equal to the pathway being tested. Thirdly, we computed the SAMSTAT score of the randomly sampled ΔlnL values. We repeated the second and third processes one million times. Fourthly, we approximated a normal distribution for SAMSTAT score of the interested pathway. Finally, the *p*-value was calculated as the probability that the expected SAMSTAT score is higher than the observed SAMSTAT score.

Removing redundancy in overlapping pathways ("pruning")

Because some pathways share high ΔlnL value genes, the identified significant pathways might be partially redundant. In other words, shared genes among several pathways can drive all these pathways to score significant. We therefore removed the overlap between pathways with a "pruning" method (Daub et al. 2013; Daub et al. 2017). Firstly, we inferred the *p*-value of each pathway with the randomization test. Secondly, we removed the genes of the most significant pathway from all the other pathways. Thirdly, we ran the randomization test on these updated gene sets. Finally, we repeated the second and third procedures until no pathways were left to be tested.

With this "pruning" method, the randomization tests are not independent and only the high scoring pathways will remain, so we need to estimate the False Discovery Rate (FDR) empirically. To achieve this, we applied the "pruning" method to pathways with permuted ΔlnL scores and repeated it for 300 times. So, for each pathway, we obtained one observed *p*-value (*p**) and 300 empirical *p*-values. The FDR was calculated as follow:

$$\hat{FDR}(p^*) = \frac{\pi_0 \hat{V}(p^*)}{R(p^*)},$$

where π_0 represents the proportion of true null hypotheses, $\hat{V}(p^*)$ represents the estimated number of rejected true null hypotheses and $R(p^*)$ represents the total number of rejected hypotheses. For π_0, we conservatively set it equal to 1 as in Daub et al. (2017). For $\hat{V}(p^*)$, in each permutation analysis, we firstly calculated the proportion of *p*-value (from permutation analysis) $\leq p^*$. Then, the value of $\hat{V}(p^*)$ was estimated by the mean proportion of *p*-value (from permutation analysis) $\leq p^*$ for the 300 permutation tests. For $R(p^*)$, we defined it as the number of *p*-value *(from original analysis)* $\leq p^*$. For *q*-value, we determined it from the lowest estimated FDR among all *p*-values (from original analysis) $\geq p^*$.

Multiple test correction

All the *p*-values reported in this study were adjusted with multiple test correction using the Benjamini–Hochberg method.

Acknowledgements

We thank Sébastien Moretti for help with Selectome data retrieval; the Bgee team for help with expression data retrieval; Barbara Piasecka for help with Iterative Signature Algorithm cluster; and Josephine T. Daub for help with polygenic selection analysis. We thank Julien Roux, Elsa Guillot, Iakov Davydov and all 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.

Table 1

Candidate pathways enriched with signal of positive selection. We reported all pathways with *q-value* <0.2 after removing overlapping genes ("pruning") for *D. rerio*, *D. melanogaster and M. musculus*.

Species	Rank	Pathway	Pathway size before /after pruning	*p* value before pruning	*q* value before pruning	*p* value after pruning	*q* value after pruning
D. rerio	**1**	Laminin interactions	12/12	2.00E-06	1.32E-03	2.00E-06	0.00E+00
	2	Phenylalanine metabolism	10/10	7.70E-05	1.51E-02	7.80E-05	8.78E-03
	3	Visual phototransduction	33/33	9.10E-05	1.51E-02	8.30E-05	8.78E-03
	4	Metabolism of carbohydrates	119/118	2.02E-04	2.23E-02	2.37E-04	2.97E-02
	5	Gamma carboxylation, hypusine formation and arylsulfatase activation	18/18	1.46E-03	8.05E-02	1.24E-03	1.16E-01
	6	Extracellular matrix organization	75/61	2.00E-05	6.62E-03	1.50E-03	1.16E-01
	7	Acyl chain remodelling of PE	10/10	5.12E-03	1.79E-01	3.79E-03	1.66E-01
	8	Base excision repair	24/24	4.94E-03	1.79E-01	3.82E-03	1.66E-01
	9	Aminoacyl-tRNA biosynthesis	30/30	6.23E-03	1.97E-01	3.93E-03	1.66E-01
	10	Phase II conjugation	37/30	1.08E-03	6.92E-02	4.00E-03	1.66E-01
D. melanogaster	**1**	Triglyceride Biosynthesis	59/59	7.60E-05	1.57E-02	7.60E-05	3.32E-02
	2	Glycosaminoglycan degradation	16/16	6.99E-04	4.70E-02	6.37E-04	9.44E-02
	3	Metabolism of porphyrins	12/12	1.57E-03	7.31E-02	1.47E-03	9.62E-02
	4	Detoxification of Reactive Oxygen Species	17/17	1.45E-03	7.31E-02	1.48E-03	9.62E-02
	5	Longevity regulating pathway	43/28	2.83E-02	3.12E-01	3.37E-03	1.54E-01
	6	ECM-receptor interaction	10/10	4.54E-03	1.66E-01	4.10E-03	1.54E-01
	7	Lysine degradation	25/15	1.84E-02	2.86E-01	5.15E-03	1.54E-01
	8	Metabolic pathways	813/767	3.04E-04	3.11E-02	5.23E-03	1.54E-01
	9	Glutathione metabolism	55/32	2.59E-02	3.12E-01	5.66E-03	1.54E-01
M. musculus	**1**	Pantothenate and CoA biosynthesis	16/16	9.10E-05	7.20E-02	9.10E-05	5.01E-02
	2	Mineralocorticoid biosynthesis	10/10	1.52E-04	7.20E-02	1.40E-04	5.01E-02
	3	Mitochondrial translation *	72/72	2.91E-04	7.86E-02	2.73E-04	5.25E-02
	4	Cytokine-cytokine receptor interaction	100/100	3.41E-04	7.86E-02	2.78E-04	5.25E-02

References

Ackermann M, Strimmer K. 2009. A general modular framework for gene set enrichment analysis. BMC Bioinformatics 10:47.

Amundson R. 2007. The changing role of the embryo in evolutionary thought : roots of evo-devo. Cambridge University Press.

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.

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.

Berg JJ, Coop G, Fisher R, Turelli M, Barton N, Slate J, Kingsolver J, Hoekstra H, Hoekstra J, Berrigan D, et al. 2014. A Population Genetic Signal of Polygenic Adaptation. PLoS Genet. 10:e1004412.

Bergmann S, Ihmels J, Barkai N. 2003. Iterative signature algorithm for the analysis of large-scale gene expression data. Phys Rev E Stat Nonlin Soft Matter Phys 67:31902.

Burga A, Wang W, Ben-David E, Wolf PC, Ramey AM, Verdugo C, Lyons K, Parker PG, Kruglyak L. 2017. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. Science 356:eaal3345.

Cai JJ, Petrov DA. 2010. Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes. Genome Biol. Evol. 2:393–409.

Canal L. 2005. A normal approximation for the chi-square distribution. Comput. Stat. Data Anal. 48:803–808.

Carroll SB. 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell 134:25–36.

Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher C a., Holland T a., Keseler IM, Kothari A, Kubo A, et al. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 42:459–471.

Castillo-Davis CI, Hartl DL. 2002. Genome evolution and developmental constraint in Caenorhabditis elegans. Mol. Biol. Evol. 19:728-735.

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.

Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, et al. 2014. The Reactome pathway knowledgebase. Nucleic Acids Res. 42:1–6.

Daub JT, Dupanloup I, Robinson-Rechavi M, Excoffier L. 2015. Inference of Evolutionary Forces Acting on Human Biological Pathways. Genome Biol. Evol. 7:1546–1558.

Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, Excoffier L. 2013. Evidence for polygenic adaptation to pathogens in the human genome. Mol. Biol. Evol. 30:1544–1558.

Daub JT, Moretti S, Davydov II, Excoffier L, Robinson-Rechavi M. 2017. Detection of pathways affected by positive selection in primate lineages ancestral to humans. Mol. Biol. Evol. 34:1391-1402.

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.
Fletcher W, Yang Z. 2010. The Effect of Insertions, Deletions, and Alignment Errors on the Branch-Site Test of Positive Selection. Mol. Biol. Evol. 27:2257–2267.
Futschik M, Carlisle B. 2005. Noise-robust soft clustering of gene expression time-course data. J. Bioinform. Comput. Biol. 3:965–988.
Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. 2009. The NCBI BioSystems database. Nucleic Acids Res. 38:492–496.
Gharib WH, Robinson-Rechavi M. 2013. The Branch-Site Test of Positive Selection Is Surprisingly Robust but Lacks Power under Synonymous Substitution Saturation and Variation in GC. Mol. Biol. Evol. 30:1675–1686.
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.
Hawkins DM, Wixley RAJ. 1986. A Note on the Transformation of Chi-Squared Variables to Normality. Am. Stat. 40:296.
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.
Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. Evolution. 61:995–1016.
Hurst LD. 2002. The Ka/Ks ratio: Diagnosing the form of sequence evolution. Trends Genet. 18:486–487.
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.
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.
Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: Back to metabolism in KEGG. Nucleic Acids Res. 42:199–205.
Kelder T, Van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, Pico AR. 2012. WikiPathways: Building research communities on biological pathways. Nucleic Acids Res. 40:1301–1307.
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. 2016. Tissue-Specificity of Gene

Expression Diverges Slowly between Orthologs, and Rapidly between Paralogs. PLOS Comput. Biol. 12:e1005274.
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.
Liu J, Robinson-Rechavi M. 2017. Developmental constraints on genome evolution in four bilaterian model species. bioRxiv. doi:10.1101/161679.
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.
Necsulea A, Kaessmann H. 2014. Evolutionary dynamics of coding and non-coding transcriptomes. Nat. Rev. Genet. 15:734–748.
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.
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. 1999. Vertebrate evolution: the developmental origins of adult variation. Bioessays 21:604–613.
Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M, Keller L. 2014. Patterns of positive selection in seven ant genomes. Mol. Biol. Evol. 31:1661–1685.
Roux J, Robinson-Rechavi M. 2008. Developmental constraints on vertebrate genome evolution. PLoS Genet. 4:e1000311.
Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. 2009. PID: The pathway interaction database. Nucleic Acids Res. 37:674–679.
Tintle NL, Borchers B, Brown M, Bekmetjev A. 2009. Comparing gene set analysis methods on single-nucleotide polymorphism data from Genetic Analysis Workshop 16. BMC Proc. 3(Suppl 7):S96.
Vitti JJ, Grossman SR, Sabeti PC. 2013. Detecting natural selection in genomic data. Annu. Rev. Genet. 47:97–120.
Von-Baer KE. 1828. Über Entwickelungsgeschichte der Tiere: Beobachtung und Reflexion. Königsberg: Gebrüder Bornträger.
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.
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.
Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19:908–917.
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.

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.