Developmental constraints, innovations and robustness

Liu Jialin

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Département d'écologie et evolution

Developmental constraints, innovations and robustness

Thèse de doctorat ès sciences de la vie (PhD)

présentée a la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Jialin Liu

Master en génomique, Institut de génomique de Beijing, Académie chinoise des sciences

Jury

Prof. Philippe Reymond, Président
Prof. Marc Robinson-Rechavi, Directeur de thèse
Dr. David Garfiled, expert
Prof. Günter Wagner, expert

Lausanne, July 2019

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Philippe	**Reymond**
Directeur·trice de thèse	Monsieur	Prof.	Marc	**Robinson-Rechavi**
Expert·e·s	Monsieur	Dr	David	**Garfield**
	Monsieur	Prof.	Günter	**Wagner**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Jialin Liu

Master of Science, University of Chinese Academy of Sciences (UCAS), Chine

intitulée

Developmental constraints, innovations and robustness

Lausanne, le 27 septembre 2019

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Philippe Reymond

Table of contents

Abstract

During my PhD, I have been working on Evo-Devo patterns (especially the debate around the hourglass model) in transcriptomes, with an emphasis on adaptation. I have characterized patterns in model organisms in terms of constraints and especially in terms of positive selection. I found that the phylotypic stage (a stage in mid-embryonic development) is an evolutionary lockdown, with stronger purifying selection and less positive selection than other stages in terms of the evolution of protein sequences and of regulatory elements. To study the adaptive evolution of gene regulation during development, I have developed a machine leaning based in silico mutagenesis approach to detect positive selection on regulatory elements.

In addition to transcriptome evolution, I have been working on the tension between precision and stochasticity of gene expression during development. More precisely, I have shown that expression noise follows an hourglass pattern, with lower noise at the phylotypic stage. This pattern can be explained by stronger histone modification mediated noise control at this stage. In addition, I propose that histone modifications contribute to mutational robustness in regulatory elements, and thus to conserved expression levels. These results provide insight into the role of robustness in the phenotypic and genetic patterns of evolutionary conservation in animal development

Résumé de thèse

Au cours de ma thèse, j'ai travaillé sur les modèles Evo-Devo (en particulier le débat autour du modèle *hourglass*) dans les transcriptomes, en mettant l'accent sur l'adaptation. J'ai mis en évidence les tendances dans des organismes modèles en termes de contraintes et en particulier de sélection positive. J'ai trouvé que le stade phylotypique (stade du développement au milieu de développement embryonnaire) est un évolutivement verrouillé, avec une sélection purifiante plus forte et moins de sélection positive que d'autres stades, en termes d'évolution des séquences protéiques et des éléments régulateurs. Pour étudier l'évolution adaptative de la régulation des gènes au cours du développement, j'ai développé une approche de mutagénèse in silico basée sur le *machine learning* afin de détecter la sélection positive sur des éléments régulateurs.

En plus de l'évolution du transcriptome, j'ai travaillé sur la tension entre précision et stochasticité de l'expression des gènes au cours du développement. Plus précisément, j'ai montré que le bruit d'expression suit le modèle *hourglass*, avec un bruit plus faible au stade phylotypique. Ce schéma peut s'expliquer par un contrôle du bruit plus intense induit par les modifications des histones à ce stade. De plus, je propose que les modifications d'histones contribuent à la robustesse mutationnelle des éléments régulateurs, et donc aux niveaux d'expression conservés. Ces résultats permettent de mieux comprendre le rôle de la robustesse dans les schémas phénotypiques et génétiques de conservation évolutive dans le développement des animaux.

Acknowledgements

I would like to thank my advisor, Prof. Marc Robinson-Rechavi, for supporting every aspect of my graduate study. His nice personality and strong sense of responsibility make the work with him a very enjoyable and fruitful experience. Thanks to his expertise and inspiration, I can always get constructive feedbacks when I face with difficulties in my research. As a very open-minded person, he always encourages and helps me to develop my own projects. He is an ideal PhD advisor.

I would like to thank my collaborators from EPFL: Bart Deplancke, Michael Frochaux, and Vincent Gardeux; from EMBL Heidelberg: Eileen Furlong, Charles Girardot, and James Reddington. They helped me significantly in my research. Thanks a lot for all the support and insightful discussions.

I thank my committee members, Dr. David Garfield and Prof. Günter Wagner, for accepting to review my PhD thesis, and for their insightful and critical comments on my research. Thanks a lot for the time they spent on traveling to Lausanne.

I also thank current and former members of the Robinson-Rechavi Group. Particularly, I would like to thank Julien Roux, for helping me on programing in the first year of my graduate study. I would also like to thank David Laloum, Sacha Laurent, and Julien Wollbrett, for helping me to understand French.

Finally, I would like to thank my wife, Xuemei, for following me to Lausanne. This is not easy for us when we come to Switzerland from a different country with different culture and language. She takes care of me in almost every aspect of my daily life. Thanks a lot for her patient encouragement and enormous support over the years.

1 Introduction

1.1 Embryonic Conservation: the early conservation model and the hourglass model

Embryological development has long been characterized by deep conservation. Animals that belong to the same phylum are considered to share a group of structural and developmental characteristics, the so called basic body plan (Wallace 2000; Valentine 2004; Irie and Kuratani 2014). For example, some anatomic features such as segmented trunk muscles, vertebrae along the anteroposterior axis and dorsally located central nervous system, were shared by all vertebrates (Irie and Kuratani 2014). Similarly, various other phyla also have their own conserved basic body plans. For example, arthropods share a set of anatomic structures such as jointed legs, exoskeleton made of chitin and segmented bodies (Zrzavy and Stys' 1997).

Based on the dynamic of morphological conservation throughout development, two main models, "early conservation model" or "funnel-like model" (Figure 1A) and "hourglass model" (Figure 1B), have been proposed to link the interplay between developmental processes and evolutionary changes. The early conservation model (Riedl 1978; Irie and Kuratani 2014), modified from the "third law" of Von Baer (1828), suggests that the earlier embryos, among species from the same phylum, have higher morphological similarities than the later embryos. In the 1990s, based on renewed morphological observations and on the finding of conserved expression profile of Hox genes along the embryonic anteroposterior axis, Duboule (1994) and Raff (1996) proposed the hourglass model . This model suggests that embryos from the "phylotypic period" (Richardson 1995), a group of stages in middle development, have higher morphological similarities than embryos from early or late development. In vertebrate, this period is corresponding to organogenesis. Notably, however, the hourglass model was not supported by one of the most comprehensive morphological studies on vertebrate embryos (Bininda-Emonds et al. 2003).

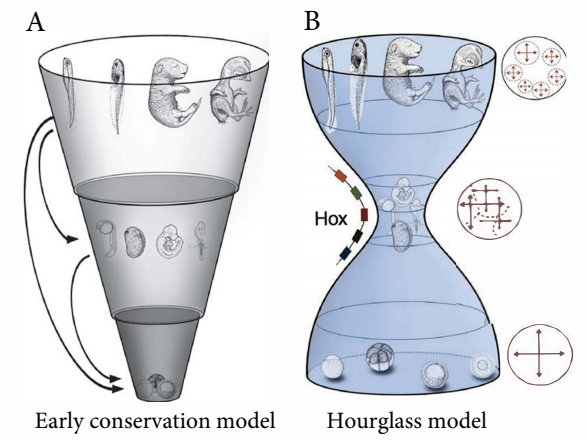

Figure 1: The two major Evo-Devo models.
In both models, the development proceeds from the bottom to the top; the width represents the similarities of embryonic morphology among species within the same phylum. Figure adapted from Irie and Kuratani (2011)

A. *The early conservation model suggests that the highest morphological similarities occurred at the early development, followed by a progressive evolutionary divergence over ontogeny. The arrows indicate the development of later stages depend on earlier stages, so earlier stage is more conserved.*
B. *The hourglass model suggests the highest morphological similarity occurred at the middle development (the "phylotypic period": organogenesis). Circles on the right of the model indicate the interaction of signalling network in early, middle and late development. the highest conservation in the middle development is caused by highly interacted signalling network and by co-linear Hox cluster gene expression in time and space.*

1.2 Both models have been supported at a molecular level

Since there are no consistent results from morphological studies in terms of embryonic conservation (Von-Baer 1828; Denis Duboule 1994; Richardson et al. 1997; Bininda-Emonds et al. 2003), many studies have tried to find evidences from molecular level. This is because evolutionary changes in the genome can cause changes in development and to morphology, which are subject to natural selection. This leads developmental processes to constrain the evolution of genome. Overall, both models have been supported by recent molecular level studies based on different properties (such as sequence divergence, duplication, phyletic age, or expression divergence), different species, and different analysis methods.

1.2.1 Conservation of genome evolution across development

With the advent of post-genomic era, studies in various model species (nematode worm *Caenorhabditis elegans*, fly *Drosophila melanogaster*, zebrafish *Danio rerio*, and mouse *Mus musculus*) have shown that the timing of expression of genes in development influences their

evolution (sequence divergence, duplication, or phyletic age) (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). Roux and Robinson-Rechavi (2008) conducted one of the first systematic study in mouse and zebrafish. They found genes expressed at earlier stage have a higher proportion of essential genes, of slowly evolving genes and of singletons reverted after whole genome duplication, supporting the early conservation model. However, among these studies, the results are inconclusive based on different analysis methods. For example, Domazet-Loso and Tautz (2010) reported that transcriptomes of middle development stages of zebrafish have higher proportion of old genes than transcriptomes of early and late development stages, supporting the hourglass model. In this study, they performed a "Transcriptome Age Index" analysis: use each gene's expression level as a weight to calculate the weighted mean value of gene age. Since highly and lowly expressed genes can span several orders of magnitude (Piasecka et al. 2013), patterns identified by using expression signal as weight can be driven by a subset of very highly expressed genes, or even by a few outliers. For example, with a standard log transformation of microarray signal intensities, Piasecka et al. (2013) re-analyzed the same data and found that the highest proportion of old genes was in transcriptomes of early development stages, supporting the early conservation model. For the details of different methods used to study developmental constraints on genome evolution, please check the introduction part of chapter 2.

Because different methodologies were adapted in previous studies, their conclusions cannot be compared consistently, making a biological conclusion concerning developmental constraints across species and features difficult. In chapter 2, based on the log-transformed "Transcriptome Age Index" approach, I tested the conservation patterns (early and hourglass) of three evolutionary traits of protein coding genes (strength of purifying selection on protein sequences, phyletic age, and duplicability) in four species: nematode worm, fly, zebrafish and mouse. Generally, I found that different genomic properties supporting different models. In all studied species, the sequence evolution follows the hourglass model, but the evolution of phyletic age and of duplicability follow the early conservation model.

1.2.2 Conservation of transcriptome evolution across development

In terms of transcriptome evolution, however, all studies in different phyla agree that middle development (the "phylotypic period") has lower expression divergence than in early and late development (Kalinka et al. 2010; Irie and Kuratani 2011; Levin et al. 2012; Hu et al. 2017; Zalts and Yanai 2017), supporting the hourglass model. One of the pioneer study was conducted by Kalinka et al. (2010), quantifying expression divergence at different stages during embryogenesis in six *Drosophila* species. They found the expression divergence follows an hourglass pattern with the minimal divergence at the extended germband stage (8-10 hours after laying egg), generally regarded as the phylotypic stage of arthropod (Sander 1976). Then, in vertebrate, Irie and Kuratani (2011) identified pharyngula as the most conserved stage in terms of gene expression divergence. In *Caenorhabditis*, Levin et al. (2012) identified ventral enclosure is the stage with most conserved expression divergence. Since the expression divergence recapitulates the morphological hourglass, it highlights the hypothesis that changes in gene expression play a central role in the morphological differences between species (King and Wilson 1975; Carroll 2008).

Although transcriptomic comparisons throughout developmental time across species underwent widespread studies, the roles of gene regulation on the evolution of expression during embryogenesis remains to be elucidated. Specifically, are regulatory elements involved in the phylotypic stage are also more conserved than in other stages? Few studies have tried to study this question (Nord et al. 2013; Piasecka et al. 2013), but there is no direct evidence supporting stronger conservation at the phylotypic stage. For example, based on the findings that the neighbourhood of genes specifically expressed in the phylotypic stage to be enriched with transposon free regions and long conserved non-coding elements, Piasecka et al. (2013) suggested that phylotypic stage has higher conservation of gene regulation. However, there are several potential *caveats* for this study. First, they didn't identify regulatory elements, they used neighbourhood of gene as an approximation instead. Second, they didn't directly compare the regulatory elements turn over between different species, they compared the sequence conservation of orthologous regions instead.

To systematically answer this question, I worked in collaboration with the Furlong lab, on DNase-seq across multiple matched embryonic development stages in two distant *Drosophila* species: *D. melanogaster* and *D. virilis*. Based on the DNase-seq dataset, I identified putative

regulatory elements (enhancers and promoters) and performed downstream evolutionary analyses. I found the phylotypic stage has higher proportion of conserved regulatory elements and less proportion of stage specific gain regulatory elements, indicating stronger conservation of gene regulation at the phylotypic stage. This study is presented in chapter 3.

1.3 Molecular conservation can be caused by strong developmental constraints and/or by weak positive selection

Molecular conservation may be the result of developmental constraints, defined as "a bias on the production of variant phenotypes or a limitation on phenotypic variability caused by the structure, character, composition, or dynamics of the developmental system"(Smith et al. 1985). Both early conservation and hourglass models were interpreted as a result developmental constraints. For early conservation model, Garstang (1922) and Riedl (1978) suggested that the development of later stages is dependent on earlier stages (Figure 1A), so higher constraints should be found in the earlier stages of development, so more conserved. In the hourglass model, Raff (1996) suggested a high inter-dependence in signalling among developmental modules in middle development (Figure 1B), so higher constraints in the middle development, so more conserved. However, conservation can also be the result of variations being less visible to positive selection (Garfield and Wray 2009). For example, Darwin interpreted Von Baer's "third law" (Von-Baer 1828) via a "selection opportunity" framework (Darwin 1871; Artieri et al. 2009). He proposed that highly divergent late development could also be driven by adaptive evolution (positive selection), at least in part. This could be due to the greater diversity of challenges to which natural selection needs to respond in juvenile and adult life than in early and mid-development.

1.3.1 Sequence evolution across development: development constraints and positive selection

For sequence evolution, development constraints can be manifested by purifying selection. Although there are many studies which have investigated the interplay between development and sequence evolution (Castillo-Davis and Hartl 2002; Cutter and Ward 2005; Davis et al. 2005; Hazkani-Covo et al. 2005; Roux and Robinson-Rechavi 2008; Artieri et al. 2009; Piasecka et al. 2013; Drost et al. 2015), there is still an open question that whether higher sequence conservation in some stages is caused by stronger development constraints or by weaker positive selection. By using branch site models (see Methodology introduction) (Yang

and Nielsen 2002; Zhang et al. 2005), I distinguished the relative contribution of purifying selection and positive selection on sequence evolution. I found genes expressed at the phylotypic stage under stronger purifying selection, indicating stronger developmental constraints on sequence evolution in the phylotypic stage. This part of results is presented in chapter 2.

Although the phylotypic stage is under stronger purifying selection in terms of sequence evolution, it could still accumulate more adaptive nonsynonymous substitutions than other stages of development (higher adaptive evolution). Richardson (1999) proposed the "adaptive penetrance hypothesis", suggesting that changes occurring in the "phylotypic period" have a higher probability to generate adaptive traits in adult. This is because the "phylotypic period" is when many adult traits are specified. So, I tested whether there are differences not just in purifying selection but also in the rates of adaptive substitutions between stages. Overall, there is a consistent signal that positive selection mainly affects genes and pathways expressed in late embryonic development and in adult, implying that the evolution of embryogenesis is mostly conservative, with most adaptive evolution affecting some stages of 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. This part of results is presented in chapter 4.

1.3.2 Expression evolution across development: development constraints and positive selection

For expression evolution, development constraints can be manifested by purifying selection or by mutational robustness (Zalts and Yanai 2017). The conserved expression can be explained by strong purifying selection on regulation programme or by strong buffering on gene regulatory variations. In order to test whether the hourglass expression divergence can be explained by developmental constraints or not, Zalts and Yanai (2017) compared the expression variations of 20 *C. elegans* mutation accumulation strains across embryonic development and found that the phylotypic stage has lower expression variation. Since mutation accumulation experiments mostly remove the effect of positive selection, this study indicates that developmental constraints (both purifying selection and mutational robustness) can contribute to the hourglass model of expression evolution.

However, the contribution of positive selection on the hourglass expression divergence is still unknown. This is because of the challenges of detecting positive selection on gene regulation (Zhen and Andolfatto 2012) (see Methodology introduction). One of the major challenges is the lack of reliable neutral reference sites to infer positive selection. To, answer this question, I first developed an in silico mutagenesis based method to detect positive selection on regulatory elements. This part of results is presented in chapter 5. And then, I applied this method to detect signatures of positive selection on stage specific enhancers in fly, and found that the phylotypic stage has a lower proportion of enhancers with evidence of positive selection. These results suggest that the lower expression divergence at the phylotypic stage could also be explained by lacking of adaptive evolution on gene regulation. This part of results is presented in chapter 3.

1.4 Developmental robustness and stochastic gene expression

The phenotypic differences between individuals mainly result from variations in underlying genetic and environmental factors. However, phenotypes can also vary among isogenic individuals in homogenous environments, suggesting contributions of stochastic effects in generating phenotypic diversity (Raj et al. 2010; Li et al. 2019). Gene expression noise (Elowitz et al. 2002; Blake et al. 2003; Raser and O'Shea 2005), variability of gene expression among genetically identical individuals under uniform conditions, is one of the most important stochastic processes in the mapping of genotype to phenotype.

Gene expression noise consists of intrinsic noise and extrinsic noise (Elowitz et al. 2002). The intrinsic noise mainly originates from the 'open' and 'closed' chromatin transitions associated with the 'transcriptional burst' process (Figure 2), and from the inherent randomness of biochemical reactions with low molecule numbers (Becskei et al. 2005; Raj and van Oudenaarden 2008). The extrinsic noise is generally caused by the propagation of intrinsic noise from upstream genes in the gene regulatory network, and by variations in cellular environment (Elowitz et al. 2002; Pedraza and van Oudenaarden 2005), such as random distribution of molecules at cell division. Although noise can be beneficial in some cases (Raj and van Oudenaarden 2008; Zhang et al. 2009; Wolf et al. 2015; Duveau et al. 2018), it is more generally harmful (Fraser et al. 2004; Batada and Hurst 2007; Lehner 2008; Chen and Zhang 2016), thus subject to purifying selection.

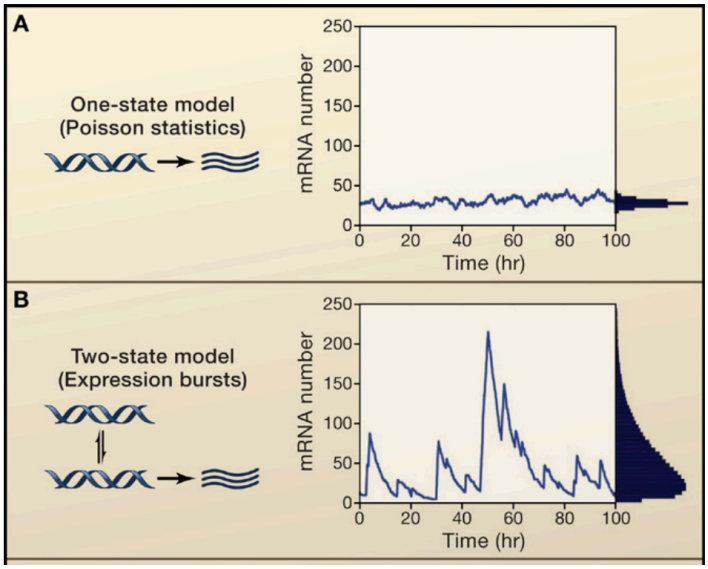

Figure 2: The contribution of transcriptional bursts to expression noise
A. Transcription without bursts with a relatively small amount of expression noise.
B. Transcription with bursts can cause significantly higher expression noise, even when the mean transcript number is similar with Transcription without bursts. Figure adapted from Raj and van Oudenaarden (2008)

Expression noise can be minimized in many ways, generally falling into two categories: transcriptional based regulation and post-transcriptional based regulation. Based on the "transcriptional burst" model, the expression level is decided by both burst size and burst frequency, while expression noise is mainly decided by burst frequency (Hornung et al. 2012; Dey et al. 2015; Wu et al. 2017). Thus, transcriptional based regulation of noise functions by increasing burst frequency. For example, low noise can be achieved by the clustering of genes in nucleosome depleted regions (Batada and Hurst 2007; Dey et al. 2015; Chen and Zhang 2016), the use of nucleosome-disfavoured promoters (Hornung et al. 2012; Dadiani et al. 2013; Sharon et al. 2014), or an increase in euchromatic histone modifications (Weinberger et al. 2012; Benayoun et al. 2014; Wu et al. 2017; Nicolas et al. 2018). The post-transcriptional based mechanisms function by buffering or filtering noise. For example, noise can be buffered by some specific regulatory motifs, such as negative feedback loops and incoherent feed-forward loops (Becskei and Serrano 2000; Alon 2007; Chalancon et al. 2012), or by microRNAs (Hornstein and Shomron 2006; Wu et al. 2009; Schmiedel et al. 2015; Schmiedel et al. 2017), or filtered by using cellular compartmentalization (Stoeger et al. 2016).

Precise regulation of gene expression appears to be notably important during animal development (Waddington 1942), since the underlying gene expression programs must ensure that different cells appear at the correct time and location to make a functional embryo. However, the development process inevitably has to deal with inherent randomness (Oates 2011), especially expression noise. This conflict between precision and stochasticity raises questions, such as how is expression noise controlled or its effects buffered during development? Is expression noise equally distributed across development or are some stages more robust to gene expression stochasticity? This is especially interesting in relation to the "phylotypic" stage. Whether the low expression divergence in the phylotypic stage is partially driven by mutational robustness is still an open question. Since genes that confer robustness to stochastic fluctuations can also buffer the effects of genetic variation (Meiklejohn and Hartl 2002; Lehner 2010), mutational robustness could be a side effect of selection against stochastic variation of gene expression at a key stage of development.

To answer these questions, I designed an experiment, carried out in collaboration with the Deplancke lab, to build a single embryo transcriptome time series of 150 embryos across fly embryonic development. This uniquely high number of isogenic replicates allowed me to find that expression noise recapitulates an hourglass model (D Duboule 1994; Raff 1996), with the minimum of noise at the phylotypic stage. I also shown that gene expression at this stage is robust both to stochastic variations and to genetic variation. Finally, I explained this robustness by the role of histone modifications over development. These results provide insight into the interplay between chromatin environment, stochastic robustness, mutational robustness, and purifying selection in the context of animal development. The results of this part are presented in chapter 6.

1.5 Methodology introduction

1.5.1 Methods for detecting positive selection on gene sequence

Detecting positive selection is about rejecting evolutionary neutrality as an explanation for change (Vitti et al. 2013). For detecting positive selection at the macro-evolutionary molecular level, one of the best established methods is using the ratio ω of dN (the number of non-synonymous substitutions per non-synonymous site) to dS (the number of synonymous substitutions per synonymous site) (Yang and Nielsen 1998; Hurst 2002). Because

synonymous changes are assumed to be functionally neutral, their evolution rate can be used as an approximation of neutral evolution rate. So, $\omega>1$ indicates evidence of positive selection; $\omega<1$ indicates evidence of purifying selection; $\omega=1$ indicates evidence of neutral evolution. Since adaptive changes can happen only at a few codon sites and at a few phylogenetic lineages, branch-site models were proposed to allow the ω ratio to vary both among codon sites and among lineages (Yang and Nielsen 2002; Zhang et al. 2005). In case of detecting positive selection among sites, the branch-site test contrasts two hypotheses: the null hypothesis is that no positive selection occurred in the phylogenetic branch of interest, and the alternative hypothesis is that at least some codons experienced positive selection. Additionally, the branch site models can also estimate the dN/dS ratio of the subset of codons which have evolved under purifying selection. So, it can separate both purifying selection and positive selection on sequence evolution. For detecting positive selection at the micro-evolution level, polymorphism based methods such as frequency spectrum, linkage disequilibrium and population differentiation were developed (Vitti et al. 2013). The main idea is that, under positive selection, a beneficial allele and its surrounding variants will reach to high frequency together (Smith and Haigh 1974). So, regions with population-wide reduction in genetic diversity can be interpreted as the signal of positive selection. 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). For detecting polygenic adaptation, Daub et al. (2013; 2017) developed a gene set enrichment test. Briefly, the evidence of polygenic adaptation is inferred by comparing the sum of adaptive signals across a gene set with a null distribution generated by random sampling.

1.5.2 Methods for detecting positive selection on gene regulation

Generally, detecting positive selection for noncoding regions is still a big challenge (Zhen and Andolfatto 2012) . A major reason is the lack of a null model of neutral evolution for non-coding regions. Currently, the methods used to detect positive selection on regulatory elements can be divided into two categories. The first category is looking for conserved noncoding regions with lineage-specific accelerated evolutionary rates by scanning multiple genomes (Pollard et al. 2006; Prabhakar et al. 2006; Gittelman et al. 2015). A major caveat for this method is that these identified accelerated regions may result from biased gene conversion (Galtier and Duret 2007) rather than from positive selection. The second category is under the McDonald–Kreitman test framework (McDonald and Kreitman 1991; Ludwig and Kreitman

1995; Andolfatto 2005; Arbiza et al. 2013; Gronau et al. 2013), which originally developed for detecting positive selection on protein sequences by integrating both divergence and polymorphism information. Since mutations under positive selection can spread through a population rapidly, they don't contribute to polymorphism but do increase divergence. So, a higher divergence relative to polymorphism ratio in regulatory elements than in neutrally evolving sites (such as synonymous sites or the fastest evolving sites of short introns) can indicate adaptive evolution. However, this approach is expected to lack power on individual elements, since many regulatory elements are too short to have enough number of variations(Andolfatto 2005). In addition, it can also be biased by the choice of appropriate neutral reference sites (Zhen and Andolfatto 2012).

To overcome these disadvantages, I developed an in silico mutagenesis approach to detect positive selection on regulatory elements based on regulatory activity changes. The effects of substitutions on regulatory activity change can be accurately predicted by deltaSVM (Lee et al. 2015), a machine leaning based method to predict the effects of regulatory variations de novo from sequence (Figure 3). My approach does not assume a subset of neutral sites and does not compare sequence evolution rate, but instead considers the effects of variations on regulatory activities (Berg et al. 2004; Moses 2009; Smith et al. 2013). I compared the observed regulatory activity change to the expectation under a no-selection model to assess when neutrality can be rejected. This part of results is presented in chapter 5.

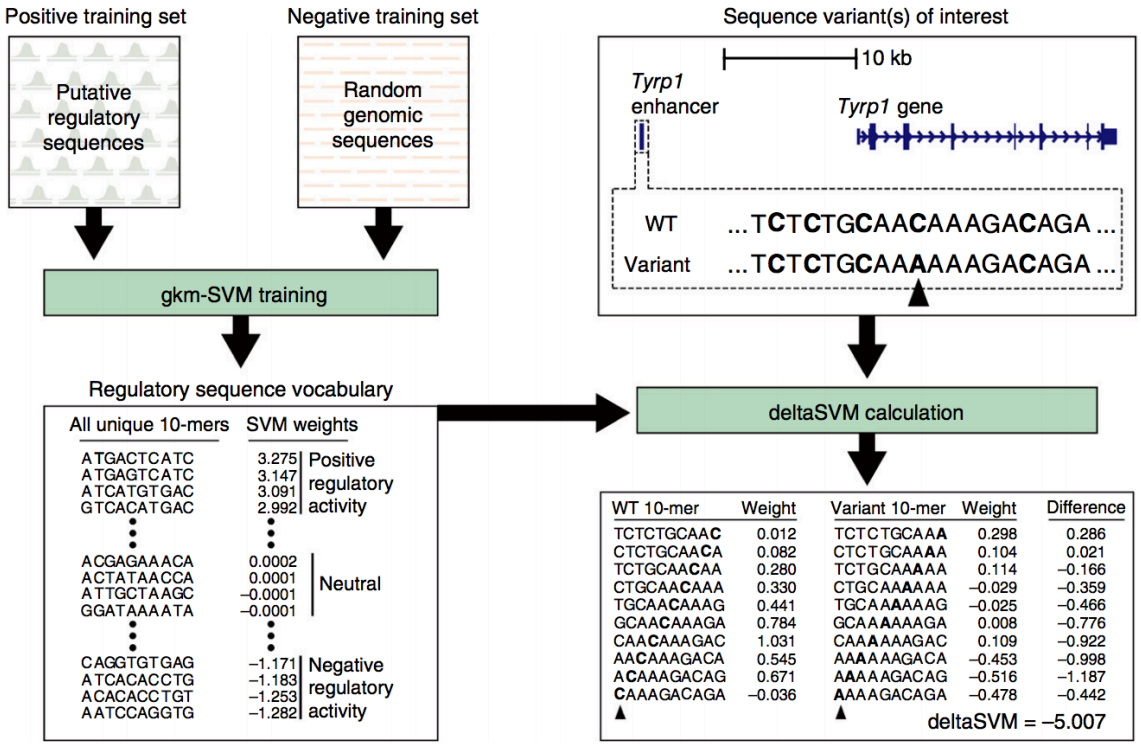

Figure 3 Overview of the deltaSVM method
The procedures of calculating deltaSVM. First, training a gapped k-mer support vector machine (gkm-SVM) classifier by using putative regulatory sequences (transcription factor biding sites, for example) as a positive training set and using random sequences as a negative training. Second, using the gkm-SVM to generate the weights of all possible 10-mers (the regulatory sequence vocabulary). Each 10-mer's weight represents its contribution to the prediction of regulatory function. Third, based on the regulatory sequence vocabulary, the impact of a single-nucleotide substitution in a melanocyte enhancer of Tyrp1 can be predicted (the right part of the figure). WT, wild type. Figure adapted from Lee et al. (2015)

In the following chapters, I will give several examples to show the interplay between developmental processes and evolutionary changes. Chapter 2 presents how development constrains the evolution of protein-coding gene sequence, duplicability, and phyletic age in four model species: nematode worm, fly, zebrafish, and mouse. Chapter 3 presents the evidence of positive selection on protein-coding genes in relation to their timing of expression over development in fly, zebrafish, and mouse. Chapters 2 and 3 are published. Chapter 4 presents the evolutionary conservation, divergence and positive selection on gene regulation under the context of *Drosophila* embryogenesis. Chapter 5 presents the in silico mutagenesis approach to detect positive selection on regulatory elements which I developed, and applied it to study the regulatory adaptive evolution on human brain. Chapter 6 reports how expression noise was

controlled during fly embryogenesis, and the contribution of mutational robustness on the hourglass expression divergence.

1.6 References

Alon U. 2007. Network motifs: theory and experimental approaches. Nat. Rev. Genet. 8:450–461.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature 437:1149–1152.

Arbiza L, Gronau I, Aksoy BA, Hubisz MJ, Gulko B, Keinan A, Siepel A. 2013. Genome-wide inference of natural selection on human transcription factor binding sites. Nat. Genet. 45:723–729.

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.

Batada NN, Hurst LD. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. Nat. Genet. 39:945–949.

Becskei A, Kaufmann BB, van Oudenaarden A. 2005. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. Nat. Genet. 37:937–944.

Becskei A, Serrano L. 2000. Engineering stability in gene networks by autoregulation. Nature 405:590–593.

Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, Devarajan K, Daugherty AC, Kundaje AB, Mancini E, et al. 2014. H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell 158:673–688.

Berg J, Willmann S, Lässig M. 2004. Adaptive evolution of transcription factor binding sites. BMC Evol. Biol. 4:42.

Berg JJ, Coop G. 2014. A Population Genetic Signal of Polygenic Adaptation. PLoS Genet. 10:e1004412.

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.

Blake WJ, KÆrn M, Cantor CR, Collins JJ. 2003. Noise in eukaryotic gene expression. Nature 422:633–637.

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

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

Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R, Babu MM. 2012. Interplay between gene expression noise and regulatory network architecture. Trends Genet. 28:221–232.

Chen X, Zhang J. 2016. The Genomic Landscape of Position Effects on Protein Expression Level and Noise in Yeast. Cell Syst. 2:347–354.

Cheng X, Hui JHL, Lee YY, Wan Law PT, Kwan HS. 2015. A “developmental hourglass” in fungi. Mol. Biol. Evol. 32:1556–1566.

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.
Dadiani M, van Dijk D, Segal B, Field Y, Ben-Artzi G, Raveh-Sadka T, Levo M, Kaplow I, Weinberger A, Segal E. 2013. Two DNA-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise. Genome Res. 23:966–976.
Darwin C. 1871. The descent of man, and selection in relation to sex. Princeton University Press
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.
Davis JC, Brandman O, Petrov DA. 2005. Protein evolution in the context of Drosophila development. J. Mol. Evol. 60:774–785.
Dey SS, Foley JE, Limsirichai P, Schaffer D V, Arkin AP. 2015. Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol. Syst. Biol. 11:806.
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 Denis. 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.
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.
Duveau F, Hodgins-Davis A, Metzger BP, Yang B, Tryban S, Walker EA, Lybrook T, Wittkopp PJ. 2018. Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. Elife 7.
Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic Gene Expression in a Single Cell. Science (80-.). 297:1183–1186.
Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. 2004. Noise Minimization in Eukaryotic Gene Expression.Ken Wolfe, editor. PLoS Biol. 2:e137.
Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. 23:273–277.
Garfield DA, Wray GA. 2009. Comparative embryology without a microscope: using genomic approaches to understand the evolution of development. J. Biol. 8:65.
Garstang W. 1922. The Theory of Recapitulation: A Critical Re-statement of the Biogenetic Law. J. Linn. Soc. London, Zool. 35:81–101.
Gittelman RM, Hun E, Ay F, Madeoy J, Pennacchio L, Noble WS, Hawkins RD, Akey JM. 2015. Comprehensive identification and analysis of human accelerated regulatory DNA. Genome Res. 25:1245–1255.
Gronau I, Arbiza L, Mohammed J, Siepel A. 2013. Inference of Natural Selection from Interspersed Genomic Elements Based on Polymorphism and Divergence. Mol. Biol. Evol. 30:1159–1171.
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.
Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. Nat. Genet. 38:S20–S24.
Hornung G, Bar-Ziv R, Rosin D, Tokuriki N, Tawfik DS, Oren M, Barkai N. 2012. Noise-mean relationship in mutated promoters. Genome Res. 22:2409–2417.
Hu H, Uesaka M, Guo S, Shimai K, Lu T-M, Li F, Fujimoto S, Ishikawa M, Liu S, Sasagawa Y, et al. 2017. Constrained vertebrate evolution by pleiotropic genes. Nat. Ecol. Evol. 1:1722–1730.
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, 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.
King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107–116.
Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, Beer MA. 2015. A method to predict the impact of regulatory variants from DNA sequence. Nat. Genet. 47:955–961.
Lehner B. 2008. Selection to minimise noise in living systems and its implications for the evolution of gene expression. Mol. Syst. Biol. 4:170.
Lehner B. 2010. Genes Confer Similar Robustness to Environmental, Stochastic, and Genetic Perturbations in Yeast.Polymenis M, editor. PLoS One 5:e9035.
Levin M, Hashimshony T, Wagner F, Yanai I. 2012. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev. Cell 22:1101–1108.
Li X, Zhao Z, Xu W, Fan R, Xiao L, Ma X, Du Z. 2019. Systems Properties and Spatiotemporal Regulation of Cell Position Variability during Embryogenesis. Cell Rep. 26:313-321.e7.
Ludwig MZ, Kreitman M. 1995. Evolutionary dynamics of the enhancer region of even-skipped in Drosophila. Mol. Biol. Evol. 12:1002–1011.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654.
Meiklejohn CD, Hartl DL. 2002. A single mode of canalization. Trends Ecol. Evol. 17:468–473.
Moses AM. 2009. Statistical tests for natural selection on regulatory regions based on the strength of transcription factor binding sites. BMC Evol. Biol. 9:286.
Nicolas D, Zoller B, Suter DM, Naef F. 2018. Modulation of transcriptional burst frequency by histone acetylation. Proc. Natl. Acad. Sci. U. S. A. 115:7153–7158.
Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, et al. 2013. Rapid and Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development. Cell 155:1521–1531.
Oates AC. 2011. What's all the noise about developmental stochasticity? Development

138:601–607.
Pedraza JM, van Oudenaarden A. 2005. Noise propagation in gene networks. Science 307:1965–1969.
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.
Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, Siepel A, Pedersen JS, Bejerano G, Baertsch R, et al. 2006. Forces Shaping the Fastest Evolving Regions in the Human Genome. PLoS Genet. 2:e168.
Prabhakar S, Noonan JP, Paabo S, Rubin EM. 2006. Accelerated Evolution of Conserved Noncoding Sequences in Humans. Science (80-.). 314:786–786.
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. University of Chicago Press.
Raj A, van Oudenaarden A. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135:216–226.
Raj A, Rifkin SA, Andersen E, van Oudenaarden A. 2010. Variability in gene expression underlies incomplete penetrance. Nature 463:913–918.
Raser JM, O'Shea EK. 2005. Noise in gene expression: origins, consequences, and control. Science 309:2010–2013.
Richardson MK. 1995. Heterochrony and the phylotypic period. Dev. Biol. 172:412–421.
Richardson MK. 1999. Vertebrate evolution: the developmental origins of adult variation. Bioessays 21:604–613.
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.
Sander K. 1976. Specification of the Basic Body Pattern in Insect Embryogenesis. Adv. In Insect Phys. 12:125–238.
Schmiedel J, Marks DS, Lehner B, Bluthgen N. 2017. Noise control is a primary function of microRNAs and post-transcriptional regulation. bioRxiv:doi.org/10.1101/168641.
Schmiedel JM, Klemm SL, Zheng Y, Sahay A, Blüthgen N, Marks DS, van Oudenaarden A. 2015. Gene expression. MicroRNA control of protein expression noise. Science 348:128–132.
Sharon E, van Dijk D, Kalma Y, Keren L, Manor O, Yakhini Z, Segal E. 2014. Probing the effect of promoters on noise in gene expression using thousands of designed sequences. Genome Res. 24:1698–1706.
Smith JD, McManus KF, Fraser HB. 2013. A Novel Test for Selection on cis-Regulatory Elements Reveals Positive and Negative Selection Acting on Mammalian Transcriptional Enhancers. Mol. Biol. Evol. 30:2509–2518.
Smith JM, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D, Wolpert L. 1985. Developmental Constraints and Evolution: A Perspective from the Mountain Lake Conference on Development and Evolution. Q. Rev. Biol. 60:265–287.
Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23–35.
Stoeger T, Battich N, Pelkmans L. 2016. Passive Noise Filtering by Cellular Compartmentalization. Cell 164:1151–1161.
Valentine JW. 2004. On the origin of phyla. University of Chicago Press

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.
Waddington CH. 1942. Canalization of development and genetic assimilation of acquired characters. Nature 150:563–565.
Wallace A. 2000. The origin of animal body plans : a study in evolutionary developmental biology. Cambridge University Press
Weinberger L, Voichek Y, Tirosh I, Hornung G, Amit I, Barkai N. 2012. Expression Noise and Acetylation Profiles Distinguish HDAC Functions. Mol. Cell 47:193–202.
Wolf L, Silander OK, van Nimwegen E. 2015. Expression noise facilitates the evolution of gene regulation. Elife 4.
Wu C-I, Shen Y, Tang T. 2009. Evolution under canalization and the dual roles of microRNAs: a hypothesis. Genome Res. 19:734–743.
Wu S, Li K, Li Y, Zhao T, Li T, Yang Y-F, Qian W. 2017. Independent regulation of gene expression level and noise by histone modifications.Przytycka TM, editor. PLOS Comput. Biol. 13:e1005585.
Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. 46:409–418.
Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19:908–917.
Zalts H, Yanai I. 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.
Zhang Z, Qian W, Zhang J. 2009. Positive selection for elevated gene expression noise in yeast. Mol. Syst. Biol. 5:299.
Zhen Y, Andolfatto P. 2012. Methods to detect selection on noncoding DNA. Methods Mol. Biol. 856:141–159.
Zrzavy J, Stys' P. 1997. The basic body plan of arthropods: insights from evolutionary morphology and developmental biology. Birkhluser Verlag.

2 Developmental constraints on genome evolution in four bilaterian model species

Jialin Liu, Marc Robinson-Rechavi

This article was published in Genome Biology and Evolution (2018) 10(9): 2266-2277.

2.1 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 strongly diverges between species, but debate on which developmental period is the most conserved. Here, based on a modified “Transcriptome Age Index” approach, i.e. weighting trait measures by expression level, 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: nematode worm *Caenorhabditis elegans*, fly *Drosophila melanogaster*, zebrafish *Danio rerio*, and mouse *Mus musculus*. In general, we found that both models can be supported by different genomic properties. Sequence evolution follows an hourglass model, but the evolution of phyletic age and of duplicability follow an early conservation model. Further analyses indicate that stronger purifying selection on sequences in the middle development are driven by temporal pleiotropy of these genes. In addition, we report evidence that expression in late development is enriched with retrogenes, which usually lack efficient regulatory elements. This implies that expression in late development could facilitate transcription of new genes, and provide opportunities for acquisition of function. Finally, in *C. elegans*, we suggest that dosage imbalance could be one of the main factors that cause depleted expression of high duplicability genes in early development.

2.2 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 evolution of 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 law” of Von Baer (1828) (as cited in Kalinka and Tomancak 2012), suggests that the highest morphological similarities among species from the same phylum occur 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 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). Of note, the higher divergence of late development could also be due to stronger adaptive selection, the "Darwin hypothesis" (Artieri et al. 2009). This question is independent of the pattern of constraints (early or hourglass) which are the focus of this study, and we explore it in a companion study (Liu and Robinson-Rechavi 2017).

Both main 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 and Yanai 2017). All of them suggested that middle development has the highest transcriptome conservation, i.e. the hourglass pattern. On the other hand, when animals are compared between different phyla, middle development has been reported to have the highest divergence (Levin et al. 2016), although this conclusion has been criticized on methodological grounds (Dunn et al. 2018). 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 genomic 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.

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 the hourglass model, Drost et al. (2015) proposed a "reductive hourglass test" which focuses on testing the presence of an hourglass pattern of divergence: high-low-high. For this, development can be divided into three periods (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 period is significantly lower than in early and late periods.

Overall, the transcriptome index analysis should be the best method to measure developmental constraints on the whole transcriptome, if care is taken to properly treat the expression values. 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 a 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 (ω_0), phyletic age, and duplicability (paralog number)), with three transformations of expression values (non-transformed, $\log_2$ transformed, and square root transformed). For *C. elegans*, the strength of purifying selection on coding sequences 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*. In general, we found results consistent with the hourglass model for sequence evolution, but with early conservation for phyletic age and paralog number, in the four species. In addition, $\log_2$ transformed transcriptome indexes are always consistent with square root transformed transcriptome indexes but not with non-transformed transcriptome indexes.

2.3 Results and Discussion

2.3.1 Effect of expression value transformation on transcriptome indexes

As mentioned in the Introduction, the pattern from a transcriptome index analysis may not reflect the global behavior of the transcriptome, but that of a small fraction of very highly expressed genes, or even of a few outliers. In order to systematically test this issue, we calculated 95% confidence intervals of transcriptome indexes based on $\log_2$ transformed, square root transformed, and non-transformed expression values (see Methods). Then, for the purpose of comparing the range of confidence intervals in the same scale, we plotted the ratio of upper to lower confidence interval boundary across development. Clearly, at a given confidence level (95% here), we can see that the ratio of non-transformed transcriptome indexes is much higher and more variable than transformed transcriptome indexes (Figure S2), indicating that the transcriptome indexes estimated from transformed expression are more stable. The most stable pattern comes from $\log_2$ transformed transcriptome indexes, although it is quite similar with square root transformation. We note that while log-transformation is

routine in most application of transcriptomics, many analyses of "hourglass" patterns use non transformed expression data.

In summary, although a subset of genes with dramatically different expression values in different stages could be interesting in some sense, when the goal is to investigate the general tendency of the transcriptome, log- or square-root-transformation for expression value is necessary and efficient to reach a stable estimation.

2.3.2 Variation of evolutionary transcriptome indexes across development

Here, based on $\log_2$ transformed expression values, we calculated transcriptome indexes for strength of purifying selection on coding sequences (ω_0), phyletic age, and duplicability (paralog number). 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, and there are many more stages sampled from late development, so we only compared the difference between early and middle 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, and hence we use a one-sided test. Notably, for early development, we did not consider the stages before the start of maternal to zygote transition (MZT, see Methods), because these stages are dominated by maternal transcripts.

For the transcriptome index of purifying selection on coding sequence (Transcriptome Divergence Index: TDI), we found that genes with stronger purifying selection tend to be more expressed at middle developmental stages, suggesting an hourglass pattern (Figure 1). However, for the transcriptome indexes of phyletic age (Transcriptome Age Index: TAI) and of paralog number (Transcriptome Paralog Index: TPI), we observed that genes with higher duplicability and younger phyletic age trend to be expressed at later developmental stages, which corresponds to early conservation (Figure 1). In addition, we also repeated these analyses based on square root transformed expression values (Figure S3) and on non-transformed expression values (Figure S4). In general, the results from square root transformation are highly consistent with those from $\log_2$ transformation, but not with those

from non-transformation. For example, with non-transformed expression data, the TPI in *C. elegans* became very noisy; the TAI in *D. rerio* changed from early conservation to hourglass pattern; or the TDI in *M. musculus* changed into an unexpected early divergence pattern. Finally, we confirmed our observations with other datasets (Figure S5). For *C. elegans, D. melanogaster* and *D. rerio*, we used a high resolution time series single embryo RNA-seq dataset (Levin et al. 2016). Since this dataset is without replicates, and is generated from single embryo, the transcriptome indexes are noisy and present extreme values in some time points. However, generally, all the results from the new dataset are consistent with the results from our previous datasets except the TDI in *D. rerio*. For the latter, we only observed the first two-thirds of an hourglass pattern in the new dataset. This is because the new dataset only covers embryo development, whereas the increased TDI in late development is driven by post-embryonic development stages. For *M. musculus*, we also confirmed our results based on a microarray dataset (Irie and Kuratani 2011) (Figure S5).

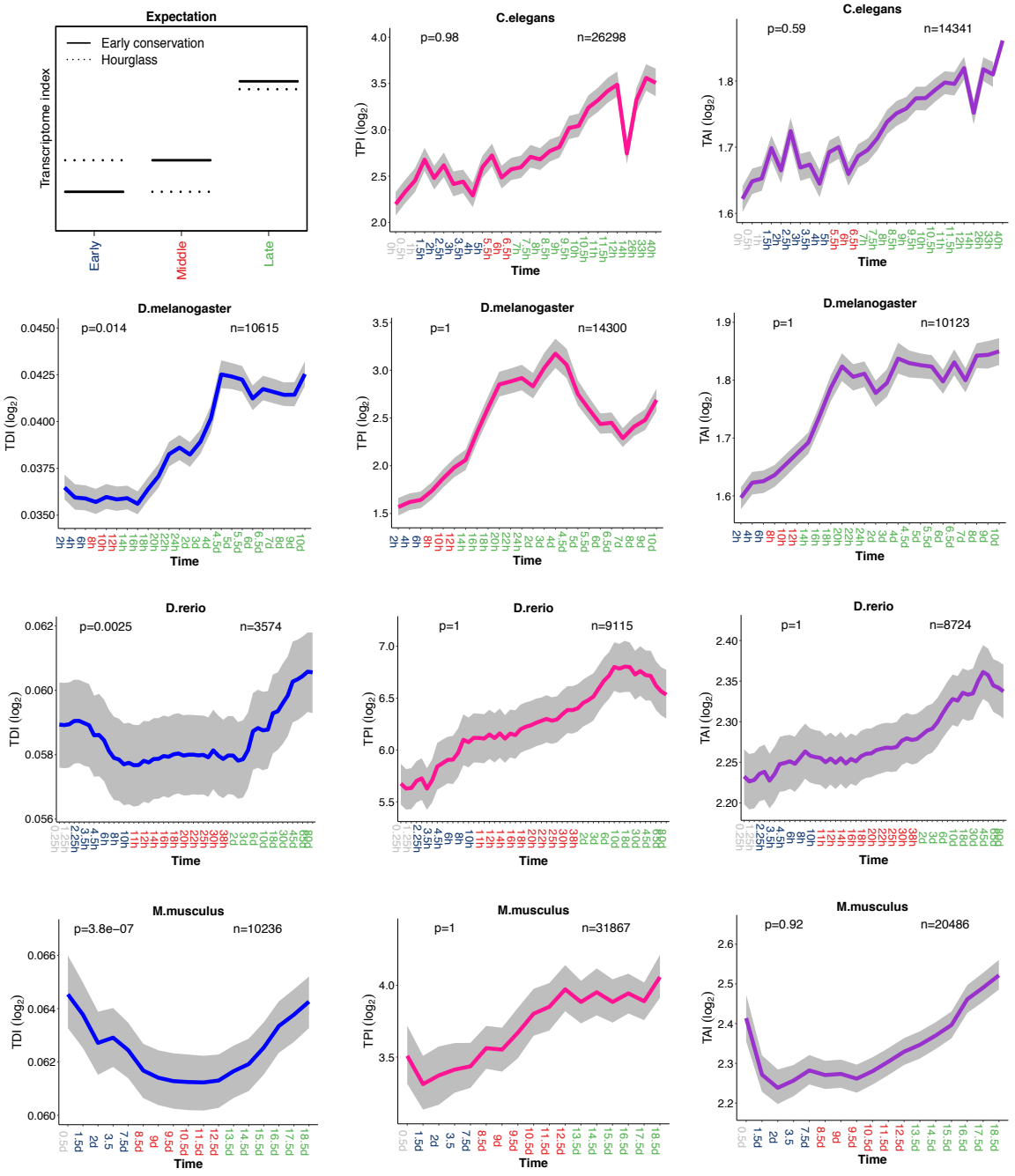

Figure 1: Evolutionary transcriptome indexes based on log_2 transformed expression values
The top-left plot schematically shows the expected transcriptome index patterns of the early conservation and hourglass models. For the other plots, grey, dark blue, red, and green marked time points in the x-axis represent stages before the start of MZT, early developmental stages, middle developmental stages, and late developmental stages, respectively. Transcriptome index of divergence (TDI): blue line; transcriptome index of paralog number (TPI): pink line; transcriptome index of phyletic age (TAI): purple line. The grey area indicates 95% confidence interval estimated from bootstrap analysis. The p-values for supporting the hourglass model (permutation test, early vs. middle development) are indicated in the top-left corner of each plot. The numbers of genes analyzed are noted in the top-right corner of each plot.

In *D. melanogaster*, we did not confirm the results of Drost et al. (2015) for phyletic age. After $\log_2$ transformation of expression data, we found an early conservation pattern instead of the hourglass pattern which they reported (Figure S6B). 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 transformation (Figure S6C). Of note, that fly phyletic age hourglass with uncorrected expression was also reported earlier (Domazet-Loso and Tautz 2010).

Overall, these results suggest that genes under strong purifying selection on their protein sequence trend to be expressed in middle development; it remains to be seen how much these observations extend to more arthropods or chordates. They also extend our previous observations that genes expressed earlier have a lower duplicability and an older age (Roux and Robinson-Rechavi 2008; Piasecka et al. 2013). In addition, it poses the question whether a pattern driven by the minority of very highly expressed genes is relevant to understanding Evo-Devo, which is generally driven by regulatory genes (Carroll 2008), such as transcription factors, with typically not very high and rather tissue-specific expression.

2.3.3 Expression of temporal pleiotropy genes across development

Several models have been proposed to explain why some developmental stages are more conserved than others, as presented in the Introduction. 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 middle developmental stages is caused by higher pleiotropy of genes expressed in these stages, pleiotropy being 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 is expected to be under stronger evolutionary constraints (Wagner and Zhang 2011).

For *C. elegans*, *D. melanogaster and M. musculus*, we defined FPKM>1 as expressed. For *D. rerio* we set genes with microarray signal rank in top 70% as expressed.

We calculated the proportion of potentially pleiotropic genes as expressed in more than 50% of development stages. In all the species, interestingly, we found pleiotropic genes enriched in middle development (Figure 2). In *D. melanogaster*, the evidence is weaker, because of the low sampling of early and middle development stages in the main dataset; but the pattern was clear in the high resolution single embryo RNA-seq dataset (Figure S7). We also found similar patterns when we define pleiotropic genes as expressed in more than 70% of development stages (Figure S8). Since the late development of *D. melanogaster* can clearly be divided into two periods with distinct patterns from the pleiotropy analysis, we removed the second period of late development, and found the same overall trend (Figure S9). 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 S10). Similar observations of higher temporal pleiotropy for genes in middle development in vertebrates were recently reported by Hu et al. (2017).

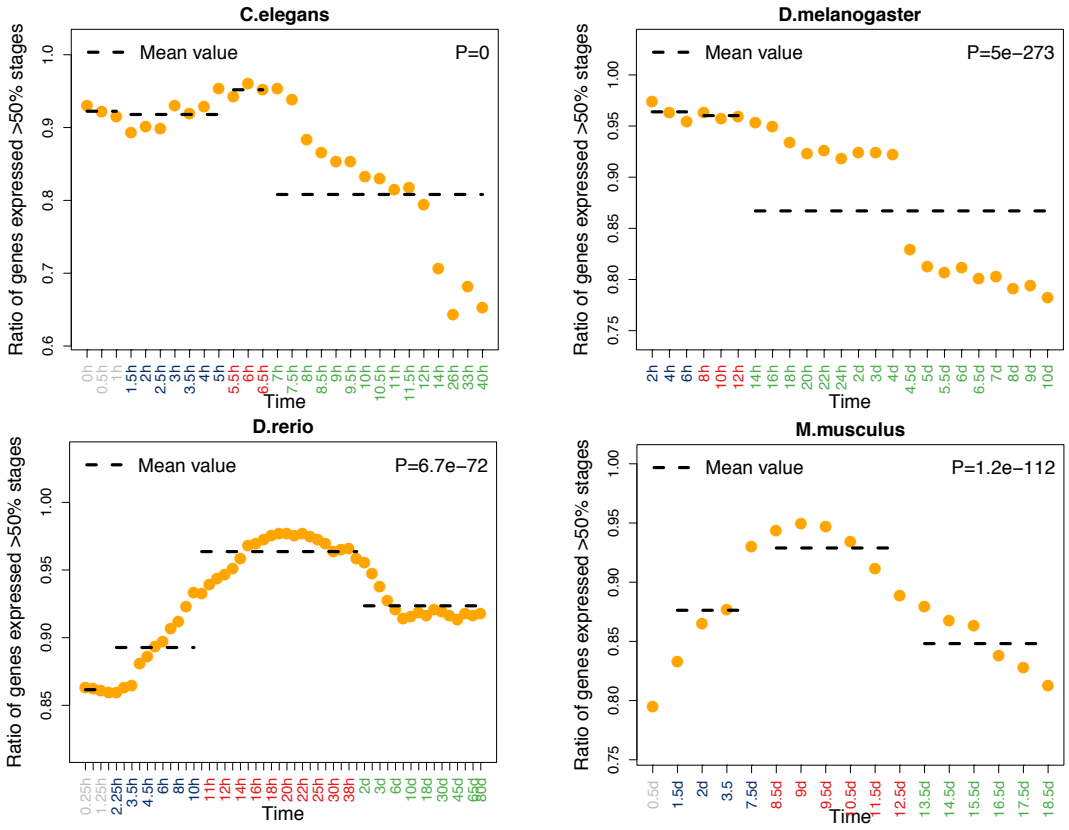

Figure 2: Proportion of temporal pleiotropic genes across development.
Grey, dark blue, red, and green marked time points in the x-axis represent stages before the start of MZT, early developmental stages, middle developmental stages and late developmental stages respectively. The proportion of temporal pleiotropic genes is plotted as orange circles. 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. The proportion of pleiotropic genes is defined as the number of pleiotropic genes divided by the number of all genes expressed in the corresponding stage.

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 ω_0 than non-pleiotropic genes (Figure 3).

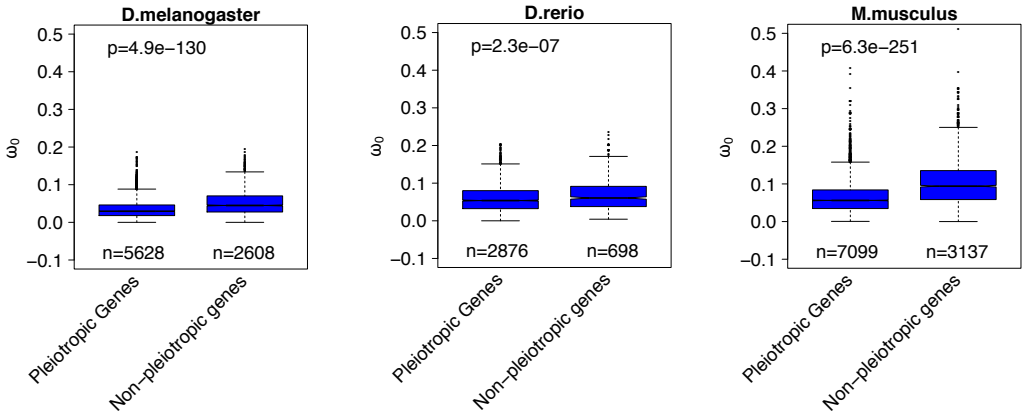

Figure 3: Comparison of ω_0 between temporal pleiotropic genes and non-pleiotropic genes.
The number of genes in each category is indicated below each box. The p-values from a Wilcoxon test comparing categories are reported above boxes. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range, or the maximum (respectively minimum) if no points are beyond 1.5 IQR (default behaviour of the R function boxplot).

In summary, we found that middle development stages with a higher proportion of broadly expressed genes are under stronger pleiotropic constraint on sequence evolution.

2.3.4 Higher expression of retrogenes in later development stages

In adult anatomy, young genes are mainly enriched for expression in testis (Kaessmann 2010). Two main factors have been proposed to explain this pattern. Firstly, permissive chromatin in testis facilitates the transcription of most genes, including new genes (Soumillon et al. 2013). The widespread expression in testis appears related to regulation of gene evolution rates based on transcription coupled repair (Xia et al. 2018). Secondly, as the most rapidly evolving organ at genomic level, there is least purifying selection acting on new genes expressed in testis (Kaessmann 2010). Is there a similar explanation for the ontogenic pattern of young genes

tending to be expressed in late development stages? As 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 to increased positive selection (Liu and Robinson-Rechavi 2017). Thus, we suggest that expression in late development might, like in testis, promote the fixation and functional evolution of new genes.

In order to test this, we analyzed the expression of retrogenes across development. Since retrogenes usually lack regulatory elements, most of them fail to acquire transcription and achieve function (Kaessmann et al. 2009). So, if late development, like testis, can facilitate the transcription of new genes, promoting their fixation, we should observe higher expression of retrogenes in later developmental stages. Because retrogenes have higher expression in testis, and testis is already differentiated after middle development, we excluded testis genes in our analyses for *D. melanogaster* and *M. musculus,* where the information of testis gene expression was available. As expected, the median expression of retrogenes is higher in late development (Figure 4), with a significant positive correlation. Generally, in *C. elegans*, *D. rerio and M. musculus*, the median expression progressively increases; in *D. melanogaster*, all the median values are 0 until stage 4 days, and then it progressively increases.

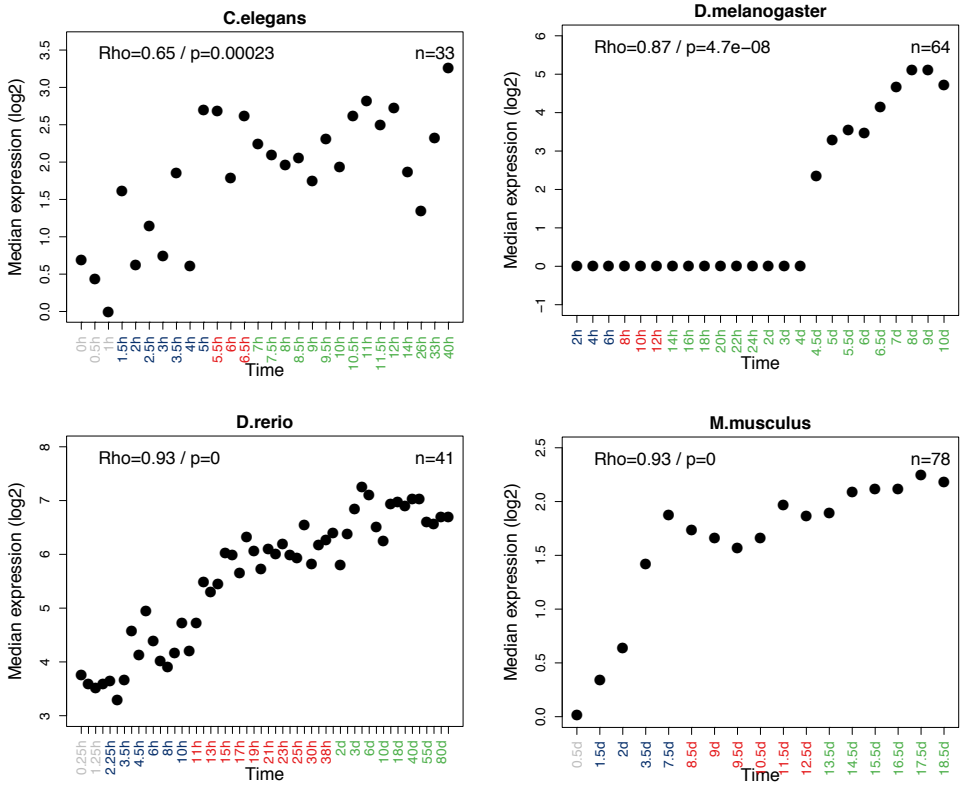

Figure 4: Expression of retrogenes in development.
Grey, dark blue, red, and green marked time points in the x-axis represent stages before the start of MZT, early developmental stages, middle developmental stages and late developmental stages respectively. A Spearman correlation was computed between development and median expression. The correlation coefficient (Rho) and p-value are indicated in the top-left corner of each plot. The numbers of genes analyzed are noted in the top-right corner of each plot.

These results confirm that late development could allow more transcription of new gene copies, which usually lack efficient regulatory elements and transcriptional activity. 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 adaptive functions in late development, 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 by these stages favoring the fixation of new genes.

2.3.5 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 *D. melanogaster* and in *D. rerio*, there is a non-monotonous pattern (increasing first, and then decreasing), but the overall trend is more connectivity with less duplicability. In *M. musculus*, however, we did not observe a significant relationship between connectivity and duplicability. Secondly, we calculated a transcriptome index of connectivity (Transcriptome Connectivity Index: TCI). In *C. elegans* and *M. musculus*, earlier developmental stages have higher TCI, which means that these stages trend to have higher expression of more connected genes (Figure 5). In *D. melanogaster* and in *D. rerio*, there is no clear pattern based on individual stages, but the mean TCI of each developmental period (early, middle and late development) also gradually decreases.

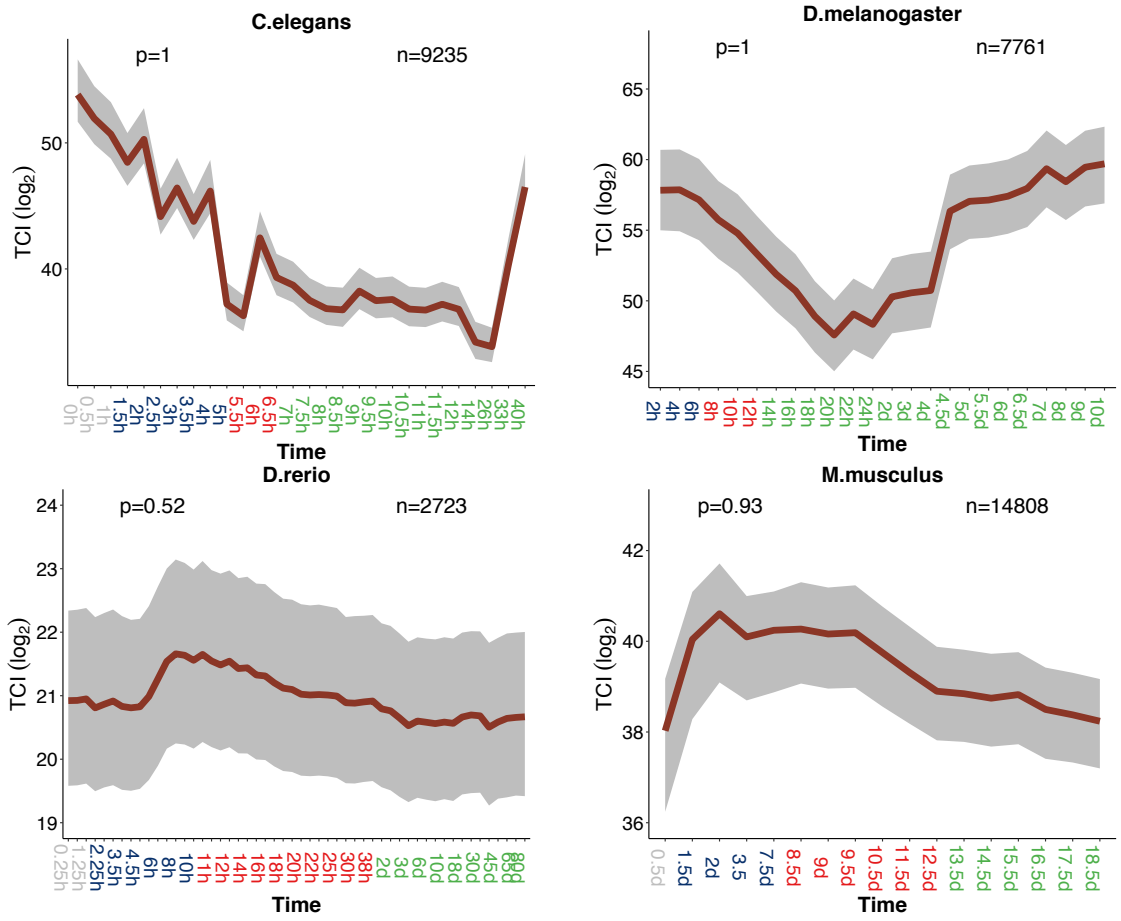

Figure 5: Transcriptome index of connectivity (TCI) across development.
Grey, dark blue, red, and green marked time points in the x-axis represent stages before the start of MZT, early developmental stages, middle developmental stages and late developmental stages respectively. TCI is plotted in dark red line. The grey area indicates 95% confidence interval estimated from bootstrap analysis. The p-values for supporting the hourglass model (permutation test, early vs. middle development) are indicated in the top-left corner of each graph. The numbers of genes analyzed are noted in the top-right corner of each plot.

These results indicate that, at least 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. 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.

2.4 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. The strength of purifying selection on protein sequence supports the hourglass model. Genes under stronger purifying selection are more expressed at middle development stages. Both duplicability and phyletic age support the early conservation model. Less duplicated genes and phyletically older genes are more expressed at earlier stages.

For processes, we investigated the potential causes of the observed patterns. The hourglass pattern of sequence evolution appears to be driven by temporal pleiotropy of gene expression. Genes expressed at middle development evolve under stronger temporal pleiotropic constraints. The enrichment in young phyletic age genes in late development might be related to a testis-like role of late development that facilitates the expression of retrogenes. Finally, in *C. elegans*, connectivity appears to be the main force explaining higher duplicability of genes expressed in later development.

2.5 Materials and Methods

Data files and analysis scripts are available on our GitHub repository: https://github.com/ljljolinq1010/developmental_constraints_genome_evolution

2.5.1 Expression data sets

Main datasets

For *C. elegans* and *D. melanogaster*, we downloaded processed (non-transformed but normalized) RNA-seq data from http://www.stat.ucla.edu/~jingyi.li/software-and-data.html (Li et al. 2014), which originally comes from (Gerstein et al. 2010; Graveley et al. 2011). For *D. rerio*, we used the processed (log-transformed and normalized) microarray data from our previous study (Piasecka et al. 2013). This data originally comes from Domazet-Loso and Tautz (2010).For *M. musculus*, we obtained processed (non-transformed but normalized) RNA-seq data from Hu et al. (2017).

Supplementary datasets

For *C. elegans, D. melanogaster* and *D. rerio*, we downloaded the processed (non-transformed and non-normalized) RNA-seq data from Levin et al. (2016). This dataset was generated by CEL-Seq (Hashimshony et al. 2012), a technique for multiplexed single cell RNA-seq. Because CEL-Seq retains only the 3′ end of the transcript, we performed sample normalization based on transcripts per million, but without transcript length normalization. Since only one embryo was sequenced in each stage, there could be larger technical error among lowly expressed genes, like in single cell RNA-seq, especially in early development. So, after normalization, we removed genes with mean expression across samples <1. For *M. musculus*, the processed (log-transformed and normalized) microarray data was retrieved from Bgee (release 13.1, July 2015; Bastian et al., 2008), a database for gene expression evolution. This data originally comes from (Irie and Kuratani 2011).

The detail information of the two datasets listed in Table S1.

2.5.2 Omega0

The ω_0 values were downloaded from Selectome (Moretti et al. 2014), a database of positive selection based on the branch-site model (Zhang et al. 2005). Selectome excludes ambiguously aligned regions before model fitting, using Guidance bootstrapping and M-Coffee consistency. ω_0 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 ω_0 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 ω_0 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.

2.5.3 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 definition of phylostrata is dependent on the available genome sequences in related lineages, hence the differences between species.

2.5.4 Number of paralogs

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

2.5.5 Retrogene data

For *C. elegans*, we retrieved 33 retrogenes from Zou et al. (2012). For *D. melanogaster*, we retrieved 72 retrogenes from retrogeneDB (Kabza et al. 2014). For *D. rerio* we retrieved 113 retrogenes from Fu et al. (2010). For *M. musculus* we retrieved 134 retrogenes from Potrzebowski et al. (2008).

2.5.6 Connectivity data

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

2.5.7 Testis specific genes

We first retrieved processed (normalized and log-transformed) RNA-seq data of 22 *M. musculus* tissues and 6 *D. melanogaster* tissues from Kryuchkova-Mostacci and Robinson-Rechavi (2016a).

Then, we calculated tissue specificity based on Tau (Yanai et al. 2005; Kryuchkova-Mostacci and Robinson-Rechavi 2016b):

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

where n is the number of tissues, and x_i is the expression of the gene in tissue i. This index ranges from zero (broadly expressed genes) to one (genes specific to one tissue). All genes that were not expressed in at least one tissue were removed from the analysis.

Finally, we defined genes with highest expression in testis and with tissue specificity value ≥ 0.8 as testis specific genes.

2.5.8 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 (ω_0, paralog number, phyletic age, stage specificity, or protein connectivity) of gene i, n is the total number of genes, and e_{is} is the expression level of gene i in developmental stage s; by default we use log-transformed expression levels for e_{is}.

2.5.9 Confidence interval analysis

Firstly, we randomly sampled gene IDs from each original data set 10,000 times with replacement. Then, we computed transcriptome indexes for the 10,000 samples. Finally, the 95% confidence interval is defined as the range from quantile 2.5% to quantile 97.5% of the 10,000 transcriptome indexes. This approach was integrated into myTAI (Drost et al. 2017), a R package for evolutionary transcriptome index analysis.

2.5.10 Stages before the start of maternal to zygote transition

The stages before the start of the maternal to zygote transition (MZT) were defined from Tadros and Lipshitz (2009). For C. elegans and *D. melanogaster*, it's the first 1 hour of embryo development; for *D. rerio*, it's the first 2 hours of embryo development; for *M. musculus*, it's the first day of embryo development.

2.5.11 Phylotypic period

From both morphological and genomic studies, we defined the phylotypic period of each species as follows: for *C. elegans*, the phylotypic period is defined as the ventral enclosure stage (Levin et al. 2012); for *D. melanogaster*, the phylotypic period is defined as an extended germband stage (Sander 1983; Kalinka et al. 2010); for *D. rerio*, the phylotypic period is defined as the segmentation and pharyngula stages (Ballard 1981; Wolpert 1991; Slack et al. 1993; Domazet-Loso and Tautz 2010); for *M. musculus*, the phylotypic period is defined as Theiler stages 13 to 20 (Ballard 1981; Wolpert 1991; Slack et al. 1993; Irie and Kuratani 2011).

2.5.12 Permutation test

We first assigned all development stages to three broad development periods (early: after the start of MZT and before the phylotypic period, middle: the phylotypic period, and late: after the phylotypic period). Next, we calculated the difference of mean transcriptome indexes between the early module and the middle module ($\Delta e\text{-}m$). Then, we permuted the values of the relevant parameter (ω_0, paralog number, phyletic age, stage specificity or protein connectivity) 10,000 times. 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 $\Delta e\text{-}m$.

2.6 Acknowledgements

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

2.7 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.

2.8 Supporting Information

Supporting materials can be downloaded from:
https://academic.oup.com/gbe/article/10/9/2266/5076813#121132694

Table S1: Expression datasets used in this study; "main": dataset used for the figures in the manuscript; "supplementary": dataset used for the figures in the supplementary materials.

Figure S1: Distributions of dN and dS for *C. elegans*

Figure S2: Comparison of 95% confidence intervals from transformed and non-transformed expression values

Grey, dark blue, red, and green marked time points in the x-axis represent stages before the start of MZT, early developmental stages, middle developmental stages and late developmental stages respectively. Y-axis represents the ratio of upper to lower 95% confidence interval boundary. The ratio from non-transformed expression values is plotted in dotted lines, while the ratio from $\log_2$ transformed expression values is plotted in solid lines, and the ratio from square root (abbreviated as "sqrt") transformed expression values is plotted in dashed lines.

Figure S3: Evolutionary transcriptome indexes based on square root transformed expression values

Legend as Figure 1, but here the indexes are based on square root transformed expression values.

Figure S4: Evolutionary transcriptome indexes based on non-transformed expression values

Legend as Figure 1, but here the indexes are based on non-transformed expression values.

Figure S5: Evolutionary transcriptome indexes based on supplementary datasets

Legend as Figure 1, but here the indexes are based on supplementary datasets.

Figure S6: Comparison of transcriptome phyletic age indexes (TAI)

Dark blue, red, and green marked time points in the x-axis represent early developmental stages, middle developmental stages and late developmental stages respectively. TAI is plotted in purple line. The grey area indicates 95% confidence interval estimated from bootstrap analysis. The *p*-values for supporting the hourglass model (permutation test, early *vs*. middle development) are indicated in the top-left corner of each graph.

A: TAI based on non-transformed expression values.

B: TAI based on $\log_2$ transformed expression values.

C: TAI based on non-transformed expression values, excluding the top 10% highest expressed genes.

Figure S7: Proportion of temporal pleiotropic genes for *D. melanogaster* with the supplementary dataset

Legend as Figure 2, but here the result comes from the supplementary dataset of *D. melanogaster*.

Figure S8: Proportion of temporal pleiotropic genes defined as expressed in more than 70% stages

Legend as Figure 2, but here the temporal pleiotropic genes are defined as expressed in more than 70% of stages.

Figure S9: Proportion of temporal pleiotropic genes for *D. melanogaster* after removal of the second period of late development

Legend as Figure 2.

Figure S10: Proportion of temporal pleiotropic genes for *D. rerio* based on expressed genes defined as microarray signal rank in top 90% or 50%

Legend as Figure 2. In upper panel graphs, expressed genes defined as microarray signal rank in top 90%. In lower panel graphs, expressed genes defined as microarray signal rank in top 50%.

Figure S11: Relation of protein connectivity and duplicability

Genes were split into 10 bins according to their connectivity. The duplicability in each bin was measured by the number of genes with paralogs divided by the number of all genes. The duplicability was fit by regression (the first degree of polynomial for *C. elegans* and *M. musculus,* while the second degree of polynomial for *D. melanogaster* and *D. rerio*), whose R^2 and *p*-value are indicated in the top-left corner of each graph. The median connectivity of each bin was plotted on the x-axis (in $\log_2$ scale).

2.9 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.

Ballard WW. 1981. Morphogenetic Movements and Fate Maps of Vertebrates. Am. Zool. 21:391–399.

Bastian F, Parmentier G, Roux J, Moretti S, Laudet V, Robinson-Rechavi M. 2008. Bgee: Integrating and comparing heterogeneous transcriptome data among species. Data Integr. Life Sci. 5109:124–131.

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.

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

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

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.
Drost H-G, Gabel A, Liu J, Quint M, Grosse I. 2017. myTAI: evolutionary transcriptomics with R. Bioinformatics 34:1589–1590.
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. 2018. Pairwise comparisons across species are problematic when analyzing functional genomic data. Proc. Natl. Acad. Sci. U. S. A. 115:E409–E417.
Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary Rate in the Protein Interaction Network. Science (80-.). 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 (80-.). 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.
Hashimshony T, Wagner F, Sher N, Yanai I. 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2:666–673.
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.

Hu H, Uesaka M, Guo S, Shimai K, Lu T-M, Li F, Fujimoto S, Ishikawa M, Liu S, Sasagawa Y, et al. 2017. Constrained vertebrate evolution by pleiotropic genes. Nat. Ecol. Evol. 1:1722–1730.
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. Tissue-Specificity of Gene Expression Diverges Slowly between Orthologs, and Rapidly between Paralogs. PLOS Comput. Biol. 12:e1005274.
Kryuchkova-Mostacci N, Robinson-Rechavi M. 2016b. A benchmark of gene expression tissue-specificity metrics. Brief. Bioinform. 18:205–214.
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. 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. Adaptive evolution of proteins expressed in late and post-embryonic development in animals. bioRxiv:doi:10.1101/161711.
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. 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.
Sander K. 1983. Development and evolution: the sixth Symposium of the British Society for Developmental Biology. Cambridge University Press.
Slack JMW, Holland PWH, Graham CF. 1993. The zootype and the phylotypic stage. Nature 361:490–492.
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.
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 Borntträger.
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 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.
Wolpert L. 1991. The Triumph of the Embryo. Oxford University Press.
Xia B, Baron M, Yan Y, Wagner F, Kim SY, Keefe DL, Alukal JP, Boeke JD, Yanai I. 2018. Widespread transcriptional scanning in testes modulates gene evolution rates. bioRxiv:doi.org/10.1101/282129.

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. 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.

Zou M, Wang G, He S. 2012. Evolutionary patterns of RNA-based gene duplicates in Caenorhabditis nematodes coincide with their genomic features. BMC Res. Notes 5:398.

3 Adaptive evolution of animal proteins over development: support for the Darwin selection opportunity hypothesis of Evo-Devo

Jialin Liu, Marc Robinson-Rechavi

This article was published in Molecular Biology and Evolution (2108) 35(12): 2862-2872.

3.1 Abstract

A driving hypothesis of Evo-Devo is that animal morphological diversity is shaped both by adaptation and by developmental constraints. Here we have tested Darwin's "selection opportunity" hypothesis, according to which high evolutionary divergence in late development is due to strong positive selection. We contrasted it to a "developmental constraint" hypothesis, according to which late development is under relaxed negative selection. Indeed, the highest divergence between species, both at the morphological and molecular levels, is observed late in embryogenesis and post-embryonically. To distinguish between adaptation and relaxation hypotheses, we investigated the evidence of positive selection on protein-coding genes in relation to their expression over development, in fly *Drosohila melanogaster*, zebrafish *Danio rerio*, and mouse *Mus musculus*. First, we found that genes specifically expressed in late development have stronger signals of positive selection. Second, over the full transcriptome, genes with evidence for positive selection trend to be expressed in late development. Finally, genes involved in pathways with cumulative evidence of positive selection have higher expression in late development. Overall, there is a consistent signal that positive selection mainly affects genes and pathways expressed in late embryonic development and in adult. Our results imply that the evolution of embryogenesis is mostly conservative, with most adaptive evolution affecting some stages of 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.

3.2 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 group progressively diverges across development (Von Baer 1828); such groups are usually understood to by phyla in a modern context. 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 (Roux and Robinson-Rechavi 2008; Artieri et al. 2009), while most recent ones support the hourglass model (Kalinka et al. 2010; Irie and Kuratani 2011; Levin et al. 2012; Quint et al. 2012; Drost et al. 2015; Hu et al. 2017; Zalts and Yanai 2017). And in fact the two models may not be mutually exclusive (Piasecka et al. 2013; Liu and Robinson-Rechavi 2018).

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 (1922) and Riedl (1978) 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 2018). An alternative explanation, however, known as Darwin's "selection opportunity" hypothesis (Darwin 1871)(cited in Artieri et al. 2009), proposed that highly divergent late development could also be driven by adaptive evolution (positive selection), at least in part. This could be due to the greater diversity of challenges to which natural selection needs to respond in juvenile and adult life than in early and mid-development. Notably, 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 stronger 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 (dS) substitutions (Yang and Nielsen 1998; Hurst 2002). Because synonymous changes are assumed to be functionally neutral, ω>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. 2017). This gene set enrichment analysis allows to detect weak but consistent adaptive signals from whole genome scale, unlike traditional enrichment tests which only consider top scoring genes with an arbitrary significance threshold.

In order to estimate the contribution of positive selection to the evolution of highly divergent late development, 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; Levin et al. 2016), and compared the signal of positive selection across modules. Second, we applied a modified "transcriptome 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 from *D. rerio*, *M. musculus*, and *D.*

melanogaster and to results of the branch-site test for positive selection in lineages leading to these species. All the analyses found a higher rate of adaptation in late and in some stages of post-embryonic development, including adult.

3.3 Results

In order to characterize 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) and 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.

3.3.1 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 a stage specific impact of natural selection. In addition, since the modules decompose the genes into different meta development stages, they allow to avoid the potential bias caused by imbalanced time points in each meta development stage from our transcriptome datasets; e.g., many more "late development" samples in fly than in the other two species studied. For *D. rerio*, we obtained seven modules from our previous study (Piasecka et al. 2013) (Figure S1). For *M. musculus* and *D. melanogaster*, we identified three and six modules respectively (see Methods; Figure S1).

Because not all genes have any evidence for positive selection, we first compared the proportion of genes either with strong evidence (q-value < 0.2) or with weak evidence (no threshold for q-value; ΔlnL > 0) of positive selection across modules. For strong evidence, the proportion is not significantly different across modules in *M. musculus* and *D. melanogaster* (Figure S2). In *D. rerio*, however, there is a higher proportion in the juvenile and adult modules. For the weak evidence, *D. melanogaster* has a higher proportion in pupae and adult modules, but there is no significant difference in *D. rerio* and *M. musculus* (Figure S3).

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 *D. rerio*, we detected an hourglass pattern of ΔlnL, at its highest in late modules. Specifically, in the juvenile module, the mean ΔlnL is significantly higher than the mean ΔlnL for all genes (*p*-values reported in Table 1). We note that the adult module also has higher mean ΔlnL, even though it's not significant. In the pharyngula module, the mean ΔlnL is significantly lower than the mean ΔlnL for all genes, as expected under the hourglass model. In the other modules, the mean ΔlnL is not significantly different from the mean for all genes. In *M. musculus*, similarly, we found an hourglass pattern of ΔlnL. The late embryo module has a higher mean ΔlnL than all genes, while the middle embryo module has a lower mean ΔlnL than all genes. In *D. melanogaster*, however, we observed an early conservation pattern of ΔlnL. Specifically, in the early embryo module, the mean ΔlnL is lower than the mean ΔlnL for all genes. In the adult module, the mean ΔlnL is higher than the mean ΔlnL for all genes. There is no significant difference for the other modules. It should be noted that the patterns reported in this modularity analysis are relatively weak, especially in *D. melanogaster*. After multiple test correction, some of the reported differences are not significant anymore (Table S1).

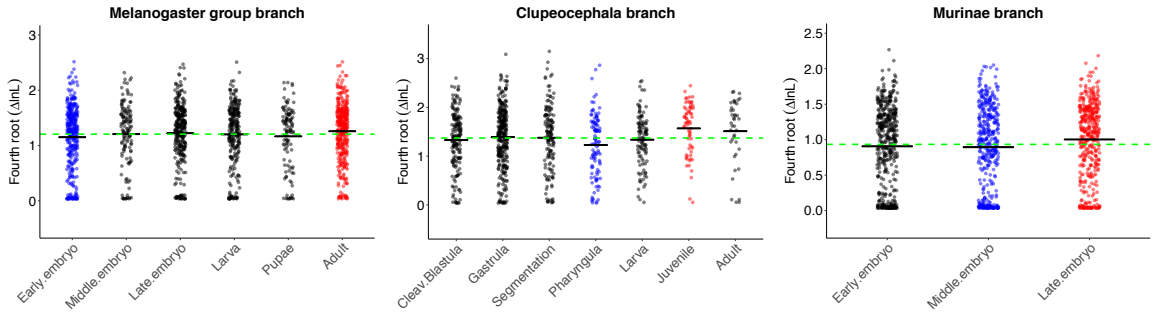

Figure 1: Variation of ΔlnL in different modules.
For each module, dots are values of ΔlnL for individual genes and the black line is the mean of ΔlnL. Red (respectively blue) dots indicate modules for which the mean of ΔlnL is significantly ($p<0.05$) higher (respectively lower) than the mean of ΔlnL from all modules. The green dashed line denotes the mean value of ΔlnL from all modular genes.

Table 1: P-values of randomization test for modular analysis.

D. melanogaster	Early embryo		Middle embryo		Late embryo	Larva	Pupae	Adult
p-value	0.014		0.484		0.263	0.435	0.213	0.018
D. rerio	Cleavage/ Blastula	Gastrula	Segmentation	Pharyngula		Larva	Juvenile	Adult
p-value	0.166	0.238	0.448	0.005		0.273	0.003	0.066
M. musculus	Early embryo		Middle embryo		Late embryo			
p-value	0.094		0.043		0.001			

Overall, these findings suggest that positive selection is stronger on genes expressed in late development or in adult than in early and middle development. It also indicates that ΔlnL on gene modules in different phyla supports different Evo-Devo models (hourglass vs. early conservation).

3.3.2 Transcriptome index analysis

Although modularity analysis guarantees independence between the sets of genes which are compared, it only considers a subset of genes. This leaves open whether the higher adaptive evolution in late development and adult holds true for the whole transcriptome as well, or just for these modular genes. Additionally, while trends were detected, significance is weak. To consider the composition of the whole transcriptome and to increase our power to detect a signal of positive selection in development, we used a modified "Transcriptome Age Index" (Domazet-Loso and Tautz 2010) to calculate the weighted mean of ΔlnL for the transcriptome. Notably, all expression levels were log-transformed before use, unlike in Domazet-Loso and Tautz (2010). See discussion in Piasecka et al. (2013) and Liu and Robinson-Rechavi (2018), but briefly log-transformation provides insight on the overall transcriptome rather than a small number of highly expressed genes. We named this modified index "Transcriptome Likelihood Index" (TLI). A higher index indicates that the transcriptome has higher expression of transcripts from genes with high ΔlnL between models with and without positive selection.

In *D. rerio*, generally, the pattern resembles an hourglass like pattern (Figure 2). The TLI first decreases and reaches a minimum in the late stage of gastrula (8h), and then progressively

increases until adult (ninth month), with finally a slight decline. In addition, in the adult stage, female has higher TLI than male, although the difference is weak. To test whether TLIs are different between developmental periods, we compared the mean TLI of all stages within a period, between each pair of periods (see Methods). We found that middle development has low TLI, early development has medium TLI, late development and maternal stage have very similar high TLI, and adult has the highest TLI. Except late development and maternal stage (p=0.24), all pairwise comparisons are significant: p<5.7e-07. In *M. musculus*, we observed a clear hourglass-like pattern of TLI. For the mean TLI comparison, we found low TLI in middle development, medium TLI in early development, high TLI in late development, and the highest TLI in maternal stage (all pairwise comparisons are significant: p<2e-16). Of note, unlike in *D. rerio*, the "late development" here only contains late embryo stages, but no post embryo stages. This may explain why late development has lower TLI than the maternal stage in this dataset. In *D. melanogaster*, we found the TLI progressively increasing over development, suggesting an early conservation model. Unlike in *D. rerio*, we found that male has higher TLI than female in the adult stage. For the mean TLI comparison, early development has low TLI, middle development has medium TLI, late development has high TLI, and adult has the highest TLI (all pairwise comparisons are significant: p<2e-16).

As in the modularity analysis, but with much stronger signal, both *D. rerio* and *M. musculus* support the hourglass model, while *D. melanogaster* follows an early conservation 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. Interestingly, the maternal stage has a comparable high TLI to late development. This could be related to the maternal stage being dominated by adult transcripts (Tadros and Lipshitz 2009). In this respect (transcriptome evolution), the maternal stage should maybe be regarded as a special adult stage rather than as an early embryonic stage.

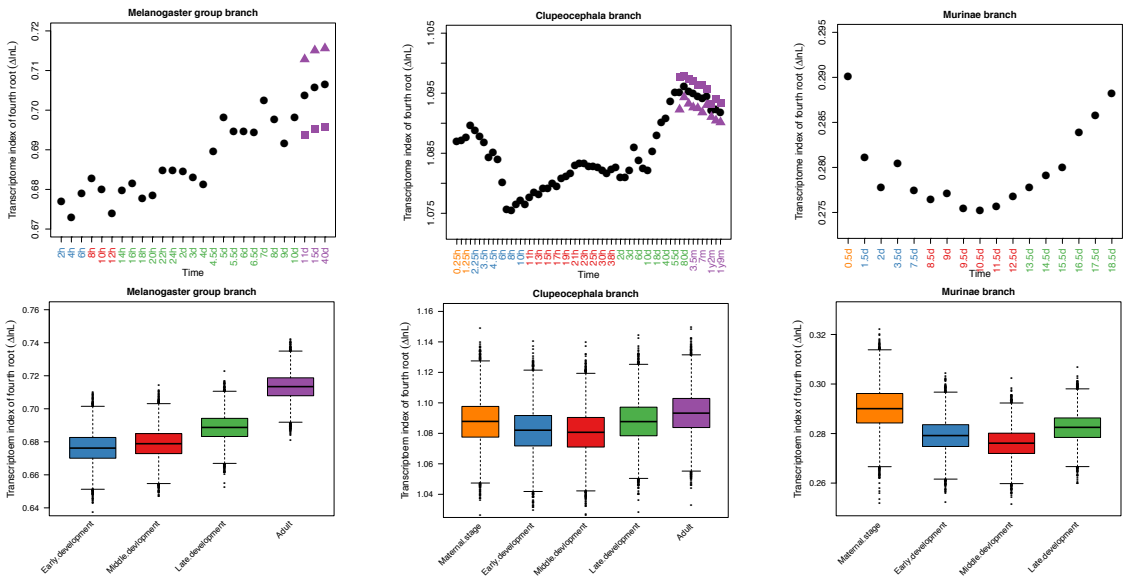

Figure 2: Transcriptome index of ΔlnL (TLI) across development.
For sub-figure A, B and C:
Orange, blue, red, green and purple time points represent stages within the developmental periods of maternal stage, early development, middle development, late development, and adult, respectively. For the adult stage, the black solid circle represents TLI from average expression between male and female; the purple solid triangle and square represent TLI from only males or females, respectively.
For sub-figure D, E and F:
Comparison of the TLI (mean TLI of all stages within a period) between any two different periods. Each period has 10,000 pseudo-TLIs which come from random resampling with replacement.

3.3.3 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). Thus, we searched for such signals of positive selection on pathways. Briefly, we calculated the sum of ΔlnL (SUMSTAT statistic) 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 lineages leading to *D. rerio*, *M. musculus* and *D. melanogaster* (q-value<0.2, Table2).

Table 2: 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

The function of these pathways, while not our primary focus, is consistent with adaptive evolution of juvenile or adult phenotypes. First, we found metabolism related pathways in all three species, suggesting pervasive adaptation, possibly related to diet; this is consistent with previous results in primates (Daub et al. 2017). Second, in *D. rerio* and *D. melanogaster*, several pathways are involved in morphogenesis and remodelling of organs (e.g., laminin interactions, extra cellular matrix, ECM-receptor interaction), suggesting potential adaptive evolution of morphological development. Third, there are several pathways involved in aging in *D. melanogaster* and *M. musculus* (e.g., reactive oxygen detoxification, longevity regulation, mitochondrial translation), suggesting potential role of natural selection on modulating lifespan or on metabolic activity. Forth, in *D. rerio,* we detected one pathway related to environmental adaptation: visual phototransduction; adaptations in vision are expected for aquatic species which under a wide variety of visual environments (Sabbah et al. 2010).

If late development and adult are 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 these periods. Thus we computed the ratio of median expression between positively selected pathway genes and genes included in pathways not positively selected. Since the median expression in the first time point of *M. musculus* is 0, we removed it from our analysis. In *D. rerio*, the ratio of median expression keeps increasing until the juvenile stage. Then, it slightly decreases (Figure 3). In *M. musculus*, except the first time point, the ratio of median expression also progressively increases. In *D. melanogaster*, there is a small peak in the first time point, but it quickly decreases to minimum within the same developmental period. Then, it keeps increasing until the middle of the larval stage. Finally, for the last development stages, it resembles a wave pattern: decrease, increase and decrease again. Again, we also tested the difference between male and female in adult stages for *D. rerio* and *D. melanogaster*. Unlike the observation in the transcriptome index analysis, here we found that male has higher ratio of median expression than female in both species.

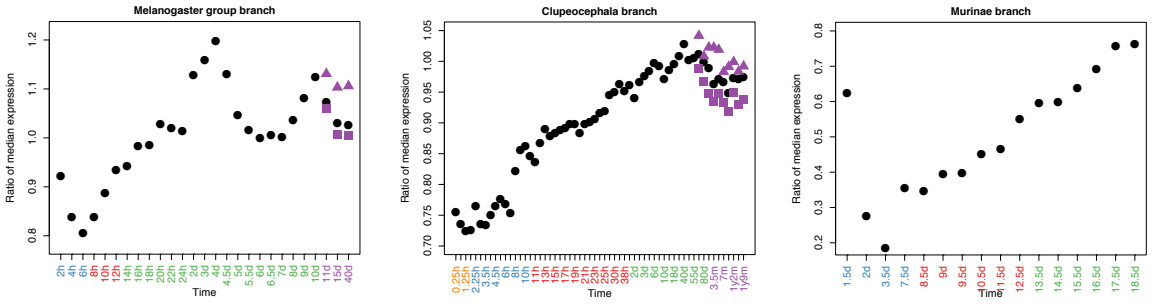

Figure 3: Expression in development for genes involved in pathways enriched with signal of positive selection.
Each solid circle represents the ratio of the median expression for genes involved in pathways enriched with signal of positive selection to the median expression for genes involved in pathways without signal of positive selection. Orange, blue, red, green and purple time points represent stages within the developmental periods maternal stage, early development, middle development, late development, and adult, respectively. In adult samples, black solid circles represent ratios generated from average expression of males and females; purple solid triangles and squares represent ratios generated from only males or only females, respectively.

Overall, consistent with previous results, we found that late development and adult tend to express genes involved in pathways enriched for signal of positive selection, indicating that adaptive evolution at the pathway level mainly affects these stages. 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 (Müller 2007; Wray 2007; Tickle and Urrutia 2017), this suggests stronger phenotypic adaptation in late development and adult.

3.4 Discussion

3.4.1 Correcting confounding factors

Since some non-adaptive factors (such as gene length, tree size (number of branches), 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 S4). 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 checked 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 check further the correlation between ΔlnL and branch length. In order to investigate whether 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 of gene length, relative to ΔlnL. For modularity analysis, the modules with higher ΔlnL have significantly lower mean gene length than all genes (Figure S5). For transcriptome index analysis, the stages with higher TLI trend to have lower transcriptome index for gene length (Figure S6), suggesting that these stages trend to express shorter genes. These findings imply that the detection of higher positive selection in late development is not driven by gene length. Immune system genes can bias positive selection analyses, since they evolve under pervasive positive selection (Flajnik and Kasahara 2010). To control for this, we also confirmed our findings after removing immune genes from our analysis (Figure S7).

3.4.2 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 diverse animal species (Roux and Robinson-Rechavi 2008; Domazet-Loso and Tautz 2010; Kalinka et al. 2010; Irie and Kuratani 2011; Levin et al. 2012; Piasecka et al. 2013; Drost et al. 2015; Liu and Robinson-Rechavi 2018), the underlying evolutionary forces driving such a pattern remain obscure. The “developmental constraint” hypothesis (Raff 2000; Brakefield 2006) suggests that this high divergence is due to relaxed purifying selection, whereas Darwin's "selection opportunity" hypothesis proposes stronger positive selection (as discussed in Artieri et al. 2009; Kalinka and Tomancak 2012).

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, and Uchida et al. (2018) found strong embryonic lethality from random mutations in early development. 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 2018). This provides direct evidence of relaxed purifying selection in late development.

To the best our knowledge, there has been no direct test of Darwin's "selection opportunity" hypothesis. One such study, in *D. melanogaster*, was proposed by Artieri et al. (2009). Unfortunately, they only had relatively poor expression data (ESTs) and limited time points (embryonic, larval/pupal and adult), and 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 rapid evolution observed in late development could be due to specific selective pressures such as sexual selection. A recent study, in *D. melanogaster*, provides indirect evidence: using in situ expression data and population genomic data to map positive selection to different embryonic anatomical structures, Salvador-Martínez et al. (2018) found larva stage enriched with signal of positive selection. Our results clearly provide a quantitative test which supports a role of positive selection in the high divergence of late development. While our sampling is very far from covering the diversity of developmental modes of animals, we show consistent patterns in a placental mammal, a direct development ray-finned fish, and a holometabolous insect. While it is possible that other patterns will be found in species with different development, this shows that adaptation in late development is not limited to one model. We show that this is not due to testis-expressed genes (Figure S8). In addition, in vertebrates, we also found some evidence of adaptive evolution in early development on single genes. This indicates that some changes in early development might be adaptive consequences to diverse ecological niches, as proposed by Kalinka and Tomancak (2012). 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).

3.4.3 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 often 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 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) and 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 was 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 might also make a significant contribution to the evolution of morphology, as suggested by Hoekstra and Coyne (2007) and Burga et al. (2017). This is also supported by the observation of tissue-specific positive selection in *D. melanogaster* development (Salvador-Martínez et al. 2018). 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.

3.5 Materials and Methods

Data files and analysis scripts are available on our GitHub repository: https://github.com/ljljolinq1010/Adaptive-evolution-in-late-development-and-adult

3.5.1 Expression data sets

For *D. rerio*, the log-transformed and normalized microarray data was 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*, the processed RNA-seq (normalized but non-transformed) data was retrieved from Hu et al. (2017). This data includes 17 stages from 2cells to E18.5. We further transformed it with $\log_2$. For *D. melanogaster*, we obtained processed (normalized but non-transformed) RNA-seq data from http://www.stat.ucla.edu/~jingyi.li/software-and-data.html 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. We further transformed it with $\log_2$.

3.5.2 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*(lnLH1-lnLH0). Importantly, 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. 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.

3.5.3 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).

3.5.4 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.

3.5.5 Testis specific genes

Testis specific genes for *M. musculus* and *D. melanogaster* were obtained from a parallel study (Liu and Robinson-Rechavi 2018). The testis specific genes were defined as genes with highest expression in testis and with tissue specificity value ≥ 0.8.

3.5.6 Immune genes

To control for the impact of immune system genes, we downloaded all genes involved in the "immune response" term (GO:0006955) from AmiGO (Carbon et al. 2009) (accessed on 25.04.2018), and repeated analyses with these genes excluded.

3.5.7 Phylotypic period

The definition of phylotypic period is based on previous morphological and genomic studies. For *D. melanogaster*, the phylotypic period defined as extended germband stage (Sander 1983; Kalinka et al. 2010); for *D. rerio*, the phylotypic period defined as segmentation and pharyngula satges (Ballard 1981; Wolpert 1991; Slack et al. 1993; Domazet-Loso and Tautz 2010); for *M. musculus*, the phylotypic period defined as Theiler Stage 13 to 20 (Ballard 1981; Wolpert 1991; Slack et al. 1993; Irie and Kuratani 2011).

3.5.8 Module detection

For D. rerio, we obtained seven modules from our previous study (Piasecka et al. 2013). This

is based on the Iterative Signature Algorithm, which identifies modules by an iterative procedure (Bergmann et al. 2003). Specifically, it was initialized with seven artificial expression profiles, similar to presented in Figure S11. Each profile corresponds to one of the zebrafish meta developmental stages. Next, the algorithm will try to find genes with similar expression profiles to these artificial ones through iterations until the processes converges. This method has proven to be very specific, but lacks power with medium or small datasets (<30 time points). For mouse and fly, the sample size is not enough, so we used the method introduced by Levin et al. (2016). Firstly, we generated standardized gene expression for each gene by subtracting its mean (across all stages) and dividing by its standard deviation. Next, we calculated the first two principal components of each gene based on the standardized expression across development. Since the expression was standardized, the genes form a circle with scatter plot (Figure S9). Then, we computed the four-quadrant inverse tangent for each gene based on its principal components, and sort these values to get gene expression order from early to late (Figure S10). Next, we performed Pearson correlation of the standardized expression and idealized expression profile of each module (Figure S11). Finally, for each module, we defined genes with correlation coefficient rank in top 10% as modular genes. Clearly, the genes in earlier modules have higher gene orders (Figure S9).

3.5.9 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.

3.5.10 Transcriptome index of log-likelihood ratio (TLI)

The TLI is calculated as:

$$TLI_s = \frac{\sum_{i=1}^{n} \sqrt[4]{\triangle \ln L_i}\, e_{is}}{\sum_{i=1}^{n} e_{is}},$$

s is the developmental stage, $\Delta\ln L_i$ 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.

3.5.11 Polynomial regression

For polynomial regression analysis, we keep increasing the degree of polynomial model until no further significant improvement (tested with ANOVA, p<0.05 as a significant improvement). For *M. musculus*, since the development time points in transcriptome data set are close to uniformly sampled, we used the natural scale of development time for regression. For *C. elegans*, *D. melanogaster* and *D. rerio*, however, we used the logarithmic scale, to limit the effect of post-embryonic time points.

3.5.12 Bootstrap approach for transcriptome index of ΔlnL (TLI) comparison between developmental periods

Firstly, we randomly sampled the same size of genes from original gene set (with replacement) for 10,000 times. In each time, we calculated the TLI of each development stage. Then, we calculated the mean TLI (mean TLI of all stages within a period) for each developmental period (maternal stage, early development, middle development, late development, and adult). Thus, each developmental period contains 10,000 mean TLI. Finally, we performed pairwise Wilcoxon test to test the differences of mean TLI between developmental periods.

3.5.13 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:

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

where p represents a pathway, and $\Delta\mathrm{lnL}_i$ represents the value of log-likelihood ratio for gene i within pathway p. Pathways 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.

3.5.14 Empirical null distribution of SUMSTAT

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 S12). 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 SUMSTAT score of the randomly sampled ΔlnL values. We repeated the second and third processes one million times. Fourthly, we approximated a normal distribution for SUMSTAT score of the interested pathway. Finally, the p-value was calculated as the probability that the expected SUMSTAT score is higher than the observed SUMSTAT score.

3.5.15 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) $\geq p^*$.

3.6 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 programming, Josephine T. Daub for help with polygenic selection analysis, and 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.

3.7 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.

3.8 Supporting Information

Supporting materials can be downloaded from:
https://academic.oup.com/mbe/article/35/12/2862/5089241#126422796

Table S1: Multiple test corrected p-values (Benjamini–Hochberg method) of randomization test for modular analysis.

Figure S1: Expression profiles of different modules across development

The bold black line represents median expression of modular genes, the two gray lines represent 25th and 75th quantiles of expression of modular genes respectively. The blue, red and green named stages represent early, middle and late modules respectively.

Figure S2: Proportion of genes with strong evidence of positive selection in each module

The number of genes in each module is indicated below each box. The *p*-value from chi-square goodness of fit test is reported in the top-left corner of each graph.

Figure S3: Proportion of genes with weak evidence of positive selection in each module

Legend as in Figure S2.

Figure S4: Spearman's correlation between gene properties and ΔlnL

Spearman's correlation coefficient (rho) and adjusted *p*-value are indicated in the top-right corner of each graph. Loess regression lines are plotted in red.

Figure S5: Variation of gene length in different modules

Legend as in Figure 1.

Figure S6: Transcriptome index of gene length across development

Legend as in Figure 2.

Figure S7: Transcriptome index of ΔlnL (TLI) for non-immune genes

Legend as in Figure 2.

Figure S8: Transcriptome index of ΔlnL (TLI) for non-testis genes in *M. musculus* and *D. melanogaster*

Legend as in Figure 2.

Figure S9: Scatter plot of genes based on principal component analysis

Each dot represents one gene, grey dots represent genes not assigned to any modules, blue dots represent genes in early embryo module, red dots represent genes in middle embryo module, blue dots represent genes in late embryo module, pink dots represent genes in larva module, and purple dots represent genes in pupae/adult module. Arrow indicates the gene expression order (from early to late).

Figure S10: Heat map of gene expression across development

Genes arranged by the order of expression (from early to late). Generally, the earlier genes have higher expression in earlier stages. Color bar represents the standardized expression value.

Figure S11: idealized expression profile for each module

Figure S12: Correction of gene length for polygenic selection

The red lines indicate the boundaries of bins which contains genes with similar length.

3.9 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.
Ballard WW. 1981. Morphogenetic Movements and Fate Maps of Vertebrates. Am. Zool. 21:391–399.
Berg JJ, Coop G. 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.
Brakefield PM. 2006. Evo-devo and constraints on selection. Trends Ecol. Evol. 21:362–368.
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.
Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, AmiGO Hub the A, Web Presence Working Group the WPW. 2009. AmiGO: online access to ontology and annotation data. Bioinformatics 25:288–289.
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 CA, Holland TA, 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.
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.
Darwin C. 1871. The descent of man, and selection in relation to sex. Princeton University Press
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.
Flajnik MF, Kasahara M. 2010. Origin and evolution of the adaptive immune system: genetic

events and selective pressures. Nat. Rev. Genet. 11:47–59.
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.
Garstang W. 1922. The Theory of Recapitulation: A Critical Re-statement of the Biogenetic Law. J. Linn. Soc. London, Zool. 35:81–101.
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.
Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. Evolution 61:995–1016.
Hu H, Uesaka M, Guo S, Shimai K, Lu T-M, Li F, Fujimoto S, Ishikawa M, Liu S, Sasagawa Y, et al. 2017. Constrained vertebrate evolution by pleiotropic genes. Nat. Ecol. Evol. 1:1722–1730.
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.
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.
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. 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. 2018. Developmental constraints on genome evolution in four bilaterian model species. bioRxiv:doi.org/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.
Müller GB. 2007. Evo–devo: extending the evolutionary synthesis. Nat. Rev. Genet. 8:943–949.
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. University of Chicago Press.
Raff RA. 2000. Evo-devo: the evolution of a new discipline. Nat. Rev. Genet. 1:74–79.
Richardson MK. 1999. Vertebrate evolution: the developmental origins of adult variation. Bioessays 21:604–613.
Riedl R. 1978. Order in living organisms. West Sussex, UK: Wiley-Interscience.
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.
Sabbah S, Laria R, Gray SM, Hawryshyn CW. 2010. Functional diversity in the color vision of cichlid fishes. BMC Biol. 8:133.
Salvador-Martínez I, Coronado-Zamora M, Castellano D, Barbadilla A, Salazar-Ciudad I. 2018. Mapping Selection within Drosophila melanogaster Embryo's Anatomy. Mol. Biol. Evol. 35:66–79.
Sander K. 1983. Development and evolution: the sixth Symposium of the British Society for Developmental Biology. Cambridge University Press.
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.
Slack JMW, Holland PWH, Graham CF. 1993. The zootype and the phylotypic stage. Nature 361:490–492.
Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. Development 136:3033–3042.
Tickle C, Urrutia AO. 2017. Perspectives on the history of evo-devo and the contemporary research landscape in the genomics era. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 372:20150473.
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.
Uchida Y, Uesaka M, Yamamoto T, Takeda H, Irie N. 2018. Embryonic lethality is not sufficient to explain hourglass-like conservation of vertebrate embryos. Evodevo 9:7.
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.
Wolpert L. 1991. The Triumph of the Embryo. Oxford University Press.
Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet.

8:206–216.
Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. 46:409–418.
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.
Zalts H, Yanai I. 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.

4 Chromatin accessibility evolution during Drosophila embryogenesis

This is a collaborative project with Eileen Furlong (EMBL, Heidelberg, Germany). My contribution to this chapter was the evolutionary analysis.

This article is in preparation and should be submitted soon.

4.1 Abstract

Transcriptome-based Evo-Devo studies revealed that a period in middle development is highly conserved across species within the same phylum. Although the developmental hourglass pattern has been observed in different phyla, the roles of gene regulation evolution on this pattern remains elusive. Here, using DNase-seq to identify accessible chromatin regions, we assessed the evolutionary conservation, divergence, and positive selection of gene regulation throughout several stages of embryogenesis in two distant *Drosophila* species: *D. melanogaster* and *D. virilis*. We found that the phylotypic stage has a higher proportion of conserved regulatory elements and a lower proportion of species specific gain of regulatory elements. Moreover, based on an in silico mutagenesis approach, we detected signatures of positive selection on developmental enhancers, and found that the phylotypic stage enhancers have less evidence of positive selection. These results suggest that both stronger positive selection and less purifying selection on regulatory sequences contribute to the lower expression divergence at the phylotypic stage.

4.2 Introduction

Embryological development has long been characterized by deep conservation. Animals that belong to the same phylum are considered to share a group of structural and developmental characteristics, the so called basic body plan or bauplan (Wallace 2000; Valentine 2004; Irie and Kuratani 2014). For example, arthropods share a set of anatomic structures such as jointed legs, exoskeleton made of chitin, and segmented bodies (Zrzavy and Stys' 1997). Based on morphological conservation, Duboule (1994) and Raff (1996) proposed the hourglass model, suggesting, within a phylum, embryos in middle development (the phylotypic period (Richardson 1995)) is morphologically more conserved than embryos in both early and late development. However, this model was not supported by later morphological studies (Richardson et al. 1997; Bininda-Emonds et al. 2003). To overcome difficulties of comparing morphological features across species, recent studies have been focusing on comparative transcriptomics (Yanai 2018). This is based on the hypothesis that changes in gene expression play a central role in the morphological differences between species (King and Wilson 1975; Carroll 2008).

Transcriptome comparisons in different phyla (Kalinka et al. 2010; Irie and Kuratani 2011; Levin et al. 2012; Hu et al. 2017) all show that expression divergence is lower in the phylotypic period than in early and late development, supporting the hourglass model. One of the pioneer studies was conducted by Kalinka et al. (2010), quantifying expression divergence at different stages during embryogenesis in six *Drosophila* species. They found that expression divergence follows an hourglass pattern with the minimal divergence at the extended germband stage (8-10 hours after laying egg), generally regarded as the phylotypic stage of arthropods (Sander 1976). Notably, this expression hourglass pattern also extends to population level (Zalts and Yanai 2017), and even to the level of variation between isogenic individuals (Liu et al. 2019). Based on a single embryo transcriptome time series of fly embryonic development with a high number of isogenic replicates, we found that the extended germband stage also has lower non-genetic expression variability (Liu et al. 2019). On the other hand, when animals are compared between different phyla, the phylotypic period has been reported to have the highest divergence (Levin et al. 2016).

Although transcriptomic comparisons throughout developmental time across species have been repeatedly studied, the role of gene regulation on the evolution of expression during

embryogenesis remains to be elucidated. What's more, the contribution of positive selection on the hourglass expression divergence is still unknown, despite recent evidence of a role of positive selection in the hourglass of protein divergence (Liu and Robinson-Rechavi 2018). To answer these questions, we performed DNase-seq to identify active regulatory elements across multiple matched embryonic development stages in *D. melanogaster* and *D. virilis*. Overall, we found that gene regulation in the phylotypic stage is more conserved than in the earlier stages. In addition, it less accessible to adaptive evolution. Thus, the phylotypic stage can be regarded as an evolutionary lockdown.

4.3 Results

4.3.1 Interspecies DNase-seq across three stages of embryogenesis

To study the evolution of regulatory regions in the context of development, we collected embryos from both *D. melanogaster* and *D. virilis* at three equivalent developmental stages, referred to as TP1, TP2, and TP3 (Figure 1A). TP3 corresponds to the phylotypic stage. Active regulatory regions were then identified by using DNase I hypersensitive sites sequencing (DNase-seq). For every stage, we performed two independent biological replicates. The high confidence peaks (active regulatory regions) for both biological replicates were called at a 5% Irreproducible Discovery Rate (IDR, a measure ensuring equivalent reproducibility between replicas (Li et al. 2011)) threshold. Replicates are highly concordant, with a median Spearman's correlation coefficient of 0.948 (Table S1). On average, we identified 17'549 peaks in each stage in *D. virilis*, and 15'745 peaks in *D. menanogaster* (Figure 1B). The higher number of regulatory elements in *D. virilis* might be related with its larger non-coding genome (Khoueiry et al. 2017). In addition, in both species, we found that TP3 has more peaks than TP1 or TP2, (Figure 1B), potentially reflecting higher tissue and cell type diversity in TP3. We divided all peaks into either putative enhancers or putative promoters based on their distance to transcriptional state sites (Figure S1, see Materials and Methods). For the sake of simplicity, we refer to these putative enhancers or putative promoters as simply enhancers or promoters in the rest of the paper. The peaks identified in *D. virilis* were translated onto *D. melanogaster* orthologous coordinates by using pslMap (Zhu et al., 2007) (see Materials and Methods), resulting in 91.8% of peaks translated on average at each stage (Figure 1C).

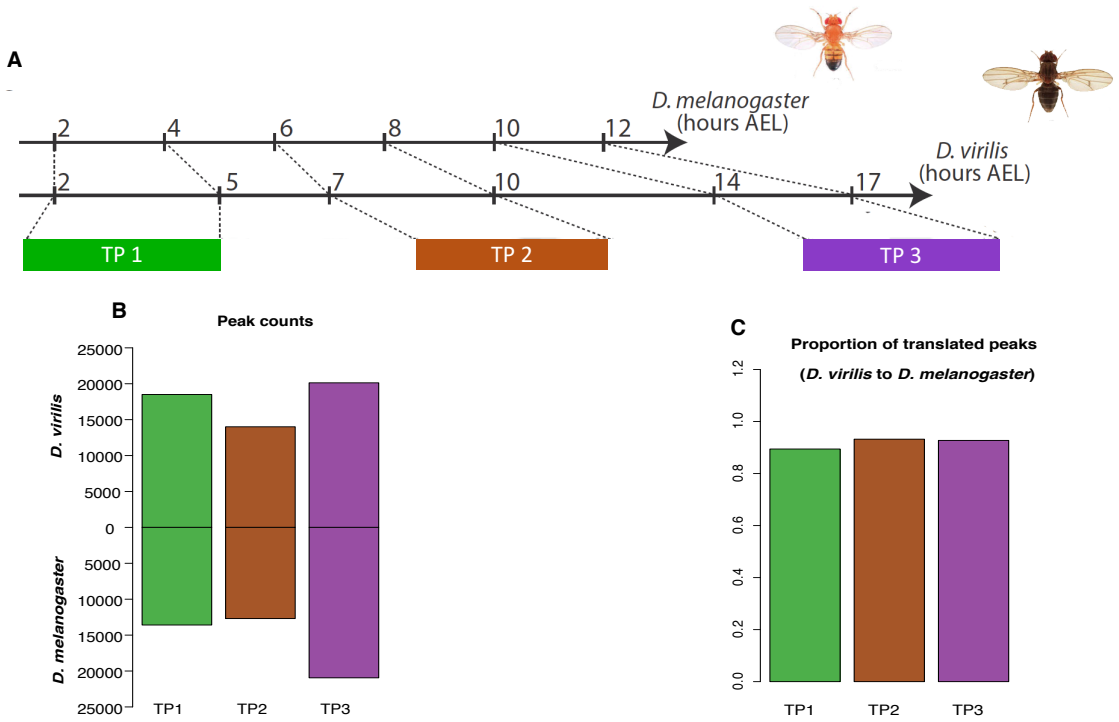

Figure 1: Studying regulatory elements evolution throughout embryogenesis

A. We performed DNase-seq at three matched embryonic development stages in D. melanogaster and D. virilis. The corresponding time points (TP1, TP2 and TP3) for the three development stages are shown on the developmental axis. Different color bars represent different time points sampled. The developmental axes are scaled in hours after egg laying (AEL).

B. Number of significant DNase peaks in the three embryonic development stages in each species.

C. The proportion of peaks identified in D. virilis that can be translated onto D. melanogaster coordinates (i.e., have an ortholog).

4.3.2 Enhancer conservation over developmental stages

To compare evolutionary conservation of enhancers between stages, we first identified stage specific enhancers for each stage in each species (Figure 2A). For example, a *D. melanogaster* TP1 stage specific enhancer was defined as a DNase peak in this stage without any overlap with peaks at other stages in *D. melanogaster* (Figure 2A). From these, we identified conserved stage specific enhancers. For example, if a *D. melanogaster* TP1 stage specific enhancer overlaps (by orthologous translation) with a *D. virilis* TP1 stage specific enhancer, we defined this as a conserved TP1 stage specific enhancer. Finally, we calculated the proportion of conserved stage specific enhancers among all *D. melanogaster* stage specific enhancers, in each stage. We found that TP3 has a significantly higher proportion of conserved enhancers than both TP1 and TP2 (Figure 2B). However, there is no significant difference between TP1 and TP2. Given the larger size of non-coding genome in *D. virilis*, the genome coordinates of

orthologous enhancers maybe shifted by insertions and deletions. To account for this, we repeated the analysis with a relaxed definition of conserved enhancer: the distance between stage specific enhancers in the two species was defined to be smaller than 1kb, but not necessarily overlap. With this definition, we found a very similar pattern, with higher conserved enhancer proportion at TP3 than at earlier stages (Figure S2). Finally, we repeated the same analyses for promoters, and found very similar patterns of stronger conservation at TP3 (Figure 2C, Figure S2).

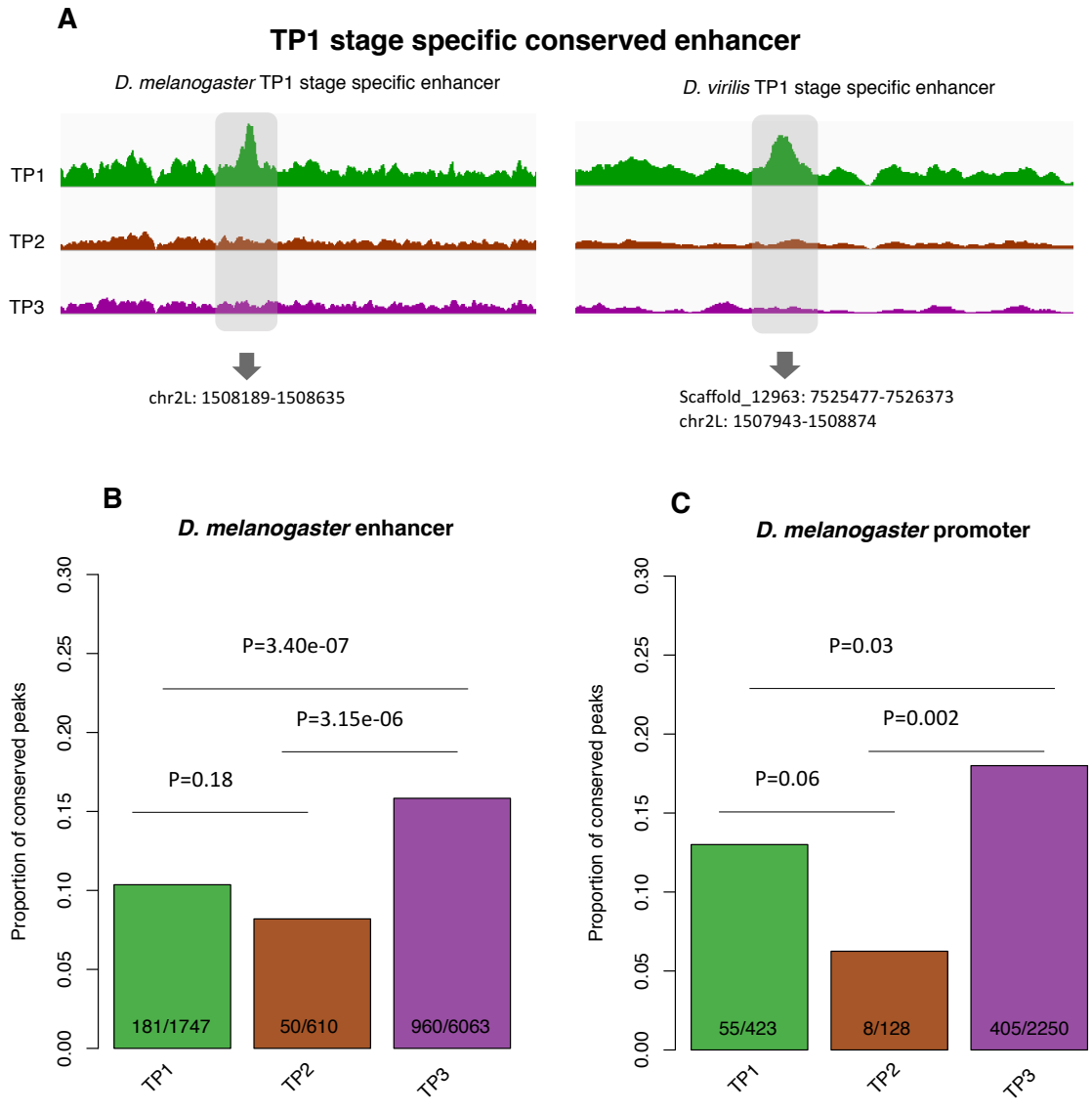

Figure 2: The phylotypic stage has a higher proportion of conserved regulatory elements

A. Genome browser visualization of a D. melanogaster TP1 stage specific conserved enhancer. The left panel is the DNase-seq signal in different development stages (TP1, TP2 and TP3) for a D. melanogaster TP1 stage specific enhancer, covered by the grey area. The coordinates of this enhancer are indicated below the grey arrow. The right panel is the DNase-seq signals in different development stages (TP1, TP2 and TP3) for a D. virilis

TP1 stage specific enhancer, covered by the grey area. The coordinates of this enhancer in D. virilis, and the orthologous coordinates in D. melanogaster, are illustrated below the grey arrow. Since there is overlap between the D. melanogaster TP1 stage specific enhancer and the orthologous position of the D. virilis TP1 stage specific enhancer, we define the D. melanogaster TP1 stage specific enhancer to be conserved.

B. Proportion of conserved stage specific enhancers at each development stage. The number of stage specific enhancers and the number of conserved stage specific enhancers at each development stage is indicated inside each bar of the plot. The p-values from Fisher's exact test are reported above the bars.

C. Proportion of conserved stage specific promoters at each development stage. The number of stage specific promoters and the number of conserved stage specific promoters at each development stage is indicated inside each bar. The p-values from pairwise Fisher's exact tests are reported above the bars.

4.3.3 Enhancer gain over developmental stages

We first inferred stage specific enhancer gains in each *D. melanogaster* developmental stage (Figure 3A). For example, a TP1 stage specific enhancer gain was defined as a DNase peak in TP1 without any overlaps with peaks in other *D. melanogaster* stages, and without a peak in any *D. virilis* stage (Figure 3A). Then, we calculated the proportion between *D. melanogaster* stage specific enhancer gains and all *D. melanogaster* stage specific enhancers, in each stage. Both TP3 and TP2 have significantly lower proportions of enhancer gain than TP1 (Figure 3B). However, there is no significant difference between TP2 and TP3 (Figure 3B). Again, we repeated the same analyses for promoters, and found very similar patterns (Figure 3C).

Overall, we found the evolution of both enhancers and promoters to be more conserved in TP3, the phylotypic stage, than in earlier stages. These results suggest a contribution of stronger purifying selection on regulatory elements evolution to the lower expression divergence in the phylotypic stage.

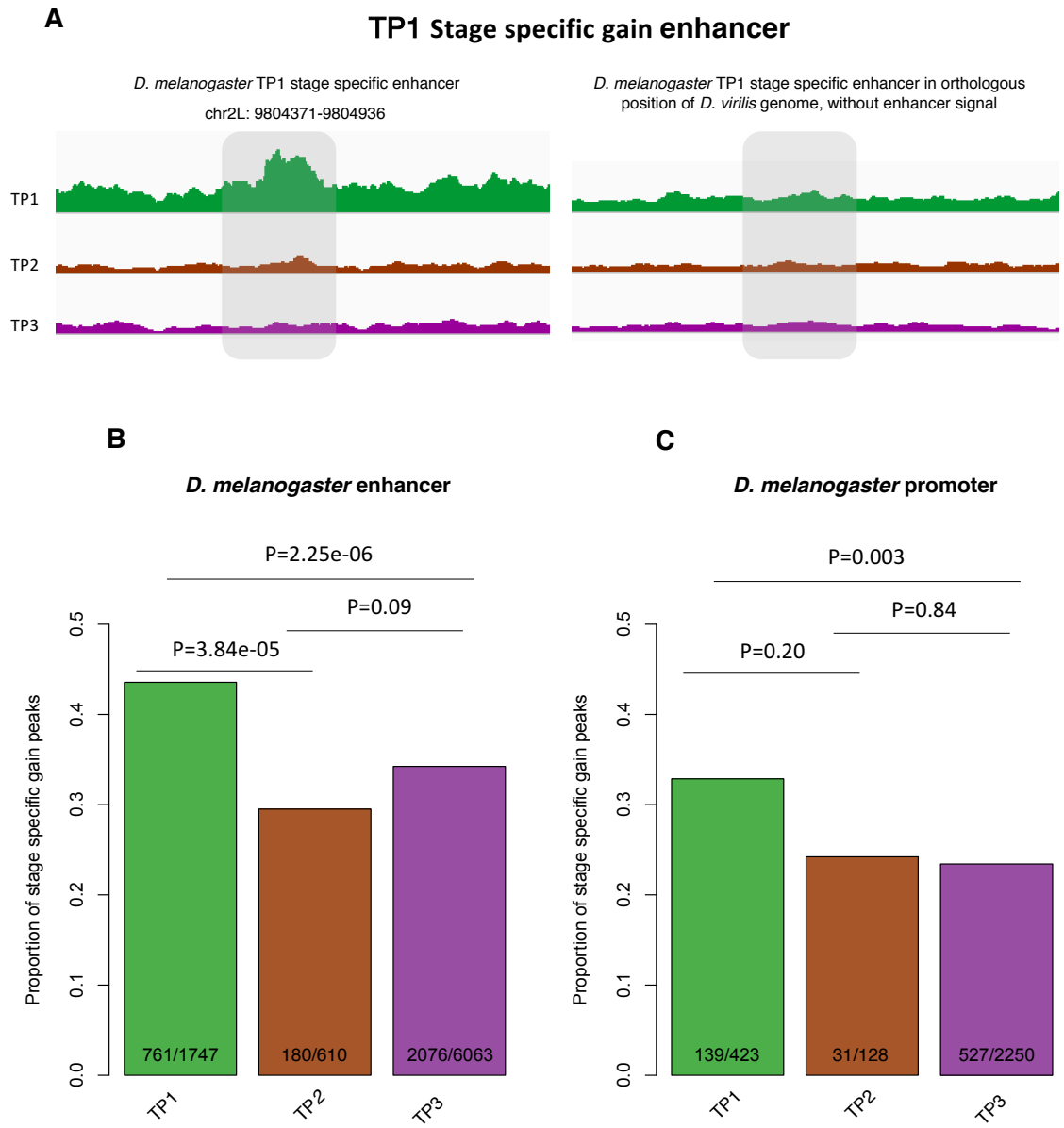

Figure 3: The phylotypic stage has a lower proportion of gain of regulatory elements

A. Genome browser visualization of a TP1 stage specific gain of enhancer. The left panel is the DNase-seq signal in different development stages (TP1, TP2 and TP3) for a D. melanogaster TP1 stage specific enhancer, covered by the grey area. The coordinates are indicated above the grey area. The right panel is the DNase-seq signals in different development stages (TP1, TP2 and TP3) of the orthologous position of in D. virilis, with no signal of enhancer in the DNase-seq.

B. Proportion of stage specific gain of enhancers in each development stage. The number of stage specific enhancers and the number of stage specific gain of enhancers in each development stage is indicated inside each bar. The p-values from Fisher's exact test are reported above the bars.

C. Proportion of stage specific gain of promoters in each development stage. The number of stage specific promoters and the number of stage specific gains of promoters in each development stage is indicated inside each bar. The p-values from Fisher's exact test are reported above the bars.

4.3.4 Detecting positive selection on enhancers

To test whether conservation at the phylotypic stage is also characterized by lower positive selection, we scanned for signatures of positive selection in TP1 and TP3 stage specific enhancers; TP2 has too few enhancers for the method used (see below). We tested substitutions on the *D. melanogaster* branch after divergence from *D. simulans* (Figure 4A). Our approach considers the effects of substitutions on enhancer regulatory activity. It is derived from a new method using in silico mutagenesis to detect positive selection in transcription factor binding site evolution (see Chapter 6). The method is based on machine learning of the sequence patterns associated to function, and testing for substitutions which increase or decrease this function. Here the function is enhancers defined as DNase I–hypersensitive sites (DHSs), and we first defined whether there were sequence features which differ between enhancers of different stages.

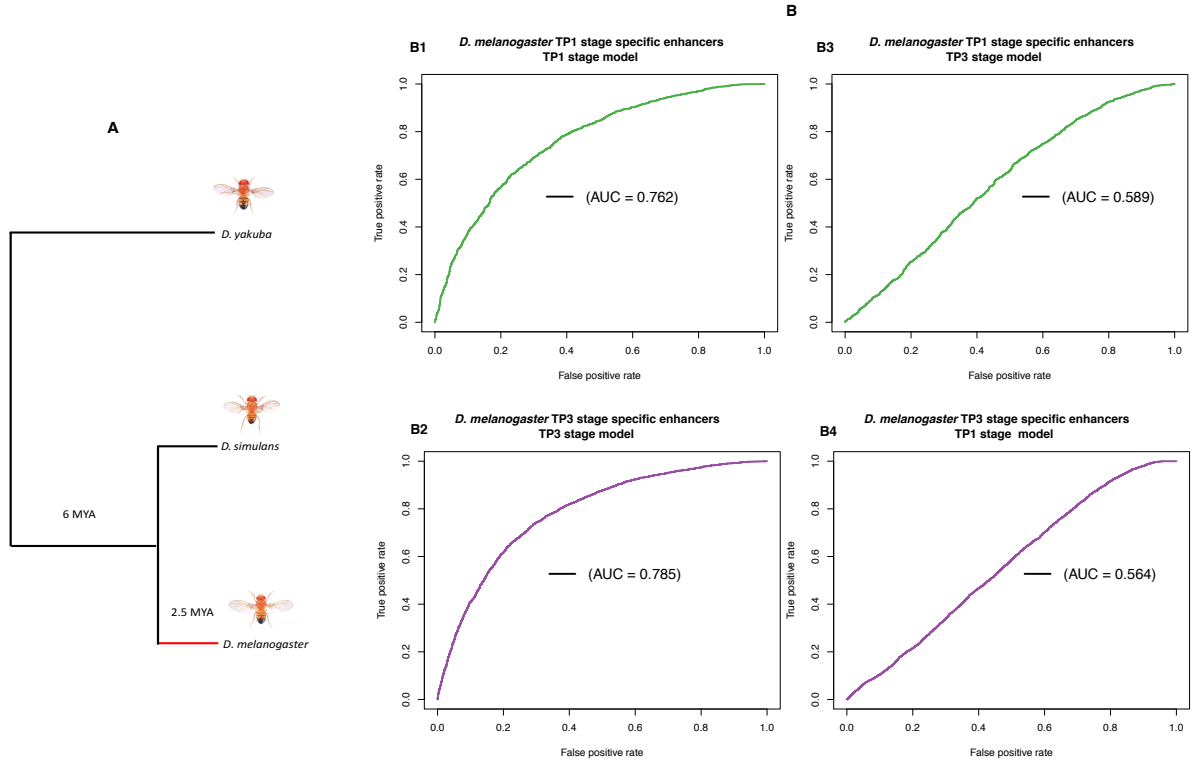

Figure 4: Gapped k-mer support vector machine (gkm-SVM) can predict stage specific enhancers

A. Topological illustration of the phylogenetic relationships between the three Drosophila species used to detect positive selection on enhancers. We want to detect positive selection which occurred on the lineage of D. melanogaster after divergence from D. simulans, as indicated by the red branch. D. yakuba is the outgroup used to infer enhancer sequence in the ancestor of D. melanogaster and D. simulans.

B. Receiver operating characteristic (ROC) curves for gkm-SVM classification performance on stage specific enhancers. AUC values represent areas under the ROC curve and provide an overall measure of predictive power.
B1. The results of a 5-fold cross validation on TP1 stage specific enhancers and matched random sequences.
B2. The results of a 5-fold cross validation on TP3 stage specific enhancers and matched random sequences.
B3. The gkm-SVM trained in TP3 used to predict TP1 stage specific enhancers and matched random sequences.
B4. The gkm-SVM trained in TP1 used to predict TP3 stage specific enhancers and matched random sequences.

To identify sequence features within enhancer regions that determine their developmental stage dependent regulatory activity, we first separately trained a gapped *k*-mer support vector machine (gkm-SVM) for 1'747 TP1 stage specific enhancers and for 6'063 TP3 stage specific enhancers (see Materials and Methods). This gkm-SVM identifies sequence features that determine chromatin accessibility in the corresponding developmental context. We did not consider TP2, because its low number of enhancers (610) limits our power to train a gkm-SVM (Lee et al. 2015). Interestingly, a gkm-SVM trained in one stage can accurately distinguish enhancers of the same stage from random sequences, but has almost no power to distinguish enhancers of the other stage from random sequences (Figure 4B, see Materials and Methods). This shows that the gkm-SVM 's predictions are not only accurate, but also developmental stage specific.

A list of weights quantifying the contribution of each possible 10-mer to chromatin accessibility in each stage was generated through a scoring function from the trained gkm-SVM (see Materials and Methods). We derived the substitutions accumulated on the *D. melanogaster* lineage for each enhancer from the whole genome pairwise alignments of *D. melanogaster* to *D. simulans* and *D. yakuba* (see Materials and Methods). We only kept enhancers with at least two substitutions, leading to 1'625 and 5'855 enhancers for TP1 and TP3 respectively. For each enhancer, we calculated deltaSVM, the effect of substitutions on the gkm-SVM function score. This can be interpreted as the effect of these substitutions on chromatin accessibility change. To infer significance, for each enhancer, we also generated a null distribution of deltaSVM based on 10'000 times random in silico mutagenesis (see Materials and Methods). This procedure results in a p-value for each enhancer, reflecting the probability that the observed deltaSVM could arise by chance. Thus, we can identify enhancers whose substitution pattern on the *D. melanogaster* branch have effects on chromatin

accessibility which are inconsistent with neutrality, and therefore imply positive selection on enhancers.

The distributions of p-values for all stage specific enhancers in TP1 and TP3 show a strong skew toward low p-values (Figure 5A and B), indicating evidence for positive selection. Since mutations under positive selection will spread through a population rapidly, they decrease polymorphisms if they were recent while increasing substitutions (McDonald and Kreitman 1991). Based on this reasoning, we predict that positive selection enhancers should have higher substitutions to polymorphisms ratios than non-positive selection enhancers. For this test and the downstream analyses, we use *p*-value 0.01 as a threshold to define an enhancer under positive selection. To test our prediction, we counted the numbers of substitutions between *D. melanogaster* and *D. simulans*, and of polymorphisms among 205 *D. melanogaster* inbred lines, separately for positive selection and non-positive selection enhancers (see Materials and Methods). As expected, in both TP1 and TP3, positive selection enhancers have a significant excesses of fixed nucleotide changes relative to non-positive selection enhancers (Figure 5C and D). This confirms that our method detects positive selection in enhancer evolution.

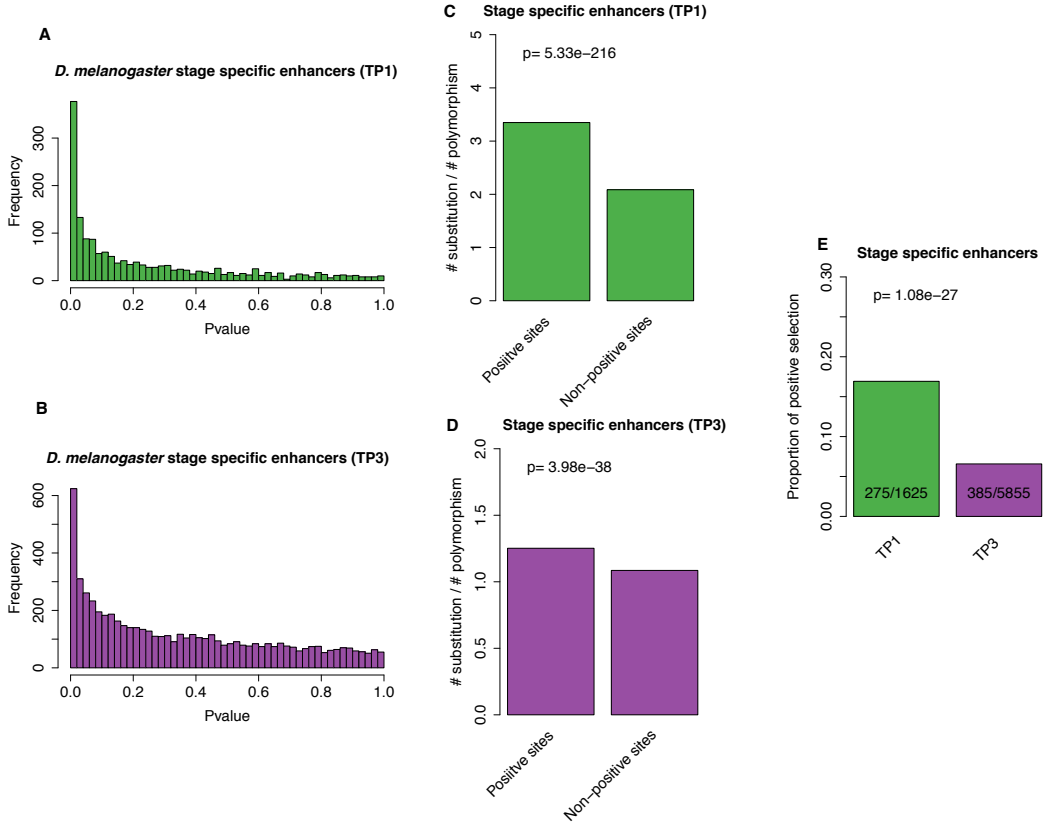

Figure 5: The phylotypic stage has a lower proportion of enhancers with evidence of positive selection

A. The distribution of deltaSVM p-values (test for positive selection) for TP1 stage specific enhancers.

B. The distribution of deltaSVM p-values (test for positive selection) for TP3 stage specific enhancers.

C. The ratio between the number of substitutions and the number of polymorphisms (SNPs) for TP1 stage specific enhancers. Positive sites are enhancers with evidence of positive selection (deltaSVM p-value < 0.01), non-positive sites are enhancers without evidence of positive selection. The p-value from Fisher's exact test is reported above thebars.

D. The ratio between the number of substitutions and the number of polymorphisms for TP3 stage specific enhancers. Legend as in panel C.

E. The proportion of enhancers with evidence of positive selection in the two stages. The number of stage specific enhancers and the number of stage specific enhancers with evidence of positive selection in each development stage is indicated inside each bar. The p-value from Fisher's exact test is reported above the bars.

Finally, we compared the proportion of positive selection enhancers between TP1 and TP3, and found a significantly higher proportion in TP1 (Figure 5E). For TP1, there are 16.6 % of stage specific enhancers with evidence of positive selection, vs. only 6.6% in TP3. Results are consistent using *p*-value 0.05 as a threshold (Figure S3). These results suggest that positive selection is an important driver of the evolution of enhancers, and that conservation in the phylotypic stage is not only characterized by stronger purifying selection on enhancers, but also less positive selection.

4.4 Discussion

4.4.1 Stronger purifying selection on regulatory elements at phylotypic stage

We performed DNase-seq in two distant *Drosophila* species across multiple matched embryonic development stages. Based on this unique dataset, we identified a set of highly conserved regulatory elements in different development stages of *D. melanogaster*. We found that there is a higher proportion of conserved regulatory elements and a lower proportion of stage specific gain regulatory elements in the phylotypic stage than other stages. These results suggest that there is stronger purifying selection on gene regulation at this stage. Why has the phylotypic stage underwent stronger purifying selection? Based on the observation that genes expressed at the phylotypic stage are also trend to be expressed at other stages, Hu et al. (2017) proposed that the evolution of gene expression in this stage was constrained by pleiotropic effects. However, since here we only focused on stage specific regulatory elements, this cannot explain our observations. One possibility, as suggested by Raff (1996), is that the genes expressed at the phylotypic stage are very crucial for the high interdependence of

different developmental modules at this stage. For example, in *C. elegans*, Zalts and Yanai (2017) found the phylotypic stage enriched with 'integration genes', their expression resulting from interactions among the cell types across germ layers. These integration gene are enriched with gene ontology (GO) terms such as cell-communication and signalling functions. In line with these findings, we also found the target genes of phylotypic stage (TP3) specific conserved promoters enriched with the GO terms signaling (GO:0023052) and cell communication (GO:0007154) (Table S2). In contrast, we did not find the same GO terms for the target genes of TP1 specific conserved promoters (Table S3). We did not perform the same analysis on the phylotypic stage specific conserved enhancers, since it's still a challenge to link enhancers to their target genes. The future work should focus on identifying the target genes of conserved enhancers, e.g. by using Hi-C (genome-wide chromosome conformation capture assays) technology (Lieberman-Aiden et al. 2009).

4.4.2 Widespread positive selection on developmental enhancers

Our results indicate that widespread positive selection shaped the evolution of developmental enhancers, especially of early development enhancers (TP1, 2–4 h after egg laying). There are 17% of early stage specific enhancers with evidence of positive selection (p-value < 0.01). Intuitively, since the later development stages is built on the earlier stages, the early development should not be so accessible to adaptation. In addition, many studies indicated that the evolution of cis-regulatory elements is the main driver of morphologic diversity (Gompel et al. 2005; McGregor et al. 2007; Wray 2007; Jeong et al. 2008). For example, the origin of male-specific wing pigmentation spot in *Drosophila biarmipes* was caused by the gain of multiple transcription factor binding sites for the yellow pigmentation gene (Gompel et al. 2005). However, for most changes in the early development, there are no clear consequences on the adult morphology (Kalinka and Tomancak 2012). This raises the question of why there are so many adaptive enhancer changes in early development? One explanation, proposed by Kalinka and Tomancak (2012), was that many variations in early development result from adaptation to diverse ecological circumstances. This can be illustrated by the evolution of insect segmentation patterns (Liu and Kaufman 2005), either short or long germ segmentation. *Drosophila* adopts the long germ segmentation pattern. From anterior to posterior, all of the body segments are simultaneously specified before gastrulation. *Oncopeltus*, however, adopts the short germ segmentation pattern, only the anterior segments established before gastrulation. Notably, the different germ segmentation patterns seem without a clear influence on the body

plan evolution between *Drosophila* and *Oncopeltus*. Instead, a major role of the long germ segmentation is shortening the embryonic development time, which is likely an adaptive strategy to a particular ecological niches (Kalinka and Tomancak 2012). These results highlight the importance of adaptive cis-regulatory evolution on the evolution of life history.

4.4.3 Expression divergence at the phylotypic stage shaped by negative selection, positive selection, and mutational robustness

We provide evidence that the lower expression divergence at the phylotypic stage can be explained by both stronger purifying selection and less adaptive evolution on gene regulations. Interestingly, recent studies found evidences that the phylotypic stage also has stronger mutational robustness on gene regulation. For example, Zalts and Yanai (2017) compared the expression variations of 20 *C. elegans* mutation accumulation strains across embryonic development and found that the phylotypic stage has lower expression variation. Since mutation accumulation experiment mostly remove the effect of positive selection, this study indicates that developmental constraints (both purifying selection and mutational robustness) can contribute to the hourglass model of expression evolution. In fly, we have recently also shown that there is stronger histone modification based mutational robustness in the phylotypic stage (Liu et al. 2019). These results suggest a contribution of mutational robustness to conserved expression divergence at the phylotypic stage. Overall, the low expression divergence at the the phylotypic stage seems to have been shaped by the interplay of purifying selection, positive selection, and mutational robustness.

4.5 Materials and Methods

4.5.1 DNase-seq sample processing

Raw paired-end reads were first trimmed with Trim Galore https://github.com/FelixKrueger/TrimGalore. Then the trimmed reads were mapped to *D. melanogaster* genome (Flybase Assembly 6 version: dm6) and to *D. virilis* genome (Flybase-R1.2 assembly version: droVir3) respectively by using Bowtie2 (Langmead and Salzberg 2012) with standard parameters. For peak calling, we used MACS2 (Zhang et al. 2008) with standard parameters. We derived peaks using 5% Irreproducible Discovery Rate (IDR) threshold leading to a unique highly confident and consistent peak sets for biological replicates.

4.5.2 Genome coordinates translation

Since *D. melanogaster* and *D. virilis* are highly diverged, the *D. virilis* peak coordinates were translated to *D. melanogaster* genome coordinates by using the pslMap (Zhu et al. 2007), as suggested by Khoueiry et al. (2017). Notably, liftover is more suitable for translation between different genome assemblies, while pslMap is more suitable for translation between different species.

4.5.3 Definition of enhancers and promoters

All peaks were assigned into either putative enhancer category or putative promoter category based on their distance to transcriptional state sites (TSSs). If a peak had at least 1bp overlap with a 500bp region around a TSS, we assigned it into the putative promoter category. Otherwise, we assigned it into the putative enhancer category.

4.5.4 Sequence alignment files

The pairwise whole genome alignments between *D. melanogaster* and *D. simulans* or *D. yakuba* were downloaded from http://hgdownload.soe.ucsc.edu/downloads.html (accessed in December, 2018).

4.5.5 Single Nucleotide Polymorphism (SNP) data

Over 4.8 million SNPs for 205 *D. melanogaster* inbred lines were downloaded from the Drosophila Genetic Reference Panel (DGRP) http://dgrp2.gnets.ncsu.edu/ (Huang et al. 2014, accesed in December, 2018).

4.5.6 In silico mutagenesis for detecting positive selection on enhancers

1. Train the gapped k-mer support vector machine (gkm-SVM)

gkm-SVM is a method for predicting regulatory DNA sequence by using *k*-mer frequencies (Ghandi et al. 2014) . For the gkm-SVM training, we followed the same approaches as Lee et al. (2015). Firstly, we defined a positive training set and its corresponding negative training set. The positive training set is stage specific enhancers. The negative training set is an equal number of sequences randomly sampled from the genome with matched length, GC content and repeat fraction of the positive training set. This negative training set was generated by using "genNullSeqs", a function of the gkm-SVM R package (Ghandi et al. 2016). Then, we trained a gkm-SVM with default parameters except -l=10 (meaning we use 10-mer as feature

to distinguish positive and negative training sets). The classification performance of the trained gkm-SVM was measured by using receiver operating characteristic (ROC) curves with fivefold cross-validation. The gkm-SVM training and cross-validation were achieved by using the "gkmtrain" function of "LS-GKM" (Lee 2016) (also see https://github.com/Dongwon-Lee/lsgkm).

2. Test the stage specificity of gkm-SVM

To test the performance of gkm-SVM trained in TP1 in predicting TP3 stage specific enhancers, we first scored both positive and negative training sets in TP3 by using the TP1 gkm-SVM. This is achieved by using the "gkmpredict" function of "LS-GKM". Then, the ROC curve was used to evaluate the prediction performance of TP1 gkm-SVM in TP3 data. In a similar way, we also tested the performance of gkm-SVM trained in TP3 in predicting TP1 stage specific enhancers.

3. Generate SVM weights of all possible 10-mers

The SVM weights of all possible 10-mers were generated by using the "gkmpredict" function of "LS-GKM". A positive value means increasing chromatin accessibility, a negative value means decreasing accessibility, and value close to 0 means no impact on chromatin accessibility (i.e. function measured in this case).

4. Infer ancestor sequence

With the sequence alignment files between *D. melanogaster* and *D. simulans* or *D. yakuba*, we derived substitutions in the focal species lineage. The ancestor sequence was inferred from sequence alignment between *D. melanogaster* and *D. simulans* by using *D. yakuba* as an outgroup.

5. Calculate deltaSVM

We calculated the sum of weights of all 10-mers for ancestor sequence and focal sequence respectively. The deltaSVM is the sum of weights of the focal sequence minus the sum of weights of the ancestor sequence. A positive deltaSVM indicates substitutions increasing the chromatin accessibility in the focal sequence, and vice versa.

6. Generate Empirical Null Distribution of deltaSVM

Firstly, we counted the number of substitutions between each ancestor sequence and focal sequence. Then, we generated a random pseudo-focal sequence by randomly introducing the same number of substitutions to the ancestor sequence. Finally, we calculated the deltaSVM

between the pseudo-focal sequence and the ancestor sequence. We repeated this process 10'000 times to get 10'000 expected deltaSVMs.

7. Calculate *p*-value of deltaSVM

The *p*-value was calculated as the probability that the expected deltaSVM is higher than the observed deltaSVM. The *p*-value can be interpreted as the probability that the observed deltaSVM could arise by chance under the assumptions of the randomisation.

4.5.7 Gene ontology (GO) enrichment analysis

We performed GO enrichment analysis for hourglass expression variability genes by using the topGO (Alexa et al. 2006) R package with the "elim" algorithm.

4.6 Acknowledgements

We thank 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) centre for high-performance computing of the SIB Swiss Institute of Bioinformatics. This work was supported by Swiss National Science Foundation grant 31003A_153341 / 1.

4.7 Supporting Information

Supporting materials can be downloaded from:

https://github.com/ljljolinq1010/Chromatin-accessibility-evolution-during-Drosophila-embryogenesis/tree/master/supplementary_figures_tables

Table S1: Spearman's correlation coefficients for bigwig signals of replicates.

Table S2: Top 10 enriched gene ontology terms (ranked by *p*-value) for the target genes of TP3 specific conserved promoters.

Table S3: Top 10 enriched gene ontology terms (ranked by *p*-value) for the target genes of TP1 specific conserved promoters.

Figure S1: The number of putative enhancers and promoters across development

Figure S2: The proportion of relaxed definition conserved regulatory elements across development

Here the conservation means the distance between stage specific enhancers in the two species must be smaller than 1kb, not necessarily overlap. The figure legend is the same as Figure 2B and 2C.

Figure S3: The proportion of enhancers with relaxed evidence of positive selection

Here the evidence of positive selection means the deltaSVM p-value < 0.05. The figure legend is the same as Figure 5E.

4.8 References

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22:1600–1607.

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.

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

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.

Ghandi M, Lee D, Mohammad-Noori M, Beer MA. 2014. Enhanced Regulatory Sequence Prediction Using Gapped k-mer Features.Morris Q, editor. PLoS Comput. Biol. 10:e1003711.

Ghandi M, Mohammad-Noori M, Ghareghani N, Lee D, Garraway L, Beer MA. 2016. gkmSVM: an R package for gapped-kmer SVM. Bioinformatics 32:2205–2207.

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. 2005. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature 433:481–487.

Hu H, Uesaka M, Guo S, Shimai K, Lu T-M, Li F, Fujimoto S, Ishikawa M, Liu S, Sasagawa Y, et al. 2017. Constrained vertebrate evolution by pleiotropic genes. Nat. Ecol. Evol. 1:1722–1730.

Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF, et al. 2014. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. 24:1193–1208.

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.

Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. 2008. The Evolution of Gene Regulation Underlies a Morphological Difference between Two Drosophila Sister Species. Cell 132:783–793.

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.

Khoueiry P, Girardot C, Ciglar L, Peng P-C, Gustafson EH, Sinha S, Furlong EE. 2017. Uncoupling evolutionary changes in DNA sequence, transcription factor occupancy and enhancer activity. Elife 6.
King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107–116.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.
Lee D. 2016. LS-GKM: a new gkm-SVM for large-scale datasets. Bioinformatics 32:2196–2198.
Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, Beer MA. 2015. A method to predict the impact of regulatory variants from DNA sequence. Nat. Genet. 47:955–961.
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. 2012. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev. Cell 22:1101–1108.
Li Q, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5:1752–1779.
Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293.
Liu J, Frochaux M, Gardeux V, Deplancke B, Robinson-Rechavi M. 2019. An hourglass pattern of inter-embryo gene expression variability and of histone regulation in fly embryogenesis. bioRxiv:700997.
Liu J, Robinson-Rechavi M. 2018. Adaptive Evolution of Animal Proteins over Development: Support for the Darwin Selection Opportunity Hypothesis of Evo-Devo.Zhang J, editor. Mol. Biol. Evol. 35:2862–2872.
Liu PZ, Kaufman TC. 2005. Short and long germ segmentation: unanswered questions in the evolution of a developmental mode. Evol. <html_ent glyph="@amp;" ascii="&"/> Dev. 7:629–646.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654.
McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, Payre F, Stern DL. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature 448:587–590.
Raff RA. 1996. The shape of life : genes, development, and the evolution of animal form. 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.
Sander K. 1976. Specification of the Basic Body Pattern in Insect Embryogenesis. Adv. In Insect Phys. 12:125–238.
Valentine JW. 2004. On the origin of phyla. University of Chicago Press
Wallace A. 2000. The origin of animal body plans : a study in evolutionary developmental biology. Cambridge University Press
Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8:206–216.

Yanai I. 2018. Development and Evolution through the Lens of Global Gene Regulation. Trends Genet. 34:11–20.
Zalts H, Yanai I. 2017. Developmental constraints shape the evolution of the nematode mid-developmental transition. Nat. Ecol. Evol. 1:0113.
Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9:R137.
Zhu J, Sanborn JZ, Diekhans M, Lowe CB, Pringle TH, Haussler D. 2007. Comparative Genomics Search for Losses of Long-Established Genes on the Human Lineage. PLoS Comput. Biol. 3:e247.
Zrzavy J, Stys' P. 1997. The basic body plan of arthropods: insights from evolutionary morphology and developmental biology. Birkhluser Verlag

5 A robust method for detecting positive selection on regulatory sequences detects recent adaptation in human brain enhancers

Jialin Liu, marc Robinson-Rechavi

This article is in preparation and should be submitted soon.

5.1 Abstract

There is a long standing hypothesis that divergent phenotypic traits between human and chimpanzee divergent traits might have been driven by regulatory level adaptions rather than protein sequence adaptations. Especially for cognitive abilities related traits which make our species unique and are plausibly adaptive, suggesting regulatory level adaptions may play an important role in the evolution of the human brain. A limitation to testing this has been the lack of a reliable way to detect positive selection on regulatory sequences. We present an in silico mutagenesis based approach to transcription factor binding sites evolution, based on a machine learning model of binding. Unlike other methods, this requires neither defining a priori neutral sites, nor detecting accelerated evolution, which have both been shown to be biased metrics for non-coding elements. The method is validated by a McDonald-Kreitman-like measure of polymorphism vs. divergence, by experimental binding site gains and losses, and by changes in expression levels. We scanned the signals of positive selection for CTCF binding sites in 29 human and 11 mouse tissues or cell types. We found that human brain related cell types have the highest proportion of positive selection. This is consistent with the importance of adaptive evolution on gene regulation in the evolution of the human brain.

5.2 Introduction

Changes in gene regulation have been suggested to constitute an important part of human evolution, maybe considerably more important than changes in protein sequences (King and Wilson 1975). This is because the divergence of orthologous proteins between human and chimpanzee are small relative to the apparently large anatomical and behavioral differences between the two species (King and Wilson 1975; Anon 2005). Although abundant gene expression differences exist between human and chimpanzee (Enard et al. 2002), a central question which is not well addressed is which of these changes on regulatory elements are adaptive, as opposed to neutral? A major reason for this is the lack of a null model of neutral evolution for regulatory elements.

Genome wide studies of sequence level adaptive evolution in human genome have reported that genes with evidences of positive selection are mainly related with functional classes such as: immune response, spermatogenesis, and olfaction (Clark et al. 2003; Bustamante et al. 2005; Nielsen et al. 2005). Yet, many human and chimpanzee divergent traits (Varki and Altheide 2005) cannot be explained by protein sequence adaptations. For example, there is little evidence to link protein sequence adaptations to cognitive abilities related traits, such as the size of brain, the number of neurons, and cortical connections (Franchini and Pollard 2015). This that indicates regulatory level adaptions may play an important role in shaping human brain evolution. The discovery of human accelerated regions (HARs), DNA sequences with increased substitution rates in the human lineage but conserved among other vertebrates or primates, has provided several suggestive evidences (Pollard et al. 2006; Prabhakar et al. 2006; Gittelman et al. 2015). For example, Gittelman et al. (2015) found the putative target genes of HARs enriched with many developmental related gene ontology terms, including brain and neuron development. In addition, Haygood et al. (2007) found that promoters with appreciably faster evolution rate than their associated introns are linked with genes involved in neurogenesis and glucose metabolism. Although these findings are promising, they may be biased by several factors. For example, the study of Gittelman et al. (2015) was restricted to highly conserved regions, which only constitute a small part of regulatory elements. What's more, the methods they applied might not be detecting actual signal of positive selection. For example, the high evolution rate of promoters found in Haygood et al. (2007) may be due to high mutation rate, not positive selection.

To address the previous limitations and systematically assess the role of regulatory adaptations on human brain evolution, we first developed a method to detect positive selection of transcription factor binding sites (TFBSs) evolution based on binding affinity changes. This is achieved by comparing the observed binding affinity changes in evolution to a null distribution. The effects of substitutions on binding affinity change can be accurately predicted by deltaSVM (Lee et al. 2015), a machine leaning based method to predict the effects of regulatory variations de novo from sequence. As a proof-of-principle, we applied this method to several well-known transcription factors (such as CEBPA and CTCF) in species triples focused on human, mouse, or fly. We validated it by three approaches: TFBSs with evidence of positive selection have higher empirical binding affinity, higher substitution to polymorphism ratio in sequence, and lower variance in expression of neighboring genes. Then, we used this method to detect positive selection of CTCF binding sites in 29 human tissues or cell types. We found the highest positive selection in brain samples, followed by male reproductive system. For comparative purposes, we performed the same analysis in a more limited mouse dataset, where the highest positive selection is in digest system, with no special signal in the brain. Thus, we combine machine leaning and quantitative genetics to detect positive selection on regulatory sequences in a powerful and robust manner, and provide strong evidence for adaptive evolution in the human brain.

5.3 Results

5.3.1 Detecting positive selection on TFBSs

Here we developed a computational model to detect positive selection on transcription factor biding sites (TFBSs), or any other elements for which we have experimental evidence similar to ChIP-seq (Figure 1) (see Methods). Briefly, a gapped k-mer support vector machine (gkm-SVM) classifier was trained by using TFBSs (narrow peaks from ChIP-seq) as a positive training set and randomly sampled sequences from the genome as a negative training set (Figure 1I). Then, SVM weights of all possible 10-mers, the contributions of prediction transcription factor binding affinity (TFBA), were generated from the gkm-SVM. Based on this weight list, we can predict the impact of any substitution on TFBA by calculating deltaSVM, the difference of sum weights between a focal sequence (species A) and its ancestor sequence (Figure 1II B). The ancestor sequence was inferred from sequence alignment with a sister species and an outgroup (Figure 1II A). We use the principle of Orr's sign test, namely that adaptive changes should be consistently in one phenotypic direction, to propose that

TFBSs with a large deltaSVM evolved adaptively. The significance of the observed deltaSVM was evaluated by comparing it with a null distribution of deltaSVM, constructed by scoring the same number of random substitutions 10'000 times (Figure 1II C and D). This procedure results in a *p*-value reflecting the probability that the observed deltaSVM could arise by chance. Thus, by comparing to a null model of deltaSVM, our method can identify sets of substitutions whose effects on binding affinity change are inconsistent with random drift, and therefore imply the action of natural selection, pushing the TFBS either to a new fitness peak of either higher (positive deltaSVM) or lower (negative deltaSVM) affinity than its ancestral state.

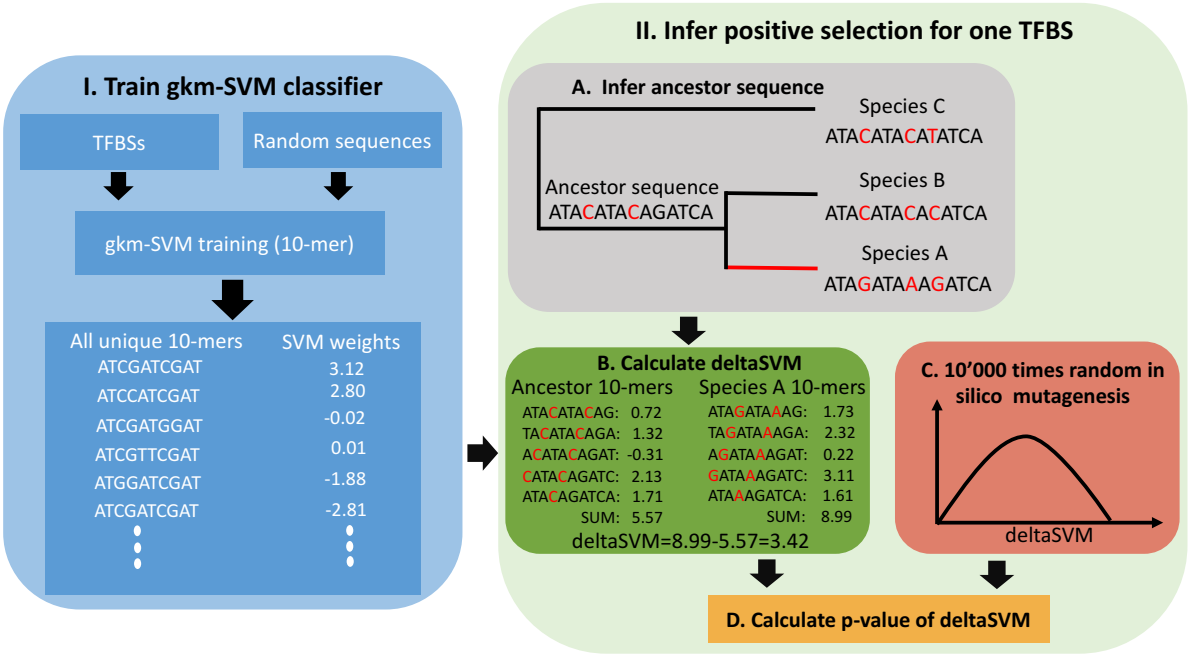

Figure 1: Illustration of the procedure for inferring positive selection
The method includes two parts. Part I (left) is the gapped k-mer support vector machine (gkm-SVM) model training. The gkm-SVM classifier was trained by using TFBSs as a positive training set and randomly sampled sequences from the genome as a negative training set. Then, SVM weights of all possible 10-mers, the contributions of prediction transcription factor binding affinity, were generated from the gkm-SVM. Part II (right) part is the positive selection inference. The ancestor sequence was inferred from sequence alignment with a sister species (species B) and an outgroup (species C). Then, the binding affinity change (deltaSVM) of the two substitutions accumulated in the red branch leading to species A were calculated based on the weight list. The significance of the observed deltaSVM was evaluated by comparing it with a null distribution of deltaSVM, constructed by scoring the same number of random substitutions 10'000 times.

5.3.2 Detecting positive selection on liver TFBSs in *Mus musculus*

To demonstrate its application to real data, we first took advantage of a large set of TFBSs in liver of three mouse species (*Mus musculus domesticus* C57BL/6J, *Mus musculus castaneus*

CAST/EiJ, and *Mus spretus* SPRET/EiJ), identified by ChIP-seq for three liver-specific transcription factors, CEBPA, FOXA1, and HNF4A (Stefflova et al. 2013). Here, we want to detect the positive selection occurred on the lineage leading to C57BL/6J after divergence from CAST/EiJ (Figure 2A). For the sake of simplicity, we only present the results of CEBPA in the main text. We also performed the same analyses for the other two transcription factors (FOXA1 and HNF4A), and results were consistent. We first trained a gkm-SVM on 41'945 CEBPA binding sites in C57BL/6J (see Methods). The gkm-SVM can very accurately separate CEBPA binding sites and random sequences (Figure 2B). Then, a list of weights quantifying the contribution of each possible 10-mer to CEBPA biding affinity in liver was generated through a scoring function from the trained gkm-SVM. Among all C57BL/6J CEBPA binding sites, we focused on three categories: conserved binding site (conserved, 24'280 sites), lineage specific gain binding site (gain, 6'304 sites), and lineage specific loss binding site (loss, 6'692 sites). They were identified based on the experimental ChIP-seq peaks in the three species, using SPRET/EiJ as an outgroup (see Methods). Based on the whole genome pairwise alignments of C57BL/6J to CAST /EiJ and to SPRET/EiJ, we derived the substitutions accumulated on the C57BL/6J lineage for each CEBPA binding site (see Methods). We only kept binding sites with at least two substitutions, leading to 5'114, 1'445, and 1'497 sites left for conserved, gain and loss categories respectively. For each binding site, we calculated deltaSVM, the effect of all accumulated substitutions on the CEBPA binding affinity change. To infer the significance, for each biding site, we also generated a null distribution of deltaSVM based on 10'000 times random in silico mutagenesis (see Methods). We performed higher-tail test for both conserved and gain, and lower-tail test for loss.

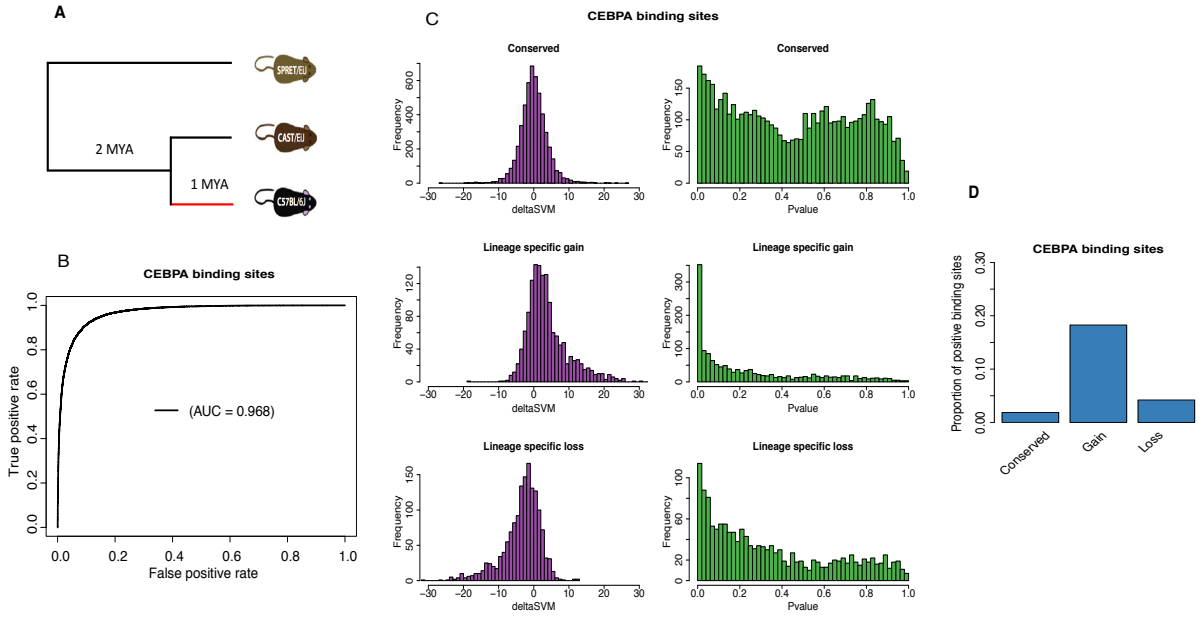

Figure 2: Mouse CEBPA biding sites study

A. Topological illustration of the phylogenetic relationships between the three mouse species used to detect positive selection on CEBPA binding sites. We want to detect positive selection which occurred on the lineage of C57BL/6J after divergence from CAST/EiJ, as indicated by the red branch. SPRET/EiJ is the outgroup used to infer binding site sequence in the ancestor of C57BL/6J and CAST/EiJ.

B. Receiver operating characteristic (ROC) curve for gkm-SVM classification performance on CEBPA binding sites (5-fold cross validation). The AUC value represents the area under the ROC curve and provides an overall measure of predictive power.

C. The left hand graphs are the distributions of deltaSVM for conserved, gain, and loss binding sites. The right hand graphs are the distributions of deltaSVM p-values (test for positive selection) for conserved, gain, and loss binding sites.

D. The proportion of CEBPA binding sites with evidence of positive selection in conserved, gain, and loss binding sites.

We plot the distributions of deltaSVMs and of its corresponding p-values for each binding site category (Figure 2C). As expected, the distribution of deltaSVMs is symmetric for conserved, but asymmetric for both gain and loss. Also as expected, for gain most binding sites have a positive deltaSVM, whereas, for loss most binding sites have a negative deltaSVM. These results confirm that the gkm-SVM based approach can accurately predict the effect of substitutions on transcription factor binding affinity change. For the distribution of p-values, in all binding site categories, there is a skew of p-values near zero, indicating that each category has some signal of positive selection. The gain has the most skewed distribution of p-values towards zero, followed by loss then conserved. Here after we will use 0.01 as a significant threshold to define positive selection, but results are robust to different thresholds. We found

almost 20% of gain experienced positive selection (Figure 2D), relative to 4% of loss and 2% of conserved.

In summary, we found widespread positive selection driving the gain and loss of CEBPA binding sites. In addition, for conserved binding sites, we also found that some sites experienced positive selection to increase binding affinity. For the other two transcription factors (FOXA1 and HNF4A), we found highly consistent patterns (Figure S1, S2).

5.3.3 Validation based on Chip-seq binding intensity

For both conserved and gain, the sites which evolved under positive selection are predicted to have increased binding affinity. Thus the positive binding sites (PBSs) should have higher biding affinity than non-positive selection binding sites (NPBSs). To valid this prediction, we compared the ChIP-seq intensity in C57BL/6J between PBSs and NPBSs. Indeed, we found that the PBSs have significantly higher intensity than NPBSs (Figure 3A and B). In addition, the conserved generally have higher binding intensity than gain, indicating that recently evolved TFBSs have lower activity than conserved ones. For loss, however, the expectation is slightly more complex. Loss driven by positive selection is expected to have affected sites which were functionally important (if they were not important, they might have been lost by drift), so the ancestor binding affinity of positive loss is expected to be higher than of non-positive loss. We used the ChIP-seq intensity from CAST/EiJ as an approximation for ancestor binding affinity. Indeed, we found the positive loss ones have higher intensity than non-positive loss in CAST/EiJ (Figure 3C). We performed the same validations in FOXA1 and HNF4A, with consistent results (Figure S3).

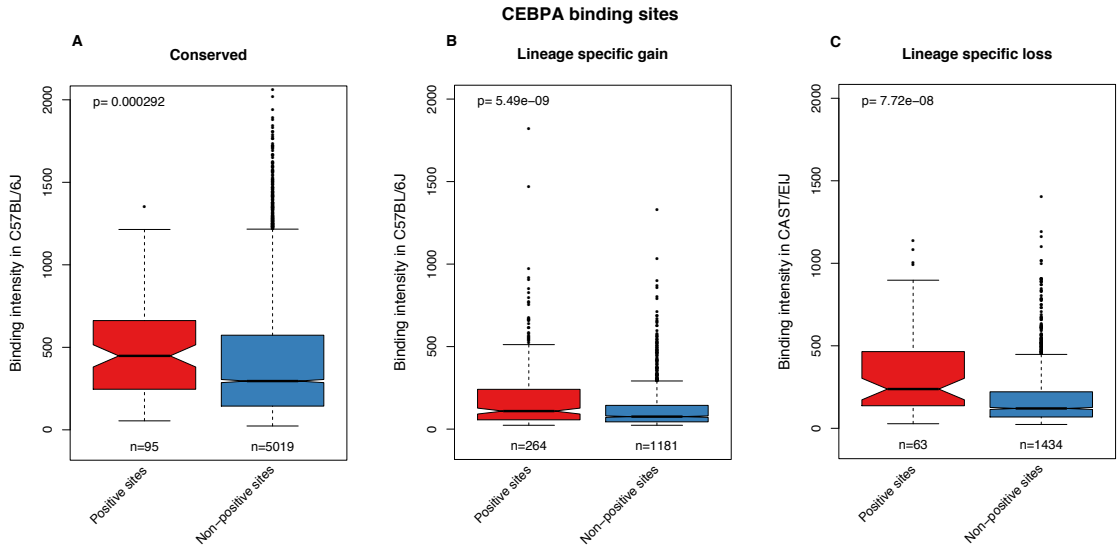

Figure 3: Comparison of biding intensity between positive sites and non-positive sites for mouse CEBPA.

The number of binding sites in each category is indicated below each box. The p-values from a Wilcoxon test comparing categories are reported above boxes. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range, or the maximum (respectively minimum) if no points are beyond 1.5 IQR (default behaviour of the R function boxplot). Positive sites are binding sites with evidence of positive selection (deltaSVM p-value < 0.01), non-positive sites are binding sites without evidence of positive selection.

A. Conserved binding sites.

B. Lineage specific gain binding sites.

C. Lineage specific loss binding sites. We compare the binding intensity from CAST/EiJ, as an approximation for ancestral binding intensity, between positive loss binding sites and non-positive loss binding sites.

5.3.4 Validating the inference of positive selection with human liver TFBSs

To further validate our method, we took advantage of the abundant population genomics data and population transcriptomics data in humans. Here, we only have ChIP-seq in one species, human, so we cannot distinguish experimentally gain or loss of TFBSs. We inferred positive selection of CEBPA binding sites in the human lineage after divergence from chimpanzee (Figure 4A). As in mouse, the gkm-SVM trained from 15'806 CEBPA binding sites in human can very accurately separate TFBSs and random sequences (Figure 4B). For selection inference, we used all CEBPA binding sites with at least two substitutions (5'807 sites). We derived substitutions by using gorilla as an outgroup, and performed in silico mutagenesis approach to test positive selection. The distribution of deltaSVMs is slightly asymmetric, with a higher proportion of positive values (Figure 4C). This is because the binding sites in here contain both conserved and gain, but obviously no loss (since we detect only in the focal species). Based on

the distribution of p-values, we infer that 7.5% of CEBPA binding sites experienced positive selection on the human lineage.

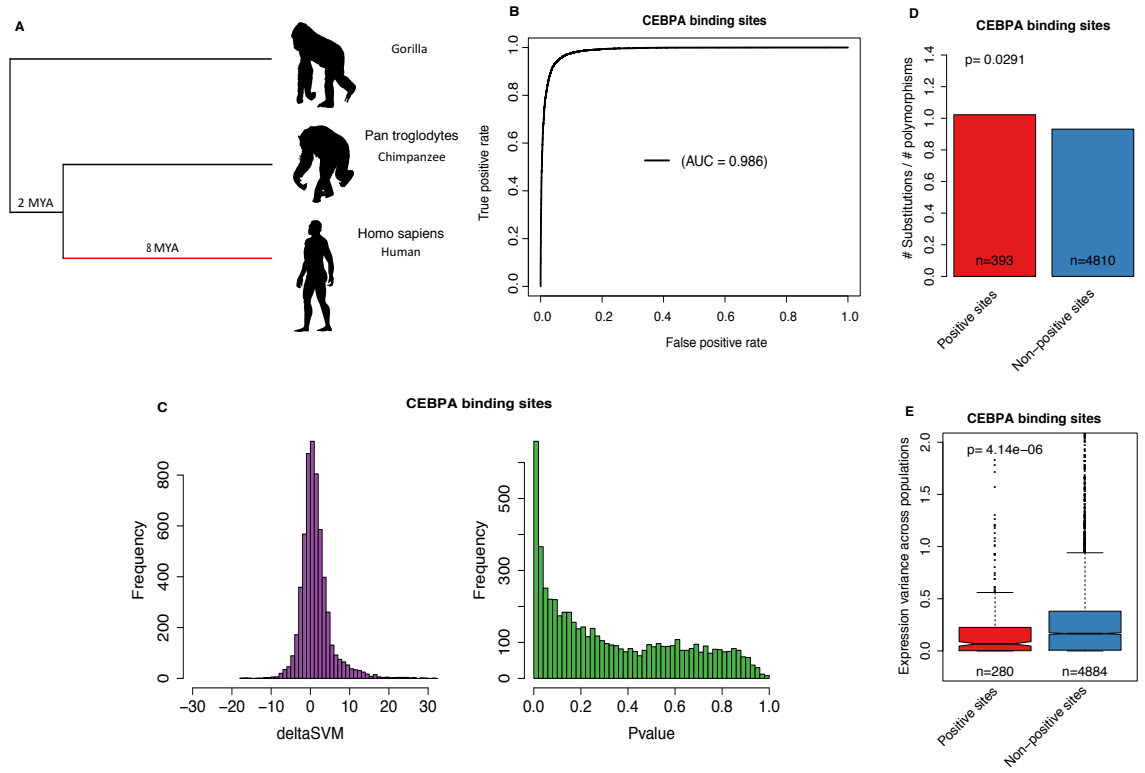

Figure 4: Human CEBPA biding sites study

A. Topological illustration of the phylogenetic relationships between human, chimpanzee and gorilla. We detected positive selection which occurred on the lineage of human after divergence from chimpanzee, as indicated by the red branch. Gorilla is the outgroup used to infer binding site sequence in the ancestor of human and chimpanzee.

B. Receiver operating characteristic (ROC) curve for gkm-SVM classification performance on CEBPA binding sites (5-fold cross validation). The AUC value represents the area under the ROC curve and provides an overall measure of predictive power.

C. The left graph is the distribution of deltaSVM. The right graph is the distribution of deltaSVM p-values (test for positive selection).

D. The ratio between the number of substitutions and the number of polymorphisms (SNPs) for CEBPA binding sites. Positive sites are binding sites with evidence of positive selection (deltaSVM p-value < 0.01), non-positive sites are binding sites without evidence of positive selection. The p-value from Fisher's exact test is reported above the bars.

E. Comparison of expression variance of putative target genes (closest gene to a TFBS) between positive sites and non-positive sites. The number of binding sites in each category is indicated below each box. The p-values from a Wilcoxon test comparing categories are reported above boxes. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range, or the maximum (respectively minimum) if no points are beyond 1.5 IQR (default behaviour of the R function boxplot). Positive sites are binding sites with evidence of positive selection (deltaSVM p-value < 0.01), non-positive sites are binding sites without evidence of positive selection.

Mutations under positive selection will spread through a population rapidly, so they don't contribute to polymorphisms but do increase substitutions (McDonald and Kreitman 1991). Based on this theory, we predict that PBSs should have higher substitutions to polymorphisms ratios than NPBSs. To test this prediction, we counted the number of substitutions between human and chimpanzee, and of polymorphisms among human populations (see Methods), for PBSs and NPBSs respectively. As expected, we found that the PBSs have significant excesses of fixed nucleotide changes (Figure 4D), indicating that positive selection has shaped CEBPA binding sites evolution. This is an external validation of our method, as the input did not contain any information about polymorphism.

Besides higher substitutions to polymorphisms ratio, we also expect that the expression of PBSs putative target genes (see Methods) should be under stronger constraints across human populations than of NPBSs putative target genes. Because the expression of PBSs target genes is an adaptive trait to the given environment for human, any changes in expression will reduce fitness. Moreover, recent adaptive sweeps might have reduced functional variability for the regulation of these genes. To measure evolutionary constraints of gene expression, we calculated the expression variance across human populations for PBSs target genes and NPBSs target genes separately. As expected, we found that PBSs target genes have significantly lower expression variance than NPBSs target genes, indicating stronger constraints.

In summary, we found that PBSs have higher interspecies to population divergence in sequence, and lower variance in expression of putative target genes. In addition, we performed the same analyses in HNF4A, and results are consistent (Figure S4). These results strongly suggest that our method is detecting real adaptive evolution signals, not just statistically significant p-values due to some data bias.

5.3.5 Detecting positive selection of TFBSs in *Drosophila melanogaster*

By using a McDonald-Kreitman test framework (McDonald and Kreitman 1991), Ni et al. (2012) detected signatures of adaptive evolution on CTCF binding sites in *D. melanogaster*. They reported that positive selection has shaped CTCF binding evolution, and that newly gained binding sites show a stronger signal of positive selection than conserved sites. To further valid our method, we applied it to the same data as used in Ni et al. (2012). We detected positive selection in the *D. melanogaster* lineage after divergence from *D. simulans* (Figure 5A and

5B). Consistent with the findings of Ni et al. (2012), we observed widespread positive selection for both conserved and gain (Figure 5C). Their method does not report selection on specific sites, so that we cannot compare directly our results. In addition, the gain have a higher proportion of positive selection than conserved (Figure 5D). For lineage specific loss binding sites, however, we did not detect any signal of positive selection (Figure 5C). Interestingly, the proportion of positive selection in *D. melanogaster* is much higher than in *Mus musculus*. For example, we find almost 40% of gain under positive selection in *D. melanogaster*, twice the proportion in *Mus musculus*.

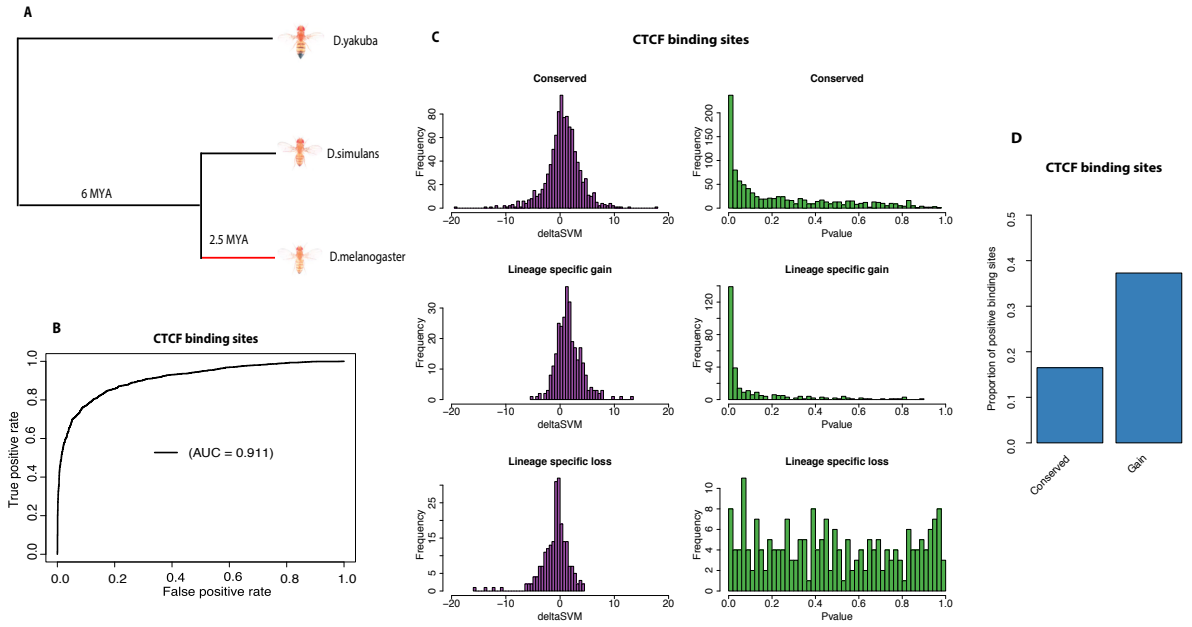

Figure 5: D. melanogaster CTCF biding sites study

A. Topological illustration of the phylogenetic relationships between the three Drosophila species used to detect positive selection on CTCF binding sites. We want to detect positive selection which occurred on the lineage of D. melanogaster after divergence from D. simulans, as indicated by the red branch. D. yakuba is the outgroup used to infer binding site sequence in the ancestor of D. melanogaster and D. simulans.

B. Receiver operating characteristic (ROC) curve for gkm-SVM classification performance on CEBPA binding sites (5-fold cross validation). The AUC value represents the area under the ROC curve and provides an overall measure of predictive power.

C. The left hand graphs are the distributions of deltaSVM for conserved, gain, and loss binding sites. The right hand graphs are the distributions of deltaSVM p-values (test for positive selection) for conserved, gain, and loss binding sites.

D. The proportion of CEBPA binding sites with evidence of positive selection in conserved and gain binding sites.

5.3.6 Adaptive evolution of CTCF binding sites across tissues in human

To study the adaptive evolution of TFBSs between different human tissues, especially test whether there is stronger adaptive evolution on human brain, we applied our method to 80074 CTCF binding sites across 29 adult tissues or primary cell types (see Table S2). CTCF is well known as a transcriptional repressor, but it also involved in transcriptional insulation and chromatin architecture remodelling (Phillips and Corces 2009). We chose CTCF because it was the factor with the largest number of tissues or primary cell types studied in a consistent manner, by the ENCODE consortium (The ENCODE Project Consortium 2012). Moreover CTCF is highly conserved in higher eukaryotes and is involved in diverse cellular processes (Phillips and Corces 2009), giving us the opportunity to compare positive selection on TFBSs of the same factor in different tissues and species. For the sake of simplicity, we use cell type not tissue hereafter.

The gkm-SVM model trained from one cell type can accurately predict the biding sites in another cell type, and the model trained with all CTCF binding sites has better performance than the model trained with cell type specific binding sites (Figure S5), so we did not train different models for different cell types, but we used a general gkm-SVM instead. Based on the *p*-value distribution of deltaSVM (Figure S6A), we detected 3.52% of binding sites having evolved under positive selection on the human lineage after divergence from chimpanzee. We found PBSs associated with a lower number of active cell types (Figure S7A) than NPBSs, consistent with the prediction that pleiotropy can limit adaptive evolution (Wagner and Zhang 2011).

We next ranked the cell types according to the proportion of binding sites that exhibit statistically significant evidence of positive selection (Figure 6A). Generally, we found nervous system has the highest rank. Specifically, among all the nervous related cell types, only the brain related cell types (choroid plexus epithelial cell, brain microvascular endothelial cell, retinal pigment epithelial cell, astrocyte of the cerebellum and astrocyte of the spinal cord) have higher proportion of positive selection than other system related cell types. For example, we did not find that tibial nerve has higher adaptation than other cell types. Notabally, we did not find high adaptation for in vitro differentiated neural cell, this may reflect that in vitro differentiated cells do not preserve the signal of specific in vivo

differentiated cells. Overall, these results suggest that, at gene regulatory evolution level, adaptive evolution mainly targeted brain related cell types.

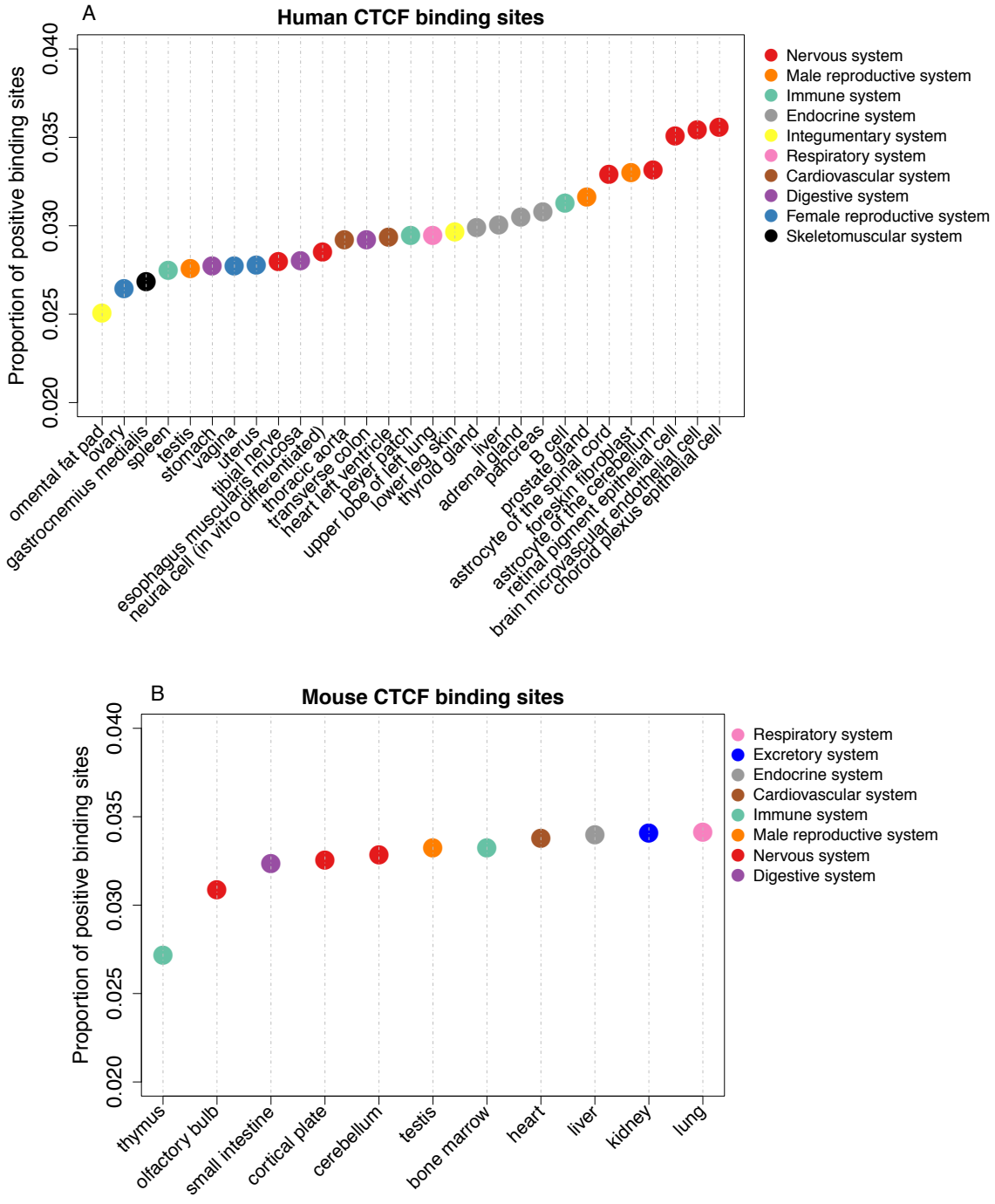

***Figure 6*: Proportion of positive CTCF binding sites in different tissues or cell types.** *Positive binding sites are binding sites with evidence of positive selection (deltaSVM p-value < 0.01). Colours correspond to broad anatomical systems.*
A. CTCF binding sites in 29 human tissues or cell types.
B. CTCF binding sites in 11 mouse tissues.

To test whether the high regulatory adaptive evolution in brain is typical of human, we trained gkm-SVM (Figure S8) on comparable CTCF binding data from 11 mouse adult tissues (Table S2). We investigated adaptive evolution in *M. musculus* branch after divergence from *M. spretus*. The extent of evolutionary divergence between *M. musculus* and *M. spretus* is in the same order of magnitude as that between human and chimpanzee (Enard et al. 2002). Similarly to human, we detected 3.54% binding sites which evolved under positive selection (Figure S6B). However, we did not find that brain related tissues (olfactory, cortical plate and cerebellum) have higher adaptive regulatory evolution than others in mouse, unlike human (Figure 6B).

5.4 Discussion

5.4.1 A robust test for positive selection on regulatory elements

Detecting positive selection on regulatory sequences has long been a difficult problem (Zhen and Andolfatto 2012). Inferring positive selection is generally based on rejecting a null model of neutral evolution, so a suitable neutral reference is crucial. Generally, for the current sequence based methods, nearby non-coding regions of interested regulatory elements have been chosen as neutral reference (Andolfatto 2005; Haygood et al. 2007; Arbiza et al. 2013). But this might be problematic. For example, to infer positive selection on promoters, Haygood et al. (2007) compared the sequence evolution rates between promoters and its adjacent introns. This is based on the assumption that introns are selectively neutral, and thus represent the background mutation rate. However, since introns can perform various important functions (Fedorova and Fedorov 2003), the higher evolution rate of some promoters than its associated introns may reflect negative selection on these introns, rather than positive selection on promoters.

Our approach does not assume a subset of neutral sites and not compare sequence evolutionary rates, but instead consider the effects of variations on regulatory activities (Berg et al. 2004; Moses 2009; Smith et al. 2013). In this study, specifically, we consider the effects of variations on transcription factor binding affinity. We compared the observed binding affinity change to the expectation under a no-selection model to assess when neutrality can be rejected. In addition, our approach should have higher power than sequence based methods even if appropriate neutral reference sites be used in this type of method (Smith et al. 2013). Indeed, positive selection on a background of negative selection might not elevate the the evolutionary

rate of a TFBS above the neutral expectation, yet consistent changes in binding affinity can still be detectable (Figure S9). Moreover, this method could also be applied to other cis-regulatory related regions such as open chromatin regions or histone modification regions. The only need is to have experimental ChIP-seq or similar data, on which the gkm-SVM model should be trained.

To go beyond providing a *p*-value for neutrality rejection, we validated our method in several ways. Firstly, with the ChIP-seq based empirical measurements, we validated the accuracy of deltaSVM for predicting transcription factor biding affinity change. For example, for lineage specific gain TFBSs, we found positive selection TFBSs have higher binding affinity than non-positive selection TFBSs (Figure 3B). This is consistent with our prediction that the positive gain TFBSs are the ones with higher intensity increase than neutral expectation. Secondly, with a substitution to polymorphism MK-like-test, we found that the positive selection TFBSs have higher substitution to polymorphism ratio than non-positive selection TFBSs (Figure 4D). This is consistent with the prediction that positive selection will increase substitutions but remove polymorphism (McDonald and Kreitman 1991). Thirdly, with gene expression data across populations, we found that the nearby genes of positive selection TFBSs have lower expression variance than of non-positive selection TFBSs (Figure 4E). This is consistent with the prediction that positive selection will increase the fitness of organism, so the corresponding phenotype will be under stronger constraints among populations. Fourthly, we found the positive selection TFBSs have lower number of active cell types than non-positive selection TFBSs (Figure S7). This is consistent with the prediction that pleiotropy can limit adaptation (Wagner and Zhang 2011). Finally, we found the high adaptive cell types do not always have high number of substitutions (Figure S9). This indicates that our method is not just detecting fast evolving cell types in terms of regulatory sequence evolution, but those where the evolution is putatively adaptive.

Despite its advantages, our method can still be improved. For example, in the null model of sequence evolution, we assume independent mutation patterns at each base-pair site and a uniform mutation rate over all sites. But both mutation rate and pattern can depend on neighbouring nucleotides (Krawczak et al. 1998). In addition, our analysis focuses only on point mutations, ignoring insertions and deletions. It will be of interest in future work to integrate both context dependent information and structural variations into the null model.

5.4.2 The importance of regulatory adaptation on human brain evolution.

There are several cases of genes with signatures of sequence level positive selection in human lineage which have a function in brain (Vallender and Lahn 2004). For example, ASPM, a gene function in regulating brain size, has been reported to experienced positive selection after the split of human and chimpanzee (Zhang 2003). However, such genes are not prominent in genome wide studies of sequence level adaptive evolution in the human genome (Clark et al. 2003; Bustamante et al. 2005; Nielsen et al. 2005; Daub et al. 2017).

Our results are consistent with the hypothesis of importance of adaptive regulatory evolution in human brain evolution (King and Wilson 1975). Appling our in-silico mutagenesis framework to transcription factor binding sites of CTCF, we tested positive selection on increasing binding affinity, on the human lineage after divergence from chimpanzee. We compared the proportion of CTCF binding sites with evidence of positive across 29 cell types, and found brain related cell types have higher proportion of positive selection than other cell types. These brain related cell types are functionally related with cognitive abilities. For example, for astrocyte, abnormal astrocytic signalling can cause synaptic and network imbalances, leading to cognitive impairment (Santello et al. 2019). In addition, for choroid plexus epithelial cell, its atrophy has been reported to related with Alzheimer disease, a neurodegenerative disorder with progressive memory and cognitive impairment (Kaur et al. 2016). While we did not find a similar pattern by applying the same analysis to mouse, it isn't possible yet to conclude to a human or primate specific pattern. Indeed, the mouse analyses have two potential *caveats*. First, for the olfactory bulb and cortical plate in the mouse *analyses*, there is no corresponding anatomical structures in the human analyses. It is an open question whether the human olfactory bulb and cortical also have high adaptation. Second, the human analyses were based on ChIP-seq at cell type level but the mouse analyses were based on ChIP-seq at tissue level. In mouse, the astrocyte in cerebellum may also have high adaptation like the astrocyte in human, but the signal might be diluted by other cell types in cerebellum.

These findings refine many previous studies (Enard et al. 2002; Pollard et al. 2006; Prabhakar et al. 2006; Haygood et al. 2007; Gittelman et al. 2015). Firstly, for example, the evidences of positive selection of previous studies are based on human accelerated regions (HARs). As we

discussed in the introduction, these regions may not reflect positive selection but high mutation rate. However, based on all the validations performed in this study, the positive selection signals detected by our method appear quite robust. Secondly, to link the HARs to brain, previous studies generally performed gene ontology (GO) enrichment analyses on putative target genes of these HARs. In this way, Gittelman et al. (2015) found HARs linked with some developmental related GO terms, including brain and neuron development. However, based on comparing the rate of positive selection between regulatory elements in different tissues, we think our results provide more direct and clear evidences for human brain adaptation.

These results also extend the study from Enard et al. (2002), investigated the gene expression evolution in different tissues (blood, liver and brain) in the three primates, and three rodents. For primate analyses, they found brain experienced accelerated gene expression evolution in the human evolutionary lineage relative to the chimpanzee, whereas such an acceleration is not observed in liver and blood. For rodent, they did not found brain has stronger acceleration than others. Our results are remarkably consistent with these, and suggest that this acceleration in human brain might be driven by enhanced positive selection not relaxed purifying selection.

5.4.3 Regulatory adaptation in other tissues

Generally, we found that male reproduction system (prostate and foreskin) has higher adaptive regulatory evolution than female reproduction system (ovary, uterus and vagina). This is consistent with the observation of high adaptive sequence evolution in human male reproduction (Wyckoff et al. 2000; Nielsen et al. 2005), and probably caused by sexual selection related selective pressures, such as sperm competition. However, for testis, we found very similar proportion of adaptive evolution with female reproduction system. This suggests that the high expression divergence in testis (Brawand et al. 2011) is mainly caused by relaxed purifying selection. But why does testis not have high adaptive evolution of TFBSs like prostate and foreskin? This might be related with a recent finding that the transcription in testis serves to correct germline mutations through 'transcriptional scanning' (Xia et al. 2018). In addition, we also observed high adaptive regulatory evolution in immune and endocrine systems. Like in male reproduction system, these systems also have high adaptive sequence evolution (Clark et al. 2003; Bustamante et al. 2005; Nielsen et al. 2005). These results suggest that for most cell types, excepting brain related cell types, the adaptive evolution patterns on

sequence level and on regulatory level are very similar, with high adaptive regulatory evolution usually coupled with high adaptive sequence evolution, although the magnitude of adaptive sequence evolution appears lower than regulatory adaptive evolution. This also raises a very interesting question of why brain has high adaptive regulatory evolution but not adaptive sequence evolution? This might be related to selective pressure against misfolding in brain proteins (Drummond and Wilke 2008; Roux et al. 2017).

5.5 Materials and Methods

5.5.1 Mutagenesis for positive selection

1. Training of the gapped k-mer support vector machine (gkm-SVM)

gkm-SVM is a method for regulatory DNA sequence prediction by using *k*-mer frequencies (Ghandi et al. 2014) . For the gkm-SVM training, we followed the approach of Lee et al. (2015). Firstly, we defined a positive training set and its corresponding negative training set. The positive training set is ChIP-seq narrow peaks of transcription factors. The negative training set is an equal number of sequences which randomly sampled from the genome with matched the length, GC content and repeat fraction of the positive training set. This negative training set was generated by using "genNullSeqs", a function of gkm-SVM R package (Ghandi et al. 2016). Then, we trained a gkm-SVM with default parameters except -l=10 (meaning we use 10-mer as feature to distinguish positive and negative training sets). The classification performance of the trained gkm-SVM was measured by using receiver operating characteristic (ROC) curves with fivefold cross-validation. The gkm-SVM training and cross-validation were achieved by using the "gkmtrain" function of "LS-GKM : a new gkm-SVM software for large-scale datasets" (Lee 2016). For details, please check https://github.com/Dongwon-Lee/lsgkm.

2. Generate SVM weights of all possible 10-mers

The SVM weights of all possible 10-mers were generated by using the "gkmpredict" function of "LS-GKM". The positive value means increasing binding affinity, the negative value means decreasing binding affinity, the value close to 0 means functionally neutral.

3. Infer ancestor sequence

The ancestor sequence was inferred from sequence alignment with a sister species and an outgroup.

4. Calculate deltaSVM

We calculated the sum of weights of all 10-mers for ancestor sequence and focal sequence respectively. The deltaSVM is the sum weights of focal sequence minus the sum weights of

ancestor sequence. The positive deltaSVM indicating substitutions increased the binding affinity in the focal sequence, vice versus.

5. Generate Empirical Null Distribution of deltaSVM

Firstly, we counted the number of substitutions between the ancestor sequence and the focal sequence. Then, we generated a random pseudo-focal sequence by randomly introducing the same number of substitutions to the ancestor sequence. Finally, we calculated the deltaSVM between the pseudo-focal sequence and the ancestor sequence. We repeated the above processes 10'000 times to get 10'000 expected deltaSVMs.

6. Calculate *p*-value of deltaSVM

For both conserved and lineage specific gain TFBSs, the *p*-value was calculated as the probability that the expected deltaSVM is higher than the observed deltaSVM. For lineage specific lose TFBSs, the *p*-value was calculated as the probability that the expected deltaSVM is lower than the observed deltaSVM. The *p*-value can be interpreted as the probability that the observed deltaSVM could arise by chance.

5.5.2 Mouse validation analysis

1. ChIP-seq data

The narrow ChIP-seq peaks and their corresponding intensity (normalized read count) datasets of three liver specific transcription factors (CEBPA, FOXA1 and HNF4A) in three mouse species (C57BL/6J, CAST/EiJ, SPRET/EiJ) were downloaded from https://www.ebi.ac.uk/research/flicek/publications/FOG09 (Stefflova et al. 2013, accessed in May, 2018). Peaks were called with SWEMBL (http://www.ebi.ac.uk/~swilder/SWEMBL). To account for both technical and biological variabilities of peak calling, Stefflova et al. (2013) carried out the following approaches. For each transcription factor in each species, they first called three sets of peaks: one for each replicate (replicate peek), and one for a pooled dataset of both replicates (pooled peak). Then, the peaks detected from the pooled dataset were used as a reference to search for overlaps in the two other replicates. When a pooled peak overlapped with both replicate peeks (at least one base pair overlap), it was kept for downstream analyses. For the number of peaks and average peak length, please check Table S1.

2. Peak coordinates transfer

Based on pairwise genome alignments between C57BL/6J and CAST/EiJ or SPRET/EiJ, Stefflova et al. (2013) transferred the coordinates of ChIP-seq peaks in both CAST/EiJ and SPRET/EiJ to its corresponding coordinates in C57BL/6J.

3. Sequence alignment files

The sequence alignment files between C57BL/6J and CAST/EiJ or SPRET/EiJ were downloaded from https://www.ebi.ac.uk/research/flicek/publications/FOG09 (Stefflova et al. 2013, accessed in May, 2018).

4. Define different types of binding sites

1) Conserved binding sites

The conserved binding sites were defined as peaks in C57BL/6J which have overlapping peaks (at least one base pair overlap) in the other two species by genome alignment.

2) Lineage specific gain binding sites

The lineage specific gain binding sites defined as peaks in C57BL/6J with no overlapping peaks (at least one base pair overlap) in the other two species.

3) Lineage specific loss binding sites

The lineage specific loss binding sites defined as peaks in CAST/EiJ which overlapping peaks in SPRET/EiJ but not in C57BL/6J.

5.5.3 Human validation analysis

1. ChIP-seq data

The narrow ChIP-seq peaks datasets of two liver specific transcription factors (CEBPA and HNF4A) in human were downloaded from https://www.ebi.ac.uk/research/flicek/publications/FOG01 (Schmidt et al. 2010, accessed in October, 2018). Peaks were called with SWEMBL (http://www.ebi.ac.uk/~swilder/SWEMBL). Negligible variation was observed between the individuals in terms of peak calling, so Schmidt et al. (2010) pooled replicates into one dataset for peak calling.

2. Sequence alignment files

The pairwise whole genome alignments between human and chimpanzee or gorilla were downloaded from http://hgdownload.soe.ucsc.edu/downloads.html (accessed in December, 2018).

3. Single nucleotide polymorphism (SNP) data

Over 36 million SNPs for 1,092 individuals sampled from 14 populations worldwide were downloaded from phase I of the 1000 Genomes Project ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/integrated_call_sets/ (1000 Genomes Project Consortium 2012, accesed in December, 2018). As suggested by Luisi et al.

(2015), we only used SNPs of a subset of 270 individuals from YRI, CEU, and CHB populations.

4. Liver expression data

The library site normalized expression data of 175 livers was downloaded from downloaded from The Genotype Tissue Expression (GTEx) project https://gtexportal.org/home/ (GTEx Consortium 2017, Release V7, accessed in December, 2018). We further transformed it with $\log_2$.

5. Putative target genes of TFBSs

We assigned the nearest gene to each TFBS as its putative target gene.

5.5.4 Fly validation analysis

1. ChIP-seq data

The narrow ChIP-seq peaks of transcription factor CTCF in three drosophila species (*D. melanogaster*, *D. simulans* and *D. yakuba*) were downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24449 (Ni et al. 2012, accessed in January, 2019). Peaks were called with QuEST(Valouev et al. 2008) at a False Discovery Rate (FDR) <1%. We obtained 2'182, 2'197 and 2'993 peaks with average length of 243bp, 240bp and 201bp for *D. melanogaster*, *D. simulans* and *D. yakuba* respectively.

2. Peak coordinates transfer

The peaks identified in *D. simulans* and *D. yakuba* were translated onto *D. melanogaster* coordinates by using pslMap (Zhu et al. 2007).

3. Sequence alignment files

The pairwise whole genome alignments between *D. melanogaster* and *D. simulans* or *D. yakuba* were downloaded from http://hgdownload.soe.ucsc.edu/downloads.html (accessed in January, 2019).

4. Define different types of binding sites. These were defined as in human, using *D. melanogaster* vs. *D. simulans* and *D. yakuba*.

5.5.5 Human CTCF analysis

1. ChIP-seq data

The narrow ChIP-seq peaks of transcription factor CTCF across 29 tissues or cell types (Table S2) were downloaded from ENCODE (The ENCODE Project Consortium 2012). We did not use ChIP-seq datasets from cell lines, and only kept ChIP-seq datasets from tissues and primary

cells. Briefly, peaks were called with MACS (Zhang et al. 2008) separately for each replicate. Irreproducible Discovery Rate (IDR) analysis was then performed (Li et al. 2011). Final peaks are the set of peak calls that pass IDR at a threshold of 2%. Peaks identified in different tissues or cell types were integrated by intersecting all peaks across datasets, with at least one base pair overlap used as the merge criteria. Overall we obtained 118970 merged peaks.

2. Sequence alignment files

The pairwise whole genome alignments between human and chimpanzee or gorilla were downloaded from http://hgdownload.soe.ucsc.edu/downloads.html (accessed in December, 2018).

5.5.6 Mouse CTCF analysis

1. CHiP-seq data

The narrow ChIP-seq peaks of transcription factor CTCF across 11 tissues (Table S2) were downloaded from ENCODE (The ENCODE Project Consortium 2012). Briefly, peaks were called with MACS (Zhang et al. 2008) separately for each replicate. Irreproducible Discovery Rate (IDR)) analysis was then performed. Final peaks are the set of peak calls that pass IDR at a threshold of 2%. Peak identified in different tissues/cell types were integrated by intersecting all peaks across data sets, with at least 1 base pair overlap used as the merge criteria. Overall we obtained 112657merged peaks.

2. Sequence alignment files

The sequence alignment file between C57BL/6J and SPRET/EiJ, please check "Mouse validation analysis" part of Materials and Methods. The sequence alignment file between C57BL/6J and Caroli/EiJ were downloaded from https://www.ebi.ac.uk/research/flicek/publications/FOG09 (Stefflova et al. 2013, accessed in May, 2018).

5.6 Acknowledgements

We thank 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) centre for high-performance computing of the SIB Swiss Institute of Bioinformatics. This work was supported by Swiss National Science Foundation grant 31003A_153341 / 1.

5.7 Supporting Information

Supporting materials can be downloaded from:
https://github.com/ljljolinq1010/A-robust-method-for-detecting-positive-selection-on-regulatory-sequences/tree/master/supplementary_figures_tables

Table S1: The number of peaks and average peak length for CEBPA, FOXA1 and HNF4A in three mouse species C57BL/6J, CAST/EiJ and SPRET/EiJ.

Table S2: The CTCF ChIP-seq datasets information for human and mouse.

Figure S1: mouse FOXA1 biding sites study

The figure legend is the same as Figure 2.

Figure S2: mouse HNF4A biding sites study

The figure legend is the same as Figure 2.

Figure S3: Comparison of biding intensity between positive sites and non-positive sites for mouse FOXA1 and HNF4A

The figure legend is the same as Figure 3.

Figure S4: human HNF4A biding sites study

The figure legend is the same as Figure 4.

Figure S5: Receiver operating characteristic (ROC) curves for gkm-SVM classification performance on human CTCF binding sites.

AUC values represent areas under the ROC curve and provide an overall measure of predictive power.

A. The results of a 5-fold cross validation on neural CTCF binding sites and matched random sequences.

B. The gkm-SVM trained in B cell used to predict neural CTCF binding sites and matched random sequences.

C. The gkm-SVM trained in all 29 tissues/cell types used to predict neural CTCF binding sites and matched random sequences.

D. The results of a 5-fold cross validation on B cell CTCF binding sites and matched random sequences.

E. The gkm-SVM trained in neural cell used to predict B cell CTCF binding sites and matched random sequences.

F. The gkm-SVM trained in all 29 tissues/cell types used to predict B cell CTCF binding sites and matched random sequences.

Figure S6: The distribution of deltaSVM and deltaSVM *p*-values for CTCF biding sites in human and mouse

Figure S7: Comparison of the number of active tissues/cell types for CTCF biding sites between positive sites and non-positive sites.

The *p*-values from a Wilcoxon test comparing categories are reported above boxes. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range, or the maximum (respectively minimum) if no points are beyond 1.5 IQR (default behaviour of the R function boxplot). Positive sites are binding sites with evidence of positive selection (deltaSVM *p*-value < 0.01), non-positive sites are binding sites without evidence of positive selection.

Figure S8: Receiver operating characteristic (ROC) curve for gkm-SVM classification performance on mouse CTCF binding sites.

The result comes from a 5-fold cross validation on all CTCF binding sites from 11 tissues and matched random sequences. The AUC value represents areas under the ROC curve and provides an overall measure of predictive power.

Figure S9: Average number of substitutions for non-positive CTCF binding sites in different tissues/cell types.

Non-positive binding sites are binding sites with evidence of positive selection (deltaSVM *p*-value >= 0.01). The number of substitutions measured by The pairwise whole genome alignments between human and chimpanzee.

A. CTCF binding sites in 29 human tissues/cell types.

B. CTCF binding sites in 11 mouse tissues.

5.8 References

1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature 437:1149–1152.

Anon. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.

Arbiza L, Gronau I, Aksoy BA, Hubisz MJ, Gulko B, Keinan A, Siepel A. 2013. Genome-

wide inference of natural selection on human transcription factor binding sites. Nat. Genet. 45:723–729.

Berg J, Willmann S, Lässig M. 2004. Adaptive evolution of transcription factor binding sites. BMC Evol. Biol. 4:42.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343–348.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Todd Hubisz M, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, et al. 2005. Natural selection on protein-coding genes in the human genome. Nature 437:1153–1157.

Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, et al. 2003. Inferring Nonneutral Evolution from Human-Chimp-Mouse Orthologous Gene Trios. Science (80-.). 302:1960–1963.

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.

Drummond DA, Wilke CO. 2008. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. Cell 134:341–352.

Enard W, Khaitovich P, Klose J, Zöllner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, et al. 2002. Intra- and interspecific variation in primate gene expression patterns. Science 296:340–343.

Fedorova L, Fedorov A. 2003. Introns in gene evolution. Genetica 118:123–131.

Franchini LF, Pollard KS. 2015. Can a few non-coding mutations make a human brain? Bioessays 37:1054–1061.

Ghandi M, Lee D, Mohammad-Noori M, Beer MA. 2014. Enhanced Regulatory Sequence Prediction Using Gapped k-mer Features.Morris Q, editor. PLoS Comput. Biol. 10:e1003711.

Ghandi M, Mohammad-Noori M, Ghareghani N, Lee D, Garraway L, Beer MA. 2016. gkmSVM: an R package for gapped-kmer SVM. Bioinformatics 32:2205–2207.

Gittelman RM, Hun E, Ay F, Madeoy J, Pennacchio L, Noble WS, Hawkins RD, Akey JM. 2015. Comprehensive identification and analysis of human accelerated regulatory DNA. Genome Res. 25:1245–1255.

GTEx Consortium. 2017. Genetic effects on gene expression across human tissues. Nature

550:204–213.

Haygood R, Fedrigo O, Hanson B, Yokoyama K-D, Wray GA. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. Nat. Genet. 39:1140–1144.

Kaur C, Rathnasamy G, Ling E-A. 2016. The Choroid Plexus in Healthy and Diseased Brain. J. Neuropathol. Exp. Neurol. 75:198–213.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107–116.

Krawczak M, Ball E V., Cooper DN. 1998. Neighboring-Nucleotide Effects on the Rates of Germ-Line Single-Base-Pair Substitution in Human Genes. Am. J. Hum. Genet. 63:474–488.

Lee D. 2016. LS-GKM: a new gkm-SVM for large-scale datasets. Bioinformatics 32:2196–2198.

Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, Beer MA. 2015. A method to predict the impact of regulatory variants from DNA sequence. Nat. Genet. 47:955–961.

Li Q, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5:1752–1779.

Luisi P, Alvarez-Ponce D, Pybus M, Fares MA, Bertranpetit J, Laayouni H. 2015. Recent positive selection has acted on genes encoding proteins with more interactions within the whole human interactome. Genome Biol. Evol. 7:1141–1154.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654.

Moses AM. 2009. Statistical tests for natural selection on regulatory regions based on the strength of transcription factor binding sites. BMC Evol. Biol. 9:286.

Ni X, Zhang YE, Nègre N, Chen S, Long M. 2012. Adaptive Evolution and the Birth of CTCF Binding Sites in the Drosophila Genome. PLoS Biol 10:1001420.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, et al. 2005. A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees.Tyler-Smith C, editor. PLoS Biol. 3:e170.

Phillips JE, Corces VG. 2009. CTCF: master weaver of the genome. Cell 137:1194–1211.

Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, Siepel A, Pedersen JS, Bejerano G, Baertsch R, et al. 2006. Forces Shaping the Fastest Evolving Regions in the

Human Genome. PLoS Genet. 2:e168.

Prabhakar S, Noonan JP, Paabo S, Rubin EM. 2006. Accelerated Evolution of Conserved Noncoding Sequences in Humans. Science (80-.). 314:786–786.

Roux J, Liu J, Robinson-Rechavi M. 2017. Selective Constraints on Coding Sequences of Nervous System Genes Are a Major Determinant of Duplicate Gene Retention in Vertebrates. Mol. Biol. Evol. 34:2773–2791.

Santello M, Toni N, Volterra A. 2019. Astrocyte function from information processing to cognition and cognitive impairment. Nat. Neurosci. 22:154–166.

Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science 328:1036–1040.

Smith JD, McManus KF, Fraser HB. 2013. A Novel Test for Selection on cis-Regulatory Elements Reveals Positive and Negative Selection Acting on Mammalian Transcriptional Enhancers. Mol. Biol. Evol. 30:2509–2518.

Stefflova K, Thybert D, Wilson MD, Streeter I, Aleksic J, Karagianni P, Brazma A, Adams DJ, Talianidis I, Marioni JC, et al. 2013. Cooperativity and Rapid Evolution of Cobound Transcription Factors in Closely Related Mammals. Cell 154:530–540.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74.

Vallender EJ, Lahn BT. 2004. Positive selection on the human genome. Hum. Mol. Genet. 13:R245–R254.

Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S, Myers RM, Sidow A. 2008. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat. Methods 5:829–834.

Varki A, Altheide TK. 2005. Comparing the human and chimpanzee genomes: searching for needles in a haystack. Genome Res. 15:1746–1758.

Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. Nat. Rev. Genet. 12:204–213.

Wyckoff GJ, Wang W, Wu C-I. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature 403:304–309.

Xia B, Baron M, Yan Y, Wagner F, Kim SY, Keefe DL, Alukal JP, Boeke JD, Yanai I. 2018. Widespread transcriptional scanning in testes modulates gene evolution rates. bioRxiv:doi.org/10.1101/282129.

Zhang J. 2003. Evolution of the human ASPM gene, a major determinant of brain size. Genetics 165:2063–2070.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9:R137.

Zhen Y, Andolfatto P. 2012. Methods to detect selection on noncoding DNA. Methods Mol. Biol. 856:141–159.

Zhu J, Sanborn JZ, Diekhans M, Lowe CB, Pringle TH, Haussler D. 2007. Comparative Genomics Search for Losses of Long-Established Genes on the Human Lineage. PLoS Comput. Biol. 3:e247.

6 An hourglass pattern of inter-embryo gene expression variability and of histone regulation in fly embryogenesis

Jialin Liu, Michael Frochaux, Vincent Gardeux, Bart Deplancke, Marc Robinson-Rechav

This article was submitted to Nature and available at BioRxiv : doi.org/10.1101/700997.

6.1 Abstract

Gene expression is in part a stochastic process [1–3], which is at odds with the need for precise expression levels in biological processes such as embryonic development. To study the tension between precision and stochasticity of gene expression during development, we quantified expression variability of single embryo transcriptomes throughout fly *Drosophila melanogaster* embryogenesis, from 150 samples over eight developmental stages. We found that expression variability follows an hourglass pattern, with lower variability at extended germband, the phylotypic stage. We explain this pattern by stronger histone modification mediated transcriptional noise control at this stage. In addition, we propose that histone modifications contribute to mutational robustness in regulatory elements, and thus to conserved expression levels. These results provide insight into the role of robustness in the phenotypic and genetic patterns of evolutionary conservation in animal development.

6.2 Introduction

Phenotypes can vary even among isogenic individuals in homogenous environments, suggesting that stochastic effects contribute to phenotypic diversity [4,5]. Gene expression variability among genetically identical individuals under uniform conditions, hereafter "variability", is one of the most important stochastic processes in the mapping of genotype to phenotype. It is caused by a combination of molecular noise (stochastic biochemical effects, e.g., transcriptional burst process based transcriptional noise) and other effects (variation in cells and their environment, e.g., distribution of molecules at cell division) [2,3,6]. Precise regulation of gene expression is notably important during development [7], where it controls the exact timing and level of gene expression to make a functional embryo. However, this process inevitably has to deal with stochasticity [8].

This tension between precision and stochasticity in development raises questions, such as how expression variability is controlled or buffered during development. And whether some stages are more robust to gene expression stochasticity. This is especially interesting in relation to the "phylotypic" stage, in mid-embryonic development, and the corresponding "hourglass" model of Evo-Devo [9,10]. The phylotypic stage is characterized by conserved morphology inside a phylum, which extends to the molecular level, notably lower expression divergence between species or strains [11–15]. To answer these questions, we investigated expression variability across fly embryonic development.

6.3 Results

We generated 288 single embryo 3' end transcriptomes using BRB-seq [16], at eight developmental stages covering the whole fly embryogenesis, with 3h intervals (Figure 1A). After quality control, 239 samples were kept (Figure S1, S2). On average, we obtained over 5 million uniquely mapped reads of protein coding genes per embryo. Based on multidimensional scaling analysis (MDS), 150 embryos follow the developmental trajectory, while there is a small cluster of 89 embryos collected at different time points mixed together (Figure 1B). The samples in this cluster appear to be unfertilized eggs (Methods and Figure S3). All further analysis was performed only on the 150 fertilized embryos.

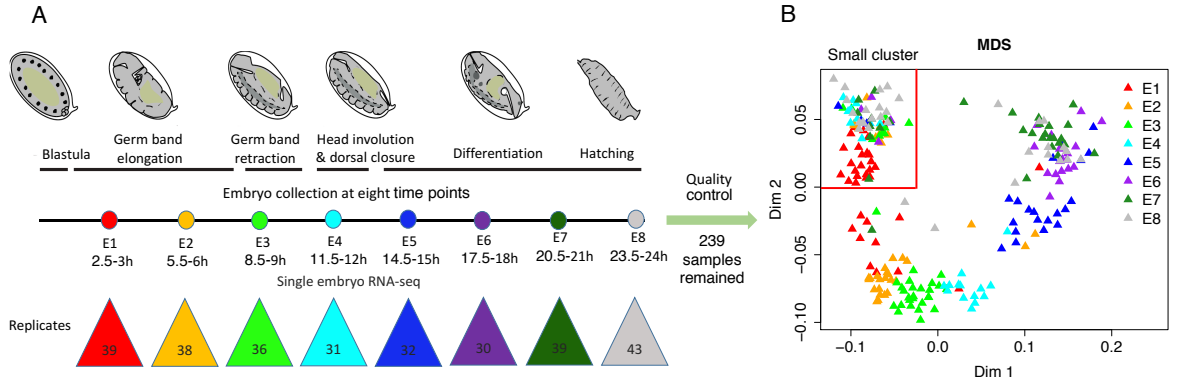

Figure 1: Studying expression variability throughout embryogenesis

A. Method outline. We performed single embryo BRB-seq[16] at eight developmental stages, indicated by different colored dots. The number of samples collected at each stage is indicated in the colored triangles. Embryo images adapted from [17] and used with permission from Springer Nature (License Number: 4547630238607) and from the authors.

B. Multidimensional scaling analysis (MDS) of 239 high quality samples. Different colors indicate different stages. The samples can be split into two groups: a small cluster in the top-left delimited by two red lines; and the remaining samples, which are organized

according to embryonic stage. Only the 150 samples which follow embryonic stages were used for further analysis.

We measured expression variability as Adjusted SD, standard deviation (SD) of expression between replicates corrected for expression level (Methods and Figures S4-6). This expression variability follows an hourglass pattern overall, with a global minimum at E3 (Figure 2A), corresponding to the phylotypic stage of fly [14]. There is also a local minimum at E6. This is consistent with the pattern of transcriptome divergence between fly and mosquito *Anopheles gambiae*, with the global minimum at E3, and a local minimum at E6 [18]. Our observations are robust to the use of different variability metrics (Figure S7), and to sampling (bootstrap analysis, Figure S8). Bootstrap results also suggest that the minimum of variability extends over E3 to E4. The embryo transcriptome is dominated by zygotic transcripts 2.5h after egg laying[19], so the high variability in E1 and E2 is not directly caused by maternal transcripts. We didn't find any significant functional enrichment for genes which follow the hourglass variability pattern. Overall, expression variability is not equally distributed throughout embryogenesis, and gene expression at the phylotypic stage appears more robust to stochastic factors than at other stages.

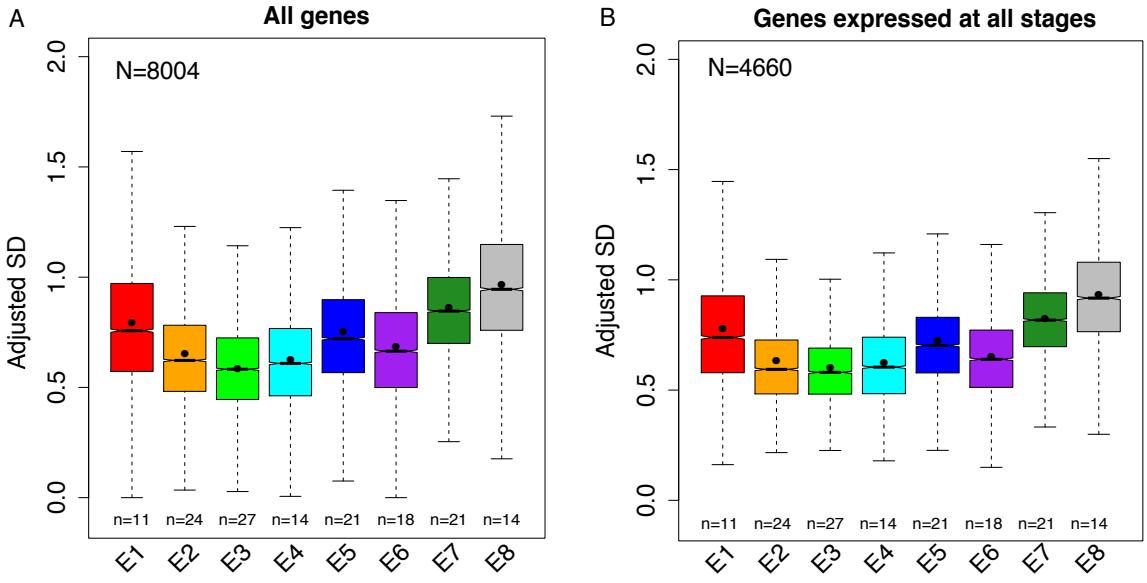

Figure 2: The phylotypic stage (E3) has lower expression variability
The number of individual samples used in each development stage is indicated below each box. The number of genes analyzed is indicated in the top-left corner of each plot. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range (IQR), and the box shows the lower and upper intervals of IQR together with the median. The black dot in each box indicates the mean.

A. *Expression variability pattern of all genes which passed quality control. We performed pairwise Wilcoxon tests between any two stages to test the significance. The multiple test corrected p-values (Benjamini–Hochberg method) are shown in Table S1; they are all* $< 10^{-7}$.
B. *Expression variability pattern of genes expressed at all stages. We performed pairwise Wilcoxon tests between any two stages to test the significance. The multiple test corrected p-values (Benjamini–Hochberg method) are shown in Table S2; they are all* $< 10^{-5}$ *except E2 vs. E4, for which p-values = 0.24.*

The variation in expression variability could either be due to changes in the set of active genes, with genes differing in their intrinsic variability levels, or to genome-wide changes in the regulation of variability. To test this, we first reproduced our results restricted to the subset of genes which are expressed at all stages. Under the first explanation, we would expect to lose the hourglass variability pattern, but the pattern is maintained (Figure 2B). We performed additional tests: restricting to genes with constant expression level over development (Figure S9A); restricting to transcription factors (Figure S9B); and contrasting genes with dispersed or precise promoters (Figure S10), following Schor et al [20]. Dispersed promoters seem to be more robust to mutations, which might also translate into robustness to noise. Despite a loss of power with fewer genes, there remains an hourglass pattern of expression variability in all cases. Interestingly, the precise promoter genes have higher variability than the dispersed promoter genes except at E3, thus a strongest hourglass pattern. Overall, these results suggest that the lower variability at E3 is due to genome-wide regulation mechanisms more than to changes in the gene set.

Histone modifications can regulate transcriptional noise [21–25], notably through the modulation of transcriptional burst frequency [22–24]. For example, high levels of histone modifications can increase chromatin accessibility, leading to an increase in transcriptional burst frequency, which leads to minimizing noise. To check this role of histone modifications, we analyzed four available euchromatin histone modifications at six developmental stages [26]. For each gene, we calculated the mean modification signal (background-subtracted tag density) separately for proximal promoters and for gene bodies [23]. Higher modification signal genes tend to have lower variability for all histone modifications (Figure 3A). This supports a role in minimizing transcriptional noise, and is consistent with previous studies in yeast and mammals [22,23].

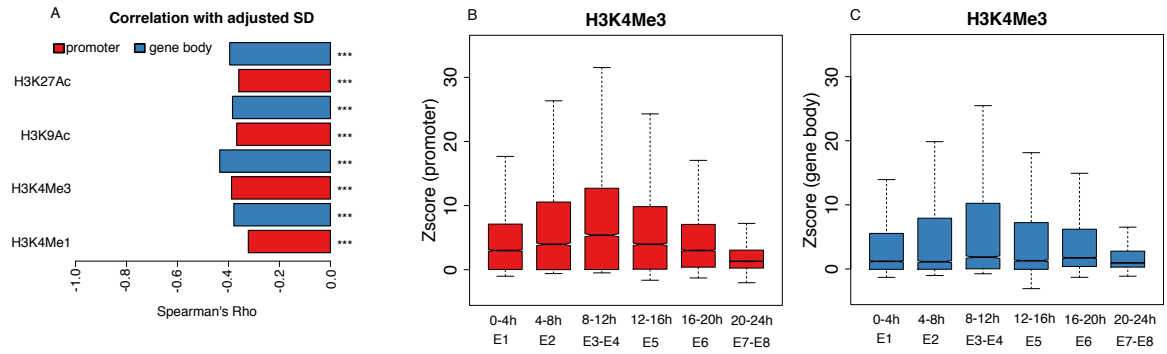

Figure 3: Histone modification signal and expression variability

Red and blue represent histone modification signals calculated on the proximal promoter (4kb around the transcription start site – TSS) and the gene body, respectively.

*A. Spearman's correlation coefficient between histone modification signal (background-subtracted tag density) and expression variability. Here, for each gene, both its variability and its histone modification signal are the mean value across stages. ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$, NS, $P \geq 0.05$.*

B. Proximal promoter H3K4Me3 signal (Z score relative to intergenic signal) in different stages. Corresponding stages of our expression variability data are indicated below. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range (IQR), and the box shows the lower and upper intervals of IQR together with the median. We performed pairwise Wilcoxon tests between any two stages to test the significance. The multiple test corrected p-values (Benjamini–Hochberg method) are shown in Table S3; they are all $<10^{-7}$ except 4-8h vs. 12-16h, for which p-value = 0.68.

C. Gene body H3K4Me3 signal (Z score relative to intergenic signal) in different stages. Corresponding stages of our single embryo BRB-seq data are indicated below. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range (IQR), and the box shows the lower and upper intervals of IQR together with the median. We performed pairwise Wilcoxon tests between any two stages to test the significance. The multiple test corrected p-values (Benjamini–Hochberg method) are shown in Table S4; they are all $< 10^{-5}$ except 0-4h vs. 20-24h, for which p-value = 0.26; and 8-12h vs. 16-20h, for which p-value = 0.26.

The gene-level relation between histone modifications and expression variability raises the possibility that the pattern of expression variability across development could be driven by changes in histone modification signal. To compare histone modification signal between stages, we normalized gene and promoter signal by that on intergenic regions (Methods), which are not expected to change histone modification signal between stages. All histone marks present an hourglass-like pattern, with the highest signal at 8-12h (except for H3K4Me1 on gene body, where it is a local but not global maximum), corresponding to E3 and E4, i.e. the lowest expression variability, for both promoters and gene body (Figures 3B-C, S11). Moreover, for all histone marks on gene body, as well as H3K4Me1 on promoters, there is another local maximum at 16-20h, corresponding to E6. Generally, these results support changes in histone

modification signal over development, with a correspondence between stronger histone modification signal and lower expression variability.

Several studies have suggested that mechanisms which confer robustness to stochastic variation can also buffer the effects of genetic variation [7,27,28]. If histone modifications can buffer the effect of genetic variation on gene expression, we should observe that genes with higher histone modification signal are less sensitive to mutations in their regulatory regions, and are thus less conserved. Indeed, genes with higher histone modification signal tend to have less conserved core promoter sequences [29] (49 bp upstream TSS and 10 bp downstream from the TSS) between species (phastCons score; Figure 4A). They are also less conserved within a population (promoter nucleotide diversity π; Figure S12). The phastCons pattern remains using 200 bp or 400 bp regions, but disappears using 1 kb regions (Figure S13), indicating a relatively narrow region around the TSS under this balance of selection and robustness.

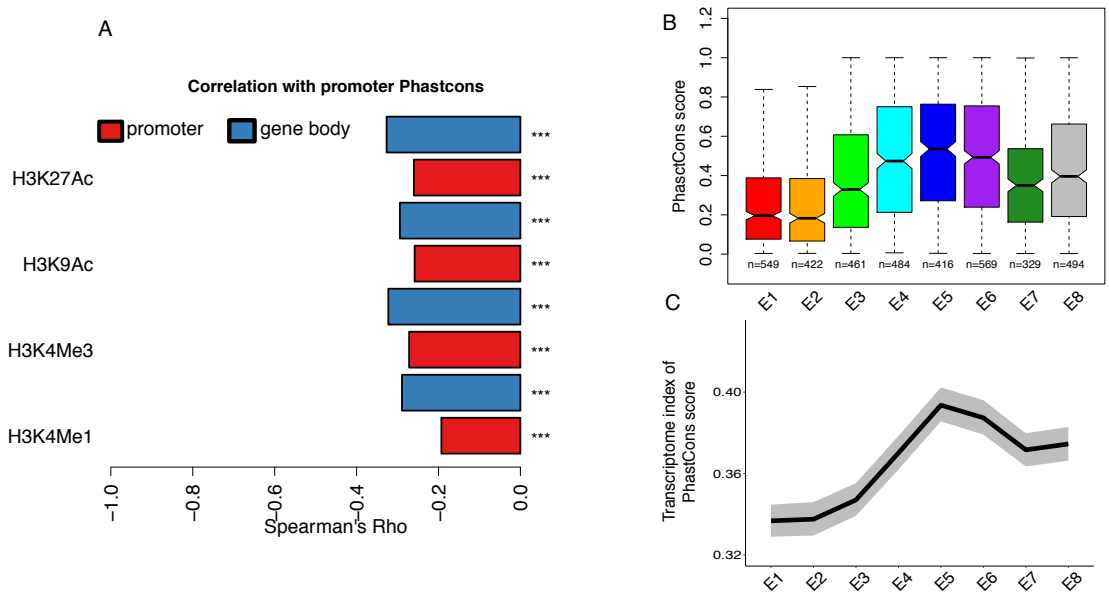

Figure 4: Histone modification signal and promoter sequence conservation
The promoter sequence conservation is the mean of the phastCons score over experimentally identified core promoter regions (49 bp upstream TSS to 10 bp downstream of the TSS) [29].

*A. Spearman's correlation coefficient between histone modification signal (background-subtracted tag density) and promoter sequence conservation. Red and blue represent histone modification signals calculated from the proximal promoter (4 kb around the TSS) and gene body respectively. Here, for each gene, the histone modification signal is the mean value across stages. ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$, NS, $P \geq 0.05$.*

B. Variation of promoter sequence conservation for stage specific genes. The number of genes in each development stage is indicated below each box. We performed pairwise

Wilcoxon test between any two stages to test the significance. The multiple test corrected p-values (Benjamini–Hochberg method) are shown in Table S5.

C. *Transcriptome index of promoter phastCons score across development. The grey area indicates 95% confidence interval estimated from bootstrap analysis.*

Since histone modifications appear to buffer genetic variation in gene expression, and since the E3 stage has stronger modification signals, the lower expression divergence in E3 between species [14] might be caused either by stronger purifying selection on mutations in regulatory regions, or by histone modifications buffering the consequences of mutations in these regions. In the first case, we expect genes specifically expressed at E3 to have higher sequence conservation on promoters. In the second case, we expect the opposite pattern, since mutations that are buffered would behave nearly neutrally. To test this, we identified genes specifically expressed in each stage and compared their promoter sequence conservation. We found that genes specific of E3 have a relatively weak promoter sequence conservation (Figure 4B), supporting a stronger buffering mechanism rather than stronger purifying selection on sequences. The transcriptome indexes of conservation and of diversity (mean promoter sequence conservation and mean π, respectively, weighted by expression) extend this observation to the full transcriptome (Figure 4C; Figure S14). These results support a role of buffering effects on regulatory mutations in the hourglass pattern of expression divergence in fly embryogenesis.

6.4 Discussion

We have found an uneven distribution of variability, and thus of robustness of the process of gene expression, across development, which mirrors the hourglass Evo-Devo model [9,10]. Stage E3 is the most robust to stochastic variation on gene expression, with lower expression variability, and is the phylotypic stage of fly, with conservation between species [14]. Our results support a role of increased gene and promoter histone modification signal in the lower variability in the phylotypic stage.

Although mutational robustness can evolve under natural selection theoretically [30], the conditions are too restrictive to be relevant in practice. Thus, we propose that the mutational buffering effect of histone modifications is a by-product of selection for minimizing transcriptional noise. The mutational buffering analyses have two potential caveats. First, our results were based solely on promoter analysis, and it remains to be seen how much these observations extend to other regulatory elements. Second, we only considered nucleotide

substitution changes, but not insertion, deletion and turnover. For these large effect mutations, purifying selection might be more efficient than mutational robustness to keep conserved expression level. Thus it is possible that both purifying selection and mutational robustness together shape the lower expression divergence in the fly phylotypic stage. Overall, our results support an important role for the control of expression variability by histone modifications in embryonic gene expression and in the evolution of development.

6.5 Materials and Methods

6.5.1 Availability of code

Data files and analysis scripts are available on GitHub:
https://github.com/ljljolinq1010/expression-noise-across-fly-embryogenesis.

6.5.2 Availability of data

Expression datasets have been deposited to the Gene Expression Omnibus with accession number GSE128370.

6.5.3 Embryo collection and RNA extraction

Fly lines (w^{1118}) were obtained from the Bloomington stock center and reared at room temperature on a standard fly medium with 12 hours light dark cycle. The fly medium we used is composed of: 6.2 g Agar powder (ACROS N. 400400050), 58.8 g Farigel wheat (Westhove N. FMZH1), 58.8 g yeast (Springaline BA10), 100 mL grape juice; 4.9 mL Propionic acid (Sigma N. P1386), 26.5 mL of Methyl 4-hydroxybenzoate (VWR N. ALFAA14289.0) solution (400 g/L) in 95% ethanol, 1 L Water. 100 to 150 flies were transferred to cages, which were sealed to a grape agar plate (1:1 mixture of 6% agar and grape juice). We used 4 separate cages to collect the embryos. The adults were kept overnight on this plate before being transferred to a new plate supplemented with yeast paste. Synchronization of eggs on this plate lasted for 2 hours before being swapped with a new plate supplemented with yeast paste. We let the adults lay eggs for 30 min before removing the plate and letting the eggs incubate for the desired time. Eggs were harvested using the following protocol. First a 1:1 bleach (Reactolab 99412) 1x PBS mix was poured on the plate and incubated for 2 min. During this incubation, we used a brush to lightly scrape the surface to mobilize the embryos. We then poured the PBS-bleach mixture through a sieve, washed the plate with 1x PBS, and poured the wash on the same sieve. We washed the sieve several time with 1x PBS until the smell of bleach disappeared. Single

embryos were then manually transferred to Eppendorf containing 50 μL beads and 350 μL Trizol (lifetechnologies AM9738). The tubes were homogenized in a Precellys 24 Tissue Homogenizer at 6000 rpm for 30 seconds. Samples were then transferred to liquid nitrogen for flash freezing and stored at –80°C. For RNA extraction, tubes were thawed on ice, supplemented with 350 μL of 100% Ethanol before homogenizing again with the same parameters. We then used the Direct-zol™ RNA Miniprep R2056 Kit, with the following modifications: we did not perform DNAse I treatment, we added another 2 min centrifugation into an empty column after the RNA Wash step, finally elution was performed by adding 8 μL of RNAse-free water to the column, incubation at room temperature for 2 min and then centrifugation for 2 min. RNA was transferred to a low-binding 96 well plate and stored at -80°C.

6.5.4 Bulk RNA Barcoding and sequencing (BRB-seq)

The BRB-seq is a technique for multiplexed RNA-seq [16] which is able to provide high-quality 3' transcriptomic data at a low cost (e.g. 10-fold lower than Illumina Truseq Stranded mRNA-seq). The data (fastq files) generated from BRB-seq are multiplexed and asymmetrical paired reads. Read R1 contains a 6 bp sample barcode, while read R2 contains the fragment sequence to align to the reference genome.

1. Library preparation

RNA quantity was assessed using picogreen (Invitrogen P11496). Samples were then grouped according to their concentration in 96-well plates and diluted to a final concentration determined by the lowest sample concentration on the plate. RNA was then used for gene expression profiling using BRB-seq. In short, the BRB-seq protocol starts with oligo-dT barcoding, without TSO for the first-strand synthesis (reverse transcription), performed on each sample separately. Then all samples are pooled together, after which the second-strand is synthesized using DNA PolII Nick translation. The sequencing library is then prepared using cDNA augmented by an in-house produced Tn5 transposase preloaded with the same adapters (Tn5-B/B), and further enriched by limited-cycle PCR with Illumina compatible adapters. Libraries are then size-selected (200 - 1000 bp), profiled using High Sensitivity NGS Fragment Analysis Kit (Advanced Analytical, #DNF-474), and measured using Qubit dsDNA HS Assay Kit (Invitrogen, #Q32851). In total, we generated four libraries. For details of library information, please check Table S20.

2. Sequencing

Libraries were mixed in equimolar quantities and were then sequenced on an Illumina Hi-Seq 2500 with pair-end protocol (read R2 with 101 bp) at the Lausanne Genomic Technologies Facility.

6.5.5 RNA-seq analysis

1. Generating expression matrix

The fastq files were first demultiplexed by using the "Demultiplex" tool from BRB-seqTools suite (available at https://github.com/DeplanckeLab/BRB-seqTools). Then, we trimmed the polyA sequences of the demultiplexed files by using the "Trim" tool. Next, the STAR aligner [31] was used to map the trimmed reads to the reference genome of fly *Drosophila melanogaster* (BDGP6, Ensembl release 91 [32]). Finally, the read count of each gene was obtained with HTSeq [33].

2. Filtering samples and genes

Low-quality samples need to be filtered out, since they might bias results of downstream analyses. In order to assess sample quality, we calculated the number of uniquely mapped reads and of expressed genes for each sample [34]. We removed samples with <0.3 million uniquely mapped reads or with <4500 expressed genes (Figure S1). We confirmed that these filtered samples are indeed outliers in a multidimensional scaling analysis (MDS) (Figure S15). Since lowly expressed genes have larger technical error, to minimize the technical noise, we need to remove lowly expressed genes as well. We first calculated counts per million (cpm) with the edgeR package [35] for each gene. Then we removed genes with mean cpm across samples ≤1, as suggested by Lun et al.[34]. Finally, for the remaining genes, we re-transformed their cpm values to the original count values for the downstream normalization analysis. After filtering, we obtained an expression count matrix with 239 samples (Figure S2) and 8004 protein coding genes.

3. Normalization and batch effect correction

Because BRB-Seq retains only the 3′ end of the transcript, we performed sample normalization by using quantile normalization with log transformation in the voom package [36], but without transcript length normalization. To remove potential batch effects across the four libraries, we applied the ComBat function in the sva package [37] to the normalized and log2 transformed expression data. For genes with expression values less than 0 after Combat, or with original expression values equal to 0, we change its values to 0 after Combat correction as suggested by Kolodziejczyk et al [38].

6.5.6 Multidimensional scaling analysis (MDS)

A number of factors could be invoked to explain the two groups observed in our multidimensional scaling analysis (MDS) (Figure 1B), but they should also explain that only one group is structured according to developmental time. The obvious hypothesis that they correspond to male and female embryos does not explain that structure, and is also not supported by examining X/autosome gene expression ratios (Figure S16). An alternative hypothesis is that samples in the small cluster are unfertilized eggs. If an egg is not fertilized, after completion of meiosis, development will be arrested [39], but they are visually indistinguishable. This hypothesis is confirmed by at least two lines of evidence, in addition to the lack of developmental time structure. First, for expression correlation, all samples in the small cluster are highly correlated with unfertilized egg, while the correlations in the other samples gradually decrease with development (Figure S3A). Second, all the samples from the small cluster are enriched with meiosis related genes (Figure S3B). Thus we excluded the small cluster for downstream analyses, i.e. we used 150 embryos with an average of 18 individuals per developmental stage.

6.5.7 Metrics of expression variability

Expression variability is generally measured by the coefficient of variation (CV) [40]. However, a gene's CV is strongly dependent on its RNA abundance (Figure S4). While this is an inherent property of time-interval counting processes (such as a Poisson process), it makes the comparison of variability between different conditions difficult [38,41]. Distance to median (DM, the distance between the squared CV of a gene and its running median) has been proposed as a variability metric that is independent of expression level [38,41,42]. However, the DM is still strongly negatively correlated with the mean expression level in our data (Figure S5). To avoid this dependency, we developed another variability measure, the adjusted standard deviation (adjusted SD), by calculating the ratio between observed SD and expected SD. Following the same approach as Barroso et al. [43], we performed polynomial regressions to predict the expected SD from mean expression. Since the adjusted SD metric works much better than DM in terms of accounting for the confounding effects of mean expression (Figure S6), we adopted it as a measure of expression variability in our study. As observed in yeast [42,44], we found that essential genes and hubs (proteins in the center of protein-protein interaction network) have lower expression variability than other genes (Figure S17), indicating selection to reduce it.

This observation provides a control that we are indeed measuring biologically relevant expression variability.

Detailed calculation of expression variability:

1. Adjusted SD.

For each gene, we computed standard deviation (SD) in each stage and over all stages. Then we fitted a polynomial model to predict the global (across all stages) SD from the global mean expression. We increased the degrees of the model until there was no more significant improvement (tested with ANOVA, $p<0.05$ as a significant improvement). Then, based on this best fitting model, for each gene, we computed its predicted global SD based on its global mean expression. Finally, the adjusted SD of a gene in one stage is this gene's SD in its corresponding stage divided by its predicted global SD. This method is derived from Barroso et al. [43], but computing adjusted SD rather than adjusted variance.

2. Distance to median: the distance between the squared coefficient of variation (CV) of a gene and its running median.

For each gene, we computed its squared CV in each stage and over all stages. Then, we ordered genes based on their global (across all stages) mean expression. Next, we defined series of sliding windows of 50 genes with 25 genes overlap, starting from lowest global mean expression. Finally, the distance to median of a gene in one stage is the stage specific log10 squared CV minus the median of global log10 squared CV in this gene's corresponding window. R code was modified from the DM function of the scran package [34].

6.5.8 Bootstrap analysis

For each stage, we randomly sampled the same number of samples. Then, we computed the adjusted SD based on these random samples. We repeated the first two steps 500 times. Each time, we only kept the median of the adjusted SD for each stage. Thus in each stage we obtained 500 medians. Finally, we performed a Wilcoxon test to test the significance of the difference between the bootstrapped values of different stages.

6.5.9 ChIP-Seq data analysis

1. Histone modification signal datasets

The signal data files of four euchromatin histone modification marks (H3K4me1, H3K4me3, H3K9ac, and H3K27ac) at six developmental stages (0-4h, 4-8h, 8-12h, 12-16h. 16-20h, 20-24h) were downloaded from modENCODE [26] (NCBI GEO: GSE16013) (March, 2018). The

signal is smoothed, background-subtracted tag density. The signal was precomputed along the genome in 35-bp windows.

2. Histone modification signal for promoter and gene body

For each gene, as suggested by Nicolas et al. [23], we separately calculated the mean signal of its proximal promoter (2 kb upstream to 2 kb downstream for transcription start site (TSS)) and of its gene body (TSS to transcription end site (TES)) by using the bedtools "map" command [45]. The TSS and TES information was retrieved from Ensembl release 91 [32]. For a gene with several TSS and TES, we use its mean coordinates.

3. Histone modification signal Z score transformation

For each modification mark in each stage, the signal value was transformed into a Z score by subtracting the mean signal across intergenic regions and dividing by the standard deviation signal of the intergenic regions. The intergenic region were defined by removing all proximal promoter regions and gene body regions with the bedtools "subtract" command [45]. Our assumption is that on average such intergenic regions are not the target of active histone modification signal, and thus allow to normalize between libraries. Then, for each gene, we re-calculated the mean signal (Z score) of its proximal promoter (2 kb upstream to 2 kb downstream for transcription start site (TSS)) and of its gene body (TSS to transcription end site (TES)) by using the bedtools "map" command [45].

6.5.10 Identification of stage specifically expressed genes

Following the same approach as previously [46], we first defined 8 stage specific expressed artificial expression profile (Figure S18A). Then, for each gene, we performed Pearson's correlation between its real expression and this artificial expression. Finally, for each artificial expression profile, we kept genes with top 10% correlation coefficient as the corresponding stage specifically expressed genes (Figure S18B).

6.5.11 Identification of hourglass expression variability genes

Similar to the stage specifically expressed gene identification approach, we correlated each gene's variability profile with the median across all genes. Then, we kept genes with the top 10% correlation coefficient as the ones following the global hourglass variability profile.

6.5.12 Identification of genes expressed at all stages

For each gene, we calculated the average expression across replicates in each stage. Then, we defined the average expression > 1 as expressed.

6.5.13 Identification of genes with constant expression across all stages

For each gene, we first preformed one-way analysis of variance (ANOVA) to compare the means of expression in different stage. Then, we calculated the *q-values* for multiple test correction. Finally, the constantly expressed genes were defined as genes with *q-values* > 0.05.

6.5.14 Gene ontology (GO) enrichment analysis

We performed GO enrichment analysis for hourglass expression variability genes by using the topGO [47] R package with the "elim" algorithm.

6.5.15 Single Nucleotide Polymorphism (SNP) data

The SNP data for 205 *D. melanogaster* inbred lines were downloaded from the Drosophila Genetic Reference Panel (DGRP [48]) (December, 2018).

6.5.16 Nucleotide diversity (π) calculation

We calculated nucleotide diversity of promoters with vcftools [49].

6.5.17 Transcriptome index analysis

A "transcriptome index" [50,51] is a weighted mean of a feature over all genes, where the weights are the expression levels of the genes at each condition (e.g., developmental stage). The transcriptome index of phastCons was calculated as:

$$TPI_s = \frac{\sum_{i=1}^{n} phastCons_i * e_{is}}{\sum_{i}^{n} e_{is}},$$

where s is the developmental stage, $phastCons_i$ is the promoter sequence conservation score of gene i, n is the total number of genes, and e_{is} is the expression level (log transformed) of gene i in developmental stage s. For the transcriptome index of nucleotide diversity (π) the same formula is used, replacing $phastCons_i$ by π_i.

6.5.18 Meiosis related genes and transcription factors

The Meiosis related genes and transcription factors were downloaded from AmiGO [52] (May, 2018).

6.5.19 Individual unfertilized eggs RNA-seq data

The normalized and log transformed expression matrix of individual unfertilized eggs was downloaded from NCBI GEO: GSE68062 [53] (May, 2018).

6.5.20 Dispersed and precise promoters

The annotation of genes with dispersed or precise promoters was downloaded from Schor et al [20] (June, 2019). Dispersed promoters are often associated with ubiquitously expressed genes, have more dispersed patterns of transcriptional initiation, and do not contain a TATA box. On the contrary, precise promoters are typically associated with restricted tissue-specific expression and with a TATA box, and have a single predominant TSS.

6.5.21 Essential gene annotation and protein connectivity datasets

The gene essentiality and protein connectivity datasets were downloaded from OGEE v2 [54] (March, 2018).

6.5.22 PhastCons score

The pre-computed sequence conservation score phastCons [55] of fly genome was downloaded from http://hgdownload.soe.ucsc.edu/goldenPath/dm3/phastCons15way/ (February, 2018). Higher value means higher conservation.

6.5.23 Experimentally validated core promoters

Experimentally validated transcription start sites (TSSs) were downloaded from the Eukaryotic Promoter Database (EPD) [29] (May, 2018). For a gene with several TSSs, we selected the most representative one (the TSS that has been validated by the largest number of samples). The core promoter region was defined as 49 bp upstream TSS to 10 bp downstream of the TSS [29]. We used EPD defined TSSs here because they are more accurate for defining core promoters, whose function is expected to be related to sequence conservation. Whereas for histone modification signal for promoter and gene body we used

Ensembl defined TSSs to be consistent with the source of TES information, and precision was less important in defining broader proximal promoters.

6.6 Acknowledgements

We thank Virginie Braman for help with embryo collection and library preparation. We thank Daniel Alpern for help with the BRB-seq technology. We thank Richard Benton, David Garfield and Laurent Keller for critically reading and commenting on the manuscript. We thank Andrea Komljenovic and other 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 and MRR are supported by Swiss National Science Foundation grants 31003A_153341 / 1 and 31003A_173048. MRR and BD are supported by SystemsX.ch grant AgingX.

6.7 Author contributions

JL designed the work with input from MRR, MF and BD. MF led all experiments. JL performed all computational analyses. VG and BD contributed expertise in the BRB-seq experiments. JL and MRR interpreted the results with input from all the other authors. JL wrote the first draft of the paper. JL and MRR finalized the paper with input from all the other authors.

6.8 Supporting Information

Supporting materials can be downloaded from:
https://www.biorxiv.org/content/10.1101/700997v1.supplementary-material

Table S1: variability (adjusted SD) comparison between any two stages (all genes).
Table S2: variability (adjusted SD) comparison between any two stages (genes expressed at all stages).
Table S3: proximal promoter H3K4Me3 signal (Z score) comparison between any two stages.
Table S4: gene body H3K4Me3 signal (Z score) comparison between any two stages.
Table S5: promoter sequence conservation comparison between any two stages.
Table S6: variability (adjusted SD) comparison between any two stages, the E1 stage also contains samples from the small clusters.
Table S7: variability (coefficient of variation) comparison between any two stages.

Table S8: variability (distance to median) comparison between any two stages.

Table S9: bootstrapped median variability comparison between any two stages.

Table S10: variability (adjusted SD) comparison between any two stages (Genes with constant expression level over development).

Table S11: variability (adjusted SD) comparison between any two stages (transcription factors).

Table S12: variability (adjusted SD) comparison between any two stages (dispersed promoter genes).

Table S13: variability (adjusted SD) comparison between any two stages (precise promoter genes).

Table S14: proximal promoter H3K4Me1 signal (Z score) comparison between any two stages.

Table S15: gene body H3K4Me1 signal (Z score) comparison between any two stages.

Table S16: proximal promoter H3K27Ac signal (Z score) comparison between any two stages.

Table S17: gene body H3K27Ac signal (Z score) comparison between any two stages.

Table S18: proximal promoter H3K9Ac signal (Z score) comparison between any two stages.

Table S19: gene body H3K9Ac signal (Z score) comparison between any two stages.

Table S20: RNA-seq library information.

Figure S1: Relationship between uniquely mapped reads and expressed genes

Each dot represents one sample. The black dots indicate low quality samples with <4500 expressed genes or with <0.3 million uniquely mapped reads. The 239 orange colored samples were retained for downstream analysis ("high quality samples").

Figure S2: Proportion of retained samples in each development stage

The number of retained samples and of total samples in each stage is indicated in the bottom of each bar.

Figure S3: Evidence that the samples from the small cluster are unfertilized eggs

A. Boxplot of Spearman's correlation coefficients (rho) of expression between individual unfertilized eggs and each sample from the small cluster or from the large cluster, showing that the small cluster has an expression profile of unfertilized eggs. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile

range (IQR), and the box shows the lower and upper intervals of IQR together with the median.

B. Expression heat map of meiosis related genes across all samples, showing that their expression decreases over development for the large cluster, but is high in all samples of the small cluster, consistent with unfertilized eggs.

For testing of an alternative explanation of the two clusters as being males and females, see Figure S16.

Figure S4: Relationship between average expression and coefficient of variation at each stage

Pearson's correlation between average expression and coefficient of variation in each development stage is indicated in the top left of each subfigure.

Figure S5: Relationship between average expression and distance to median at each stage

Pearson's correlation between average expression and distance to median in each development stage is indicated in the top left of each subfigure.

Figure S6: Relationship between average expression and adjusted SD at each stage

Pearson's correlation between average expression and adjusted SD in each development stage is indicated in the top left of each subfigure.

Figure S7: Variation of expression variability across development using alternate measures of variability

A. Variability measured by adjusted SD; unlike in Figure 2, the variability in E1 was calculated using all samples from both small and large clusters.

B. Variability measured by coefficient of variation (CV).

C. Variability measured by distance to median (DM).

The legend is the same as for Figure 2. We performed pairwise Wilcoxon test between any two stages to test the significance. The multiple test corrected *p*-values (Benjamini–Hochberg method) are shown in Tables S6, S7 and S8.

Figure S8: Bootstrap analysis of the variability calculation

We performed pairwise Wilcoxon test between any two stages to test the significance. The multiple test corrected *p*-values (Benjamini–Hochberg method) are shown in Table S9.

Figure S9: Variation of expression variability across development for different categories of genes

A. Genes with constant expression level over development.

B. Transcription factor.

The legend is the same as for Figure 2. We performed pairwise Wilcoxon test between any two stages to test the significance. The multiple test corrected *p*-values (Benjamini–Hochberg method) are shown in Tables S10 and S11.

Figure S10: Variation of expression variability across development for dispersed promoter genes and for precise promoter genes

For each stage, the first and the second box represents dispersed promoter genes and precise promoter genes respectively. The legend is the same as for Figure 2. We performed pairwise Wilcoxon test between any two stages to test the significance separately for dispersed promoter genes and for precise promoter genes. The multiple test corrected *p*-values (Benjamini–Hochberg method) are shown in Tables S12 and S13.

Figure S11: Histone modification signal across development

The legend is the same as for Figure 3B and 3C. The median signal value in each development stage is indicated above each box. We performed pairwise Wilcoxon test between any two stages to test the significance. The multiple test corrected *p*-values (Benjamini–Hochberg method) for H3K4Me1, H3K27Ac and H3K9Ac are shown in Tables S14-S19.

Figure S12: Spearman's correlation coefficient between histone modification signal and promoter nucleotide diversity (π).

The legend is the same as for Figure 4A.

Figure S13: Spearman's correlation coefficient between histone modification signal and promoter sequence conservation for different definitions of promoter width

The figure legend is the same as in Figure 4A.

A. Promoter defined as 200 bp around TSS

B. Promoter defined as 400 bp around TSS

C. Promoter defined as 1000 bp around TSS

Figure S14: transcriptome index of π across development.

The legend is the same as for Figure 4C.

Figure S15: Multidimensional scaling analysis for all samples

Different colors indicate different stages. The solid triangles represent high quality samples according to Figure S1; the hollow triangles represent low quality samples which were discarded.

Figure S16: Mapping of X/autosome gene expression ratios to the multidimensional scaling analysis plot

We calculated the ratio of mean expression between genes from the X chromosome and from the autosomes for each sample. Red represents high ratio, blue represents low ratio. For Drosophila, dosage compensation is achieved by increasing expression of X chromosome genes in males. Since the dosage compensation is still incomplete during development, females should have a higher ratio of mean expression between genes from the X chromosome and from the autosomes. Here, we found both high ratio samples and low ratio samples are well mixed in both the cluster and large clusters. Thus, we reject the hypothesis that the two different clusters are due to sex.

Figure S17: Relationship between expression variability and protein importance

We used the average variability across all development stages.

A. Spearman's correlation between expression variability and proportion of essential genes. We split genes into 10 equally sized bins based on expression variability. The Spearman's correlation is calculated by using the median variability and proportion of essential genes in the ten bins. The Spearman's correlation coefficient and *p*-value are indicated in the top-right.
B. Spearman's correlation between connectivity in a protein-protein interaction network and expression variability. The coefficient and *p*-value are indicated in the top-right.

Figure S18: Detection of stage specific genes

A. The artificial expression profile.
B. The expression of identified stage specific genes. The bold black line represents the median expression, the two gray lines represent 25th and 75th quantiles of expression, respectively.

6.9 References

1. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science*. **297**, 1183–1186 (2002).
2. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet*. **6**, 451–464 (2005).
3. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–26 (2008).
4. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–8 (2010).
5. Li, X. *et al*. Systems Properties and Spatiotemporal Regulation of Cell Position Variability during Embryogenesis. *Cell Rep*. **26**, 313-321.e7 (2019).

6. Eling, N., Morgan, M. D. & Marioni, J. C. Challenges in measuring and understanding biological noise. doi:10.1038/s41576-019-0130-6
7. Waddington, C. H. Canalization of development and genetic assimilation of acquired characters. *Nature* **150**, 563–565 (1942).
8. Oates, A. C. What's all the noise about developmental stochasticity? *Development* **138**, 601–7 (2011).
9. Raff, R. A. *The shape of life : genes, development, and the evolution of animal form*. (University of Chicago Press., 1996).
10. Duboule, D. 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 (1994).
11. Hu, H. *et al*. Constrained vertebrate evolution by pleiotropic genes. *Nat. Ecol. Evol.* **1**, 1722–1730 (2017).
12. Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* **2**, 248 (2011).
13. Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. *Dev. Cell* **22**, 1101–8 (2012).
14. Kalinka, A. T. *et al*. Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–4 (2010).
15. Zalts, H. & Yanai, I. Developmental constraints shape the evolution of the nematode mid-developmental transition. *Nat. Ecol. Evol.* **1**, 0113 (2017).
16. Alpern, D. *et al*. BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biol.* **20**, 71 (2019).
17. Levin, M. *et al*. The mid-developmental transition and the evolution of animal body plans. *Nature* **531**, 637–41 (2016).
18. Schep, A. N. & Adryan, B. A Comparative Analysis of Transcription Factor Expression during Metazoan Embryonic Development. *PLoS One* **8**, e66826 (2013).
19. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042 (2009).
20. Schor, I. E. *et al*. Promoter shape varies across populations and affects promoter evolution and expression noise. *Nat. Genet.* **49**, 550–558 (2017).
21. Benayoun, B. A. *et al*. H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* **158**, 673–88 (2014).
22. Wu, S. *et al*. Independent regulation of gene expression level and noise by histone modifications. *PLOS Comput. Biol.* **13**, e1005585 (2017).
23. Nicolas, D., Zoller, B., Suter, D. M. & Naef, F. Modulation of transcriptional burst frequency by histone acetylation. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7153–7158 (2018).
24. Weinberger, L. *et al*. Expression Noise and Acetylation Profiles Distinguish HDAC Functions. *Mol. Cell* **47**, 193–202 (2012).
25. Faure, A. J., Schmiedel, J. M. & Lehner, B. Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. *Cell Syst.* **5**, 471-484.e4 (2017).
26. Nègre, N. *et al*. A cis-regulatory map of the Drosophila genome. *Nature* **471**, 527–531 (2011).
27. Lehner, B. Genes Confer Similar Robustness to Environmental, Stochastic, and Genetic Perturbations in Yeast. *PLoS One* **5**, e9035 (2010).
28. Meiklejohn, C. D. & Hartl, D. L. A single mode of canalization. *Trends Ecol. Evol.* **17**, 468–473 (2002).
29. Dreos, R., Ambrosini, G., Périer, R. C. & Bucher, P. The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res.*

43, D92–D96 (2015).
30. Wagner, G. P., Booth, G. & Bagheri-Chaichian, H. A Population Genetic Theory of Canalization. *Evolution*. **51**, 329 (1997).
31. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
32. Zerbino, D. R. *et al*. Ensembl 2018. *Nucleic Acids Res*. **46**, D754–D761 (2018).
33. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
34. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data. *F1000Research* **5**, 2122 (2016).
35. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–40 (2010).
36. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. **15**, R29 (2014).
37. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
38. Kolodziejczyk, A. A. *et al*. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* **17**, 471–85 (2015).
39. Avilés-Pagán, E. E. & Orr-Weaver, T. L. Activating embryonic development in Drosophila. *Semin. Cell Dev. Biol **84**, 100-110* (2018).
40. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–3 (2005).
41. Tung, P.-Y. *et al*. Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep*. **7**, 39921 (2017).
42. Newman, J. R. S. *et al*. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
43. Barroso, G. V., Puzovic, N. & Dutheil, J. Y. The Evolution of Gene-Specific Transcriptional Noise Is Driven by Selection at the Pathway Level. *Genetics* **208**, 173–189 (2018).
44. Lehner, B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol. Syst. Biol*. **4**, 170 (2008).
45. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–2 (2010).
46. Liu, J. & Robinson-Rechavi, M. Adaptive Evolution of Animal Proteins over Development: Support for the Darwin Selection Opportunity Hypothesis of Evo-Devo. *Mol. Biol. Evol*. **35**, 2862–2872 (2018).
47. Alexa, A., Rahnenfuhrer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).
48. Huang, W. *et al*. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Res*. **24**, 1193–208 (2014).
49. Danecek, P. *et al*. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–8 (2011).
50. Liu, J. & Robinson-Rechavi, M. Developmental Constraints on Genome Evolution in Four Bilaterian Model Species. *Genome Biol. Evol*. **10**, 2266–2277 (2018).
51. Domazet-Loso, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010).
52. Carbon, S. *et al*. AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288–9 (2009).

53. Paris, M., Villalta, J. E., Eisen, M. B. & Lott, S. E. Sex Bias and Maternal Contribution to Gene Expression Divergence in Drosophila Blastoderm Embryos. *PLoS Genet*. **11**, e1005592 (2015).
54. Chen, W.-H., Lu, G., Chen, X., Zhao, X.-M. & Bork, P. OGEE v2: an update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res*. **45**, D940–D944 (2017).
55. Siepel, A. *et al*. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. **15**, 1034–50 (2005).

7 Outlook

The work presented in this thesis has focused on the interplay between animal development and molecular evolution. I found the phylotypic period to be more conserved than other stages in terms of both gene sequence and regulation evolution. In addition, the phylotypic period is also lacking adaptive evolution on both gene sequence and regulation. Finally, the phylotypic period is even robust to both genetic and non-genetic variations on gene expression. These results together suggest that the phylotypic period is an evolutionary lockdown, sensitive to all kinds of variations. I hope that my PhD work will contribute to providing some fundamental insights into evolution and development. Here, I would like to briefly summarize my results and discuss possible extension of my work in the field of Evo-Devo.

In chapter 2, I found that genes expressed at the phylotypic stage underwent stronger purifying selection on sequence evolution, supporting the hourglass model. Further analyses indicate that stronger purifying selection on sequences in middle development is driven by temporal pleiotropy of these genes. However, after removing this temporal pleiotropy effect by comparing the sequence evolution rate for stage specific genes, the hourglass pattern still holds true (Piasecka et al. 2013). So, excepting temporal pleiotropic effect, there should be other factors which may also constraint the sequence evolution of genes expressed at the phylotypic stage. One possibility is that genes specifically expressed at the phylotypic stage were shared by many cell different cell types in this stage, so it may under stronger functional pleiotropic constraints. The recent published mouse organogenesis single cell RNAseq dataset (Cao et al. 2019) can be used to identify gene activity in each cell type and to test this hypothesis in the future.

In chapter 3, I investigated the evidence of positive selection on protein-coding genes in relation to their expression over development. Overall, there is a consistent signal that positive selection mainly affects genes and pathways expressed in late embryonic development and in adult. These findings were further supported by using more recent positive selection signals (Coronado-Zamora et al. 2019) and by using organ level developmental transcriptomes (Cardoso-Moreira et al. 2019). Notably, I observed some evidence of positive selection in early development. For most changes in the early development, however, there are no clear consequences on the adult morphology Kalinka and Tomancak (2012). This raises a question

of why there are higher adaptive evolution in early development than in the phylotypic stage? One explanation, proposed by Kalinka and Tomancak (2012), suggests that many variations in early development often result from adaptation to diverse ecological circumstances. For future work, it would be of great interest to study which traits regulated by these early development related adaptive genes and how it contributes to the ecological adaptations.

In chapter 4, I studied the evolutionary conservation, divergence and positive selection on gene regulation. I found the phylotypic stage has higher proportion of conserved regulatory elements and less proportion of stage specific gain regulatory elements. Notably, I also found the phylotypic stage has lower proportion enhancers with evidence of positive selection, suggesting the contribution of positive selection shape the hourglass expression divergence pattern. Why are the phylotypic stage specific regulatory elements so conserved? One possibility, as suggested by Raff (1996), is that the genes expressed at the phylotypic stage are very crucial for the high interdependence of different developmental modules at this stage. For example, in *C. elegans*, Zalts and Yanai (2017) found the phylotypic stage enriched with 'integration genes', their expression resulting from interactions among the cell types across germ layers. These integration genes were enriched with gene ontology (GO) terms such as cell-communication and signalling functions. So, future work should focus on identifying the target genes of conserved enhancers by using Hi-C (genome-wide chromosome conformation capture assays) technology (Lieberman-Aiden et al. 2009). And then, with classic developmental biology approaches, to study the function of these genes on animal development. In addition, since the regulatory elements identified in this study are based on whole-embryo DNase-Seq, it still unclear whether the higher regulatory conservation in the phylotypic stage is true for all cell types and for only some specific cell types? With the rapid development of single cell ATAC-seq (Buenrostro et al. 2015), a future improvement of this work can be to focus on the evolution of regulatory elements in the single cell level.

In chapter 5, I first developed a method to detect positive selection on TFBS evolution based on binding affinity changes. My approach does not assume a subset of neutral sites and doesn not compare sequence evolution rates. Despite its advantages, my method can still be improved in the future. For example, in the null model of sequence evolution, I assume mutation pattern is independently at each base-pair site and mutation rate is uniform over all sites. But in the reality, both mutation rate and pattern will depend on its neighbouring nucleotides (Krawczak

et al. 1998). In addition, the analysis focuses only on point mutations, ignoring insertions and deletions. It will be of interest in future work to integrate both context dependent information and structural variations into the null model. Then, I adopted this method to detect positive selection of CTCF binding sites in 29 human tissues and cell types. I found the highest positive selection is in brain then male reproductive system. Since here I only studied one transcription factor, it remains to be seen how much these observations extend to other transcription factors. In addition, the method I developed in this study can be used to study the non-coding adaptive evolution of human specific morphological traits and be applied to other species.

In chapter 6, I studied how expression noise affects different developmental stages of fly embryogenesis. I found that expression noise follows an hourglass pattern, with lower noise at the phylotypic stage. One of the most important findings in this study is that histone modifications can confer the robustness to both stochastic and genetic gene expression variation. So, both lower expression noise and expression divergence in the phylotypic stage can be explained by stronger histone modification signal in this stage. Although I found some evidence to show that the phylotypic stage has higher histone modification signal, with the development of quantitative ChIP-Seq methods (Bonhoure et al. 2014; Orlando et al. 2014), future work should focus on directly comparing absolute histone modification signals between stages. In addition, I found that all of the four euchromatin histone modifications (H3KeMe1, H3K4Me3, H3K9Ac and H3K27Ac) have higher signal in the phylotypic stage, and are negatively correlated with expression noise. This raises a question: is the lower expression noise in the phylotypic stage driven by the cooperation of the four marks or is it only driven by some of them? In addition, there are other noise buffering mechanisms than histone modifications. For example, noise can be buffered by some specific regulatory motifs, such as negative feedback loops and incoherent feed-forward loops (Becskei and Serrano 2000; Alon 2007; Chalancon et al. 2012), or by microRNAs (Hornstein and Shomron 2006; Wu et al. 2009; Schmiedel et al. 2015; Schmiedel et al. 2017), or filtered by using cellular compartmentalization (Stoeger et al. 2016). What is the contribution of other factors on the lower expression noise in the phylotypic stage is also an interesting question.

In conclusion, I suggest that the higher conservation at the phylotypic period is caused by stronger purifying selection, stronger mutational robustness and less positive selection. In addition, I propose that the mutational robustness is a by-product of selection for minimizing

non-genetic expression variations. These results are important to understanding how the genotype-phenotype map is maintained in the face of stochasticity and mutation. Since all these findings are based on datasets from whole embryo level, with the development of single cell biology, I hope that we will soon be able to study Evo-Devo at single cell level. On the one hand, we can check whether the patterns observed at whole body are also hold true for all cell types or only for some cell types? On the other hand, single cell omics provide opportunities, for the first time, to systematically study the origin and evolution of cell types.

References

Alon U. 2007. Network motifs: theory and experimental approaches. Nat. Rev. Genet. 8:450–461.

Becskei A, Serrano L. 2000. Engineering stability in gene networks by autoregulation. Nature 405:590–593.

Bonhoure N, Bounova G, Bernasconi D, Praz V, Lammers F, Canella D, Willis IM, Herr W, Hernandez N, Delorenzi M, et al. 2014. Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. Genome Res. 24:1157–1168.

Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ. 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523:486–490.

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al. 2019. The single-cell transcriptional landscape of mammalian organogenesis. Nature 566:496–502.

Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, Liechti A, Ascenção K, Rummel C, Ovchinnikova S, et al. 2019. Gene expression across mammalian organ development. Nature:1.

Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R, Babu MM. 2012. Interplay between gene expression noise and regulatory network architecture. Trends Genet. 28:221–232.

Coronado-Zamora M, Salvador-Martínez I, Castellano D, Barbadilla A, Salazar-Ciudad I. 2019. Adaptation and Conservation throughout the Drosophila melanogaster Life-Cycle.Costantini M, editor. Genome Biol. Evol. 11:1463–1482.

Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. Nat. Genet. 38:S20–S24.

Kalinka AT, Tomancak P. 2012. The evolution of early animal embryos: conservation or divergence? Trends Ecol. Evol. 27:385–393.

Krawczak M, Ball E V., Cooper DN. 1998. Neighboring-Nucleotide Effects on the Rates of Germ-Line Single-Base-Pair Substitution in Human Genes. Am. J. Hum. Genet. 63:474–488.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293.

Orlando DA, Chen MW, Brown VE, Solanki S, Choi YJ, Olson ER, Fritz CC, Bradner JE, Guenther MG. 2014. Quantitative ChIP-Seq normalization reveals global modulation of

the epigenome. Cell Rep. 9:1163–1170.
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.
Raff RA. 1996. The shape of life : genes, development, and the evolution of animal form. University of Chicago Press.
Schmiedel J, Marks DS, Lehner B, Bluthgen N. 2017. Noise control is a primary function of microRNAs and post-transcriptional regulation. bioRxiv:doi.org/10.1101/168641.
Schmiedel JM, Klemm SL, Zheng Y, Sahay A, Blüthgen N, Marks DS, van Oudenaarden A. 2015. Gene expression. MicroRNA control of protein expression noise. Science 348:128–132.
Stoeger T, Battich N, Pelkmans L. 2016. Passive Noise Filtering by Cellular Compartmentalization. Cell 164:1151–1161.
Wu C-I, Shen Y, Tang T. 2009. Evolution under canalization and the dual roles of microRNAs: a hypothesis. Genome Res. 19:734–743.
Zalts H, Yanai I. 2017. Developmental constraints shape the evolution of the nematode mid-developmental transition. Nat. Ecol. Evol. 1:0113.