Mechanisms of Phenotypic Evolution at Different Levels of Biological Organization: from Molecules to Organisms

by

Daohan Jiang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in The University of Michigan
2022

Doctoral Committee:

Professor Jianzhi Zhang, Chair
Associate Professor Jeffery Kidd
Professor L. Lacey Knowles
Professor Patricia Wittkopp

Daohan Jiang

ORCID iD: 0000-0002-3524-9725

rexjiang@umich.edu

©Daohan Jiang 2022

Acknowledgements

First of all, I thank my advisor, Prof. Jianzhi "George" Zhang, for his invaluable support in the past five years. Thanks to how open-minded my advisor is, I have been able to explore diverse potential research topics, including what had been new to our lab; even though I have wasted time on studies that ended up being fruitless, he has remained supportive and ready to help. My advisor has also allowed me to work in my own pace and stay comfortable; on the other hand, he would give me timely suggestions at key moments, helping me overcome my confusions and move forward. Importantly, Prof. Zhang has not only been my advisor, but also a role model from whom I have been learning. Whether or not it could be far-fetched to reproduce his success, it is certain that interactions with him have shed light on what used to confuse me and prompted me to improve myself.

Throughout my graduate study, I have also benefitted from interactions with a number of other people. I thank all members of my thesis committee for their support and patience. I thank current and former lab members with whom I have communicated for all the inspiring discussions both about my own research and a broader range of topics in science. I must also thank Prof. Alexey Kondrashov for a series of insightful discussions about evolutionary genetics in the Winter 2019 semester, which expanded my sight and refreshed my mind.

Last but not least, I would like to thank all people that have helped me deal with mental health issues. In particular, I thank my current therapist and professionals I have previously seen at Counseling and Psychological Services and University Health Service of University of

Michigan. In the end, I would like to thank creators of anime series that have “healed” me and helped me through some of my darkest hours.

Table of Contents

Acknowledgements **ii**

List of figures **ix**

List of tables **xii**

List of appendices **xiv**

Abstract **xv**

Chapter 1: Introduction **1**

1.1 Modeling phenotypic evolution: a quantitative genetic framework 3

1.2 Phenotypic evolution on a phylogeny 6

1.3 Phenotypic evolution under natural selection 8

1.3.1 Stabilizing selection 9

1.3.2 Directional selection 10

1.4 Nonadaptive phenotypic evolution 11

1.4.1 Mutational bias 12

1.4.2 Pleiotropy 13

1.4.3 Entrenchment 14

1.5 Correlated evolution of phenotypic traits 15

1.5.1 Selection on individual trait(s) 16

1.5.2 Multivariate selection 17

1.6 Molecular errors and evolution of molecular traits 18

1.7 Thesis overview 20

1.8 References 22

Chapter 2: Evolution of Transcription Regulation: Positive Selection on Transcription Factor Binding Sites, or False Positives? **26**

2.1 Abstract 26

2.2 Introduction 27

2.3 Results 30

2.3.1 Ascertainment bias leads to elevated false positive rate 30

2.3.2 A simple correction to reduce false positive rate 31

2.4 Discussion 33

2.5 Materials and Methods 35

2.5.1 Generation of simulated sequences 35

2.5.2 Testing positive selection on TFBSs 37

2.5.3 Modeling identifiability of TFBSs 38

2.5.4 Corrected test of positive selection 39

2.6 References 40

Chapter 3: Evolution of Post-transcriptional RNA Modifications: the Preponderance of Nonsynonymous A-to-I RNA Editing in Coleoids is Nonadaptive **47**

3.1 Abstract 47

3.2 Introduction 48

3.3 Results 50

3.3.1 A nonadaptive hypothesis and its predictions 50

3.3.2 Patterns of restorative and diversifying editing 53

3.3.3 Accelerated nonsynonymous G-to-A substitutions 55

3.3.4 The potential benefit of shared editing among species 56

3.4 Discussion 59

3.5 Methods 64

3.5.1 Transcriptomes, editing sites, and ancestral sequences 64

3.5.2 Restorative and diversifying editing 65

3.5.3 Comparing median editing levels 66

3.5.4 Proportion of diversifying editing that is adaptive 67

3.5.5 Contributions of compensatory editing to F_R and L_R 67

3.6 References 68

Chapter 4: Evolution of Organismal Traits: Fly Wing Evolution Explained By a Neutral Model with Mutational Pleiotropy **80**

4.1 Abstract 80

4.2 Introduction 81

4.3 Materials and Methods 82

4.3.1 Comparison of covariance matrices 82

4.3.2 Examination of matrix comparison methods using simulation 83

4.3.3 A neutral model with pleiotropy to explain the evolution of fly wing morphologies 85

4.3.4 Simulation of evolution of neutral traits with mutational pleiotropy 87

4.4 Results 90

4.4.1 Houle *et al*.'s method of matrix comparison is biased 90

4.4.2 Unequal constraints on the fly wing traits 92

4.4.3 A neutral model with mutational pleiotropy explains patterns of fly wing evolution 93

4.5 Discussion 95

4.6 References 97

Chapter 5: Detecting Natural Selection in Trait-Trait Coevolution 105

5.1 Abstract 105

5.2 Introduction 106

5.3 Results 109

5.3.1 Evolutionary correlations differ from mutational correlations for many trait pairs 109

5.3.2 Effects of different selection regimes on trait-trait coevolution 113

5.3.3 Selection for enhanced modularity of yeast morphological traits 114

5.4 Discussion 116

5.5 Materials and Methods 120

5.5.1 Phenotypic data 120

5.5.2 Influence of the sampling error on the *M* matrix 121

5.5.3 Impact of the environmental difference on the correlational structure of the yeast traits 122

5.5.4 Comparison between mutational and evolutionary correlations 122

5.5.5 Computer simulation of trait-trait coevolution under selection 125

5.5.6 Computer simulation of trait-trait coevolution under mutational bias 126

5.6 References 127

Chapter 6: Discussions and Future Directions 142

6.1 Introduction 142

6.2 More models for evolution of molecular traits? 142

6.2.1 Evolution of post-transcriptional modifications 143

6.2.2 Coevolution of transcription and translation 144

6.3 Mechanistic basis of the M matrix 146

6.4 Comparing different levels of biological organization 147

6.5 Evolution of traits with evolving genetic architectures 149

6.6 References 150

Appendices **153**

List of Figures

Figure. 2-1. Correlation between the SVM score cutoff and false positive rate. 43

Figure. 2-2. Correlation between identifiability and false positive rate. 44

Figure 2-3. *P*-values before and after correction for human CEBPA binding sites. 45

Figure. 3-1. The harm-permitting model and a strategy to detect the harm-permitting effect. 73

Figure. 3-2. Comparison of restorative and diversifying editing with synonymous editing in coleoids. 75

Figure. 3-3. Coleoid nonsynonymous to synonymous substitution rate ratios (d_N/d_S) for various nucleotide changes. 77

Figure. 3-4. Patterns of shared editing in coleoids. 79

Figure. 4-1. Comparison between Houle *et al.*'s method and our new method. 100

Figure. 4-2. Linear regression between $\log_{10}$(*M* variance) and $\log_{10}$(*R* variance) or $\log_{10}$(*G* variance) along orthogonal directions for fly wing traits. 102

Figure. 4-3. Patterns of phenotypic evolution observed from computer simulation of 20 neutral, orthogonal, focal traits with mutational pleiotropy. 104

Figure. 5-1. Mutational (COR_M) and evolutionary (COR_E) correlations for all pairs of the 220 yeast morphological traits. 132

Figure. 5-2. Examples of yeast trait pairs with COR_E significantly different from the corresponding COR_M. 134

Figure. 5-3. Schematic illustration of predictions made by models of trait-trait coevolution. Each circle represents the equilibrium mean phenotype of two hypothetical traits (trait 1 and trait 2) of a diverging lineage. 136

Figure. A-1. Neighbor-joining tree based on brain transcriptomes of species considered in this study. 156

Figure. B-1. Respectively comparing restorative and diversifying editing with synonymous editing for species-specific editing supports the nonadaptive hypothesis of the preponderance of nonsynonymous editing in coleoids. 157

Figure. B-2. Comparison of editing levels between synonymous and nonsynonymous editing sites of different genes. 159

Figure. B-3. Ratios of editing frequencies of synonymous (F_S), diversifying (F_D), and radical (F_R) editing in various ranges of editing levels. 160

Figure C-1. Performances of Houle *et al.*'s method and the new method across a range of similarity between *M* and *R*, with *M* and *R* matrices subject to estimation error. 165

Figure C-2. Performances of Houle *et al.*'s method and the new method when the empirical *M* matrix is set to be $M_{original}$ and the Drosophilid tree is used to simulate trait evolution. 167

Figure. D-1. Neighbor-joining tree of the 16 natural yeast strains used in this study, based on 1,544,489 biallelic single nucleotide polymorphism (SNP) sites. 171

Figure. D-2. Mutational (COR_M) and evolutionary (COR_E) correlations for all pairs of the 187 yeast morphological traits. 172

Figure. D-3. Mutational (COR_M) and evolutionary (COR_E) correlations for all pairs of the 24 fly wing morphological traits. 174

List of Tables

Table. 2-1. Overview of simulated and real data. 46

Table. 5-1. Numbers of trait pairs with significantly different COR_E and COR_M in the yeast and fly data. 138

Table. 5-2. Parameters and results of simulations of trait-trait coevolution. 139

Table. 5-3. Overall phenotypic integration (V_{eigen}) and modularity (CR) at the levels of mutation and evolutionary divergence. 141

Table. B-1. Numbers of editing sites of each category. In the parentheses are numbers of species-specific editing. 161

Table. B-2. Patterns of restorative and diversifying editing in 12 different tissues of the bimac. 162

Table. B-3. Proportion of editing sites belonging to each editing level range. 163

Table. B-4. Numbers of sites that are respectively edited, unedited, and substituted in the fourth coleoid among those sites that are edited in the other three coleoids examined. 164

Table. C-1. Models examined by Houle *et al*. and their predictions regarding evolutionary rate, scaling exponent of evolutionary rate with mutational variance, and phylogenetic signal. 169

Table. C-2. Patterns of phenotypic evolution observed from computer simulation of non-neutral, orthogonal, focal traits with mutational pleiotropy. 170

Table. D-1. Pearson's correlation in eigenvalues between M and the covariance matrix estimated from a sample from M ($\widetilde{M}$). 176

Table. D-2. Parameters and results of simulations of trait-trait coevolution in the presence of mutational bias 177

List of Appendices

Appendix A: Parallel transcriptomic changes in the origins of divergent monogamous vertebrates? 153

A.1 Results and discussion 153

A.2 Methods 155

A.3 References 155

Appendix B: Supplementary figures and tables for chapter 3 157

Appendix C: Supplementary figures and tables for chapter 4 165

Appendix D: Supplementary figures and tables for chapter 5 171

Abstract

Understanding the origin of the vast phenotypic diversity among organisms is a central goal of evolutionary biology. Phenotypic traits evolve in face of mutations, genetic drift, and selection, which, alongside genetic architectures of traits, shape patterns of phenotypic variation within and between species. While the relative contributions of these processes to phenotypic evolution is of interest, teasing apart their effects has been challenging. In this dissertation, I take advantage of publicly available large-scale phenomic data to investigate mechanisms underlying the evolution of multiple phenotypic traits that represent different levels of biological organization. Combining computational and theoretical approaches, I test models of phenotypic evolution and evaluate the extent to which each trait of interest is influenced by selection. These analyses cover pre-transcriptional (Chapter 2) and post-transcriptional (Chapter 3) molecular processes as well as morphological traits (Chapter 4 and 5), representing different levels of biological organization. In Chapter 2, I focus on interaction between the transcription factor and the transcription factor binding sites and ask how selection that enhances their interaction can be detected without being confounded by intrinsically biased data. In Chapter 3, I focus on A-to-I RNA editing and show that the high abundance of nonsynonymous editing in coleoid cephalopods, which appeared to be adaptive, is instead better explained by a nonadaptive error-permitting model. In Chapter 4, I investigate the relationship between mutational variance and evolutionary rate among a set of wing morphological traits in drosophilid flies and show a formerly proposed interpretation resulted from a biased method; with scaling between mutational variance and evolutionary rate re-estimated with a corrected method, I also propose a mechanic

model that explains the observations. In Chapter 5, I study correlated evolution of different phenotypic traits and ask how frequently correlation in long-term evolution differs from correlation resulting from mutational covariances; by analyzing cell morphological traits of yeasts and wing morphological traits of flies, I identify pairs of traits that are likely subject to multivariate selection for allometric relationship between traits. I conclude that nonadaptive processes, which used to be overlooked, can play a remarkable role in shaping patterns of phenotypic variation.

Chapter 1

Introduction

Organisms that have lived on earth exhibit spectacular phenotypic diversity spanning all levels of biological organization: from microscopic processes taking place within the cell to the shape and size of an organism. To understand the origin of this diversity has naturally been a central goal of evolutionary biology.

Theories in population genetics have established a framework in which the change of allele frequencies through time in face of mutation, random genetic drift, and/or selection can be predicted (Gillespie 2004). This theoretical framework, in the light of quantitative genetics, also yields predictions regarding the evolution of phenotypic traits (Lande 1976, 1979; Lynch and Hill 1986; Harmon 2018). Fields of population genetics and molecular evolution have seen a debate between neutralists and selectionists as advancement of DNA sequencing technologies enabled more rigorous tests of selection (Kimura 1968, 1983). In contrast, at the phenotypic level, the theoretical predictions remain largely untested.

Nonetheless, the evolution of phenotypic traits has often been believed to be driven primarily by selection, such that observed phenotypes of organisms reflect adaptation to the environment. This view is backed up by numerous documented examples of adaptation observed over diverse taxa, including the famous cases of Darwin's finches and the orchid *Angraecum sesquipedale*: the former was cited as evidence for evolution by natural selection, while

coevolution between the latter and its pollinator demonstrated predictive power of the theory of natural selection (Darwin 1859, 1862). Even Kimura, who proposed the neutral theory of molecular evolution, did not believe his theory to hold at the phenotypic level (Kimura 1983). However, the view that observed phenotypes are generally adaptive has been facing difficulties as well: adaptive explanations for many observations are not supported by solid evidence and have been criticized as storytelling; sometimes, alternative explanations appear to be more reasonable or turn out to be better supported by evidence (Gould and Lewontin 1979).

Thanks to recent advancements in high-throughput phenotyping, a lot more large-scale phenotypic datasets have been generated, allowing phenotypic evolution to be studied in more quantitative ways. In my studies, I take advantage of several such datasets to quantitatively test predictions made by theoretical models of phenotypic evolution and investigate the role of different evolutionary processes in shaping phenotypic variation at different levels of biological organization.

To help understand questions being addressed in my works as well as the methods being used, I first introduce the theoretical framework under which phenotypic evolution is modeled and the expected patterns of phenotypic evolution under neutrality derived under this framework. After introducing the neutral expectation, I introduce how selection could influence phenotypic evolution and make patterns of phenotypic variation differ from neutral expectations. Then I move from selection to nonadaptive phenotypic evolution and introduce some underlying processes. The last two sections cover two special topics: first, I introduce mechanisms underlying correlated evolution of traits; second, I introduce evolution of molecular traits, as the nature of these traits could lead to distinct patterns in long-term evolution.

1.1 Modeling phenotypic evolution: a quantitative genetic framework

Phenotypic traits of potential interest are usually quantitative traits (aka complex traits) that vary in a continuous fashion among individuals. Although their exact genetic underpinnings are usually unclear, it is generally agreed that the number of loci underlying a quantitative trait is so large that any individual mutation's effect on the phenotype is very small, which manifests as predictable patterns in long-term evolution. By incorporating this feature of quantitative traits' genetic architectures into the theoretical framework of population genetics, quantitative traits' evolution can be readily modeled. Below I demonstrate how some key predictions about a quantitative trait's evolution can be derived under a highly simplified and idealized setting. Although some assumptions made under this setting are likely often violated in real phenotypic evolution, this simple and idealized model can be a good starting point for more complex ones and its predictions can serve as null hypotheses in empirical studies.

Let us consider a haploid population with effective size N_e and a quantitative trait controlled by x loci. Each locus has two possible alleles, a "null" allele A_0 and a "mutant" allele A_1, which contribute 0 and a to the trait value, respectively. An individual's phenotype is a direct sum of all loci's phenotypic values, which is zero when the individual has A_0 alleles only. Effect size of each mutation, thus, is simply $|a|$. The number of mutations that affect the trait under concern entering the population per generation is then $xN_e\mu$, where μ is mutation rate per locus affecting the trait. In the absence of selection, each mutation's fixation probability is $\frac{1}{N_e}$ (Kimura 1962), so the total number of mutations that could possibly contribute to long-term evolution is $x\mu$. Phenotypic variance introduced into the population by newly arising mutations per generation can be written as (Lynch and Hill 1986):

$$V_M = x\mu Var(a) = x\mu E(a^2) \quad (1)$$

, which is commonly known as the mutational variance. In the case of a diploid population, N_e in the derivation is replaced by $2N_e$.

Change in the population's mean phenotype $\bar{z}$ in a generation, is the sum of these $x\mu$ mutations' effects. As x is very large, the amount of the change, denoted $\Delta\bar{z}$, should be normally distributed. Its mean and variance can be respectively written as (Lynch and Hill 1986):

$E(\Delta\bar{z}) = x\mu E(a)$ (2)

, which is equal to zero when $E(a) = 0$, and

$Var(\Delta\bar{z}) = x\mu Var(a) = x\mu E(a^2) = V_M$ (3)

The equations above can be readily extended to longer time periods:

$E(\Delta\bar{z}_T) = x\mu E(a)T$ (4)

$Var(\Delta\bar{z}_T) = V_M T$ (5)

, where $\Delta\bar{z}_T$ is the total amount of change in $\bar{z}$ in T generations.

If the initial population only has A_0 alleles and thus a mean phenotype of zero, distribution of $\Delta\bar{z}_T$ is also distribution of $\bar{z}_T$, the population mean after T generations.

The probability distribution of $\bar{z}_T$ can also be viewed as its distribution among many initially identical replicate populations after T generations of independent evolution. $Var(\Delta\bar{z}_T)$, therefore, captures how fast lineages diverge phenotypically and is often referred to as the evolutionary rate parameter. Under neutrality, $Var(\Delta\bar{z}_t)$ is equal to V_M and is independent of the population's size, which can be viewed as a phenotypic-level equivalent of the fundamental principle in molecular evolution that evolutionary rate is equal to the mutation rate under neutrality (Kimura 1962).

If $Var(\bar{z})_0 = 0$, which would be the case if replicate populations with identical $\bar{z}$ are considered, there is

$$Var(\bar{z})_T = V_M T \quad (6)$$

As $\Delta\bar{z}$ is normally distributed, evolution of $\bar{z}$ can be viewed as a Brownian motion (BM) process, with rate parameter of the BM process, σ^2, equal to V_M in this case (Lande 1976). This fact, along with convenient mathematical properties of the BM process, has made BM a widely used null model in studies of phenotypic evolution.

It is worth noting that many theoretical and empirical studies focus on phenotypic variance introduced by standing genetic variations each generation, which is referred to as genetic variance and denoted V_G, instead of V_M. As mutations that are too deleterious to stay long in the population only contribute to V_M but not V_G, V_G represents what can possibly contribute to long-term evolution. In this introduction, however, only V_M is used when neutral models are being introduced, since there is $V_G = V_M$ under neutrality.

The above formulas can be readily extended to be multivariate (i.e., multiple traits considered simultaneously) (Lande 1979). If there are m traits under consideration, $\bar{z}$ and $\Delta\bar{z}$ in the above formulas become vectors of length m:

$$\bar{z} = \begin{bmatrix} \bar{z}_1 \\ \vdots \\ \bar{z}_m \end{bmatrix} \quad (7)$$

$$\Delta\bar{z} = \begin{bmatrix} \Delta\bar{z}_1 \\ \vdots \\ \Delta\bar{z}_m \end{bmatrix} \quad (8)$$

, and V_M becomes by the mutational variance-covariance (VCV) matrix, M, which is a symmetric m-by-m matrix:

$$M = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{1m} & \cdots & \sigma_m^2 \end{bmatrix} \quad (9)$$

, where diagonal element σ_i^2 is mutation variance of trait i and off-diagonal element σ_{ij} is mutational covariance between trait i and trait j. Mutational covariance refers to phenotypic covariance between traits introduced by newly arising mutations each generation and reflects pattern of mutational pleiotropy. $\Delta\bar{z}$ follows a multivariate distribution with mean $\vec{0}$ (i.e., a vector with only zeros), and (co)variances given by the M matrix. Increase of phenotypic (co)variances in T generations, therefore, can be written as:

$$P_T = MT \quad (10)$$

1.2 Phenotypic evolution on a phylogeny

In the previous section, rate of phenotypic divergence is expressed as the rate at which variance among replicate populations increase through time. Referred to as "replicates", these population are assumed to be independent of each other within the time frame under consideration. Species in the real world, however, cannot be modeled as such due to their shared evolutionary history. As a result, distribution of phenotypes among a set of species is determined jointly by the rate of phenotypic evolution and the phylogenetic relationship of these species. Hence, models and methods have been developed to study evolutionary divergence in phenotypic traits in a phylogenetic context. Note that the theoretical framework described below is applicable not only to well-defined biological species, but also populations, strains, or operational taxonomic units as long as their evolutionary relationships can be described (or approximated) with a phylogenetic tree.

For a set of n species for which a phylogenetic tree (including information of both topology and branch lengths) is available, the amount of shared evolutionary history between them can be summarized with an n-by-n phylogenetic VCV matrix C, where the diagonal entries

represent each species' distance from the most recent common ancestor (MRCA) of the n species (i.e., root of the tree) while the off-diagonal entries represent amount of evolutionary history shared by each pair of species. Specifically, $C[i,i]$ is obtained by summing up lengths of all branches leading to species i while $C[i,j]$ (and $C[j,i]$) is obtained by summing branches leading to MRCA of species i and j. If there is no selection and V_M is constant, phenotypes of the n species can be viewed as a vector drawn from a multivariate normal distribution given as (Harmon 2018):

$$P = V_M C \quad (11)$$

, where P is the phenotypic VCV matrix. Each diagonal element of P represents variance of one species' phenotype and each off-diagonal element represents phenotypic covariance between two species. If branch lengths reflect divergence time, variance for each species is proportional to the amount of time for which it has evolved, while covariance between two species is proportional to amount of time for which they evolved as one lineage before splitting. Hence, formula (11) is essentially an equivalent of formula (3).

Formula (11) can be readily extended to the case of m traits and written as Koreneker product of M and C (Harmon 2018):

$$P = M \otimes C = \begin{bmatrix} C[1,1]M & \cdots & C[1,n]M \\ \vdots & \ddots & \vdots \\ C[n,1]M & \cdots & C[n,n]M \end{bmatrix} \quad (12)$$

, where P is an mn-by-mn matrix with diagonal "blocks" representing different species and structure within each "block" representing (co)variances of traits; combination of traits' values in all species can be viewed as sampled from a multivariate distribution with (co)variances given by P; specifically, covariance between value of trait i in species k and value of trait j in species l, therefore, is a product of $M[i,j]$ and $C[k,l]$.

Formula (12) can be written in a more general form with M replace by the evolutionary covariance matrix R:

$$P = R \otimes C = \begin{bmatrix} C[1,1]R & \cdots & C[1,n]R \\ \vdots & \ddots & \vdots \\ C[n,1]R & \cdots & C[n,n]R \end{bmatrix} \quad (13)$$

Diagonal element $R[i,i]$ represents average evolutionary rate of trait i, while off-diagonal element $R[i,j]$ represents covariance between traits i and j during evolution. R is the same as M under neutrality and would be different when selection is present.

The above equations show how the phenotype would be distributed among species given the evolutionary rate parameter. Conversely, evolutionary rate can be inferred from phenotypes of a set of species with known phylogeny. A method that is commonly used is the phylogenetic independent contrast (PIC) method, which converts the observed phenotypes of n species into $n-1$ data points that are phylogenetically independent (i.e., PICs). Each contrast reflects rate of divergence for a pair of sister species or clades and their average is the PIC estimate of evolutionary rate (Felsenstein 1985).

1.3 Phenotypic evolution under natural selection

Introduced in the previous sections is a neutral model of phenotypic evolution that predicts patterns of phenotypic variation in the absence of selection. Natural selection, on the other hand, can lead to different patterns of phenotypic evolution and variation. How exactly a population would respond to selection is dependent on direction (i.e., what phenotype(s) have high fitness) and strength (i.e., how rapidly fitness changes with phenotype) of selection, which can have infinitely many possible combinations. Nevertheless, this myriad of scenarios can be

qualitatively divided into two classes based on if the trait under concern is under stabilizing selection or directional selection.

1.3.1 Stabilizing selection

Stabilizing selection happens when there exists an intermediate phenotype that maximizes fitness and mutations creating extreme phenotypes are selected against. Under stabilizing selection, the population's mean phenotype is expected to stay near the optimal phenotype but can still fluctuate due to fixation of slightly deleterious mutations (i.e., coefficient of selection $s < \frac{1}{N_e}$, or $s < \frac{1}{2N_e}$ for s diploid population). With effect of mutations normally distributed, long-term expectation of the population mean phenotype is simply the optimum. The steady-state variance of the population mean phenotype to be reached upon a mutation-drift-selection balance is determined jointly by strength of selection (i.e., shape of fitness function around the peak), mutational variance V_M, and N_e. Therefore, evolution of the population mean under stabilizing selection through time can be considered a random walk with constraint and conveniently modeled as an Ornstein-Uhlenbeck (OU) process (Lande 1976).

The model of stabilizing selection has two major predictions regarding pattern of variation among species, both of which can be tested against predictions of the neutral model with phenotypic data from multiple species and their phylogenetic relationship. First, the amount of variance among lineages should be less than what would be expected under neutrality (Lande 1976; Lynch 1990). Second, in contrast to indefinite increase of variance among lineages under neutrality, under stabilizing selection, variance among lineages saturates after an equilibrium is reached (Lande 1976; Bedford and Hartl 2009).

Ongoing stabilizing selection can also be inferred from the pattern of within-population variation: intensifying stabilizing selection causes phenotypic variance to decrease over a short period of time as extreme phenotypes are being purged by selection; stabilizing selection can also be inferred from the quadratic selection gradient, which is estimated from phenotypes of a set of individuals and their respective fitness values (Lande and Arnold 1983; Arnold and Wade 1984a, b; Kingsolver et al. 2001).

1.3.2 Directional selection

A trait is said to be under directional selection when phenotypic changes in a particular direction are selected for. As more mutations causing phenotypic changes in the desired direction are fixed by positive selection, the population mean phenotype moves in the same direction through time until an optimum is reached, in which case stabilizing selection takes over. Disruptive selection, which refers to selection for extreme phenotypes and against intermediate phenotypes in between, can be viewed as a special case of directional selection where two (sub)populations evolve towards different optima.

The rate of adaptive evolution under directional selection depends on both strength of selection and supply of adaptive mutations, as summarized by the well-known Breeder's Equation, $R = h^2 S$, which states that response to selection, R (i.e., evolutionary change in population), is a product of heritability, h^2 (i.e., proportion of phenotypic variance caused by additive genetic variation) and selection differential, S (i.e., local slope of fitness function) (Lush 1937). Strength of ongoing directional selection is commonly described using the linear selection gradient estimated from phenotypes and fitness values of a set of individuals (Kingsolver et al. 2001).

Directional selection also leaves its signature on pattern of phenotypic variation among species. In contrast to stabilizing selection, which makes among-species variance lower than the neutral expectation, directional selection would have an opposite effect if direction of selection is not the same in all species. However, looking into the overall variance does not tell us what lineage(s) underwent directional selection, making it necessary to examine the distribution of species' phenotype in greater detail. For instance, directional selection in a specific clade could lead to drastic difference between mean phenotype of this clade (with structure of phylogeny within the clade accounted for) and that of the remaining species that is unlikely to occur under neutrality.

Lineage-specific directional selection can also be tested in only a pair of species or strains using quantitative trait loci (QTLs) identified in an F_2 population from a cross of the two species or strains. Directional selection is inferred if alleles from the same parent tend to have a concordant effect on the phenotype (i.e., a particular parent's alleles have positive effect on trait value in significantly more than 50% QTLs) (Orr 1998). Another recently proposed test only utilizes phenotypes of both parents and F_2 individuals: if between-parents variance is significantly greater than variance among F_2 individuals, directional selection is inferred; if the opposite is true, stabilizing selection is inferred (Fraser 2020). These tests do not require information from many species but are only applicable when there is no strong hybrid incompatibility.

1.4 Nonadaptive phenotypic evolution

While selection optimizes the phenotype, full optimization can often be prevented by other processes. These processes can slow down adaptation, have the population mean

phenotype stay at a sub-optimal state, or cause nonadaptive evolutionary changes (i.e., evolutionary changes that do not improve the population's mean fitness).

1.4.1 Mutational bias

Mutational bias, in the context of phenotypic evolution, refers to mutations tend to cause phenotypic changes in a particular direction. As introduced in section 1.1, under neutrality, there is $E(\Delta\bar{z}) = 0$ when $E(a) = 0$. Conversely, when $E(a) \neq 0$, there is $E(\Delta\bar{z}) \neq 0$. As such directional change can occur under directional as well, mutational bias can potentially create false signal of directional selection.

Mutational bias is also known to be able to interfere with selection and prevent optimization of the phenotype. If selection and mutational bias are in opposite directions, rate of adaptation can be slowed down, and the population may be stalled at a mutation-drift-selection balance (or simply mutation-selection balance if effect of drift is negligible) before reaching the optimum.

Evidence for mutational bias has been found in empirical studies. For example, for more than 80% of yeast cell morphological traits, distribution of gene deletions' effects is significantly biased in one direction (Ohya et al. 2005). Mutational bias has also been observed for mutations in a promoter in terms of their effect on gene expression (Hodgins-Davis et al. 2019). Another trait subject to mutational bias is genomic GC content, which arises directly from mutational bias against C and G bases at molecular level. Observed correlation between GC content and effective population size indicate this trait could be in mutation-drift-selection balance in many lineages (Long et al. 2018).

1.4.2 Pleiotropy

Mutational pleiotropy refers to the phenomenon that mutations can affect multiple traits. A trait would be affected by pleiotropic mutations when it shares genetic underpinnings with other traits (e.g., controlled by the same gene, or underlying genes controlled by the same regulatory elements); its evolution would be influenced not only by selection on itself, but also by selection on those other traits.

In the presence of pleiotropy, selection on a trait could leave a signature in the pattern of variation in other traits that are not under selection *per se*. For example, if some mutations that affect a neutral trait (i.e., not under selection) also affect a trait under stabilizing selection, amount of interspecific variation in the former neutral trait would be lower than expected, as a fraction of its mutational input is purged by selection and rendered unable to contribute to long-term evolution. In such scenarios, observed signals of selection can be misinterpreted. A trait may be considered important to adaptation to the environment despite it is not directly responsible.

Pleiotropy could interfere with selection if traits affected by pleiotropic mutations are all under selection. Adaptive evolution of a focal trait can be slowed down or even stalled if mutations that are adaptive with respect to this focal trait also cause disadvantageous changes in other traits. As the number of traits affected by each mutation increases, such constraint can be very strong such that adaptation is almost impossible (Hansen et al. 2003; Hansen and Houle 2008), a phenomenon also known as “cost of complexity” (Orr 2000; Wagner et al. 2008).

As pleiotropy can make mutations more likely to be deleterious and constrain adaptation, it is hypothesized that second-order selection has been optimizing genetic architectures of traits during evolution to reduce deleterious pleiotropic effects of mutations and/or enhance

adaptability, ultimately resulting in modularity of genotype-phenotype (G-P) networks. In a modular G-P network, traits that belong to the same functional module tend to share genetic underpinnings while traits in different modules tend not to, such that mutations tend not to simultaneously affect functionally unrelated traits and thus less likely to be strongly deleterious and/or more likely to be adaptive (Wagner 1996; Wagner and Altenberg 1996; Wagner et al. 2007; Melo and Marroig 2015). Modularity of G-P networks has been frequently observed and is agreed to be rather universal (e.g., (Fabre et al. 2020; Navalon et al. 2020)), while the cause is contentious: second-order selection is generally weaker than first-order selection and may not be sufficient for modularity to evolve; on the other hand modularity may also arise spontaneously without selection or as a byproduct of selection for something else (Wagner et al. 2007; Clune et al. 2013).

1.4.3 Entrenchment

Another process that can lead to spurious signal of selection and/or adaptation is entrenchment, which is also known to constraint protein sequences' evolution (Pollock et al. 2012; Goldstein et al. 2015; Shah et al. 2015). A trait is entrenched if maintained primarily because its loss would reveal deleterious effect of other evolutionary changes it had permitted in the past. The focal trait could have evolved because of adaptation in the first place, but after the original cause of selection is withdrawn, this trait *per se* is not adaptive anymore and would be lost in evolution if the otherwise deleterious changes had not been permitted. Furthermore, since it usually takes energy and other resources to maintain a trait, the entrenched status is probably suboptimal. The optimum, instead, is the ancestral state that neither the focal trait nor what it permitted is present, which is no longer be achievable. This process is also known as

constructive neutral evolution, with "constructive" referring to increase and/or maintenance of complexity (Stoltzfus 1999; Lukes et al. 2011; Munoz-Gomez et al. 2021).

It is worth noting that an entrenched trait could nonetheless show apparent signal of adaptation if selection acts to improve fitness without having this trait lost. For example, selection may have the entrenched trait better conceal deleterious effects of evolution it previously permitted. The entrenched, unlosable trait may also be "repurposed" to serve additional function(s). Thus, signal of adaptation is not mutually exclusive with entrenchment, nor does it necessarily mean the observed state is optimal.

A prime example of entrenched complex trait is RNA editing. RNA editing in plant mitochondria is likely maintained due to entrenchment, as most editing events revert the sequence back to an ancestral state (Covello and Gray 1993; Gray and Covello 1993; Sloan 2017). As to be introduced later, RNA editing can still be entrenched even if the number of such restorative editing events is not particularly large (Jiang and Zhang 2019b).

1.5 Correlated evolution of phenotypic traits

Phenotypic traits do not evolve independently of other traits. Instead, evolutionary change in different traits can be correlated. For example, brain size and body size are strongly correlated among vertebrate species and show an approximately log-linear relationship (Lande 1979). Correlated evolution of different traits can result from mutation, selection, or linkage between underlying loci of different traits. As linkage can be broken down by recombination, linkage probably has little contribution to long-term coevolution of traits. Thus, it is mutation and selection's impact on traits' coevolution that are introduced in this section. Under neutrality, correlation between traits in long-term evolution (i.e., evolutionary correlation) should simply

reflect correlation created by mutations (i.e., mutational correlation). Selection, on the other hand, can influence the correlation in different ways.

It is worth noting that apparent correlation between traits across species can result from structure of phylogeny alone. To better understand whether pattern of correlation evolution results from mutation or selection, it is necessary to rule out this confounding factor. One method is the PIC method introduced in section 1.2: with phenotypic data converted to PICs, correlation coefficient can be estimated from these independent data points. An alternative method, phylogenetic generalized least squares (PGLS), can do the same.

In this section, I mainly introduce how selection would influence correlation between traits during evolution when two traits are under consideration. However, all conclusions are readily applicable to a multivariate phenotypic space with more dimensions.

1.5.1 Selection on individual trait(s)

As mentioned in section 1.4.2, a trait that is not under selection can exhibit signal of selection when it shares mutational input with trait(s) under selection. Thus, two traits with non-zero mutational covariance are expected to undergo correlated when one of them is under either directional or stabilizing selection. It is possible, however, that stabilizing selection is so strong that there is little evolutionary divergence and no detectable evolutionary correlation.

In contrast, when both traits are under selection, the resulting pattern of coevolution would depend on how mutational covariance and selection are aligned. If one of the two traits considered is under directional selection while the other is under stabilizing selection, their evolutionary correlation would lie somewhere between zero and their mutational correlation; as

time goes by, correlation would become weaker as more and more mutations affecting mainly the one trait under directional selection are fixed, until the optimum is reached.

When both traits under concern are under directional selection towards their respective optima and their mutational covariance is not perfectly aligned with direction of selection, their trajectory of adaptive evolution may reflect selection instead of mutations as long as there is still a reasonable amount of mutational input in the direction of selection. If mutations in this desired direction are really in scarce, trajectory of evolution could be different from the shortest one between the starting phenotype and the optimum (Hether and Hohenlohe 2014).

1.5.2 Multivariate selection

Discussions about selection in previous sections focus on scenarios where a single optimal phenotype (i.e., a fitness peak) exists. It is also possible that the optimum is not a peak, but a ridge in the fitness landscape: it is the relationship between traits rather than values of individual traits that matters.

Intuitively, with only neutral mutations with effects aligned with the fitness ridge permitted, phenotypes of different species would be distributed along this ridge, resulting in correlation across lineages. Such multivariate selection and selection on individual traits' values are not mutually exclusive: it is plausible that one trait is under direction towards an optimum, and other trait(s) evolve to follow up such that the allometric relationship is maintained (Lande 1979). The aforementioned brain-body allometry is a classic example of such selection, as it is obvious that the ratio of brain size and body size must not be too extreme for the animal to be viable, and selection on brain or body size alone is known to mediate coevolution of brain and body sizes in certain clades.

1.6 Molecular errors and evolution of molecular traits

Technological advancements in recent years have allowed quantification of many molecular phenotypes, such as gene expression levels, locations of post-transcriptional and post-translational modifications, cellular concentrations of metabolites, and intensity of inter-molecular interactions. Such traits used to be hard, if not impossible, to observe and measure compared to organismal levels traits like morphological or life history traits and therefore remained less well studied by evolutionary biologists until recently. As these traits represent a level of biological organization in between DNA sequences and organismal traits, it is of great interest whether their evolution has distinct patterns and underlying mechanisms.

With quantification of molecular phenotypes often done at genomic (or transcriptomic, proteomic etc.) scale, numerous molecular traits of with no known functions (i.e., benefit they provide to the organism is unknown) have been identified. For example, many different sites in many genes' mRNAs are subject to post-transcriptional modification, but the function of the modification is only known for a small number of sites.

It is commonly believed that these traits must serve some functions that are yet to be uncovered and the observed phenotype must represent an optimal state achieved and/or maintained as a result of selection. However, a fundamental property of molecular traits is intrinsic stochasticity. Given the stochastic nature of chemical reactions and the fact that the cell is indeed "crowded" with biomolecules, reactions that offer no fitness advantage to the organism can happen merely by chance (Guin and Gruebele 2019). Even if such an event is detrimental, selection may not be able to purge them due to small effective population size (N_e) and/or physical limits to error rate (Lynch and Hagner 2015; Lynch 2018, 2020). The observed

phenotype, thus, may represent a suboptimal state reached upon a mutation-drift-selection balance. For example, proteins that interact with DNA or RNA molecules usually bind to their substrate upon recognition of specific sequences motifs. However, a sequence may possess a motif merely by chance, which can be highly likely when the motif is short, resulting in "unintended" binding. Although selection can reduce such mis-interactions by purging motifs at "wrong" locations, it may not be possible to completely get rid of them because some mis-interactions are too weak to be effectively purged by selection given N_e and/or mutations would give rise to more motifs.

To test this possibility, one can examine the association between the molecular phenotype and strength of negative selection: if a molecular phenotype tends to be observed when negative selection is weaker, it is likely nonadaptive and probably happened by chance. For some molecular traits, such as A-to-I RNA editing, one can compare patterns of editing at nonsynonymous sites and at synonymous sites, as the latter are under much weaker, if at all, negative selection. For modifications of gene products, strength of negative selection can also be approximated by the expression level of the gene that is affected: given frequency at which an undesired modification happens, the deleterious is expected to be stronger when the gene being affected is highly expressed. In interspecific comparisons, one can examine the relationship between the phenotype and N_e of different species: when N_e is large, unfit phenotypes are purged more efficiently by selection.

This nonadaptive, "molecular error" view has been supported by studies of a variety of molecular traits, including leakage transcription, alternative transcription initiation (Xu et al. 2019), alternative polyadenylation (Xu and Zhang 2018), A-to-I RNA editing (Xu and Zhang 2014), C-to-U RNA editing (Liu and Zhang 2017), N^6-methyladenosine (m^6A) (Liu and Zhang

2018), alternative RNA splicing (Saudemont et al. 2017; Xu and Zhang 2021), stop codon readthrough (Li and Zhang 2019) and protein phosphorylation (Landry et al. 2009).

1.7 Thesis overview

In this thesis, I examine phenotypic traits representing different level of biological organization. I investigate the contribution of adaptation and nonadaptive processes in these traits' evolution by analyzing publicly available genomic and phenomic data, aided by modeling and computer simulations.

In Chapter 2, I study the evolution of binding affinity between transcription factor binding sites (TFBSs) and transcription factors, which mediates evolution of gene expression. In particular, I investigate bias in a test of positive selection on TFBSs that utilizes binding affinity prediction by machine learning (Liu and Robinson-Rechavi 2020). By analyzing simulated data, I demonstrate elevated false positive rate resulting from the bias. I also present a correction method that helps such false inference of positive selection and reanalyze a published dataset to investigate if prevalence of positive selection estimated in the previous study is valid.

In Chapter 3, I study A-to-I RNA editing, a type of post-transcriptional modification of RNA molecules that can potentially alter the protein sequences. I test whether apparent signatures of adaptative RNA editing could instead be explained by entrenchment. To this end, I examine A-to-I RNA editing sites in coleoid cephalopods, where nonsynonymous A-to-I editing is abundant and formerly suggested to be adaptive. By reanalyzing published transcriptomic data, I show the pattern of editing is better explained by a nonadaptive error-permitting model rather than an adaptive model (Jiang and Zhang 2019b).

In Chapter 4, I study the role of mutational (co)variance in long-term evolution of drosophilid flies' wing morphology. I particular, I investigate how evolutionary rates of the wing morphological traits are related to their mutational variances. By analyzing published and simulated data, I show patterns that appear to be difficult to explain in the current theoretical framework of phenotypic evolution (Houle et al. 2017) are artifacts resulting from bias in the method previously used for analysis (Jiang and Zhang 2020). I propose a mechanistic model that explains the true scaling between mutational variance and evolutionary rate and demonstrate how it can produce patterns that resemble the observations using computer simulations (Jiang and Zhang 2020).

In Chapter 5, I focus on correlation between traits and ask whether mutations or selection contributed to pattern of correlation in evolution. I analyze cell morphological traits of budding yeast *Saccharomyces cerevisiae* as well as wing morphological traits of drosophilid flies and test whether correlation between pairs of traits among divergent lineages (i.e., evolutionary correlation) can be explained by correlation among mutants (i.e., mutational correlation). I identify trait pairs with evolutionary correlations strengthened, weakened, or reversed compared to mutational correlation. Using simulations, I demonstrate that a large fraction of these significant differences needs to be explained by selection on allometric relationships between traits (Jiang and Zhang 2021).

The main chapters focus on different types of phenotypic traits yet are all centered about relative contributions of adaptive and nonadaptive processes in phenotypic evolution. In Appendix A, I study parallel evolution, a phenomenon that is widely considered to reflect adaptation. Specifically, I investigate previously proposed connection between independent origination of a behavioral trait, monogamy, and parallel evolution of transcriptomes (Young et

al. 2019). I reexamine gene expression levels in brains of vertebrate species from several divergent clades using a more rigorous method. Although the observed frequency of parallel expression evolution is confirmed to be slightly higher than random expectation, transcriptomic divergence is mostly explained by phylogeny rather than mating system, which does not support transcriptome-wide parallel evolution associated with evolution of monogamy (Jiang and Zhang 2019a).

1.8 References

Arnold, S. J. and M. J. Wade. 1984a. On the Measurement of Natural and Sexual Selection: Applications. Evolution 38:720-734.

Arnold, S. J. and M. J. Wade. 1984b. On the Measurement of Natural and Sexual Selection: Theory. Evolution 38:709-719.

Bedford, T. and D. L. Hartl. 2009. Optimization of gene expression by natural selection. P Natl Acad Sci USA 106:1133-1138.

Clune, J., J. B. Mouret, and H. Lipson. 2013. The evolutionary origins of modularity. Proc Biol Sci 280:20122863.

Covello, P. S. and M. W. Gray. 1993. On the evolution of RNA editing. Trends Genet 9:265-268.

Darwin, C. R. 1859. On the Origin of Species. John Murray, Lodon.

Darwin, C. R. 1862. Fertilisation of Orchids. John Murray, Lodon.

Fabre, A. C., C. Bardua, M. Bon, J. Clavel, R. N. Felice, J. W. Streicher, J. Bonnel, E. L. Stanley, D. C. Blackburn, and A. Goswami. 2020. Metamorphosis shapes cranial diversity and rate of evolution in salamanders. Nat Ecol Evol 4:1129-1140.

Felsenstein, J. 1985. Phylogenies and the Comparative Method. Am Nat 125:1-15.

Fraser, H. B. 2020. Detecting selection with a genetic cross. Proc Natl Acad Sci U S A 117:22323-22330.

Gillespie, J. H. 2004. Population genetics : a concise guide. Johns Hopkins University Press, Baltimore, Md.

Goldstein, R. A., S. T. Pollard, S. D. Shah, and D. D. Pollock. 2015. Nonadaptive Amino Acid Convergence Rates Decrease over Time. Mol Biol Evol 32:1373-1381.

Gould, S. J. and R. C. Lewontin. 1979. Spandrels of San-Marco and the Panglossian Paradigm - a Critique of the Adaptationist Program. Proc R Soc Ser B-Bio 205:581-598.

Gray, M. W. and P. S. Covello. 1993. RNA editing in plant mitochondria and chloroplasts. Faseb J 7:64-71.

Guin, D. and M. Gruebele. 2019. Weak Chemical Interactions That Drive Protein Evolution: Crowding, Sticking, and Quinary Structure in Folding and Function. Chem Rev 119:10691-10717.

Hansen, T. F. and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J Evolution Biol 21:1201-1219.
Hansen, T. F., C. Pelabon, W. S. Armbruster, and M. L. Carlson. 2003. Evolvability and genetic constraint in Dalechampia blossoms: components of variance and measures of evolvability. J Evol Biol 16:754-766.
Harmon, L. J. 2018. Phylogenetic comparative methods: learning from trees. EcoEvoRxiv.
Hether, T. D. and P. A. Hohenlohe. 2014. Genetic Regulatory Network Motifs Constrain Adaptation through Curvature in the Landscape of Mutational (Co)Variance. Evolution 68:950-964.
Hodgins-Davis, A., F. Duveau, E. A. Walker, and P. J. Wittkopp. 2019. Empirical measures of mutational effects define neutral models of regulatory evolution in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 116:21085-21093.
Houle, D., G. H. Bolstad, K. van der Linde, and T. F. Hansen. 2017. Mutation predicts 40 million years of fly wing evolution. Nature 548:447-+.
Jiang, D. and J. Zhang. 2019a. Parallel transcriptomic changes in the origins of divergent monogamous vertebrates? Proc Natl Acad Sci U S A 116:17627-17628.
Jiang, D. and J. Zhang. 2019b. The preponderance of nonsynonymous A-to-I RNA editing in coleoids is nonadaptive. Nat Commun 10:5411.
Jiang, D. and J. Zhang. 2021. Detecting natural selection in trait-trait coevolution. BioRxiv.
Jiang, D. H. and J. Z. Zhang. 2020. Fly wing evolution explained by a neutral model with mutational pleiotropy. Evolution 74:2158-2167.
Kimura, M. 1962. On the probability of fixation of mutant genes in a population. Genetics 47:713-719.
Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217:624-626.
Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.
Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. Am Nat 157:245-261.
Lande, R. 1976. Natural-Selection and Random Genetic Drift in Phenotypic Evolution. Evolution 30:314-334.
Lande, R. 1979. Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain - Body Size Allometry. Evolution 33:402-416.
Lande, R. and S. J. Arnold. 1983. The Measurement of Selection on Correlated Characters. Evolution 37:1210-1226.
Landry, C. R., E. D. Levy, and S. W. Michnick. 2009. Weak functional constraints on phosphoproteomes. Trends in Genetics 25:193-197.
Li, C. and J. Z. Zhang. 2019. Stop-codon read-through arises largely from molecular errors and is generally nonadaptive. Plos Genet 15.
Liu, J. and M. Robinson-Rechavi. 2020. Robust inference of positive selection on regulatory sequences in the human brain. Sci Adv 6.
Liu, Z. and J. Zhang. 2017. Human C-to-U Coding RNA Editing Is Largely Nonadaptive. Mol Biol Evol 35:963-969.
Liu, Z. and J. Z. Zhang. 2018. Most m(6)A RNA Modifications in Protein-Coding Regions Are Evolutionarily Unconserved and Likely Nonfunctional. Mol Biol Evol 35:666-675.

Long, H., W. Sung, S. Kucukyildirim, E. Williams, S. F. Miller, W. Guo, C. Patterson, C. Gregory, C. Strauss, C. Stone, C. Berne, D. Kysela, W. R. Shoemaker, M. E. Muscarella, H. Luo, J. T. Lennon, Y. V. Brun, and M. Lynch. 2018. Evolutionary determinants of genome-wide nucleotide composition. Nat Ecol Evol 2:237-240.

Lukes, J., J. M. Archibald, P. J. Keeling, W. F. Doolittle, and M. W. Gray. 2011. How a neutral evolutionary ratchet can build cellular complexity. IUBMB Life 63:528-537.

Lush, J. 1937. Animal Breeding Plans. Iowa State College Press, Ames.

Lynch, M. 1990. The Rate of Morphological Evolution in Mammals from the Standpoint of the Neutral Expectation. Am Nat 136:727-741.

Lynch, M. 2018. Phylogenetic divergence of cell biological features. Elife 7.

Lynch, M. 2020. The evolutionary scaling of cellular traits imposed by the drift barrier. P Natl Acad Sci USA 117:10435-10444.

Lynch, M. and K. Hagner. 2015. Evolutionary meandering of intermolecular interactions along the drift barrier. P Natl Acad Sci USA 112:E30-E38.

Lynch, M. and W. G. Hill. 1986. Phenotypic Evolution by Neutral Mutation. Evolution 40:915-935.

Melo, D. and G. Marroig. 2015. Directional selection can drive the evolution of modularity in complex traits. P Natl Acad Sci USA 112:470-475.

Munoz-Gomez, S. A., G. Bilolikar, J. G. Wideman, and K. Geiler-Samerotte. 2021. Constructive Neutral Evolution 20 Years Later. J Mol Evol 89:172-182.

Navalon, G., J. Marugan-Lobon, J. A. Bright, C. R. Cooney, and E. J. Rayfield. 2020. The consequences of craniofacial integration for the adaptive radiations of Darwin's finches and Hawaiian honeycreepers. Nature Ecology & Evolution 4:270-+.

Ohya, Y., J. Sese, M. Yukawa, F. Sano, Y. Nakatani, T. L. Saito, A. Saka, T. Fukuda, S. Ishihara, S. Oka, G. Suzuki, M. Watanabe, A. Hirata, M. Ohtani, H. Sawai, N. Fraysse, J. P. Latge, J. M. Francois, M. Aebi, S. Tanaka, S. Muramatsu, H. Araki, K. Sonoike, S. Nogami, and S. Morishita. 2005. High-dimensional and large-scale phenotyping of yeast mutants. P Natl Acad Sci USA 102:19015-19020.

Orr, H. A. 1998. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. Genetics 149:2099-2104.

Orr, H. A. 2000. Adaptation and the cost of complexity. Evolution 54:13-20.

Pollock, D. D., G. Thiltgen, and R. A. Goldstein. 2012. Amino acid coevolution induces an evolutionary Stokes shift. Proc Natl Acad Sci U S A 109:E1352-1359.

Saudemont, B., A. Popa, J. L. Parmley, V. Rocher, C. Blugeon, A. Necsulea, E. Meyer, and L. Duret. 2017. The fitness cost of mis-splicing is the main determinant of alternative splicing patterns. Genome Biol 18.

Shah, P., D. M. McCandlish, and J. B. Plotkin. 2015. Contingency and entrenchment in protein evolution under purifying selection. Proc Natl Acad Sci U S A 112:E3226-3235.

Sloan, D. B. 2017. Nuclear and mitochondrial RNA editing systems have opposite effects on protein diversity. Biol Letters 13.

Stoltzfus, A. 1999. On the possibility of constructive neutral evolution. J Mol Evol 49:169-181.

Wagner, G. P. 1996. Homologues, natural kinds and the evolution of modularity. Am Zool 36:36-43.

Wagner, G. P. and L. Altenberg. 1996. Perspective: Complex Adaptations and the Evolution of Evolvability. Evolution 50:967-976.

Wagner, G. P., J. P. Kenney-Hunt, M. Pavlicev, J. R. Peck, D. Waxman, and J. M. Cheverud. 2008. Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature 452:470-U479.

Wagner, G. P., M. Pavlicev, and J. M. Cheverud. 2007. The road to modularity. Nature Reviews Genetics 8:921-931.

Xu, C., J. K. Park, and J. Z. Zhang. 2019. Evidence that alternative transcriptional initiation is largely nonadaptive. Plos Biol 17.

Xu, C. and J. Z. Zhang. 2018. Alternative Polyadenylation of Mammalian Transcripts Is Generally Deleterious, Not Adaptive. Cell Systems 6:734-+.

Xu, C. A. and J. Z. Zhang. 2021. Mammalian circular RNAs result largely from splicing errors. Cell Reports 36.

Xu, G. X. and J. Z. Zhang. 2014. Human coding RNA editing is generally nonadaptive. P Natl Acad Sci USA 111:3769-3774.

Young, R. L., M. H. Ferkin, N. F. Ockendon-Powell, V. N. Orr, S. M. Phelps, A. Pogany, C. L. Richards-Zawacki, K. Summers, T. Szekely, B. C. Trainor, A. O. Urrutia, G. Zachar, L. A. O'Connell, and H. A. Hofmann. 2019. Conserved transcriptomic profiles underpin monogamy across vertebrates. Proc Natl Acad Sci U S A 116:1331-1336.

Chapter 2

Evolution of Transcription Regulation: Positive Selection on Transcription Factor Binding Sites, or False Positives?

2.1 Abstract

Evolutionary divergence in gene expression levels mediated by *cis*- regulatory divergence has been speculated as an important contributor to adaptation, yet identification of adaptive *cis*-divergence remains challenging due to difficulty in finding an appropriate null model. A new way of detecting positive selection on transcription factor binding sites (TFBSs) is made possible recently thanks to the application of machine learning in predicting TF binding affinity of DNA sequences. Given TFBS sequence of a focal species and an inferred ancestral sequence that differs from the former at *n* sites, one can predict TF binding affinity for many *n*-step mutational neighbors of the ancestral sequence to obtain a null distribution of derived binding affinity. Level of significance can then be computed by comparing binding affinity of the real derived sequence to the null distribution. Such a test, however, suffers from ascertainment bias if only TFBSs identified in one focal species are studied, as these sequences, even in the absence of selection, are on average more likely to have gained binding affinity compared to the set of all TFBSs and proto-TFBSs in the ancestor. By analyzing simulated data, we show this bias indeed inflates the false positive rate of the selection test. We also show the false positive rate can be

reduced by removing extreme values from the predicted null distribution such that it reflects bias in the data. Applying the correction to human CEBPA binding sites formerly found to be under positive selection, we found about one third of these cases are likely false positives.

2.2 Introduction

Evolution of gene expression levels has been suggested to play an important role in evolution of organismal phenotypes and adaptation (King and Wilson 1975; Wray 2007). In particular, it could be divergence in gene expression that is responsible for the substantial phenotypic difference between humans and chimpanzee, as divergence in protein sequences appears to be too little (King and Wilson 1975; Enard et al. 2002; The Chimpanzee Sequencing and Analysis Consortium 2005; Franchini and Pollard 2015). Evolutionary changes in a gene's expression level can result from changes in either *cis*- or *trans*- regulatory elements, yet it is widely believed that evolutionary changes in *cis*- elements more frequently underlie in evolution of organismal traits and adaptation because *cis*- elements, such as promoters and enhancers, are usually gene-specific while *trans*- regulators, such as transcription factors (TFs), tend not to be gene-specific. As a result, mutations in these regulatory elements are generally less pleiotropic, less likely to be strongly deleterious and thus more likely to persist as standing genetic variations compared to mutations in *trans*- regulators (Wray 2007; Carroll 2008; Jones et al. 2012; Signor and Nuzhdin 2018).

Despite their hypothesized importance in adaptation, *cis*- elements under positive selection are hard to identify. Many selection tests for coding sequences, such as various versions of the d_N/d_S (aka K_a/K_s) test (Li et al. 1985; Nei and Gojobori 1986) and McDonald-Kreitman (MK) test (McDonald and Kreitman 1991), involve comparisons of nonsynonymous

and synonymous sites with synonymous sites serving as a neutral control. For non-coding sequences, sites cannot be divided into a neutral group and a non-neutral group in the same way, making it more difficult to test selection from sequence alone.

Since many *cis*- elements regulate gene expression by binding with a *trans*- factor, it is the binding affinity that matters to expression; selection on these regulatory sequences, therefore, is essentially mediated by selection on binding affinity (Berg et al. 2004; Moses 2009). In theory, one can compare observed evolutionary change in binding affinity to a null distribution (i.e., distribution of evolutionary change in binding affinity in the absence of selection) to test whether a *cis*- element is subject to selection. Such a test is possible only if binding affinities of the numerous potential mutant sequences, which cannot be directly measured, are known. In a recent study, this problem was solved by using machine learning to predict TF binding affinity (Liu and Robinson-Rechavi 2020). Using previously established techniques (Ghandi et al. 2014; Lee et al. 2015; Ghandi et al. 2016), Liu and Robinson-Rechavi trained a gapped *k*-mer support vector machine (gkm-SVM) with TFBSs that had been identified (i.e., ChIP-seq peaks) as the training set. This trained program was used it to calculate SVM weights (i.e., contribution to overall binding affinity of TFBS) of all possible 10-mers, based on which a binding affinity score (SVM score) can be calculated for any arbitrary sequence by summing up SVM weights of all 10-mers it contains. With this tool, one can calculate affinity scores of many mutational neighbors of an ancestral sequence to obtain a null distribution of binding affinity change and compare the observed evolutionary change (deltaSVM) to this null distribution. A significant difference (i.e., the observed deltaSVM is in the right 5% or 1% tail of the null distribution) means there has been lineage-specific accelerated evolution of binding affinity and indicates positive selection on the TFBS (Prabhakar et al. 2006; Gittelman et al. 2015). Signal of selection

would further help distinguish TFBSs that are functional (i.e., provide fitness advantage) and genomic regions that merely acquired some binding motif by chance.

However, this test can suffer from ascertainment bias if sequences with high derived binding affinity (i.e., high binding affinity in the focal lineage) are preferentially included in the dataset. Even in the absence of selection, sequences with high derived affinity are more likely to have undergone increase of binding affinity than those with lower binding affinity. By focusing on a subset of sequences with high derived affinity, it becomes more likely that chance gain of binding affinity by chance is mis-identified as a result of positive selection, leading to elevated false positive rate. Therefore, if the dataset only includes TFBSs identified in one focal species, test for positive selection in the focal species' lineage could be subject to increased false positive rate because of this ascertainment bias. An analogy is testing if a coin is fair by tossing for many times and comparing the number of heads and tails. If, for some reason, heads are more likely to be recorded, there would naturally be more heads than tails in the recorded results, leading the conclusion that the coin is unfair even if it is actually a fair one.

Here we confirm, using sequences generated in simulations of neutral evolution, that ascertainment bias can indeed increase false positive rate in the test for positive selection. We also demonstrate how this problem could be mitigated with a simple method, which is to adjust the null distribution to have it reflect bias in the dataset. We applied this corrected method to binding sites of a human liver TF, CEBPA, which were previously analyzed by Liu and Robinson-Rechavi and show that about one third of TFBSs inferred as under positive selection are likely false positives, yet there are still TFBSs that show significant signal of positive selection after the correction is done, indicating line-specific positive selection did cause gain of binding affinity at many loci in human.

2.3 Results

2.3.1 Ascertainment bias leads to elevated false positive rate

To illustrate how ascertainment bias would affect test for positive selection, we simulated a dataset consisting of 10,000 pairs of ancestral TFBSs and introduced random substitutions to each of them to obtain 10,000 derived TFBSs. As substitutions introduced are all random, there is no true positive selection, and any signal of positive selection detected in the subsequent test must be false positive. For each sequence, we calculated the SVM score for binding with human's CEBPA. We confirmed derived and ancestral SVM scores of this dataset is not significantly different ($P = 0.9139$, paired Wilcoxon rank-sum test), as expected under neutrality. In contrast, the previously identified human CEBPA binding sites tend to have derived SVM scores higher than the corresponding ancestral SVM scores ($P = 4.47 \times 10^{-131}$, paired Wilcoxon rank-sum test). We conducted one-tailed test of selection for high binding affinity, in the same way as Liu and Robinson-Rechavi. We found 98 out of these 10,000 TFBSs (0.98%) showed signal of positive selection for increased binding affinity with human CEPBA ($P < 0.01$; Table 2-1), which is also consistent with the neutral expectation.

To better illustrate the effect of ascertainment bias on false positive rate, we examined different subgroups of simulated TFBSs and investigated trends associated with false positive rate. To this end, we used several different cutoffs to pick smaller subsets of simulated TFBSs with high derived affinity. As shown in Fig. 2-1, for the cutoffs examined, the false positive rate increased with the SVM score cutoff and there is a positive correlation ($\rho = 0.891$, $P < 10^{-300}$). We then shuffled *P*-values of the simulated TFBSs such that correspondence between SVM scores and *P*-values are broken and found the aforementioned correlation calculated from

500 shuffled datasets never exceeded that from unshuffled data ($P < 0.002$), confirming the correlation between the SVM score cutoff and false positive rate indeed results from association between derived SVM score and false positive rate.

We also considered a scenario where identifiability of TFBSs is not determined only by a single cutoff. That is, identification of TFBSs is subject to randomness and the probability that a genomic region is identified as a TFBS is a function of its binding affinity. To this end, we devised a model for TFBS identifiability and adjusted model parameters to simulate TFBS identification processes of different level of sensitivity and thus different level of bias (see Materials and Methods). Briefly, under this model, probability that a TFBS is successfully identified is zero when its binding affinity is below a minimum value for identifiability; probability of successful identification increases with binding affinity and asymptotically approaches one. We computed probability of identification for each simulated TFBS under different combinations of model parameters and ran in silico identification experiments to obtain subsets of "identified" TFBSs. As shown in Fig. 2-2a, for such sets of TFBSs, false positive rate is significantly correlated with median derived affinity of the set ($r = 0.943, P = 9.45 \times 10^{-211}$). Therefore, it is highly likely that previously analyzed datasets are subject to ascertainment bias and inference of selection is likely to have been affected.

2.3.2 A simple correction to reduce false positive rate

To control for the aforementioned bias, we propose a simple correction method. The main idea of this correction is to adjust the null distribution generated from simulations such that the null distribution is subject to the same level of bias as real data. By removing a proportion of sequences with the low binding affinity from the control group, one can have distribution of

deltaSVM of the remaining sequences as asymmetric as identified TFBSs. The *P*-value is then calculated as the proportion of sequences with higher derived affinity than observed among the remaining sequences of the control group. When applied to subsets of simulated TFBSs picked on bases of fixed cutoffs, both the pre-correction false positive rate and the post-correction rate increased with the cutoff, but the latter increased more slowly (Fig. 2-1). For subsets picked in the in silico identification experiments, on the other hand, the false positive rate is not significantly correlated with the median SVM score of derived sequences (Fig. 2-2b, $r = 0.0865, P = 0.0699$).

We re-analyzed the set of human CEBPA binding sites analyzed by Liu and Robinson-Rechavi previously. As these TFBSs were identified only in the focal species, this dataset is inevitably subject to ascertainment bias. Indeed, deltaSVM of this dataset is positively biased and 62.15% of these TFBSs have positive deltaSVM. We re-examined the 1,135 TFBSs in this dataset that were previously found to have $P < 0.05$, including those inferred to be under positive selection ($P < 0.01$) and estimated the proportion of false positives before and after correction. We chose this set because TFBSs that would have $P < 0.01$ in a reanalysis should have rather low *P*-values in the previous analysis as well, while those found to have $P > 0.05$ in the previous study are unlikely to have $P < 0.01$ in a reanalysis. Hence, focusing on this set would allow a reasonably accurate estimate of the false positive rate.

Before the correction is applied, 361 of these TFBSs showed signal of positive selection ($P < 0.01$). After applying the correction described in the previous section, of these TFBSs no longer show significant signal of positive selection in a one-tailed test, yet the rest remained significant (Fig. 2-3, Table 2-1). The proportion of TFBSs inferred to be under positive selection out of all TFBSs in the dataset, is reduced from 6.22% to 5.08%. The fact that the latter is still

greater than the *P*-value cutoff of 0.01 indicates positive selection has indeed played a role in gain of binding affinity in human lineage.

2.4 Discussion

In this study, we demonstrate detection of positive selection on TFBSs can suffer from elevated false positive rate caused by ascertainment bias. Such bias arises from the way in which TFBS are identified in the first place. Genome regions with high binding affinity to a TF is more like to be identified as the TF's binding site and included in the dataset, yet it is also selection on binding affinity that is of interest. One way to avoid such ascertainment bias is to include TFBSs identified in more species or lineages. For instance, mouse TFBSs identified by Stefflova *et al*. (Stefflova et al. 2013) and analyzed by Liu and Robinson-Rechavi consisted of TFBSs identified in different lineages and the distribution of deltaSVM in the focal lineage is indeed more symmetric than that of human TFBSs analyzed in the same study (Liu and Robinson-Rechavi 2020).

Although our simulated data help illustrate the bias and its effect on tests of selection, it does have limitations. Since this simulated dataset consists of a total of 10,000 TFBSs, the number of those showing false signals of positive selection, naturally, would be around 50. The absolute number of false positives is arguably too small. As a result, it is challenging to directly compare false positive rates of different subsets of simulated TFBSs because the sample size is so small that a statistical tests for difference in false positive rate has little power. Hence, it is desirable to have further increase sample size and have more simulated sequences included. For results presented in this thesis, the number of simulated TFBSs that can be included is limited by time, as calculation of SVM scores, especially those of control sequences used in the selection

test, is time-consuming. The sample size is expected to increase as more simulated sequences are included for analyses.

It is also desirable to consider a wider range of substitution numbers when generating the simulated dataset. Having the number of substitutions limited in a fixed range could have limited the absolute number of false positives in the simulated data and thus power of statistical test for different false positive rates. In addition, if it is evolutionary divergence over a longer time interval that is under concern, there would be more sequence divergence and more difference between the ancestral and derived sequences. Hence, having a broader range of substitution numbers included for analysis could be helpful to future applications of the selection test in other taxa.

Given that the dataset consists of only TFBS identified in the focal species, a correction can be applied to reduce false positive rate. The correction is done by removing mutant sequences with low binding affinity from the control set such that the null distribution of deltaSVM as positively biased as real data. Extent to which the null distribution is adjustment depends on how biased real data's deltaSVM distribution is. If the distribution is only weakly biased, adjustment to the null distribution is also minimal. Thus, our correction method, in principle, is not expected to "over-correct" when the dataset is only subject to rather weak ascertainment bias.

While it is clear the test of selection on TFBS examined in this study is subject to false positives, it is less well understood how sensitive the test is when positive selection indeed happened. Thus, it would be interesting to examine the test's performance using data simulated under a model with positive selection and how the correction process could affect its power. The key parameter of this selection model would be the fitness function, which describes how exactly

phenotype (i.e., binding affinity, or ultimately, expression level of the target gene) and fitness are related; given the fitness function, a mutation's effect on binding affinity can be readily converted to its fitness effect and fixation probability. By randomly generating mutations and fixing some of them until the total number of substitutions reaches a desired number, a derived sequence can be generated, and the test for selection can be applied.

It is worth noting the selection test examined in this study is reliable only if the reconstructed ancestral sequences are reasonably accurate. If the ancestral sequence is incorrect, the resulting null distribution of deltaSVM would be wrong as well, rendering analyses unmeaningful. While it is known that ancestral sequence reconstruction is subject to errors (Zhang and Nei 1997), this problem is probably not severe for data examined by Liu and Robinson-Rechavi and in this study, as only closely related species are considered.

In addition to test for positive selection, TF binding affinity predicted by machine learning may also be used to study other questions about evolution. For example, theoretical models have been developed for evolution of inter-molecular interactions and made predictions about expected patterns of variation across lineages given values of key parameters (Lynch and Hagner 2015). These predictions, however, are derived under specific assumptions about the genotype-phenotype map and the distribution of mutations' phenotypic effects, which is usually unclear. With a prediction model established via machine learning, effect of each mutation can be known, making it possible to test the model's predictions under a realistic setting.

2.5 Materials and Methods

2.5.1 Generation of simulated sequences

We simulated 10,000 random sequences that represent ancestral sequences and introduced random substitutions into each of them to generate their corresponding derived sequences. To make our results as comparable as possible to those from analysis of real data, we used nucleotide frequencies and mutational spectrum estimated from human CEBPA binding sites. These TFBSs were identified by Schmidt *et al*. (Schmidt et al. 2010) and ancestral sequences were inferred by Liu and Robinson-Rechavi (Liu and Robinson-Rechavi 2020). When simulating an ancestral sequence, probabilities of each base at a site was equal to its average frequencies among all ancestral sequences of the human CEBPA binding sites. Similarly, when a mutation would be introduced at a site, probability of mutating to each target base was equal to frequency of this type of mutation out of all mutations with the same starting base in the human CEBPA binding sites.

Specifically, each pair of ancestral and derived sequences were generated in the following steps: 1) pick a random TFBS from the set of human CEBPA binding sites analyzed previously by Liu and Robinson-Rechavi; 2) simulate an ancestral sequence with length equal to the TFBS picked in the previous step; 3) randomly choose n sites in the sequence, where n is the number of differences between ancestral and derived sequences chosen in the first step; 4) introduce a random mutation at each of the n sites of the ancestral sequence chosen in the previous step to generate a derived sequence. As this process involves no selection on TF binding affinity, difference in binding affinity between the ancestral and derived sequences only represents neutral evolutionary change, and signal of positive selection, if detected, can only be false positive.

The number of substitutions, n, is at least two, as Liu and Robinson-Rechavi only conducted the test for selection when ancestral and derived sequences differ at two or more sites.

Therefore, the number of differences between ancestral and derived sequences in our simulated dataset was also at least two. We did not consider multiple hits, so the number of substitutions is equal to the number of differences.

Human CEBPA binding affinity was calculated for all simulated sequence using the corresponding prediction model presented by Liu and Robinson-Rechavi. Briefly, the SVM score for binding affinity was calculated for a sequence by adding up weights of all 10-nucleotide windows within this sequence. Thus, for a sequence of length l, its SVM is a sum of weights of $l - 9$ 10-mers. Datafile containing weight of each 10-mer is available at https://raw.githubusercontent.com/ljljolinq1010/A-robust-method-for-detecting-positive-selection-on-regulatory-sequences/master/data/human_SVM_model/CEBPA/kmer_10_library_weigths.txt. For 10-mers that are not shown in this datafile, weight of its reverse complementary sequence was used. For each simulated sequence pair, deltaSVM was calculated by subtracting the ancestral sequence's SVM from the derived sequence's SVM.

We chose CEBPA as an example and focused on human CEBPA binding affinity when analyzing the simulated dataset. However, the same procedure can be readily applied to any other TF.

2.5.2 Testing positive selection on TFBSs

For each pair of simulated ancestral and derived sequences, we first conducted the test for positive selection with the previously used method (Liu and Robinson-Rechavi 2020). For a pair of ancestral sequence and a derived sequence differing from the ancestral one at *n* sites, we simulated a set of 10,000 random mutant sequences as the control group with each of these

mutants differ from the ancestral sequences at *n* randomly chosen sites. When analyzing the simulated TFBSs, we generated the control set using the mutational spectrum estimated from human CEBPA binding sites. When analyzing real data, we did not consider mutational bias. Liu and Robinson-Rechavi's previous work did not consider it for most analyses, though it was suggested that the method is rather robust to mutational bias. A one-tailed *P*-value was then calculated as the number of sequences in the control set with SVM score greater than the derived sequence divided by the total number of sequences in the control set. Significant signal of positive selection was be inferred if $P < 0.01$, as the same cutoff was used by Liu and Robinson-Rechavi.

In this study, in order to save time, the number of sequences in the control set for each TFBS was set to be 2,000 instead of 10,000 as in Liu and Robinson-Rechavi's study. To confirm whether this sample size is sufficient, we examined 200 simulated TFBSs. From the set of mutants derived from each TFBS, we randomly sampled 2,000 mutant sequences and compared *P*-values computed from 10,000 and 2,000 mutants. The sampling was repeated for 100 times and correlation between *P*-values computed from 10,000 and 2,000 mutants were computed each time. The correlation lowest coefficient is 0.9994 and the median is 0.9996, indicating 2,000 mutants are sufficient for accurately calculating the *P*-value.

2.5.3 Modeling identifiability of TFBSs

We modeled probability at a which a sequence is identified as a TFBS, P_I, as a function of binding affinity, a, which is quantified by the SVM score in this study:

$$P_I(a) = 1 - \exp[C(a_{min} - a)]$$

, where a_{min} is the minimum affinity required for a sequence to be identified as a TFBS and C is a positive number. When $a = a_{min}$, probability at which the sequence is identified as a TFBS is zero; as a increases, P_I asymptotically approaches one. C, on the other hand, influences the rate of which P_I increases with a, with greater C corresponding to more rapid increase. Different combinations of a_{min} and C represent different levels of sensitivity in TFBS identification.

Values of a_{min} examined in this study were -100, -95, -90, -85, -80, -75, -70, -65, -60, -55 and -50 and values of C considered were 0.05, 0.1, 0.5 and 1, resulting in 44 combinations. For each combination, we ran 10 independent in silico identification experiments. False positive rates before and after correction were calculated for each subset and plotted against their median derived affinities in Fig. 2-2.

2.5.4 Corrected test of positive selection

The correction is applicable if a set of TFBSs has an asymmetric distribution of deltaSVM. In this study, we focused on situations where more than 50% of deltaSVMs are positive. For each TFBS, if the proportion of sequences with positive deltaSVM in the control set is lower than the proportion of TFBSs with positive deltaSVMs in the whole dataset, a fraction of sequences with low SVM score would be removed from the control set, such that the distribution of deltaSVM of the control set is as positively biased as the entire dataset's distribution. This process is done in the following steps: 1) remove the sequence with the lowest SVM from the control set; 2) calculate the proportion of deltaSVMs that are positive in remaining sequences; 3) if the proportion is smaller than that of the entire set of TFBSs, go back

to step 1) and repeat this process with the remaining sequences; otherwise, end the process and calculate *P*-value with the remaining sequences as described in the previous section.

This corrected test is also done for human CEBPA binding sites found to be under positive selection by Liu and Robinson-Rechavi. As 62.15% out of all CEBPA binding sites included in analyses of Liu and Robinson-Rechavi (i.e., regardless of whether there is signal of positive selection) have positive deltaSVMs, null distribution of each TFBS was adjusted as described above to have 62.15% of deltaSVMs that are positive. Ancestral sequences used in the tests were inferred by Liu and Robinson-Rechavi using sequences of human (*Homo sapiens*) chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla gorilla*) and represent sequence states in a common ancestor of human and chimpanzee that had split from the lineage of gorilla. Difference between the ancestral sequence and the derived sequence, therefore, resulted from substitutions in human's lineage after human and chimpanzee split, which took place 5.5 million years ago (Glazko and Nei 2003).

All simulations and analyses were conducted in an R environment (R Core Development Team 2010).

2.6 References

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

Carroll, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134:25-36.

Enard, W., P. Khaitovich, J. Klose, S. Zollner, F. Heissig, P. Giavalisco, K. Nieselt-Struwe, E. Muchmore, A. Varki, R. Ravid, G. M. Doxiadis, R. E. Bontrop, and S. Paabo. 2002. Intra- and interspecific variation in primate gene expression patterns. Science 296:340-343.

Franchini, L. F. and K. S. Pollard. 2015. Can a few non-coding mutations make a human brain? Bioessays 37:1054-1061.

Ghandi, M., D. Lee, M. Mohammad-Noori, and M. A. Beer. 2014. Enhanced regulatory sequence prediction using gapped k-mer features. Plos Comput Biol 10:e1003711.
Ghandi, M., M. Mohammad-Noori, N. Ghareghani, D. Lee, L. Garraway, and M. A. Beer. 2016. gkmSVM: an R package for gapped-kmer SVM. Bioinformatics 32:2205-2207.
Gittelman, R. M., E. Hun, F. Ay, J. Madeoy, L. Pennacchio, W. S. Noble, R. D. Hawkins, and J. M. Akey. 2015. Comprehensive identification and analysis of human accelerated regulatory DNA. Genome Res 25:1245-1255.
Glazko, G. V. and M. Nei. 2003. Estimation of divergence times for major lineages of primate species. Mol Biol Evol 20:424-434.
Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, P. Broad Institute Genome Sequencing, T. Whole Genome Assembly, J. Baldwin, T. Bloom, D. B. Jaffe, R. Nicol, J. Wilkinson, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55-61.
King, M. C. and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107-116.
Lee, D., D. U. Gorkin, M. Baker, B. J. Strober, A. L. Asoni, A. S. McCallion, and M. A. Beer. 2015. A method to predict the impact of regulatory variants from DNA sequence. Nat Genet 47:955-961.
Li, W. H., C. I. Wu, and C. C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150-174.
Liu, J. and M. Robinson-Rechavi. 2020. Robust inference of positive selection on regulatory sequences in the human brain. Sci Adv 6.
Lynch, M. and K. Hagner. 2015. Evolutionary meandering of intermolecular interactions along the drift barrier. P Natl Acad Sci USA 112:E30-E38.
McDonald, J. H. and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652-654.
Moses, A. M. 2009. Statistical tests for natural selection on regulatory regions based on the strength of transcription factor binding sites. BMC Evol Biol 9:286.
Nei, M. and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426.
Prabhakar, S., J. P. Noonan, S. Paabo, and E. M. Rubin. 2006. Accelerated evolution of conserved noncoding sequences in humans. Science 314:786.
R Core Development Team. 2010. R: A language and environment for statistical computing.
Schmidt, D., M. D. Wilson, B. Ballester, P. C. Schwalie, G. D. Brown, A. Marshall, C. Kutter, S. Watt, C. P. Martinez-Jimenez, S. Mackay, I. Talianidis, P. Flicek, and D. T. Odom. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science 328:1036-1040.
Signor, S. A. and S. V. Nuzhdin. 2018. The Evolution of Gene Expression in cis and trans. Trends Genet 34:532-544.
Stefflova, K., D. Thybert, M. D. Wilson, I. Streeter, J. Aleksic, P. Karagianni, A. Brazma, D. J. Adams, I. Talianidis, J. C. Marioni, P. Flicek, and D. T. Odom. 2013. Cooperativity and

rapid evolution of cobound transcription factors in closely related mammals. Cell 154:530-540.
The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69-87.
Wray, G. A. 2007. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8:206-216.
Zhang, J. and M. Nei. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J Mol Evol 44 Suppl 1:S139-146.

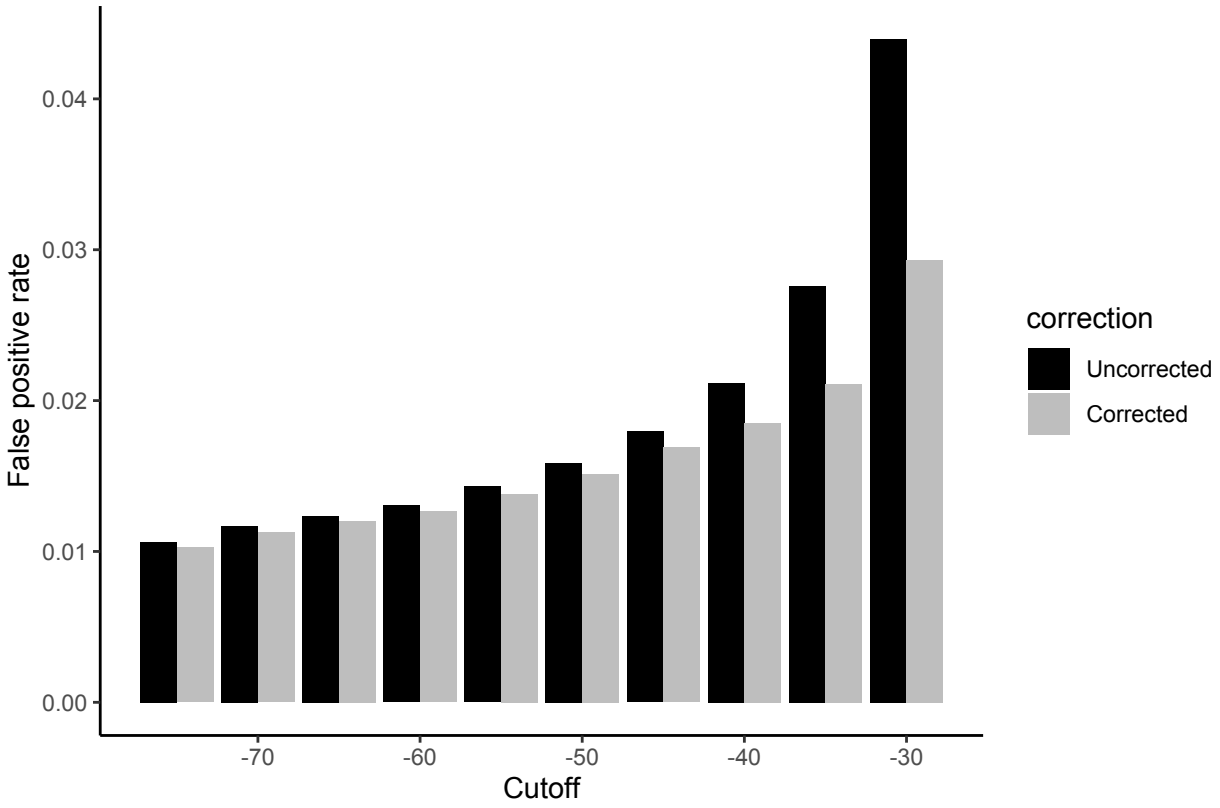


Figure. 2-1. False positive rates at different SVM score cutoffs before and after correction.

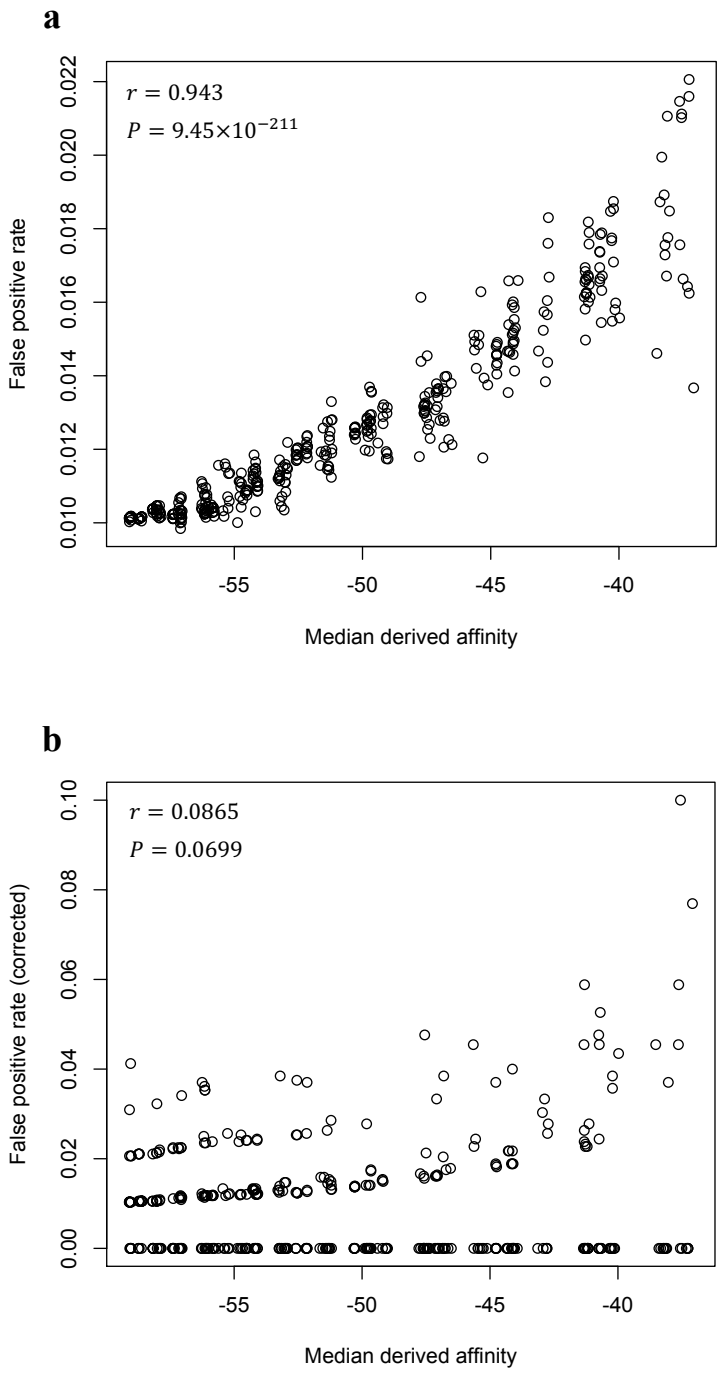


Figure. 2-2. Correlation between median derived affinity and false positive rate for subsets of simulated TFBSs identified in in silico identification experiments (see Materials and Methods). Each dot represents a subset of simulated TFBSs identified in silico. For each combination of parameter values there are ten sets of TFBS. (a) False positive rate before correction. (b) False positive rate after correction.

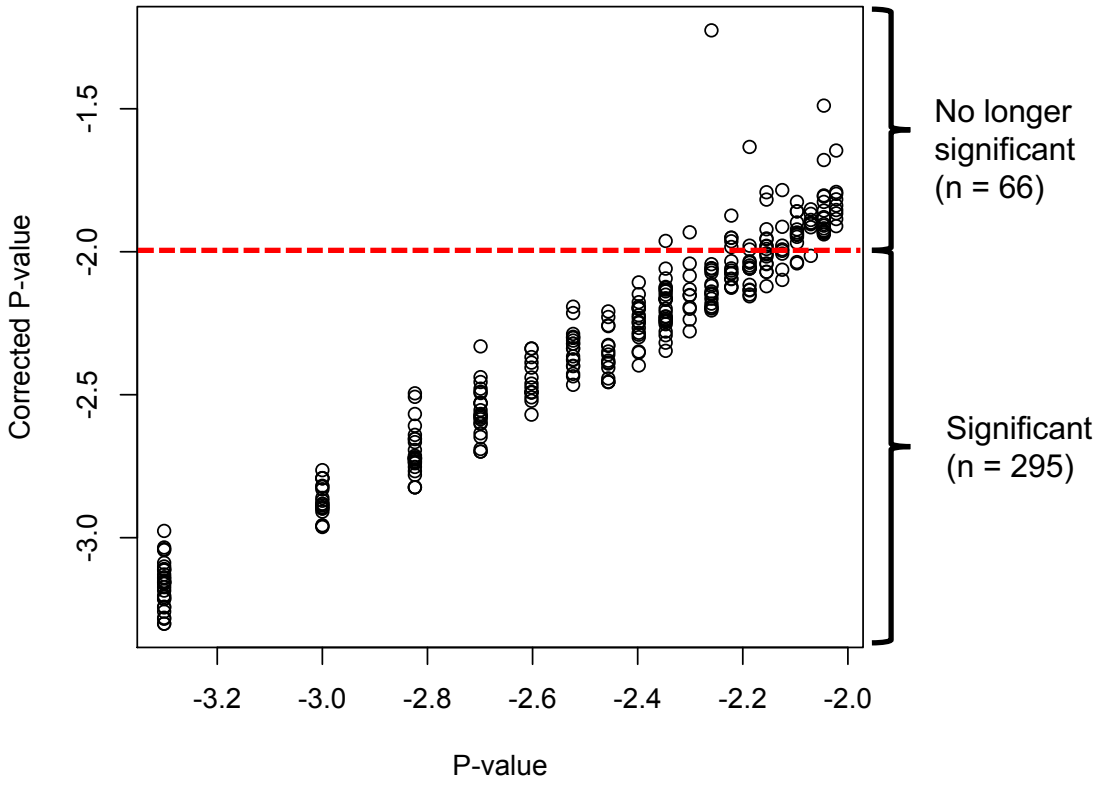


Figure. 2-3. *P*-values before and after correction for human CEBPA binding sites. Both axes are in log10 scale. Location of the red dashed line corresponds to $P = 0.01$ after correction, such that dots above the line are TFBSs identified as under positive selection before the correction but not any more after the correction.

Table. 2-1. Overview of simulated and real data.

	Simulated TFBSs	**Human CEBPA binding sites**
Number	10000	5,807
Median ancestral SVM score	-60.201	-43.162
Median derived SVM score	-60.051	-41.787
Median substitution number	2	2
Median deltaSVM	0.0228	0.784
Number (proportion) of positive deltaSVMs	5042 (50.42%)	3618 (62.30%)
Number (proportion) of positive results (before correction)	98 (0.98%)	361 (6.22%)
Number (proportion) of positive results (after correction)	NA	295 (5.08%)

Chapter 3

Evolution of Post-Transcriptional RNA Modifications: the Preponderance of Nonsynonymous A-to-I RNA Editing in Coleoids is Nonadaptive

3.1 Abstract

A-to-I editing enzymatically converts the base adenosine (A) in RNA molecules to inosine (I), which is recognized as guanine (G) in translation. Exceptionally abundant A-to-I editing was recently discovered in the neural tissues of coleoids (octopuses, squids, and cuttlefishes), with a greater fraction of nonsynonymous sites than synonymous sites subject to high levels of editing. Although this phenomenon is thought to indicate widespread adaptive editing, its potential advantage is unknown. Here we propose an alternative, nonadaptive explanation. Specifically, increasing the cellular editing activity permits some otherwise harmful G-to-A nonsynonymous substitutions because the As are edited to Is at sufficiently high levels. These high editing levels are constrained upon substitutions, resulting in the predominance of nonsynonymous editing at highly edited sites. Our evidence for this explanation suggests that

Works of this chapter have been published:
Jiang, D. and J. Zhang. 2019. The preponderance of nonsynonymous A-to-I RNA editing in coleoids is nonadaptive. Nat Commun 10:5411.

the prevalent nonsynonymous editing in coleoids is generally nonadaptive, as in species with much lower editing activities.

3.2 Introduction

RNA editing refers to a variety of posttranscriptional alterations of RNA molecules, including chemical modifications as well as insertions and deletions of nucleotides, but excluding RNA processing events such as splicing, capping, and polyadenylation (Gott and Emeson 2000; Farajollahi and Maas 2010). Transcriptome-wide profiling of each type of RNA editing and understanding its biochemical and physiological functions are a major task of molecular and genome biology and have seen rapid progress in the last decade (Gilbert et al. 2016; Li et al. 2016; Harcourt et al. 2017; Roundtree et al. 2017). Among over 100 different types of RNA editing, adenosine (A)-to-inosine (I) editing of RNAs transcribed from animal nuclear genomes is arguably best studied(Nishikura 2016; Yablonovitch et al. 2017; Eisenberg and Levanon 2018). The A-to-I conversion is catalyzed by a family of adenosine deaminase acting on RNA (ADAR), and the resultant I is recognized as guanine (G) in translation. For simplicity, we refer to A-to-I editing as A-to-G editing hereafter. If the editing takes place in protein-coding regions, it could be either nonsynonymous (also known as recoding) or synonymous, depending on whether the encoded amino acid is altered or not. A-to-G editing has been reported in multiple animal phyla (Porath et al. 2017a; Hung et al. 2018), such as many vertebrates (Li et al. 2009; Bahn et al. 2012; Danecek et al. 2012; Peng et al. 2012; Bazak et al. 2014; Chen et al. 2014; Porath et al. 2017a), as well as fruit flies (Graveley et al. 2011; Rodriguez et al. 2012; Ramaswami et al. 2013; Yu et al. 2016; Duan et al. 2017; Zhang et al. 2017), cephalopods (Garrett and Rosenthal 2012; Albertin et al. 2015; Alon et al. 2015;

Liscovitch-Brauer et al. 2017), nematodes (Zhao et al. 2015; Deffit et al. 2017), and cnidarians (Porath et al. 2017b). While an editing mechanism could emerge by chance and become fixed by genetic drift (Covello and Gray 1993), studies of functional consequences of a handful of A-to-G recoding events led to the initial belief that recoding offers an "extreme advantage" (Nishikura 2010), because disrupting recoding could be lethal (Maas et al. 2006). This view has been challenged in the last few years by transcriptome-wide analysis of RNA editing. Specifically, there is a long tradition in molecular evolutionary genetics to compare the rate of synonymous nucleotide substitution (d_S) with that of nonsynonymous substitution (d_N) in protein-coding DNA sequence evolution. Because synonymous changes are presumably neutral, while nonsynonymous changes may or may not be neutral, an observation of $d_N > d_S$ indicates overall positive selection promoting beneficial nonsynonymous substitutions, whereas $d_N < d_S$ indicates overall purifying selection hindering deleterious nonsynonymous substitutions. Although RNA editing is a molecular phenotype, similar comparisons between synonymous and nonsynonymous editing can be made (Xu and Zhang 2014). For instance, in humans, the fraction of sites subject to nonsynonymous editing is lower than that subject to synonymous editing, and the editing level (i.e., the proportion of RNA molecules edited at a site) is also lower for nonsynonymous than synonymous editing (Xu and Zhang 2014). These patterns suggest that nonsynonymous editing is generally deleterious and selectively removed and/or suppressed when compared with synonymous editing, which is presumably inconsequential to protein function. Therefore, most A-to-G coding RNA editing events appear to be nonadaptive and are probably attributable to cellular errors resulting from ADARs' limited specificity (Xu and Zhang 2014). This conclusion is compatible with the fact that only a handful of editing events have known functions (Maas et al. 2006) and that only 1.8% of ~2000 human coding RNA editing events are

shared with mouse (Pinto et al. 2014; Xu and Zhang 2015).

The trend, however, is drastically different in coleoid cephalopods, which include octopuses, squids, and cuttlefishes. Tens of thousands of coding A-to-G editing events, including a considerable proportion of recoding, have been identified in the neural tissues of coleoids (Alon et al. 2015; Liscovitch-Brauer et al. 2017). In particular, the frequency of nonsynonymous sites subject to high levels of editing exceeds that of synonymous sites, leading to the inference that nonsynonymous editing has been promoted by positive selection and is generally advantageous in coleoids(Alon et al. 2015; Liscovitch-Brauer et al. 2017). We will refer to this hypothesis as the adaptive hypothesis. Furthermore, because the high editing activity appears to be limited to their neural tissues, it was speculated that the extraordinary abundance of RNA editing in coleoids is related to their complex nervous system and behavior (Rosenthal and Seeburg 2012; Albertin et al. 2015; Alon et al. 2015; Liscovitch-Brauer et al. 2017). Nonetheless, with the exception of recoding of an octopus potassium channel that is associated with cold adaptation (Garrett and Rosenthal 2012), no benefit of the widespread editing is known in coleoids. Here we propose and provide evidence for an alternative, nonadaptive explanation of the preponderance of highly edited nonsynonymous sites in coleoids.

3.3 Results

3.3.1 A nonadaptive hypothesis and its predictions

Let us consider a genomic position in a coding region that is currently occupied by G and does not accept A (see top row in Fig. 3-1a). As the editing activity in the species rises, a G-to-A mutation at the site may become neutral and fixed if the resultant A is edited back to G in a sufficiently large proportion of mRNA molecules (see middle row in Fig. 3-1a). Upon the G-to-

A substitution, the high editing level at the site will be selectively maintained because it is G rather than A that is permissible at the mRNA level. Because the above situation applies only to nonsynonymous G-to-A substitutions and the coupled nonsynonymous A-to-G editing, it inflates the number of nonsynonymous editing sites and nonsynonymous editing levels relative to the corresponding synonymous values. Although here the nonsynonymous editing has permitted the fixation of the otherwise deleterious G-to-A mutation, the derived genotype with a genomic A that is highly edited is no fitter than the original genotype with a genomic G. Thus, the editing is nonadaptive. We assumed in the above scenario that the editing level is so high that the otherwise deleterious G-to-A mutation becomes neutral. It is also possible that the editing level is not high enough, rendering the G-to-A mutation slightly deleterious (see bottom row in Fig. 3-1a). A slightly deleterious mutation may nevertheless get fixed, and the editing level may be selectively increased in subsequent evolution. Even under this scenario, there is no net fitness gain from the original genotype with a genomic G to the derived genotype with a genomic A that is highly edited. We refer to the above nonadaptive model including both of the described scenarios as the harm-permitting model, because RNA editing permits the fixation of otherwise harmful mutations. While the possibility of harm-permitting by RNA editing has been proposed multiple times (Covello and Gray 1993; Stoltzfus 1999; Tian et al. 2008; Klinger et al. 2018), especially regarding the editing of organelle transcriptomes, empirical evidence that it is entirely or primarily responsible for creating "adaptive signals" of RNA editing is lacking.

Given the exceptionally high editing activity in coleoid neural tissues(Alon et al. 2015; Liscovitch-Brauer et al. 2017), we hypothesize that the reported preponderance of nonsynonymous editing is explained by the harm-permitting model and is nonadaptive. To test this hypothesis, we divide nonsynonymous editing into two categories: restorative and

diversifying (Sloan 2017). Restorative editing converts the amino acid state back to an ancestral state (Fig. 3-1b), whereas diversifying editing converts the amino acid state to a non-ancestral state (Fig. 3-1c). Because restorative editing but not diversifying editing can confer a harm-permitting effect, our hypothesis predicts that the reported preponderance of nonsynonymous editing in coleoids is attributable to restorative but not diversifying editing. In particular, we predict that (i) the frequency of sites edited is greater for restorative (F_R) than synonymous (F_S) editing and that (ii) the median editing level is higher for restorative (L_R) than synonymous (L_S) editing. It further predicts that (iii) the frequency of sites edited is no greater for diversifying (F_D) than synonymous (F_S) editing and that (iv) the median editing level is no higher for diversifying (L_D) than synonymous (L_S) editing. By contrast, the adaptive hypothesis does not have specific predictions about F_R and L_R, but predicts that F_D and L_D are respectively greater than F_S and L_S. Note that while only restorative editing can be harm-permitting, not all restorative editing is necessarily harm-permitting. For instance, the restorative editing would be neutral if it restores a neutral G-to-A substitution.

3.3.2 Patterns of restorative and diversifying editing

To test the nonadaptive hypothesis, we analyzed the published neural transcriptomes of six mollusk species (Liscovitch-Brauer et al. 2017), whose phylogenetic relationships are depicted in Fig. 3-2a. Among them, the four coleoids have widespread coding A-to-G editing in neural tissues, while the two outgroups have substantially fewer editing sites(Liscovitch-Brauer et al. 2017).

We identified 3979 one-to-one orthologous genes in these six species and inferred ancestral coding sequences at all interior nodes of the species tree (Fig. 3-2a). We regarded a

nonsynonymous editing event in an exterior node of the tree that modifies the amino acid state from X to Y as restorative if the inferred genomic sequence-based amino acid state is Y at any node of the tree that is ancestral to the focal exterior node (Fig. 3-1b; also see Methods), or diversifying if Y is not present at any node of the tree that is ancestral to the focal exterior node (Fig. 3-1c). Note that these definitions are based on amino acid states and are applied to nonsynonymous editing only. Synonymous editing is presumably neutral so need not be separated into restorative and diversifying editing. Furthermore, separating synonymous editing into the two categories would be less accurate because of lower reliabilities in inferring ancestral sequences at synonymous sites. Of the two categories of nonsynonymous editing sites, the number of diversifying editing sites is 15.7–20.4 times that of restorative editing sites in the four coleoids (Table B-1).

In each of the four coleoids, F_R is significantly lower than F_S when all editing sites are considered (Fig. 3-2b). However, because the harm-permitting hypothesis concerns relatively highly edited sites, we analyzed sites with editing levels exceeding 10% and found F_R to be significantly higher than F_S (Fig. 3-2b inset). We observed that L_R is significantly greater than L_S in all four coleoids (Fig. 3-2c). By contrast, F_D is significantly smaller than F_S when we considered all editing sites (Fig. 3-2b) or sites with >10% editing levels (Fig. 3-2b inset). L_D is not significantly different from L_S except in the squid (Fig. 3-2c). These results largely confirm all four predictions of the nonadaptive hypothesis and are at odds with the predictions of the adaptive hypothesis, strongly suggesting that the preponderance of nonsynonymous editing in coleoids is explained by the harm-permitting model and is nonadaptive. Fig. 3-2c shows that, while L_R is significantly higher than L_S in each coleoid, it is lower than 2.5%. One might ask whether such low median levels of restorative editing can be harm-permitting. As mentioned,

not all restorative editing is necessarily harm-permitting, which could explain why L_R is not particularly high. Nevertheless, Fig. 3-2c reveals a larger fraction of restorative editing than synonymous editing with appreciable editing levels. For example, in the squid, 29.14% and 15.65% of restorative editing sites but only 23.03% and 6.82% of synonymous editing sites have editing levels >5% and >20%, respectively. Depending on the harm of the G-to-A mutation and the relative dominance of the A and G isoforms, these appreciable levels of A-to-G editing could substantially increase the fixation probability of the G-to-A mutation. It should also be noted that the harm-permitting hypothesis is proposed as an alternative to the adaptive hypothesis. If moderate levels of nonsynonymous editing could be beneficial as asserted by the adaptive hypothesis, there is no reason why they could not be harm-permitting. Furthermore, the general trend of $L_R > L_S$ and $L_D \approx L_S$ supports the harm-permitting hypothesis relative to the adaptive hypothesis.

To examine the robustness of our results, we conducted four additional analyses. First, we respectively examined editing sites that are specific to each of the four coleoids, because species-specific editing events have similar evolutionary ages, allowing fairer comparisons. The results obtained are highly similar to those in Fig. 3-2 and are robust to potential misidentifications of species-specific editing (Fig. B-1). Second, we probed editing events identified from individual tissues in bimac. $F_D < F_S$ holds across tissues, but editing level comparisons as well as comparisons between F_R and F_S are mostly non-significant, likely due to the reduced statistical power as a result of decreased sample sizes (Table B-2). Third, because editing levels of neighboring editing sites may be co-affected by a mutation, which would reduce the statistical power in comparing synonymous with nonsynonymous editing sites, we compared synonymous editing sites in one half of the gene set with nonsynonymous editing sites in the

other half. Specifically, we ranked all genes by the d_N/d_S ratio between octopus and squid orthologs and respectively grouped genes with odd ranks into bin 1 and those with even ranks into bin 2. We then compared synonymous editing in bin 1 with nonsynonymous editing in bin 2, as well as synonymous editing in bin 2 with nonsynonymous editing in bin 1. The results (Fig. B-2) are similar to those obtained from all editing sites (Fig. 3-2). Fourth, we respectively investigated F_R/F_S and F_D/F_S in five editing level ranges (0-20%, 20-40%, 40-60%, 60-80%, and 80-100%) in each coleoid (Fig. B-3). While both F_R/F_S and F_D/F_S generally increase with the editing level, F_D/F_S is smaller than one except when the editing level exceeds 40%. It is important to stress that only a few percent of diversifying editing sites in a coleoid fall in this editing level range (Table B-3), suggesting that the vast majority of diversifying editing is nonadaptive (see below for quantitative estimates).

3.3.3 Accelerated nonsynonymous G-to-A substitutions

The harm-permitting model further predicts that the rate of nonsynonymous G-to-A substitution relative to that of synonymous G-to-A substitution (d_N/d_S for G-to-A) should be elevated, because the high editing activity renders some otherwise deleterious nonsynonymous G-to-A mutations acceptable. Furthermore, this elevation should be particularly pronounced in genes exclusively expressed in neural tissues but not in genes unexpressed in neural tissues, because the high editing activity is so far observed only in neural tissues(Alon et al. 2015; Liscovitch-Brauer et al. 2017). However, because only bimac and squid have available RNA sequencing data from several non-neural tissues and because genes unexpressed in neural tissues are not in the transcript sequence data of the octopus and cuttlefish and hence are excluded from our alignments, we had to define two groups of genes with relatively high and relatively low

specificities in neural expression, respectively. The genes with high neural expression specificities are expressed exclusively in neural tissues in the bimac or squid, whereas those with low neural expression specificities are expressed in both neural and non-neural tissues in both the bimac and squid. The harm-permitting model predicts that d_N/d_S for G-to-A is greater for genes with relatively high neural expression specificities than for those of relatively low neutral expression specificities. Because the harm-permitting effect is present only when a G-to-A mutation at a site is deleterious without editing, we focused on nonsynonymous sites that are conserved in the two outgroup species (i.e., nautilus, sea hare, and the immediately ancestral node of the focal species share the same pre-editing state) to increase the sensitivity of our test. Furthermore, the elevation in d_N/d_S should be specific to G-to-A changes, because the potential harms of other changes such as C/T-to-A and G-to-C/T cannot be alleviated by A-to-G editing.

To this end, we considered all six branches descendent from the common ancestor of the four coleoids. We computed d_N and d_S of each of these branches using the extant and inferred ancestral sequences, and then calculated d_N/d_S by dividing the total d_N by the total d_S of these branches. In support of our prediction, d_N/d_S for G-to-A changes is greater for genes of relatively high neural expression specificities than those of relatively low specificities (Fig. 3-3). By respectively bootstrapping the two group of genes 200 times, we found that the above difference is statistically significant ($P < 0.005$). By contrast, no significant difference in d_N/d_S exists between the two groups of genes when C/T-to-A changes or G-to-C/T changes are considered (Fig. 3-3). Note that $d_N/d_S < 1$ in all cases in Fig. 3-3, consistent with the harm-permitting model that does not involve positive selection.

3.3.4 The potential benefit of shared editing among species

It has been suggested that shared editing among multiple species is likely beneficial, because otherwise the editing status is unlikely to be evolutionarily conserved(Xu and Zhang 2015). In support of this suggestion was the finding that, even in mammals, where most nonsynonymous editing appears neutral or deleterious, the frequency of conserved sites subject to nonsynonymous editing in both human and mouse significantly exceeds the frequency of conserved sites subject to synonymous editing in both species (Xu and Zhang 2015). A similar phenomenon was reported in fruit flies (Yu et al. 2016). In coleoids, a sizable fraction of nonsynonymous editing is shared by at least two species and highly edited sites tend to be shared (Liscovitch-Brauer et al. 2017). To understand the potential evolutionary forces maintaining RNA editing at specific sites across multiple coleoids, we analyzed editing shared by a clade of two or more species.

A nonsynonymous editing event shared by a clade of species that modifies the amino acid state from X to Y is considered restorative if the inferred genomic sequence-based amino acid state is Y at any node of the tree that is ancestral to the most recent common ancestor of the clade, or diversifying if Y is not present at any of these ancestral nodes. In the study of shared editing, we considered the average editing level in the clade where the editing is shared. For each of the three groups of editing respectively shared between octopus and bimac, between squid and cuttlefish, and among the four coleoids, both F_R and F_D are lower than F_S (Fig. 3-4a). However, when the editing level exceeds 10%, F_R is significantly higher than F_S for shared editing between octopus and bimac and that between squid and cuttlefish, and F_D is significantly higher than F_S for shared editing between squid and cuttlefish (Fig. 3-4a inset). L_R and L_D are both higher than L_S for each group of shared editing (Fig. 3-4b). A significantly greater F_D than F_S for shared editing could be caused by (i) positive selection promoting the initial fixation of

mutations that lead to nonsynonymous editing and/or (ii) purifying selection preventing the loss of presumably beneficial nonsynonymous editing; therefore, it is a clear indicator of adaptive nonsynonymous editing. A significantly greater L_D than L_S for shared editing could be caused by (i) positive selection promoting the increase of editing levels of presumably beneficial nonsynonymous editing, (ii) purifying selection preventing the decrease of editing levels of presumably beneficial nonsynonymous editing, (iii) purifying selection preferentially preventing the loss of high-level nonsynonymous editing presumably because high editing levels are associated with larger benefits than low editing levels, and/or (iv) positive selection preferentially promoting the loss of low-level nonsynonymous editing, probably because an A-to-G substitution is favored at an edited site, especially when the editing level is low. Regardless, a significantly greater L_D over L_S also indicates adaptive nonsynonymous editing. Hence, diversifying editing shared by different coleoids shows adaptive signals, suggesting that a fraction is adaptive.

Because most nonsynonymous editing is species-specific (Table B-1), the above finding is not inconsistent with the analysis of individual species revealing the nonadaptive nature of most editing events. We estimated that, of species-specific diversifying editing sites, 0.15%, 0%, 1.20%, and 0.16% are adaptive in the octopus, bimac, squid, and cuttlefish, respectively (see Methods). Similarly, 3.50%, 3.00%, 14.98%, and 12.26% of shared diversifying editing sites are adaptive in octopus, bimac, squid, and cuttlefish, respectively. Taken together, 1.37%, 1.72%, 4.91%, and 2.18% of diversifying editing sites are adaptive in the four coleoids, respectively.

What is the general benefit of the shared editing that shows adaptive signals? Two hypotheses exist. First, editing may be beneficial because of the intra-organism protein diversity created (Gommans et al. 2009; Nishikura 2010; Alon et al. 2015; Liscovitch-Brauer et al. 2017).

That is, editing allows the existence of two protein isoforms per edited site in an organism, which may confer a higher fitness, analogous to heterozygote advantage at polymorphic sites. Alternatively, editing offers a new isoform that may be simply fitter than the unedited isoform. In this latter hypothesis, the benefit of editing is comparable to that of a nucleotide substitution. To distinguish between these two hypotheses, we focused on sites that are edited in at least three of the four coleoids, because editing should have existed at these sites in the common ancestor of the four species according to the parsimony principle (Fig. 3-2a). We then estimated the frequency of replacement of editing with an A-to-G substitution in any of the four species. Such replacements are expected to be more or less neutral for synonymous editing. For nonsynonymous editing, such replacements are deleterious under the first hypothesis due to the loss of protein diversity but are neutral under the second hypothesis. Hence, the first hypothesis predicts a lower frequency of such replacements for nonsynonymous editing than synonymous editing, whereas the second hypothesis predicts equal frequencies of such replacements for synonymous and nonsynonymous editing.

Our analysis showed that the frequency of A-to-G substitutions at nonsynonymous editing sites is significantly lower than that at synonymous editing sites (Fig. 3-4c; Table B-4). Because it is the shared diversifying editing for which the nature of the benefit is in question, we restricted the analysis to diversifying editing only, but obtained a similar result (Fig. 3-4c; Table B-4). Furthermore, A sites edited in at least three coleoid species are more likely to be edited in all four coleoids when the editing is nonsynonymous than when it is synonymous (36.6% vs. 27.6%, $P = 0.001$, chi-squared test; Table B-4), suggesting that shared nonsynonymous editing is less likely to be lost than shared synonymous editing. Together, these observations suggest that the benefit of adaptive shared editing is the provision of two protein isoforms per edited site in

an organism.

3.4 Discussion

The recent discovery of the preponderance of nonsynonymous A-to-G RNA editing among highly edited sites in coleoid neural tissues led to the assertion of widespread adaptive editing in these organisms, but the potential benefits of the editing are unknown. In this work, we proposed an alternative, nonadaptive explanation. Our reanalysis of published transcriptome data from four coleoids and two outgroup species lends strong support to the nonadaptive hypothesis. Combined with previous findings from other species, the new finding suggests a generally nonadaptive nature of coding A-to-G editing among animals. As explained in the harm-permitting model, nonadaptive editing such as some restorative editing, may, however, be selectively protected (middle row in Fig. 3-1a) or even promoted (bottom row of Fig. 3-1a). While such editing events likely originated as molecular errors due to ADARs' limited target specificity, they are no longer errors today. That a nonadaptive feature can nevertheless be under purifying selection or even be positively selected is well known in evolutionary biology (Hartl and Taubes 1996; Stoltzfus 1999).

In the harm-permitting model, A-to-G editing permits the fixation of otherwise deleterious G-to-A mutations and hence the editing is nonadaptive. In theory, it is also possible that A-to-G editing emerged in evolution after a G-to-A substitution at the same site. If the substitution is slightly deleterious, the editing would be slightly beneficial (i.e., compensatory). However, such sites have minimal contributions to F_R and L_R so do not alter our interpretation of the nonadaptive nature of restorative editing (see Methods).

The principle of our test of the nonadaptive hypothesis of RNA editing is similar to that

of the test of the adaptive hypothesis, except that the new test requires a distinction between restorative and diversifying editing, which in turn depends on ancestral coding sequences inferred for the interior nodes of a phylogeny (Fig. 3-1b, c). While ancestral sequence inference is generally reliable, it is not expected to be 100% correct (Zhang and Nei 1997). Will errors and potential biases in this inference bias our test? The answer is no. F_R is the number of edited sites with an ancestral nonsynonymous G-to-A substitution divided by the total number of sites with an ancestral nonsynonymous G-to-A substitution. Because our ancestral sequence inference is based on genomic sequences and is blind to RNA editing, any potential bias in estimating the number of sites with an ancestral G-to-A substitution is cancelled out in computing F_R. The same applies to F_D. Errors and potential biases in ancestral sequence inference only increase the stochastic errors of F_R and F_D estimates, reducing the statistical power in testing our hypothesis. Notwithstanding, the vast majority of our key statistical tests yielded significant results, suggesting that sufficient statistical power remains in these tests.

Although our study explains the preponderance of nonsynonymous editing in coleoids, we have not addressed a related question—why the editing activity was drastically elevated in neural tissues during coleoid evolution. A substantial rise in editing activity is expected to be harmful because its effect is similar to inducing A-to-G mutations. Indeed, expression of the human ADAR2 gene in the budding yeast *Saccharomyces cerevisiae*, which does not naturally possess any ADAR gene, inhibits yeast growth because of ADAR2's RNA editing activity(Eifler et al. 2013). Our observation of a significantly lower F_D than F_S in every coleoid examined (Fig. 3-2b) strongly suggests that diversifying editing is generally deleterious and has been selectively purged. Hence, it is almost certain that the pervasive coding RNA editing was not the reason for the elevation of the editing activity in coleoids but its byproduct. Whatever the reason was, the

relevant benefit must at least offset the harm from pervasive nonsynonymous editing, under the assumption that the evolutionary elevation of the editing activity was not due to genetic drift alone because the population size of ancestral coleoids was probably not small. It is worth mentioning that a number of physiological functions have been proposed for A-to-G editing, including suppressing the proliferation of transposons (Athanasiadis et al. 2004), inhibiting viral replication (Taylor et al. 2005), marking RNAs for degradation (Nishikura 2010), marking RNAs to prevent innate immunity against self-RNAs (Liddicoat et al. 2015; Chung et al. 2018), regulating alternative splicing (Nishikura 2010), and modulating nuclear retention of RNAs (Nishikura 2010). Because the primary physiological function of A-to-G editing is unknown, it is difficult to discern why the editing activity rose drastically in coleoids.

Similar to previous findings in mammals and flies (Xu and Zhang 2015; Yu et al. 2016), we observed some adaptive signals from nonsynonymous editing shared between species. Our additional analysis suggests that the benefit of these adaptive editing events lies in the protein diversity brought by editing. While this finding supports the prevailing view on the mechanism through which coding RNA editing may be adaptive, it is important to stress that, based on our estimation, only about 2.5% of nonsynonymous editing appears adaptive in an average coleoid.

Liscovitch-Brauer and colleagues noted that flanking regions of sites edited in multiple species tend to be evolutionarily conserved and asserted that coleoids "use extensive RNA editing to diversify their neural proteome at the cost of limiting genomic sequence flexibility and evolution"(Liscovitch-Brauer et al. 2017). However, the observation prompting such a conclusion is caused by an ascertainment bias. Specifically, because of the various requirements for a site to be edited, such as specific flanking sequences (Liscovitch-Brauer et al. 2017) and secondary structures (Eggington et al. 2011), a shared editing site by definition satisfies these

requirements in its neighborhood in multiple species. Thus, the site is expected to show a higher interspecific similarity in flanking sequences than a randomly picked site, regardless of whether the editing is shared because of selective constraints or not. The same ascertainment bias occurs in the comparison of intraspecific polymorphisms of flanking sequences between shared editing sites and random sites. In particular, given the flanking sequence requirement for editing, an edited site with a lower flanking sequence polymorphism is expected to be edited in a greater percentage of individuals in the species. Hence, provided that a site is found to be edited in multiple species when only one individual is examined per species, the polymorphism is expected to be low irrespective of the presence/absence of selective constraints on the editing.

The nonadaptive hypothesis we proposed is based on the harm-permitting effect of high levels of editing, which inflates the frequency and level of restorative editing, relative to those of synonymous editing. Because previous comparisons of synonymous and nonsynonymous editing in non-coleoid species never considered this effect, one wonders whether their conclusions are still valid. Ignoring the harm-permitting effect renders conclusions of nonadaptive editing more conservative. Hence, such conclusions should still hold. For claims of adaptive editing that are based on comparisons between synonymous and nonsynonymous editing frequencies and levels, a reanalysis taking into account the harm-permitting effect is warranted. In other words, a significantly greater F_D than F_S and/or a significantly greater L_D than L_S are required to demonstrate positive selection promoting nonsynonymous editing. This is especially true to the group of fungi that show pervasive A-to-G editing as in coleoids (Liu et al. 2016; Liu et al. 2017; Teichert et al. 2017).

It is worth mentioning that transcriptome-wide analyses of several other types of RNA editing such as C-to-U editing (Liu and Zhang 2018a) and m^6A modification (methylation of A

at the nitrogen-6 position) (Liu and Zhang 2018b) also suggest that most editing events are nonadaptive. In addition, variations in several steps of RNA production and processing such as alternative transcriptional initiation (Xu et al. 2019), alternative splicing (Saudemont et al. 2017), and alternative polyadenylation (Xu and Zhang 2018) have been shown to be largely molecular errors. Similarly, it is plausible that variations in the translational process such as stop-codon read-through (Li and Zhang 2019) and events of posttranslational modifications such as phosphorylation (Landry et al. 2009) and glycosylation (Park and Zhang 2011) are primarily manifestations of molecular errors. Whether it is generally true that phenotypic variations at the molecular level are less likely to be adaptive than those at the cellular, tissue, organ, and organismal levels is worth exploration (Zhang 2018).

3.5 Methods

3.5.1 Transcriptomes, editing sites, and ancestral sequences

The transcriptomes of six mollusk species and the list of A-to-I editing sites in the four coleoid species were previously published (Liscovitch-Brauer et al. 2017). We extracted coding sequences from the previously assembled transcriptomes (Liscovitch-Brauer et al. 2017) on the basis of the annotations in the dataset and converted Gs in the sequences to As at edited sites. In some genes, we observed stop codons occurring upstream of the last three nucleotides of the annotated coding sequence, possibly due to erroneous inclusions of 3'-untranslated regions. We therefore removed nucleotides downstream of the first stop codon in these sequences. All but one A-to-G editing site in the data are upstream of the first stop codons, suggesting that these annotation errors barely influenced the previous analysis of RNA editing. If a gene appeared more than once in the original dataset for a species, only the longest sequence was retained in our

analyses.

Orthologous genes among the six mollusks were previously identified (Liscovitch-Brauer et al. 2017) and a total of 3979 genes have orthologs in all six species and contain at least one A-to-G editing site in at least one coleoid. We first made a protein sequence alignment of orthologous sequences using MAFFT (Katoh and Standley 2013) and then generated a coding sequence alignment of these genes using PAL2NAL (Suyama et al. 2006). Ancestral sequences were inferred using the codeml program in PAML4 (Yang 2007) under default parameters without excluding gap sites, and the best joint inferences of all interior nodes were used in subsequence analyses. The unrooted topology of the tree in Fig. 3-2a was used in ancestral sequence inference. Subsequent analyses used in-house Perl scripts.

All reported editing sites in the 3979 genes (Liscovitch-Brauer et al. 2017) were included in our analyses unless otherwise noted. Although some editing sites may be sequencing errors, the probability of error is expected to be low given the tiny amount of other types of DNA-RNA mismatches observed (Liscovitch-Brauer et al. 2017).

3.5.2 Restorative and diversifying editing

The tree in Fig. 3-2a shows three interior nodes ancestral to each coleoid species. A coding A site in a coleoid is considered a potential site for restorative editing if changing the A to G is nonsynonymous and if the corresponding amino acid after the change becomes identical to the amino acid state at any one of the three ancestral nodes. A potential site for restorative editing becomes a restorative editing site if it is edited in the focal species. By definition, F_R is the number of sites with restorative editing divided by the number of potential sites for restorative editing, whereas L_R is the median editing level at restorative editing sites. A coding A

site in a focal species is considered a potential site for diversifying editing if changing the A to G is nonsynonymous and if the corresponding amino acid after the change differs from all amino acid states of the three ancestral nodes. A potential site for diversifying editing becomes a diversifying editing site if it is edited in the focal species. By definition, F_D is the number of sites with diversifying editing divided by the number of potential sites for diversifying editing, whereas L_D is the median editing level at diversifying editing sites. F_S is the number of sites with synonymous editing divided by the number of A sites where A-to-G editing would be synonymous, whereas L_S is the median editing level at synonymous editing sites(Xu and Zhang 2014). While the comparison between F_R (or F_D) and F_S and that between L_R (or L_D) and L_S are not entirely independent from each other, each comparison is fair.

An editing event is considered to be shared by a clade of two or more species if the event occurs in all species of the clade in the tree of Fig. 3-2a and all of these species have the same pre- and post-editing amino acid states. In studying shared editing by a clade, we followed the above procedure in distinguishing restorative from diversifying editing, except that we considered all interior nodes ancestral to the most recent common ancestor of the clade instead of all interior nodes ancestral to one species.

3.5.3 Comparing median editing levels

When the mRNA concentration is low, RNA editing cannot be detected unless the editing level is sufficiently high. This bias would make the median editing level appear higher in weakly expressed genes than strongly expressed genes even when no such difference actually exists. To alleviate this bias, we considered only those sites that are covered by at least 400 RNA sequencing reads when comparing median editing levels. Nevertheless, the bias does not affect

the comparison between synonymous and nonsynonymous editing, because their detectabilities are equally influenced by the gene expression level. For a shared editing site, the average editing level and average read number of all species in the focal clade are used to represent the site. We did not apply editing level cutoffs in the comparison of editing levels of different sites due to potential biases that may arise.

3.5.4 Proportion of diversifying editing that is adaptive

Under the presumption that the excess of F_D over F_S represents adaptive editing, we calculated F_D and F_S in each of 10 editing level intervals (0-10%, 10-20%, …, and 90-100%). For each interval exhibiting $F_D > F_S$, the number of adaptive diversifying editing sites equals $ADP = N_D(1-F_S/F_D)$, where N_D is the number of diversifying editing sites in the interval. Summing up these *ADP* numbers yields the total number of diversifying editing sites that are adaptive.

3.5.5 Contributions of compensatory editing to F_R and L_R

In the harm-permitting model, A-to-G editing permits the fixation of otherwise deleterious G-to-A mutations and hence the editing is nonadaptive. In theory, it is also possible that A-to-G editing emerged in evolution after a G-to-A substitution at the same site. If the substitution is slightly deleterious, the editing would be slightly beneficial (i.e., compensatory). For several reasons, such sites should contribute minimally to F_R and L_R. First, the probability that the G-to-A substitution occurred in the most recent common ancestor of cephalopods (the top five species in Fig. 3-2a) or more recently is small, because it could occur at any time prior to the emergence of the editing at the site, which most likely took place when the cellular editing

activity rose substantially in the branch immediately preceding the common ancestor of coleoids. Hence, the probability that the editing is classified as restorative is small and such compensatory events are unlikely to affect our analysis of restorative editing sites. Although such compensatory events are potentially included in diversifying editing sites we analyzed, diversifying editing still show lower editing frequencies and editing levels when compared with synonymous editing. Thus, our interpretation that diversifying editing is overall under purifying selection remains valid. Furthermore, even for the minority of compensatory events that are classified as restorative, the impact is small. This is because deleterious G-to-A mutations that could get fixed without editing are presumably only slightly deleterious. Hence, the benefit of A-to-G editing at such sites is also presumably small such that their editing level may not be selectively raised or selectively maintained at high levels. More importantly, there will be a comparable number of slightly beneficial G-to-A substitutions followed by slightly deleterious A-to-G editing that are included in the category of restorative editing. The effects of these two groups of events are likely cancelled out.

3.6 References

Albertin, C. B., O. Simakov, T. Mitros, Z. Y. Wang, J. R. Pungor, E. Edsinger-Gonzales, S. Brenner, C. W. Ragsdale, and D. S. Rokhsar. 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524:220-+.

Alon, S., S. C. Garrett, E. Y. Levanon, S. Olson, B. R. Graveley, J. J. C. Rosenthal, and E. Eisenberg. 2015. The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. Elife 4.

Athanasiadis, A., A. Rich, and S. Maas. 2004. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol 2:e391.

Bahn, J. H., J. H. Lee, G. Li, C. Greer, G. Peng, and X. Xiao. 2012. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res 22:142-150.

Bazak, L., A. Haviv, M. Barak, J. Jacob-Hirsch, P. Deng, R. Zhang, F. J. Isaacs, G. Rechavi, J. B. Li, E. Eisenberg, and E. Y. Levanon. 2014. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res 24:365-376.

Chen, J. Y., Z. Y. Peng, R. L. Zhang, X. Z. Yang, B. C. M. Tan, H. Y. Fang, C. J. Liu, M. M. Shi, Z. Q. Ye, Y. E. Zhang, M. H. Deng, X. Q. Zhang, and C. Y. Li. 2014. RNA Editome in Rhesus Macaque Shaped by Purifying Selection. Plos Genet 10.

Chung, H., J. J. A. Calis, X. Wu, T. Sun, Y. Yu, S. L. Sarbanes, V. L. Dao Thi, A. R. Shilvock, H. H. Hoffmann, B. R. Rosenberg, and C. M. Rice. 2018. Human ADAR1 Prevents Endogenous RNA from Triggering Translational Shutdown. Cell 172:811-824 e814.

Covello, P. S. and M. W. Gray. 1993. On the evolution of RNA editing. Trends Genet 9:265-268.

Danecek, P., C. Nellaker, R. E. McIntyre, J. E. Buendia-Buendia, S. Bumpstead, C. P. Ponting, J. Flint, R. Durbin, T. M. Keane, and D. J. Adams. 2012. High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. Genome Biol 13.

Deffit, S. N., B. A. Yee, A. C. Manning, S. Rajendren, P. Vadlamani, E. C. Wheeler, A. Domissy, M. C. Washburn, G. W. Yeo, and H. A. Hundley. 2017. The C. elegans neural editome reveals an ADAR target mRNA required for proper chemotaxis. Elife 6.

Duan, Y., S. Dou, S. Luo, H. Zhang, and J. Lu. 2017. Adaptation of A-to-I RNA editing in Drosophila. PLoS Genet 13:e1006648.

Eggington, J. M., T. Greene, and B. L. Bass. 2011. Predicting sites of ADAR editing in double-stranded RNA. Nat Commun 2:319.

Eifler, T., S. Pokharel, and P. A. Beal. 2013. RNA-Seq analysis identifies a novel set of editing substrates for human ADAR2 present in Saccharomyces cerevisiae. Biochemistry 52:7857-7869.

Eisenberg, E. and E. Y. Levanon. 2018. A-to-I RNA editing - immune protector and transcriptome diversifier. Nat Rev Genet 19:473-490.

Farajollahi, S. and S. Maas. 2010. Molecular diversity through RNA editing: a balancing act. Trends Genet 26:221-230.

Garrett, S. and J. J. C. Rosenthal. 2012. RNA Editing Underlies Temperature Adaptation in K+ Channels from Polar Octopuses. Science 335:848-851.

Gilbert, W. V., T. A. Bell, and C. Schaening. 2016. Messenger RNA modifications: Form, distribution, and function. Science 352:1408-1412.

Gommans, W. M., S. P. Mullen, and S. Maas. 2009. RNA editing: a driving force for adaptive evolution? Bioessays 31:1137-1145.

Gott, J. M. and R. B. Emeson. 2000. Functions and mechanisms of RNA editing. Annu Rev Genet 34:499-531.

Graveley, B. R., A. N. Brooks, J. Carlson, M. O. Duff, J. M. Landolin, L. Yang, C. G. Artieri, M. J. van Baren, N. Boley, B. W. Booth, J. B. Brown, L. Cherbas, C. A. Davis, A. Dobin, R. H. Li, W. Lin, J. H. Malone, N. R. Mattiuzzo, D. Miller, D. Sturgill, B. B. Tuch, C. Zaleski, D. Y. Zhang, M. Blanchette, S. Dudoit, B. Eads, R. E. Green, A. Hammonds, L. C. Jiang, P. Kapranov, L. Langton, N. Perrimon, J. E. Sandler, K. H. Wan, A. Willingham, Y. Zhang, Y. Zou, J. Andrews, P. J. Bickel, S. E. Brenner, M. R. Brent, P. Cherbas, T. R. Gingeras, R. A. Hoskins, T. C. Kaufman, B. Oliver, and S. E. Celniker. 2011. The developmental transcriptome of Drosophila melanogaster. Nature 471:473-479.

Harcourt, E. M., A. M. Kietrys, and E. T. Kool. 2017. Chemical and structural effects of base modifications in messenger RNA. Nature 541:339-346.

Hartl, D. L. and C. H. Taubes. 1996. Compensatory nearly neutral mutations: selection without adaptation. J Theor Biol 182:303-309.

Hung, L. Y., Y. J. Chen, T. L. Mai, C. Y. Chen, M. Y. Yang, T. W. Chiang, Y. D. Wang, and T. J. Chuang. 2018. An Evolutionary Landscape of A-to-I RNA Editome across Metazoan Species. Genome Biol Evol 10:521-537.

Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780.

Klinger, C. M., L. Paoli, R. J. Newby, M. Y. Wang, H. D. Carroll, J. D. Leblond, C. J. Howe, J. B. Dacks, C. Bowler, A. B. Cahoon, R. G. Dorrell, and E. Richardson. 2018. Plastid Transcript Editing across Dinoflagellate Lineages Shows Lineage-Specific Application but Conserved Trends. Genome Biol Evol 10:1019-1038.

Landry, C. R., E. D. Levy, and S. W. Michnick. 2009. Weak functional constraints on phosphoproteomes. Trends Genet 25:193-197.

Li, C. and J. Zhang. 2019. Stop-codon read-through arises largely from molecular errors and is generally nonadaptive. PLoS Genet 15:e1008141.

Li, J. B., E. Y. Levanon, J. K. Yoon, J. Aach, B. Xie, E. Leproust, K. Zhang, Y. Gao, and G. M. Church. 2009. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science 324:1210-1213.

Li, X., X. Xiong, and C. Yi. 2016. Epitranscriptome sequencing technologies: decoding RNA modifications. Nat Methods 14:23-31.

Liddicoat, B. J., R. Piskol, A. M. Chalk, G. Ramaswami, M. Higuchi, J. C. Hartner, J. B. Li, P. H. Seeburg, and C. R. Walkley. 2015. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. Science 349:1115-1120.

Liscovitch-Brauer, N., S. Alon, H. T. Porath, B. Elstein, R. Unger, T. Ziv, A. Admon, E. Y. Levanon, J. J. C. Rosenthal, and E. Eisenberg. 2017. Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. Cell 169:191-202 e111.

Liu, H. Q., Y. Li, D. P. Chen, Z. M. Qi, Q. H. Wang, J. H. Wang, C. Jiang, and J. R. Xu. 2017. A-to-I RNA editing is developmentally regulated and generally adaptive for sexual reproduction in Neurospora crassa. P Natl Acad Sci USA 114:E7756-E7765.

Liu, H. Q., Q. H. Wang, Y. He, L. F. Chen, C. F. Hao, C. Jiang, Y. Li, Y. F. Dai, Z. S. Kang, and J. R. Xu. 2016. Genome-wide A-to-I RNA editing in fungi independent of ADAR enzymes. Genome Res 26:499-509.

Liu, Z. and J. Zhang. 2018a. Human C-to-U coding RNA editing is largely nonadaptive. Mol Biol Evol.

Liu, Z. and J. Z. Zhang. 2018b. Most m(6)A RNA Modifications in Protein-Coding Regions Are Evolutionarily Unconserved and Likely Nonfunctional. Mol Biol Evol 35:666-675.

Maas, S., Y. Kawahara, K. M. Tamburro, and K. Nishikura. 2006. A-to-I RNA editing and human disease. RNA Biol 3:1-9.

Nishikura, K. 2010. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem 79:321-349.

Nishikura, K. 2016. A-to-I editing of coding and non-coding RNAs by ADARs. Nat Rev Mol Cell Biol 17:83-96.

Park, C. and J. Zhang. 2011. Genome-wide evolutionary conservation of N-glycosylation sites. Mol Biol Evol 28:2351-2357.

Peng, Z. Y., Y. B. Cheng, B. C. M. Tan, L. Kang, Z. J. Tian, Y. K. Zhu, W. W. Zhang, Y. Liang, X. D. Hu, X. M. Tan, J. Guo, Z. R. Dong, Y. Liang, L. Bao, and J. Wang. 2012. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol 30:253-+.

Pinto, Y., H. Y. Cohen, and E. Y. Levanon. 2014. Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. Genome Biol 15:R5.
Porath, H. T., B. A. Knisbacher, E. Eisenberg, and E. Y. Levanon. 2017a. Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. Genome Biol 18.
Porath, H. T., A. A. Schaffer, P. Kaniewska, S. Alon, E. Eisenberg, J. Rosenthal, E. Y. Levanon, and O. Levy. 2017b. A-to-I RNA Editing in the Earliest-Diverging Eumetazoan Phyla. Mol Biol Evol 34:1890-1901.
Ramaswami, G., R. Zhang, R. Piskol, L. P. Keegan, P. Deng, M. A. O'Connell, and J. B. Li. 2013. Identifying RNA editing sites using RNA sequencing data alone. Nat Methods 10:128-132.
Rodriguez, J., J. S. Menet, and M. Rosbash. 2012. Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. Mol Cell 47:27-37.
Rosenthal, J. J. C. and P. H. Seeburg. 2012. A-to-I RNA Editing: Effects on Proteins Key to Neural Excitability. Neuron 74:432-439.
Roundtree, I. A., M. E. Evans, T. Pan, and C. He. 2017. Dynamic RNA Modifications in Gene Expression Regulation. Cell 169:1187-1200.
Saudemont, B., A. Popa, J. L. Parmley, V. Rocher, C. Blugeon, A. Necsulea, E. Meyer, and L. Duret. 2017. The fitness cost of mis-splicing is the main determinant of alternative splicing patterns. Genome Biol 18.
Sloan, D. B. 2017. Nuclear and mitochondrial RNA editing systems have opposite effects on protein diversity. Biol Letters 13.
Stoltzfus, A. 1999. On the possibility of constructive neutral evolution. J Mol Evol 49:169-181.
Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609-W612.
Taylor, D. R., M. Puig, M. E. Darnell, K. Mihalik, and S. M. Feinstone. 2005. New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1. J. Virol. 79:6291-6298.
Teichert, I., T. A. Dahlmann, U. Kuck, and M. Nowrousian. 2017. RNA Editing During Sexual Development Occurs in Distantly Related Filamentous Ascomycetes. Genome Biology and Evolution 9:855-868.
Tian, N., X. J. Wu, Y. Z. Zhang, and Y. F. Jin. 2008. A-to-I editing sites are a genomically encoded G: Implications for the evolutionary significance and identification of novel editing sites. Rna 14:211-216.
Xu, C., J. K. Park, and J. Zhang. 2019. Evidence that alternative transcriptional initiation is largely nonadaptive. PLoS Biol 17:e3000197.
Xu, C. and J. Zhang. 2018. Alternative polyadenylation of mammalian transcripts is generally deleterious, not adaptive. Cell Syst 6:734-742.
Xu, G. X. and J. Z. Zhang. 2014. Human coding RNA editing is generally nonadaptive. P Natl Acad Sci USA 111:3769-3774.
Xu, G. X. and J. Z. Zhang. 2015. In Search of Beneficial Coding RNA Editing. Mol Biol Evol 32:536-541.
Yablonovitch, A. L., P. Deng, D. Jacobson, and J. B. Li. 2017. The evolution and adaptation of A-to-I RNA editing. PLoS Genet 13:e1007064.
Yang, Z. H. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-1591.

Yu, Y., H. Zhou, Y. Kong, B. Pan, L. Chen, H. Wang, P. Hao, and X. Li. 2016. The Landscape of A-to-I RNA Editome Is Shaped by Both Positive and Purifying Selection. PLoS Genet 12:e1006191.
Zhang, J. 2018. Neutral Theory and Phenotypic Evolution. Mol Biol Evol 35:1327-1331.
Zhang, J. and M. Nei. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J Mol Evol 44 Suppl 1:S139-146.
Zhang, R., P. Deng, D. Jacobson, and J. B. Li. 2017. Evolutionary analysis reveals regulatory and functional landscape of coding and non-coding RNA editing. PLoS Genet 13:e1006563.
Zhao, H. Q., P. Zhang, H. Gao, X. He, Y. Dou, A. Y. Huang, X. M. Liu, A. Y. Ye, M. Q. Dong, and L. Wei. 2015. Profiling the RNA editomes of wild-type C. elegans and ADAR mutants. Genome Res 25:66-75.

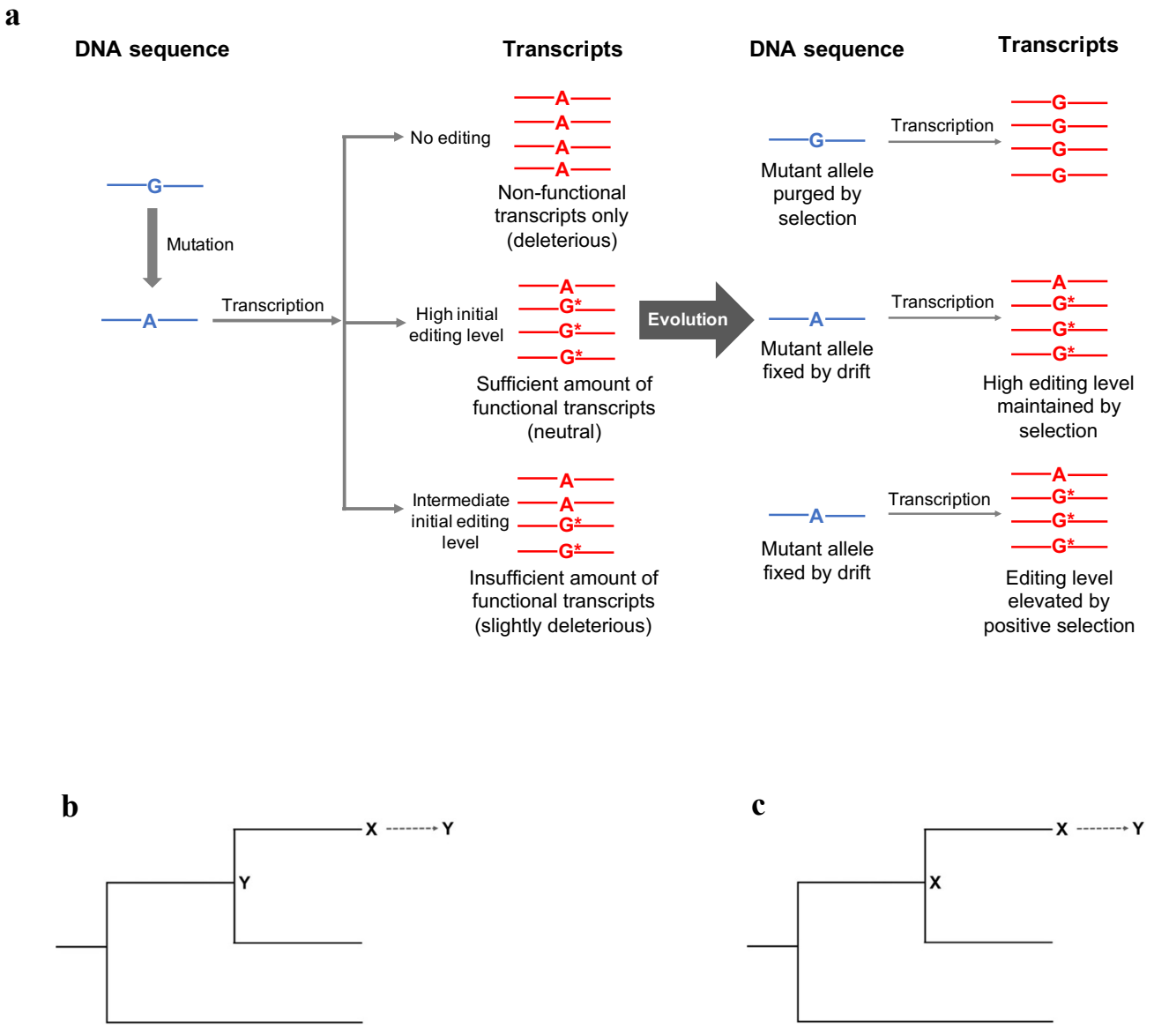
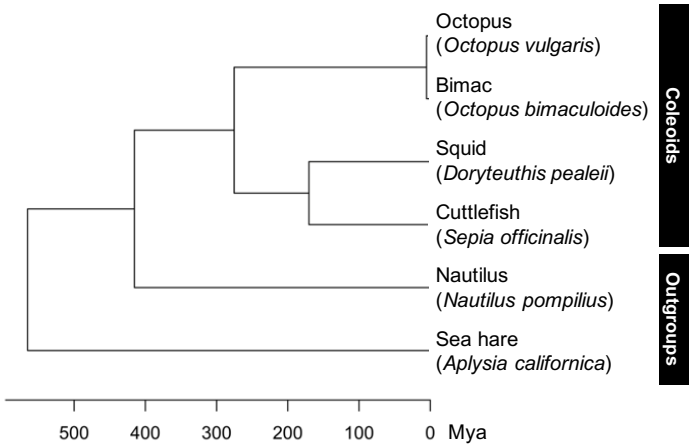


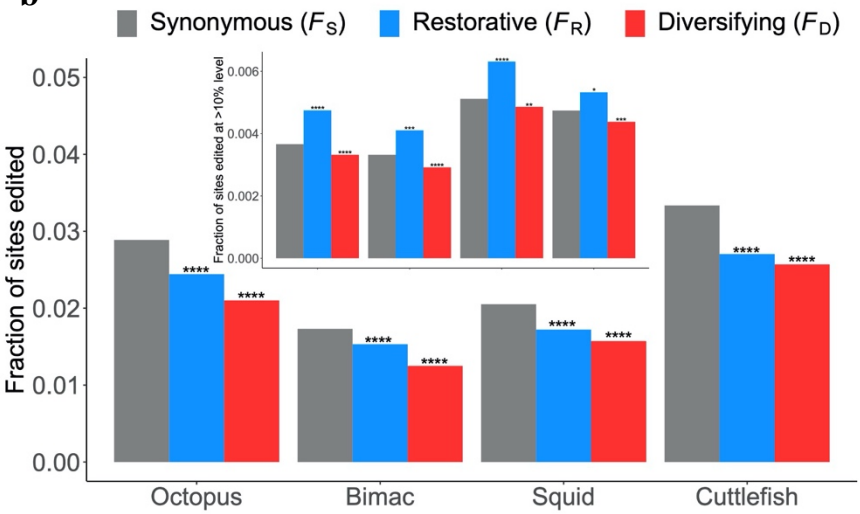
Figure. 3-1. The harm-permitting model and a strategy to detect the harm-permitting effect. (**a**) The harm-permitting effect of nonsynonymous editing. The top row shows that, when a nonsynonymous A site is not edited (or is subject to a low level of editing), a G-to-A mutation at the site is too deleterious to get fixed. The middle row shows that, when the site is highly edited, the G-to-A mutation becomes neutral and is fixed by genetic drift. The high editing level is then

selectively constrained. The bottom row shows that, when the editing level of the site is intermediate, the G-to-A mutation is slightly deleterious and fixed by genetic drift. The editing level may be further elevated by positive selection (or maintained by negative selection). Despite the relatively high nonsynonymous editing levels in the middle and bottom rows, no adaptation (i.e., no net increase in fitness) occurred when the final genotype is compared with the original genotype. DNA is shown in blue, while RNA is in red. Post-edited nucleotides are marked with stars. (**b**) Restorative editing restores an ancestral amino acid state lost upon an amino acid substitution, which may have occurred in the exterior branch as shown here or in an earlier branch. In other words, the post-editing state is identical to an ancestral pre-editing state. (**c**) Diversifying editing creates an amino acid state that differs from pre-editing states in a set of ancestors considered. Although only the state of one ancestor is shown here, the states of multiple ancestors may be considered. In (b) and (c), X and Y represent different amino acid states, whereas the arrow shows the effect of editing. Restorative but not diversifying editing can confer a harm-permitting effect.

a


Octopus
(Octopus vulgaris)
Bimac
(Octopus bimaculoides)
Squid
(Doryteuthis pealeii)
Cuttlefish
(Sepia officinalis)
Nautilus
(Nautilus pompilius)
Sea hare
(Aplysia californica)
Coleoids
Outgroups
500
400
300
200
100
0
Mya

b


Synonymous (F_S)
Restorative (F_R)
Diversifying (F_D)
Fraction of sites edited
0.05
0.04
0.03
0.02
0.01
0.00
Fraction of sites edited at >10% level
0.006
0.004
0.002
0.000
****
Octopus
Bimac
Squid
Cuttlefish

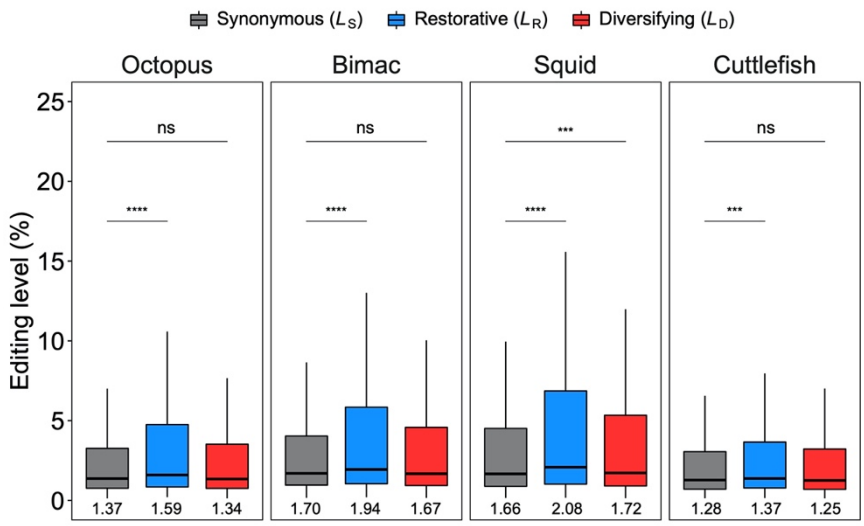


Figure. 3-2. Comparison of restorative and diversifying editing with synonymous editing in coleoids. (**a**) The phylogenetic relationship of the six mollusks studied here. Branch lengths represent divergence times based on the mid-points of the divergence time ranges in a previous study(Liscovitch-Brauer et al. 2017). (**b**) Frequencies of sites with synonymous (F_S), restorative (F_R), and diversifying (F_D) editing, respectively, in each of the four coleoids. A significant difference between F_S and F_R (or F_D) is indicated by stars above the bin of F_R (or F_D) (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant; chi-squared test). The inset shows the corresponding fractions of sites with editing levels >10%. (**c**) Synonymous (L_S), restorative (L_R), and diversifying (L_D) editing levels in each of the four coleoids. The lower and upper edges of a box represent the first (qu_1) and third quartiles (qu_3), respectively, the horizontal line inside the box indicates the median (md), and the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(qu_3\text{-}qu_1)$. The median editing levels are also given below the corresponding boxes. A significant difference between L_S and L_R (or L_D) is indicated by stars (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant; Mann-Whitney U test).

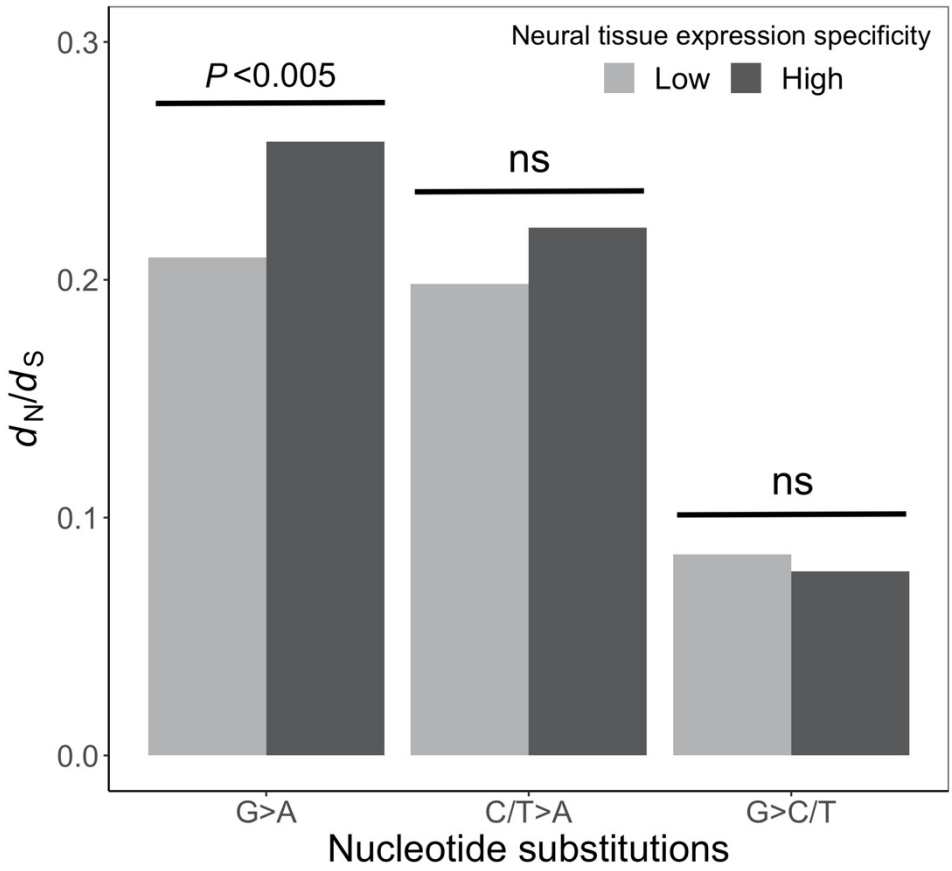
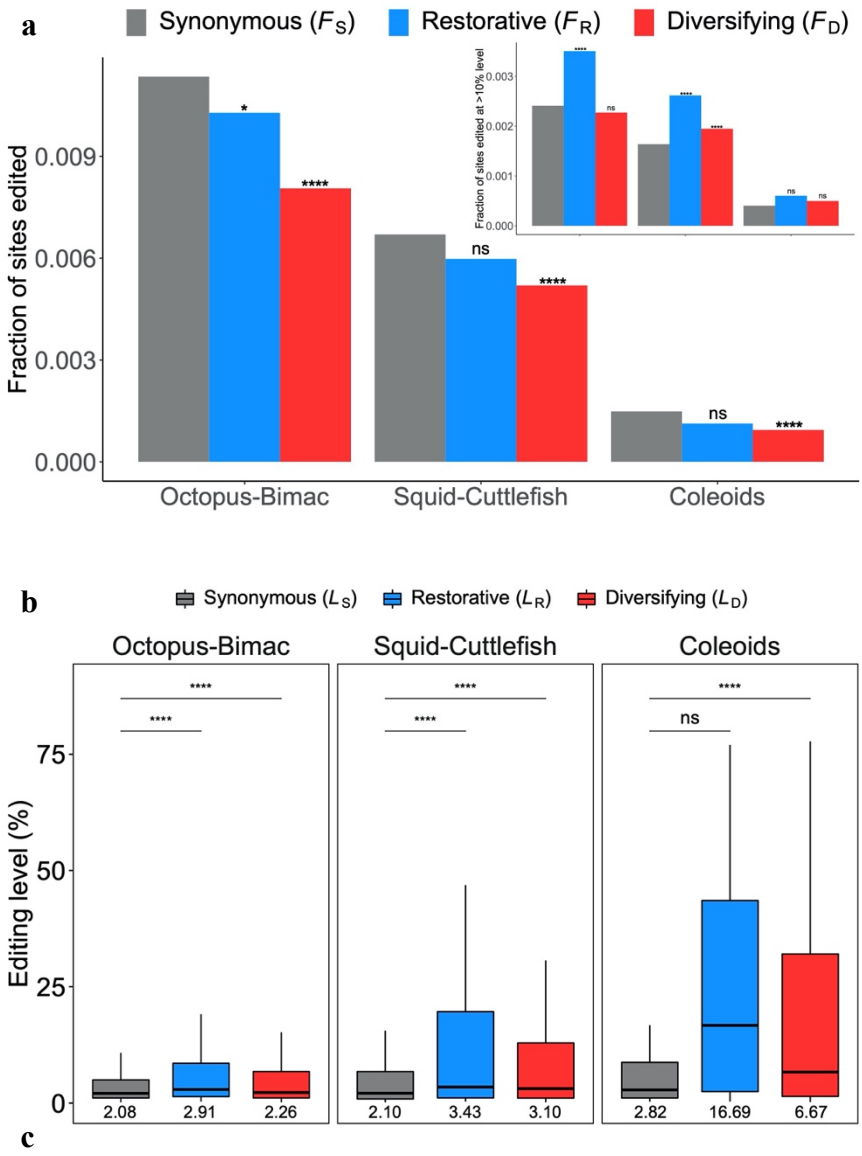


Figure. 3-3. Coleoid nonsynonymous to synonymous substitution rate ratios (d_N/d_S) for various nucleotide changes. The *P*-value is based on 200 bootstrap samples; ns, not significant.


a
Synonymous (F_S)
Restorative (F_R)
Diversifying (F_D)
Fraction of sites edited
0.009
0.006
0.003
0.000
*
****
ns
****
ns
****
Octopus-Bimac
Squid-Cuttlefish
Coleoids
Fraction of sites edited at >10% level
0.003
0.002
0.001
0.000
b
Synonymous (L_S)
Restorative (L_R)
Diversifying (L_D)
Octopus-Bimac
Squid-Cuttlefish
Coleoids
Editing level (%)
75
50
25
0
****
****
****
****
****
ns
2.08
2.91
2.26
2.10
3.43
3.10
2.82
16.69
6.67
c

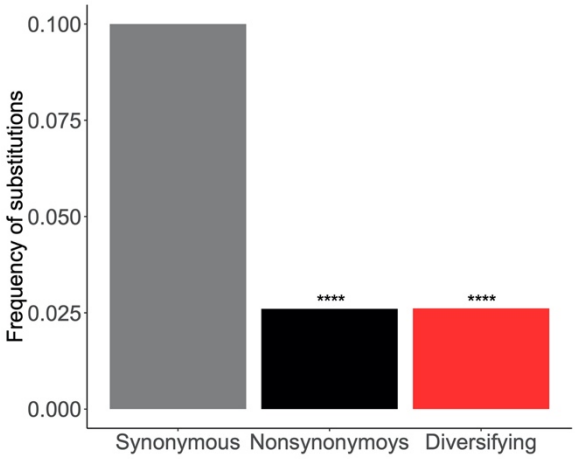


Figure. 3-4. Patterns of shared editing in coleoids. (**a**) Frequencies of sites with shared synonymous (F_S), restorative (F_R), and diversifying (F_D) editing, respectively, among sites shared between the octopus and bimac, between the squid and cuttlefish, and among all four coleoids, respectively. A significant difference between F_S and F_R (or F_D) is indicated by stars above the bin of F_R (or F_D) (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant; chi-squared test). The inset shows corresponding fractions of sites with editing levels >10%. (**b**) Synonymous (L_S), restorative (L_R), and diversifying (L_D) editing levels among the three sets of shared editing sites. The lower and upper edges of a box represent the first (qu_1) and third quartiles (qu_3), respectively, the horizontal line inside the box indicates the median (md), and the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(qu_3 - qu_1)$. The median editing levels are also given below the corresponding boxes. A significant difference between L_S and L_R (or L_D) is indicated by stars (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant; Mann-Whitney U test). (**c**) Fraction of sites edited in the common ancestor of the four coleoids that have a genomic G in a coleoid (****, $P < 0.0001$; chi-squared test).

Chapter 4

Evolution of Organismal Traits: Fly Wing Evolution Explained by a Neutral Model with Mutational Pleiotropy

4.1 Abstract

To what extent the speed of mutational production of phenotypic variation determines the rate of long-term phenotypic evolution is a central question. Houle *et al.* recently addressed this question by studying the mutational variances, additive genetic variances, and macroevolution of locations of vein intersections on fly wings, reporting very slow phenotypic evolution relative to the rates of mutational input, high phylogenetic signals, and a strong, linear relationship between the mutational variance of a trait and its rate of evolution. Houle *et al*. found no existing model of phenotypic evolution to be consistent with all these observations, and proposed the improbable scenario of equal influence of mutational pleiotropy on all traits. Here we demonstrate that the purported linear relationship between mutational variance and evolutionary divergence is artifactual. We further show that the data are explainable by a simple model in which the wing traits are effectively neutral at least within a range of phenotypic values but their

Works of this chapter have been published:
Jiang, D. and J. Zhang. 2020. Fly wing evolution explained by a neutral model with mutational pleiotropy. Evolution. 74:2158-2167.

evolutionary rates are differentially reduced because mutations affecting these traits are purged owing to their different pleiotropic effects on other traits that are under stabilizing selection. Thus, the evolutionary patterns of fly wing morphologies are explainable under the existing theoretical framework of phenotypic evolution.

4.2 Introduction

A fundamental question in evolutionary biology is the extent to which the rate of long-term phenotypic evolution is determined by the rate of production of phenotypic variation by newly arising mutations (Lande 1976; Chakraborty and Nei 1982; Hill 1982; Lynch and Hill 1986; Lynch 1990; Schluter 1996; Wagner and Altenberg 1996; Futuyma 2010). This question likely has different answers for different traits. At one extreme are purely neutral traits whose evolutionary rates are dictated by the rates with which phenotypic variations originate via mutation. At another extreme are traits subject to strong positive selection such that their evolutionary rates are primarily determined by the strength, duration, and frequency of Darwinian selection instead of mutation. The lack of empirical answers to this question is in a large part owing to the scarcity of suitable data to address this question, because such data require the information about the same phenotypic traits from mutants (e.g., mutation accumulation lines or gene deletion lines) as well as from different species.

In a recent study, Houle *et al.* addressed the above question by studying the evolution of locations of vein intersections on fly wings in the past 40 million years after inspecting over 50,000 wings from more than 100 Drosophilid species (Houle et al. 2017). After comparing the covariance matrices respectively representing the mutational inputs (M) and evolutionary rates (R), they reported that (1) the rate of phenotypic evolution is orders of magnitude lower than the

neutral expectation from the mutational variance, (2) the phylogenetic signals of most of these phenotypic traits are high, and (3) the evolutionary rate of a trait depends linearly on its mutational variance (i.e., with a log-linear slope of 1). Houle *et al*. examined nine existing models of phenotypic evolution (**Table C-1**) but found none that is consistent with all of the above features. After exhausting all existing models, Houle *et al*. suggested that their observations may be explained if "most mutations cause deleterious pleiotropic effects that render them irrelevant to adaptation, and, more importantly, the proportion of mutational variation that is deleterious is similar for all traits" (Houle et al. 2017).

Here we demonstrate that the reported linear relationship (or a log-linear slope of 1) between mutational variance and evolutionary divergence among the fly wing traits is an artifact resulting from the use of a biased method and that applying an unbiased method reveals a sublinear relationship (i.e., with a log-linear slope smaller than 1). Thus, Houle *et al.*'s model that all wing traits are equally impacted by mutational pleiotropy is not only theoretically improbable but also empirically refuted. Having re-estimated the relationship, we consider the implications of the corrected observations in the context of existing evolutionary models, and show that patterns of fly wing evolution are explainable by a simple model in which the focal wing traits are themselves effectively neutral at least within a range of phenotypic values but most mutations affecting the focal traits are purged by selection due to their pleiotropic effects on fitness-related traits.

4.3 Materials and Methods

4.3.1 Comparison of covariance matrices

To compare the covariance matrices respectively representing the mutational inputs (*M*) and evolutionary rates (*R*) of various traits along a set of orthogonal directions in the phenotypic space, Houle *et al*. first rescaled the matrices so that they have the same trace, which is the sum of diagonal elements. That is, they set $\tilde{R} = R\frac{tr(M)}{tr(R)}$, where *tr*(*M*) and *tr*(*R*) are traces of *M* and *R* matrices, respectively, and $\tilde{R}$ is the rescaled *R* matrix. They then computed $H = \frac{\tilde{R}+M}{2}$. *M* and $\tilde{R}$ were subsequently converted to K^TMK and $K^T\tilde{R}K$, respectively, where *K* denotes the matrix comprising the eigenvectors of *H*. The diagonal elements of K^TMK and $K^T\tilde{R}K$ were then compared to obtain a scaling exponent between *M* and *R*. Comparisons between *M* and *G* (the matrix of additive genetic covariances) and between *R* and *G* were performed likewise. The practice of deriving *K* from the average of *M* and $\tilde{R}$ inevitably biases the subsequent analysis to dimensions where *M* and *R* variances are similar, which would cause overestimation of the correlation coefficient and regression slope between $\log_{10}$(*R* variance) on $\log_{10}$(*M* variance).

To avoid the above problem, we used a new matrix comparison method, which is identical to Houle *et al*.'s method except that *K* is derived solely from *M* (when comparing *M* with *R* or *G*) or *G* (when comparing *G* with *R*) instead of *H*. In the reanalysis of the fly wing traits, the number of eigenvectors of *K* along which variances were compared is the same as the number of orthogonal traits plotted in the corresponding comparison of matrices in Houle *et al.* (2017). That is, regressions were performed using the first 18 eigenvectors of *K* when *M* was compared with *R* or *G* and the first 17 eigenvectors of *K* when *G* was compared with *R*.

4.3.2 Examination of matrix comparison methods using simulation

To evaluate the performances of Houle *et al*.'s method and the new method in comparing *M* with *R*, we used each method to analyze simulated covariance matrices. We first

independently generated two random, 24×24 covariance matrices that respectively represent *M* and *R*. Each matrix was obtained by first generating a correlation matrix and then converting it to the corresponding covariance matrix. Each correlation matrix was generated by the *rcorrmatrix* function of the R package *clusterGeneration* (Qiu and Joe 2015), with each correlation coefficient following a beta distribution $\beta(\alpha + \frac{d-2}{2}, \alpha + \frac{d-2}{2})$, where α was set to be 1 and *d*, the number of dimensions or traits, was set at 24. The diagonal elements of each matrix were drawn from a gamma distribution with the shape parameter $k = 0.5$; we chose this distribution because it resembles the empirical distribution of mutational variances (Houle and Fierst 2013). The simulated *M* and *R* matrices were subjected to the analysis by Houle *et al.*'s method and our new method, respectively. The above process was repeated 100 times to acquire distributions of the linear regression slope and Pearson's correlation coefficient between $\log_{10}$(*R* variance) and $\log_{10}$(*M* variance) produced by each method.

Because empirically estimated covariance matrices are subject to estimation error due to limited sample sizes, we repeated the above analyses with each matrix now replaced with an estimate of it. For each *M* (denoted M_{original}) or *R* (denoted R_{original}) matrix used in the previous analysis, we estimated a covariance matrix from 100 independent samples from the multivariate normal distribution of the original matrix. The estimates of *M* and *R* (denoted M_{obs} and R_{obs}) were then compared using Houle *et al.*'s and our methods, respectively. Additionally, these analyses were performed across a range of similarity between M_{original} and the *R* matrix. For each pair of M_{original} and R_{original}, we obtained a series of new *R* matrices and compared *M* with each of the new *R* matrices. The new *R* matrices was obtained as $R_{\text{w}} = wR_{\text{original}} + (1\text{-}w)M_{\text{original}}$, where *w* is the weight of R_{original} and a series of *w* equal to 0, 0.1, 0.2, …, 0.9, and 1 was considered.

Finally, to subject the matrices to realistic levels of error arising from the stochasticity of evolution as well as imprecise estimation, we simulated multivariate trait evolution along the phylogenetic tree of Drosophilid species used by Houle *et al*. and treated the *M* matrix estimated from the fly wing traits (Houle and Fierst 2013), M_{HF}, as $M_{original}$ in the simulation because it is presumably similar to the true *M* in terms of the structure. In each simulation, the matrix describing trait evolution (R_w) was obtained from $M_{original}$ and an independent matrix denoted $R_{original}$ using the preceding formula. Here, $R_{original}$ was generated in the same way as the random *M* and *R* matrices used in the previous steps, except that its diagonal elements were sampled from a gamma distribution with the shape parameter $k = 0.05$, such that the skewness of the distribution of the diagonal elements of R_w is similar to that observed from the empirically estimated *R* matrix of the fly wing traits (Houle, et al. 2017). Similar results were obtained when we used $k = 0.025$ or 0.1. The evolution followed a multivariate Brownian motion model. In particular, the multivariate phenotype at each node of the phylogeny was obtained by $X = X_A + l\Delta X$, where X_A is the phenotype at the node immediately ancestral to the focal node, *l* is the length of the branch connecting the two nodes, and ΔX is a vector sampled from the corresponding multivariate normal distribution of R_w. For each combination of *w* and *k*, we performed 50 simulations, each with an independently simulated $R_{original}$ as well as the corresponding R_w. After each simulation, we respectively used Houle *et al*.'s method and the new method to compare estimates of *M* and *R*, M_{obs} and R_{obs}. M_{obs} was estimated from 150 independent vectors taken from the distribution of M_{HF}; the sample size was set to be 150 because the empirical *M* matrix for the fly wing traits was estimated from 150 sublines (Houle and Fierst 2013). R_{obs} was estimated from the evolved phenotypes using the *ratematrix* function of the *geiger* package in R (Revell et al. 2007; Pennell et al. 2014), which estimates evolutionary

covariances using an independent contrast approach (Felsenstein 1985) and fits the observations to a Brownian motion model.

4.3.3 A neutral model with pleiotropy to explain the evolution of fly wing morphologies

Below we describe a neutral model with mutational pleiotropy that can almost perfectly explain the evolutionary patterns of fly wing morphologies. In our model, the focal wing traits are neutral, but mutations affecting the focal traits also influence other (unconsidered) traits that are subject to stabilizing selection (Turelli 1985; McGuigan et al. 2011). Consequently, these pleiotropic mutations are purged because of their effects on the non-focal traits; the higher the pleiotropy of a mutation (i.e., the number of non-focal traits affected by the mutation), the higher the chance that it will be purged (Estes and Phillips 2006; McGuigan et al. 2014). Importantly, the extent to which a focal trait is affected by the mutations' pleiotropic effects on non-focal traits can be correlated with its mutational variance.

Because all analyses of focal traits are conducted on the orthogonal traits, the following description of focal traits refer to focal orthogonal traits unless otherwise noted. In our model, the number (n) of non-focal traits that are genetically correlated with a focal trait is a function of the mutational variance of the focal trait (V_M) described by

$$n = A + B \log_{10} V_M. \qquad [1]$$

When $V_M < 10^{-A/B}$, n is set to 0. We model n as a logarithmic rather than linear function of V_M to slow the impact of V_M on n as V_M increases.

Given n, the number (m) of non-focal traits affected by a mutation that impacts the focal trait is a binomial random variable under the assumption that the mutation affects the non-focal traits independently. The expected value of m, $E(m)$, is equal to $\sum_{i=1}^{n} P_i$, where P_i is the

probability that the mutation affects the *i*th non-focal trait. A larger $E(m)$ means that an average mutation is more pleiotropic. The parameter *B* determines the scaling relationship between *M* and *R* variances. When $B > 0$, *M* variance has a positive effect on *n*, $E(m)$, and the pleiotropic effect of mutations on non-focal traits. By contrast, when $B < 0$, these effects become negative.

4.3.4 Simulation of evolution of neutral traits with mutational pleiotropy

To illustrate that the model described above is capable of producing evolutionary patterns similar to those of fly wing traits, we conducted a new set of simulations of trait evolution where mutational pleiotropy depends on *M* variance as described in the above section. Parameters were chosen such that the scaling of *R* variance with *M* variance resembles that observed from the fly wing traits under the new matrix comparison method. While the parameters used in the simulation presented are specific, we note that they are not unique in yielding results resembling patterns of fly wing evolution. Because the orthogonal traits were derived from the original fly wing traits by extracting the axes with the largest mutational variances, mutant phenotypes are generally concentrated along these axes and effects of mutations are expected to be largely following them. Although it is an approximation to assume that mutations strictly follow these axes, this approximation allows separate simulations of the evolution of different focal traits, which substantially simplifies the simulation that is meant to illustrate our model.

Mutational input

For a neutral focal trait, mutations affecting the trait were generated per unit of time by simulation. The number of mutations followed a Poisson distribution with the mean equal to λ_M, which is a constant for a given focal trait. The effect of a mutation on the focal trait followed a

normal distribution with a mean of 0 and a standard deviation of σ. Thus, mutational variance of the focal trait, which is defined as phenotypic variance introduced by new mutations per unit time, is $V_M = \lambda_M \sigma^2$.

Because it is unclear whether the observed variation in V_M across traits is contributed by the variation in mutation rate, effect size, or a combination of the two, we considered two extreme cases, denoted model 1 and model 2, respectively, in which one of λ_M and σ varies across focal traits while the other is constant. We sampled V_M of focal traits from a gamma distribution with the scale parameter $\theta = 0.1$ and shape parameter $k = 0.5$, because such a skewed distribution is similar to those observed for mutational variances of fly wing traits (Houle and Fierst 2013) and yeast cell morphology traits (Ho et al. 2017). In particular, when fitting the positive eigenvalues of the *M* matrix to a gamma distribution, we found the estimated shape parameters for the two datasets to be 0.41 and 0.22, respectively. When λ_M varies across focal traits (model 1), σ was set at 0.01 and λ_M of each focal trait was obtained as $\lambda_M = \frac{V_M}{\sigma^2}$; when σ varies across focal traits (model 2), λ_M was set at 500 and σ of each focal trait was $\sqrt{\frac{V_M}{\lambda_M}}$.

Before comparing mutational input and evolutionary divergence, we estimated V_M for each focal trait from 150 independent samples taken from a normal distribution with the variance equal to the corresponding true value.

Pleiotropic effect of mutations

We assume that a mutation affecting a focal trait had a probability of 0.5 to influence each non-focal trait; when it did have an effect, the effect followed a normal distribution with a mean of 0 and a standard deviation of 0.01. To identify *A* and *B* values in Eq. [1] that make our model predictions resemble actual observations, we first looked for a value of *n* that can yield the

evolutionary rate 10^{-6} to 10^{-4} times the neutral expectation, because the ratio of the observed amount of evolution of the fly wing traits and that expected from the mutational input falls in the range from 3.81×10^{-6} to 5.77×10^{-5}. Because the probability that a mutation affects a non-focal trait is 0.5 in our model, the proportion of mutations that are not pleiotropic is 0.5^n. If all pleiotropic mutations are unacceptable, $-\log_{0.5} 10^{-4} \leq n \leq -\log_{0.5} 10^{-6}$. That is, $13.9 < n < 19.9$. Because there is a small proportion of pleiotropic mutations that are benign enough to be acceptable, n may need to be slightly larger to produce the same rate of evolution. Hence, we decided that n should roughly range between 14 and 20, and accordingly set $A = 20$. Because the observed scaling exponent between *M* variance and *R* variance is smaller than 1 for the fly wing traits (see Results), *B* should be positive. We looked for a positive *B* that would make n no less than 14 for the focal trait with the smallest *M* variance (0.000401) in our simulation. Solving $n = 20 + B\log_{10} 0.000401 \geq 14$, we found $B \leq 1.77$. We therefore set *B* at 1.7. For the focal trait with the greatest *M* variance (0.222), the maximum of n is obtained as the largest integer under $20 + 1.7 \times \log_{10} 0.222 = 18.89$, which is 18, a number within the desired range. We note that the probability that a mutation affects a non-focal trait, which is 0.5 in our simulation, was arbitrarily chosen; having it been a different value, the values of *A* and *B* would be different.

Fitness function and selection

The fitness of a mutant depends on its phenotype and is assumed to be a function of its phenotypic Euclidean distance (*D*) from the optimal phenotype. *D* is computed from values of the non-focal traits and equals $\sqrt{\sum_{i=1}^{n} {d_i}^2}$, where d_i is the distance between the value of the *i*th non-focal trait and the optimal value of this trait. We also considered scenarios where the focal

trait is under selection rather than strictly neutral, in which case D is equal to $\sqrt{c{d_f}^2 + \sum_{i=1}^{n} {d_i}^2}$, where d_f is the distance between the phenotypic value of the focal trait and the optimum and c is a coefficient deciding the importance of the focal trait, relative to that of a non-focal trait, to fitness. We examined three values of c in our simulations, 10^{-4}, 10^{-3}, and 10^{-2}. We considered a Gaussian fitness function in which mutant fitness $f = \exp\left(-\frac{D^2}{2}\right)$. The selection coefficient is $s = \frac{f}{f_a} - 1$, where f_a is the ancestral fitness (i.e., fitness of the mean phenotype of the population). The fixation probability of a mutation is given by $F = \frac{1-\exp(-2s)}{1-\exp(-4N_e s)}$ (Kimura 1962). At the end of each time unit, the phenotypic effect of a mutation is added onto the population mean with a probability of $2N_eF$ (when $2N_eF > 1$, it is treated as 1).

Phenotypic divergence and phylogenetic signal

For each trait with given V_M and n, we independently simulated its phenotypic evolution 50 times, all starting from the phenotypic optimum. Each simulation lasted for t = 1,000 units of time, after which the variance among the 50 replicates (R variance) was calculated at each time unit. Pearson's correlation coefficient between time and R variance at the time was calculated to represent the phylogenetic signal. We note that the length of the simulation (t) has a negligible effect on the simulation result, because the focal trait is neutral (or nearly so) and the R variance increases with time at a constant rate, as indicated by the high phylogenetic signal observed.

4.4 Results

4.4.1 Houle *et al.*'s method of matrix comparison is biased

To investigate the scaling relationship between mutational variance and evolutionary divergence among the fly wing morphological traits under investigation, Houle *et al.* compared the covariance matrices that represent the mutational inputs (*M*) and evolutionary divergences (*R*) of various traits along a set of orthogonal directions in the phenotypic space. For this purpose, Houle *et al.* used an eigenvector matrix *K* of the mean of *M* and *R* to determine the orthogonal directions, which could create a bias towards directions shared by *M* and *R* and inflate the correlation between $\log_{10}$(*M* variance) and $\log_{10}$(*R* variance). In theory, this potential bias can be avoided if *K* is derived solely from *M* such that the set of orthogonal directions are mutationally independent. We refer to this modified method as the new method.

To examine the potential bias of Houle *et al.*'s method, we first used Houle *et al.*'s method as well as the new method to analyze *M* and *R* matrices that were completely independently generated. We analyzed 100 such pairs, obtaining distributions of the linear regression slope and Pearson's correlation coefficient between $\log_{10}$(*R* variance) and $\log_{10}$(*M* variance). Both the slope (**Fig. 4-1a**) and the correlation coefficient (**Fig. 4-1b**) tend to be positive when we used Houle *et al.*'s method (mean slope = 0.844, $P = 3.76 \times 10^{-101}$, two-sided *t*-test, degree of freedom df = 99; mean correlation coefficient = 0.840, $P = 2.72 \times 10^{-78}$, two-sided *t*-test, df = 99), although the matrices being compared are independent from each other. By contrast, distributions of the slope (**Fig. 4-1a**) and correlation coefficient (**Fig. 4-1b**) when we used the new method are centered around 0, as expected (mean slope = 0.010, $P = 0.25$, two-sided *t*-test, df = 99; mean correlation coefficient = 0.059, $P = 0.12$, two-sided *t*-test, df = 99).

Because the observed *M* and *R* matrices are subject to estimation error, we also repeated the above analyses with such errors introduced. In particular, we drew samples from each *M* (or

R) matrix used in the previous analyses, denoted $M_{original}$ (or $R_{original}$), to obtain its estimate, denoted M_{obs} (or R_{obs}) (see Materials and Methods). We estimated, using both matrix comparison methods, the regression slope and the correlation coefficient between $\log_{10}(M_{obs}$ variance) and $\log_{10}(R_{obs}$ variance) across a wide range of similarity between $M_{original}$ and R_w, which is a weighted average of $M_{original}$ and $R_{original}$. The slope estimated by the new method tends to be lower compared with that in the absence of error in matrices to be compared, especially when $M_{original}$ and R_w are similar (**Fig. C-1**). Nevertheless, the difference is rather small, suggesting that the new method is sufficiently robust to error in matrix estimation. By contrast, estimates of the slope and the correlation coefficient by Houle *et al.*'s method are positive and almost invariable with respect to how similar R_w is to $M_{original}$ (**Fig. C-1**), indicating that Houle *et al.*'s method is not only biased but also uninformative.

In reality, when the true M is given, R_{obs} is subject to variation due to the stochasticity of evolution. To ensure that our finding holds for evolutionary data, we simulated phenotypic evolution along the fly phylogenetic tree used by Houle *et al.*, with the M matrix estimated from the fly wing data (Houle and Fierst 2013), M_{HF}, treated as the $M_{original}$. Trait evolution followed a multivariate Brownian motion model described by a matrix R_w, which was obtained by taking weighted averages of M_{HF} and an independent matrix $R_{original}$ (see Materials and Methods). Results of this analysis are generally similar to those based on randomly generated matrices and no trees. In particular, when M_{obs} and R_{obs} are compared using Houle *et al.*'s method, the linear regression slope of $\log_{10}(R_{obs}$ variance) on $\log_{10}(M_{obs}$ variance) is close to 1 regardless of the similarity between M_{HF} and R_w (**Fig. C-2**). Taken together, we conclude that comparing M and R variances along eigenvectors of an average matrix of them is uninformative and sheds little light on the real scaling relationship between M and R variances.

4.4.2 Unequal constraints on the fly wing traits

Using Houle *et al*.'s method, we reproduced their result of a slope of nearly 1 in the linear regression between $\log_{10}$(*M* variance) and $\log_{10}$(*R* variance) for fly wing traits (**Fig. 4-2a**). But our simulation suggested that this estimation is unlikely to be reliable. Indeed, when the fly wing data are reanalyzed using the new method, the slope reduced to 0.54, which is significantly smaller than 1 ($t = 6.27$, $P = 1.13 \times 10^{-5}$, two-sided *t*-test, df = 16; **Fig. 4-2b**). Applying the new method also caused a similar reduction in the slope of the linear regression between $\log_{10}$(*M* variance) and $\log_{10}$(*G* variance) ($t = 8.62$, $P = 2.09 \times 10^{-7}$, two-sided *t*-test, df = 16; **Fig. 4-2c, d**), where the *G* matrix represents intraspecific phenotypic variations caused by additive genetic (co)variances (Mezey and Houle 2005). The finding of a slope of approximately 0.5 for the linear regression between $\log_{10}$(*M* variance) and $\log_{10}$(*R* variance) or $\log_{10}$(*G* variance) indicates that *R* or G variance does not scale linearly with *M* variance. Rather, the positive impact of *M* variance on *R* or *G* variance gradually diminishes as *M* variance rises.

Interestingly, *G* and *R* matrices are indeed similar in the structure. When the *G* and *R* variances are compared along the eigenvectors of *G*, the regression slope between $\log_{10}$(*G* variance) and $\log_{10}$(*R* variance) is 0.95, which is not significantly different from 1 ($t = 0.72$, $P = 0.48$, two-sided *t*-test, df = 15). This observation is consistent with the view that standing genetic variation has a profound impact on long-term evolution (Schluter 1996).

4.4.3 A neutral model with mutational pleiotropy explains patterns of fly wing evolution

Our analysis shows that Houle *et al.*'s third observation on fly wing evolution mentioned in Introduction should be replaced with a sublinear relationship between the evolutionary rate of

a trait and its mutational variance (i.e., the scaling exponent is below 1). Even with Houle *et al.*'s third observation corrected, the nine existing models of phenotypic evolution previously considered by these authors are still unable to reconcile the three observations, because none of the models predicting a high phylogenetic signal as in the fly wing data allow a scaling exponent of about 0.5 that is observed in these data (**Table C-1**). Furthermore, the model proposed by Houle *et al*. that all wing traits concerned are affected by pleiotropy to the same extent is empirically false in addition to being theoretically implausible.

Below we show that patterns of fly wing evolution are explainable by a neutral model with mutational pleiotropy. In our model, the focal wing traits are effectively neutral, but mutations affecting the focal traits also influence non-focal traits that are subject to stabilizing selection. In addition, focal traits with higher *M* variances are genetically correlated with more non-focal traits so are impacted more by their mutational correlation to non-focal traits (see Materials and Methods). Our model makes three predictions that are respectively consistent with the three patterns of fly wing evolution. First, a focal trait is expected to evolve more slowly than predicted from the *M* variance, because most mutations affecting the focal trait are selectively removed due to their deleterious effects on correlated traits. Second, because the focal trait itself is effectively neutral, its divergence is unbounded, resulting in a high phylogenetic signal. Finally, the positive correlation between *M* variance and pleiotropy means that the fraction of mutations that are acceptable declines with *M* variance, creating a slope that is lower than 1 for the linear regression between $\log_{10}$(*M* variance) and $\log_{10}$(*R* variance) or $\log_{10}$(*G* variance).

To illustrate the above model predictions on long-term phenotypic evolution, we simulated the evolution of the population mean values of 20 neutral, orthogonal, focal traits, each

genetically correlated with a set of non-focal traits that are under stabilizing selection. Mutations were randomly generated per unit time and subject to drift and selection. The simulation lasted for 1000 time units and was repeated 50 times per trait to create 50 replicate lineages. For each trait, the phenotypic variance among the 50 lineages at each of the 1000 time units was used to represent the evolutionary divergence (*R* variance) at that time, and its correlation with time is a measure of the phylogenetic signal. The simulation results showed that, for most traits, the amount of phenotypic divergence is four to five orders of magnitude lower than predicted from the total mutational input (**Fig. 4-3a**, **b**). In addition, almost all traits exhibited phylogenetic signals exceeding 0.9 (**Fig. 4-3c**, **d**). Finally, the slope of the linear regression between $\log_{10}$(*M* variance) and $\log_{10}$(*R* variance) is significantly lower than 1 ($t = 8.49$, $P = 1.03 \times 10^{-7}$ when mutation rate is variable, two-sided *t*-test, df = 18; $t = 8.44$, $P = 1.13 \times 10^{-7}$ when mutation effect size is variable, two-sided *t*-test, df = 18; **Fig. 4-3a**, **b**). These results closely matched those observed in fly wing evolution, quantitatively verifying the validity and suitability of our model. Similar patterns were observed when the focal traits are not strictly neutral but are under weak selection (**Table C-2**), indicating that the observed patterns of fly wing evolution can also be explained even if the wing traits are not strictly neutral but have not been outside the range of effective neutrality in the last 40 million years.

4.5 Discussion

In this study, we showed that the method used by Houle *et al.* to compare matrices is biased, resulting in the erroneous conclusion of a linear relationship between mutational variance and evolutionary divergence among fly wing morphologies. We demonstrated by computer simulation that a simple modification of their method yields virtually unbiased results under a

wide range of parameter values, including those reflecting the fly wing data. Using the new method, we estimated that the scaling exponent between mutational variance and evolutionary divergence is significantly smaller than 1, suggesting that the impact of the rate of mutational input on the rate of phenotypic evolution is not constant but declines with the rate of mutational input. That is, compared with traits with relatively low mutational inputs, those with relatively high mutational inputs do not evolve as rapidly as predicted linearly from their mutational inputs. With these findings, patterns of fly wing evolution are explainable by a model in which the wing traits are effectively neutral, but their evolutionary rates are differentially reduced because mutations affecting these traits are purged owing to their different pleiotropic effects on other traits that are under stabilizing selection. When the extent to which the evolution of a trait is constrained by pleiotropy increases with the mutational variance of the trait, a scaling exponent between 0 and 1 will result. Mechanistically, such a relationship may have resulted from the positive correlation between the pleiotropic level of a mutation and its effect size on individual traits, a trend present among mouse skeletal characters and yeast cell morphologies (Wagner et al. 2008; Wang et al. 2010).

Our model differs from the one proposed by Houle *et al.* in several aspects. First, Houle *et al.* did not specify whether the focal traits are neutral or not in their model, whereas they are effectively neutral in our model. The neutrality appears to be key to the reconciliation between the observations of a high phylogenetic signal and slow evolution relative to the mutational input. Second, in our model, mutational pleiotropy arises from the non-focal traits genetically correlated with the focal trait concerned, but Houle *et al.* did not state whether their mutational pleiotropy arises from non-focal traits alone or also other focal traits. Third, all focal traits are constrained to the same level by mutational pleiotropy in Houle *et al.*'s model, which is a highly

restrictive assumption that is not supported by the fly wing data. By contrast, in our model, the level of mutational pleiotropy (arising from non-focal traits) is allowed to vary among focal traits. For example, the present analysis of the fly wing data suggests a positive relationship between the mutational variance of a focal trait and the level of mutational pleiotropy that the trait is subject to. Finally, Houle *et al.*'s model remains a verbal one, while we have provided quantitative formulation of our model (see Materials and Methods).

Our evolutionary simulation under the newly proposed model is able to recapitulate all major patterns observed in the evolution of the fly wing morphologies. Nevertheless, it is possible that the fly data also fit some other models. In particular, our results suggest the plausibility but do not prove that the fly wing traits are neutral. Indeed, an expanded model in which the focal traits are effectively neutral only within a range of phenotypic values can also explain fly wing evolution, provided that 40 million years of evolution under mutational pleiotropy has not reached the boundaries of this range (**Table C-2**). Regardless, our analysis suggests that fly wing evolution is explainable under the existing theoretical framework of phenotypic evolution.

The invaluable data collected by Houle *et al.* have allowed an unprecedented population genetic analysis of macroevolution of morphologies. To the best of our knowledge, no other large phenotypic data simultaneously comprising M, G, and R from long-term evolution exist. Only when many such data become available may we test the general applicability of our model or its expanded version in explaining phenotypic evolution, and only then can one tell whether the current theoretical framework of phenotypic evolution is generally correct.

4.6 References

Chakraborty, R. and M. Nei. 1982. Genetic Differentiation of Quantitative Characters between Populations or Species .1. Mutation and Random Genetic Drift. Genet Res 39:303-314.
Estes, S. and P. C. Phillips. 2006. Variation in pleiotropy and the mutational underpinnings of the G-matrix. Evolution 60:2655-2660.
Felsenstein, J. 1985. Phylogenies and the Comparative Method. American Naturalist 125:1-15.
Futuyma, D. J. 2010. Evolutionary constraint and ecological consequences. Evolution 64:1865-1884.
Hill, W. G. 1982. Rates of Change in Quantitative Traits from Fixation of New Mutations. P Natl Acad Sci-Biol 79:142-145.
Ho, W. C., Y. Ohya, and J. Z. Zhang. 2017. Testing the neutral hypothesis of phenotypic evolution. P Natl Acad Sci USA 114:12219-12224.
Houle, D., G. H. Bolstad, K. van der Linde, and T. F. Hansen. 2017. Mutation predicts 40 million years of fly wing evolution. Nature 548:447-450.
Houle, D. and J. Fierst. 2013. Properties of Spontaneous Mutational Variance and Covariance for Wing Size and Shape in Drosophila Melanogaster. Evolution 67:1116-1130.
Kimura, M. 1962. On the probability of fixation of mutant genes in a population. Genetics 47:713-719.
Lande, R. 1976. Natural-selection and random genetic drift in phenotypic evolution. Evolution 30:314-334.
Lynch, M. 1990. The Rate of Morphological Evolution in Mammals from the Standpoint of the Neutral Expectation. Am Nat 136:727-741.
Lynch, M. and W. G. Hill. 1986. Phenotypic evolution by neutral mutation. Evolution 40:915-935.
McGuigan, K., J. M. Collet, S. L. Allen, S. F. Chenoweth, and M. W. Blows. 2014. Pleiotropic mutations are subject to strong stabilizing selection. Genetics 197:1051-1062.
McGuigan, K., L. Rowe, and M. W. Blows. 2011. Pleiotropy, apparent stabilizing selection and uncovering fitness optima. Trends Ecol Evol 26:22-29.
Mezey, J. G. and D. Houle. 2005. The dimensionality of genetic variation for wing shape in Drosophila melanogaster. Evolution 59:1027-1038.
Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216-2218.
Qiu, W. and H. Joe. 2015. clusterGeneration: Random Cluster Generation (with Specified Degree of Separation). R package version 1.3.4. https://CRAN.R-project.org/package=clusterGeneration.
Revell, L. J., L. J. Harmon, R. B. Langerhans, and J. J. Kolbe. 2007. A phylogenetic approach to determining the importance of constraint on phenotypic evolution in the neotropical lizard Anolis cristatellus. Evol Ecol Res 9:261-282.
Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766-1774.
Turelli, M. 1985. Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. Genetics 111:165-195.
Wagner, G. P. and L. Altenberg. 1996. Perspective: Complex Adaptations and the Evolution of Evolvability. Evolution 50:967-976.

Wagner, G. P., J. P. Kenney-Hunt, M. Pavlicev, J. R. Peck, D. Waxman, and J. M. Cheverud. 2008. Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature 452:470-472.
Wang, Z., B. Y. Liao, and J. Z. Zhang. 2010. Genomic patterns of pleiotropy and the evolution of complexity. P Natl Acad Sci USA 107:18034-18039.

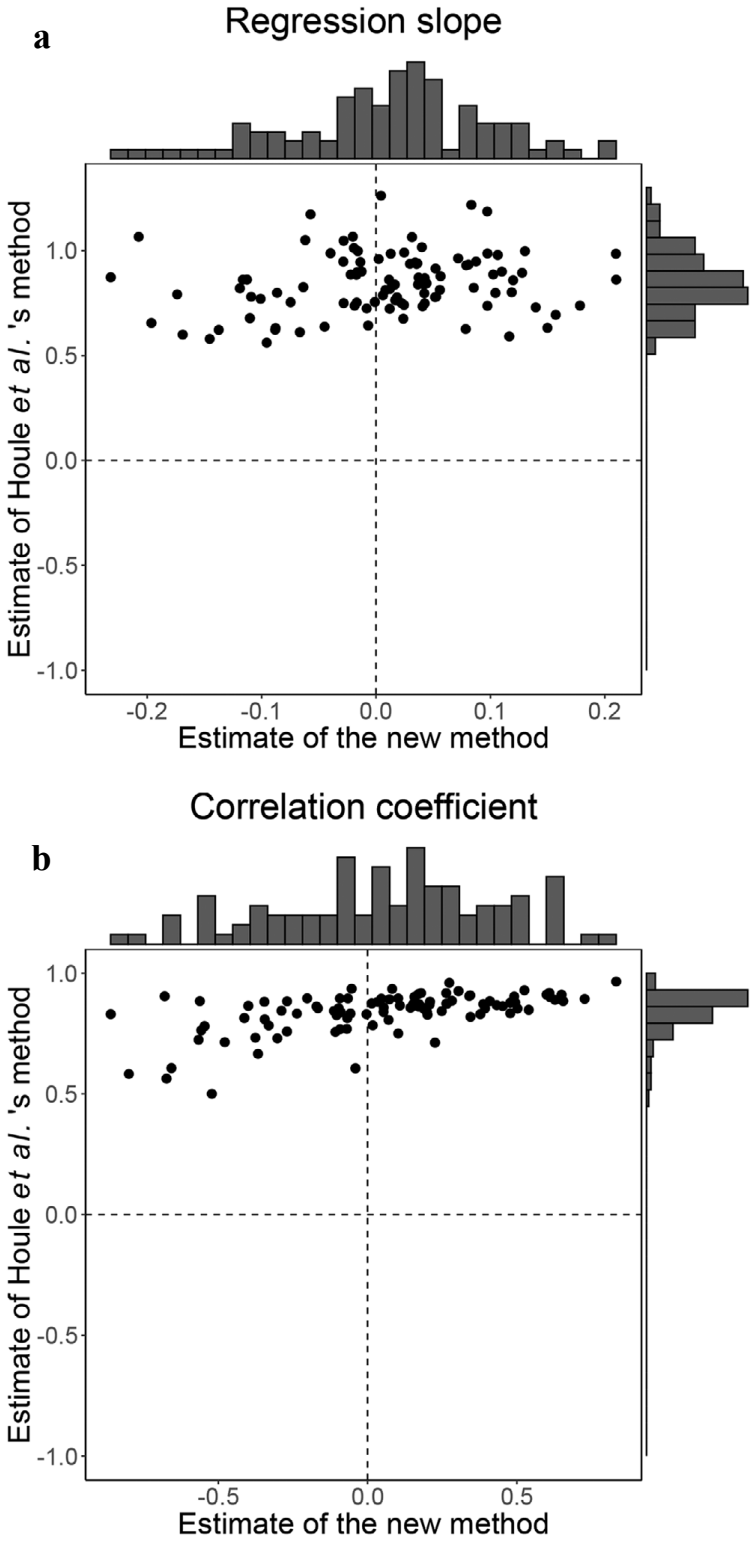

a
Regression slope
Estimate of Houle et al.'s method
1.0
0.5
0.0
-0.5
-1.0
-0.2
-0.1
0.0
0.1
0.2
Estimate of the new method
b
Correlation coefficient
Estimate of Houle et al.'s method
1.0
0.5
0.0
-0.5
-1.0
-0.5
0.0
0.5
Estimate of the new method

Figure. 4-1. Comparison between Houle *et al.*'s method and our new method. (**a**) Frequency distributions (grey bars) of the linear regression slope between $\log_{10}$(*M* variance) and $\log_{10}$(*R* variance) estimated by Houle *et al.*'s method (*y*-axis) and our new method (*x*-axis), when *M* and *R* matrices are independently generated. (**b**) Frequency distributions (grey bars) of the correlation between $\log_{10}$(*M* variance) and $\log_{10}$(*R* variance) estimated by Houle *et al.*'s method (*y*-axis) and our new method (*x*-axis), when *M* and *R* matrices are independently generated. In both panels, each dot represents a pair of uncorrelated matrices compared.

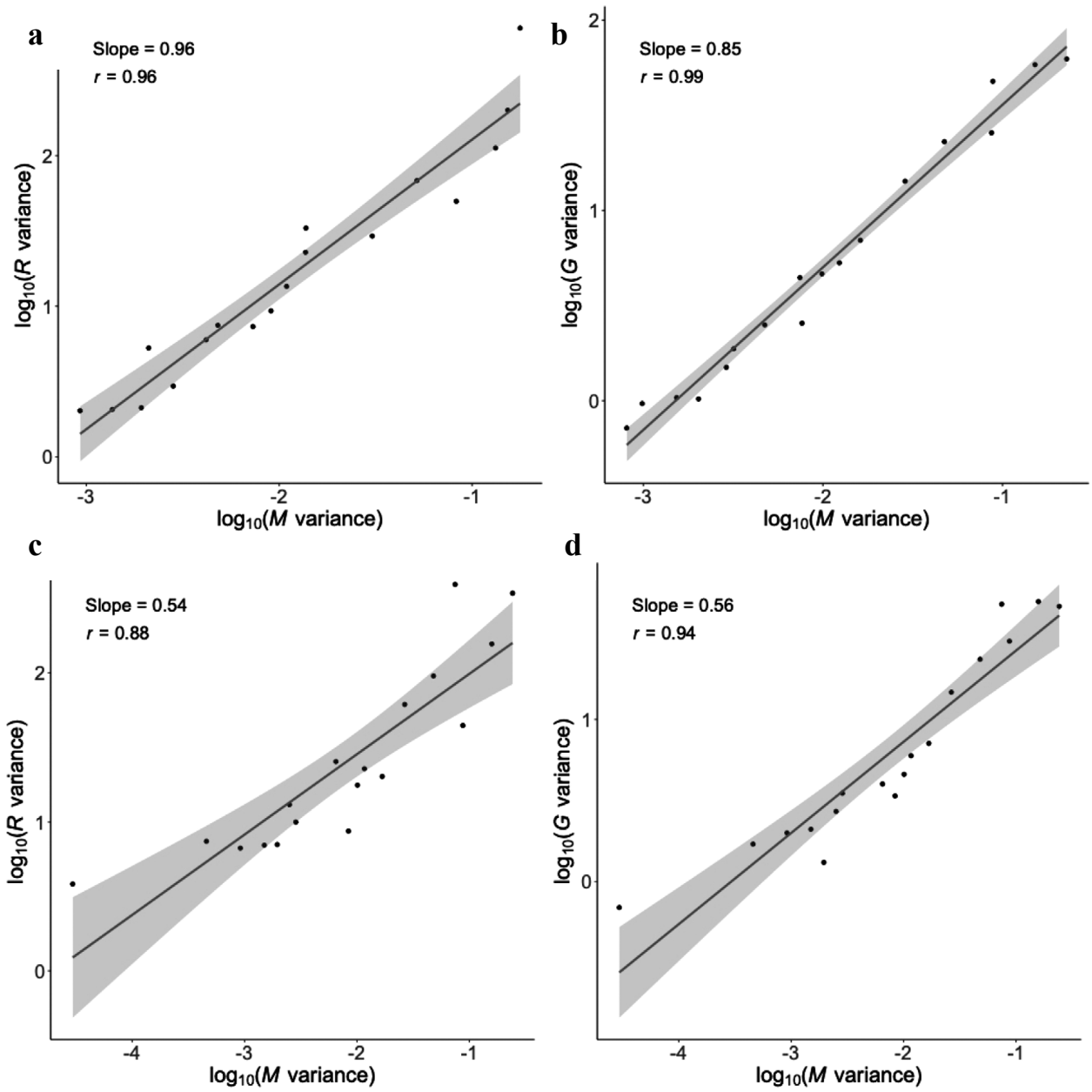


Figure. 4-2. Linear regression between $\log_{10}(M$ variance) and $\log_{10}(R$ variance) or $\log_{10}(G$ variance) along orthogonal directions for fly wing traits. (**a-b**) Linear regression between $\log_{10}(M$ variance) and $\log_{10}(R$ variance) estimated using Houle *et al.*'s method (**a**) or the new method (**b**). (**c-d**) Linear regression between $\log_{10}(M$ variance) and $\log_{10}(G$ variance) estimated using Houle *et al.*'s method (**c**) or the new method (**d**). In each panel, *r* stands for Pearson's

correlation coefficient, and the shaded region shows the 95% confidence interval of the regression. The number of orthogonal traits presented in each panel is the same as in Houle *et al*. (2017).

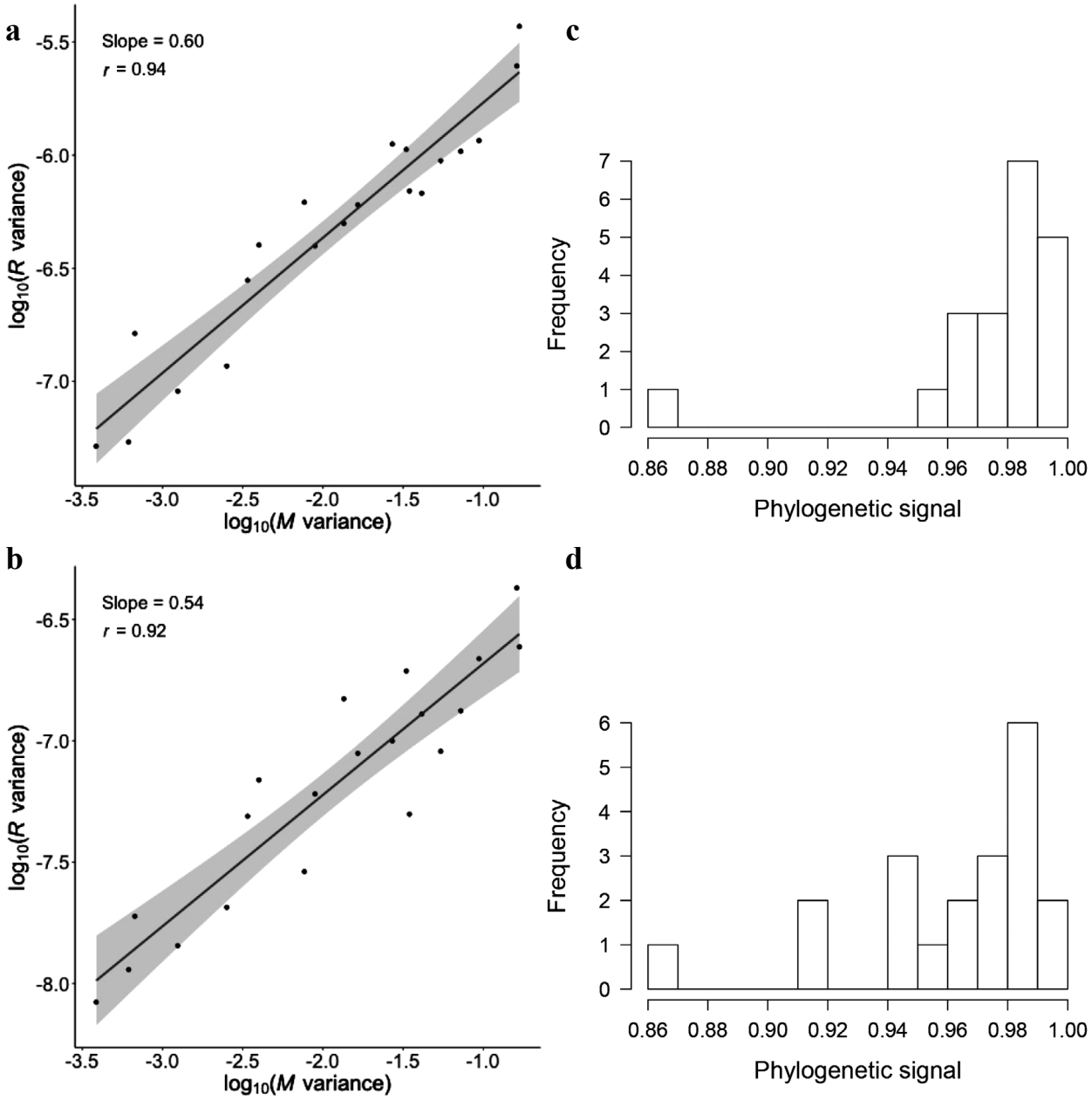


Figure. 4-3. Patterns of phenotypic evolution observed from computer simulation of 20 neutral, orthogonal, focal traits with mutational pleiotropy, when the between-trait variation in *M* variance is due to the variation in mutation rate (**a**, **c**) or effect size (**b**, **d**). (**a-b**) Linear regression between $\log_{10}$(*M* variance) and $\log_{10}$(*R* variance) upon evolution of 1000 time units. Presented are *M* and *R* variances per time unit. The shaded region shows the 95% confidence

interval of the regression. (**c-d**) Distribution of the phylogenetic signals of the focal traits. The phylogenetic signal of a trait is measured by Pearson's correlation between the evolutionary time and *R* variance at the time for the trait.

Chapter 5

Detecting Natural Selection in Trait-Trait Coevolution

5.1 Abstract

No phenotypic trait evolves independently of all other traits, but the cause of trait-trait coevolution is poorly understood. While the coevolution could arise simply from pleiotropic mutations that simultaneously affect the traits concerned, it could also result from multivariate natural selection favoring certain trait relationships. To gain a general mechanistic understanding of trait-trait coevolution, we examine the evolution of 220 cell morphology traits across 16 natural strains of the yeast *Saccharomyces cerevisiae* and the evolution of 24 wing morphology traits across 110 fly species of the family Drosophilidae, along with the variations of these traits among gene deletion or mutation accumulation lines (a.k.a. mutants). For numerous trait pairs, the phenotypic correlation among evolutionary lineages differs significantly from that among mutants. Specifically, we find hundreds of cases where the evolutionary correlation between traits is strengthened or reversed relative to the mutational correlation, which, according to our population genetic simulation, is likely caused by multivariate selection. Furthermore, we detect selection for enhanced modularity of the yeast traits analyzed. Together, these results demonstrate that trait-trait coevolution is shaped by natural selection and suggest that the

pleiotropic structure of mutation is not optimal. Because the morphological traits analyzed here are chosen largely because of their measurability, our conclusion is likely general.

5.2 Introduction

Many phenotypic traits covary during evolution. For example, the logarithm of brain weight and that of body weight show a nearly perfect linear relationship across mammals (Gould 1966; Huxley 1972; Lande 1979). In theory, four processes may explain such trait-trait coevolution. First, it could arise simply from pleiotropic mutations that simultaneously influence these traits with a more or less constant ratio of effects (Lande 1980; Wagner 1989; Wagner and Zhang 2011). Second, trait covariation could arise from the linkage disequilibrium between genes controlling these traits (Gardner and Latta 2007; Lande 2007; Wagner and Zhang 2011; Saltz et al. 2017), but such trait covariation is expected to be restricted to closely related individuals due to the deterioration of linkage disequilibrium as a result of recombination. If the linkage disequilibrium is stably maintained due to, for example, chromosomal inversion, the involved linked genes can be regarded as a supergene with mutational pleiotropy (Saltz et al. 2017). For this reason, linkage disequilibrium is negligible except for trait covariation among closely related individuals. Third, shared ancestry can also create apparent trait correlations across lineages, which, however, can be explained away when the phylogenetic relationships are taken into account in correlation analysis (Felsenstein 1985). Finally, trait covariation could be a result of natural selection for particular trait relationships that are advantageous (Lande 1979; Roff et al. 2002; Sinervo and Svensson 2002; Shoval et al. 2012; Bolstad et al. 2015; Svensson et al. 2021).

Despite a long-standing interest in trait correlation in evolution (Lande 1979; Wagner and Altenberg 1996; Saltz et al. 2017), which is also referred to as phenotypic integration in the literature (Olson and Miller 1999; Pigliucci 2003), our understanding of the roles of mutation and selection in trait-trait coevolution remains limited. Most studies on the subject focused on a small number of traits that are physiologically or ecologically important (Kingsolver et al. 2001), such as skull anatomy characters (Goswami et al. 2014; Porto et al. 2015; Simon et al. 2016; Watanabe et al. 2019; Fabre et al. 2020; Navalon et al. 2020), behavioral syndrome (i.e., sets of correlated behavioral traits) (Sih et al. 2004; Dochtermann and Dingemanse 2013), and ecological or organismal traits correlated with the metabolic rate (Martin 1981; Brown et al. 2004; Glazier 2010; Pettersen et al. 2016; White et al. 2019); hence, they may not provide a general, unbiased picture of trait-trait coevolution. Additionally, it is the trait correlation resulting from standing genetic variation and its effect on adaptation that have received the most attention (Schluter 1996; Steppan et al. 2002; Arnold et al. 2008; Agrawal and Stinchcombe 2009; Walsh and Blows 2009; Blows and Mcguigan 2015; Walter et al. 2018). But, because standing genetic variation could have been influenced by selection, the resulting trait correlation may not inform the correlation produced by mutation. Not knowing the mutational correlation hinders a full understanding of the contribution of selection.

Related to trait-trait correlation is the concept of modularity. It has been hypothesized that it is beneficial for organisms to have a modular organization such that functionally related traits belonging to the same module covary (Wagner and Altenberg 1996; Wagner 1999; Wagner et al. 2007; Goswami et al. 2014). Although modularity is a well-recognized feature of many trait correlation networks, the relative contribution of selection and mutational pleiotropy to modularity has not been assessed at the phenome scale (Wagner et al. 2007; Wang et al. 2010).

To gain a general mechanistic understanding of trait-trait coevolution, we study the phenotypic correlations for a large number of trait pairs at the levels of mutation and long-term evolution; natural selection is inferred when the evolutionary correlation between traits cannot be fully explained by the mutational correlation. Our primary data include 220 cell morphology traits of the budding yeast *Saccharomyces cerevisiae* that have been measured in 4817 single-gene deletion lines (Ohya et al. 2005), 89 mutation accumulation (MA) lines (for a subset of 187 traits) (Geiler-Samerotte et al. 2016), and 16 natural strains with clear phylogenetic relationships (Ohya et al. 2005; Yvert et al. 2013). These traits were quantified from fluorescent microscopic images of triple-stained cells and were originally chosen for study because of their measurability regardless of potential roles in evolution and adaptation (Ohya et al. 2005). Subsequent studies found that these traits are correlated with the yeast mitotic growth rate to varying degrees (Ho and Zhang 2014). Hence, these traits may be considered representatives of phenotypic traits that have different contributions to fitness. Previous analyses of these traits among natural strains unveiled signals of positive selection on individual traits (Ho et al. 2017), but their potential coevolution has not been studied. While studying these trait pairs can offer a general picture of trait-trait coevolution, we recognize that the selective agent would be hard to identify should selection be detected because the biological functions of these traits (other than correlations with the growth rate) are generally unknown (Ho et al. 2017). To verify the generality of the findings from the yeast traits, we analyze another dataset that includes 12 landmark vein intersections on the fly wings that have been measured in 150 MA lines of *Drosophila melanogaster* (Houle and Fierst 2013) and 110 Drosophilid species (Houle et al. 2017). Because each intersection is described by two coordinates, which are counted as two traits, there are 24 traits in this dataset. In both datasets, we discover that the evolutionary correlation differs significantly from the

mutational correlation for numerous trait pairs, revealing a role of natural selection in trait-trait coevolution. We also provide evidence for selection for enhanced modularity of the yeast traits.

5.3 Results

5.3.1 Evolutionary correlations differ from mutational correlations for many trait pairs

To investigate if trait correlations in evolution can be fully accounted for by the correlations generated by mutation, we examined all pairs of the 220 yeast cell morphology traits previously measured. For each pair of traits, we computed the mutational correlation COR_M, defined as Pearson's correlation coefficient across 4,817 gene deletion lines (upper triangle in **Fig. 5-1a**), and evolutionary correlation COR_E, defined as Pearson's correlation coefficient across 16 natural strains (lower triangle in **Fig. 5-1a**) with their phylogenetic relationships (**Fig. D-1**) taken into account (see Materials and Methods). COR_M computed from gene deletion lines are used as a proxy for correlation resulting from spontaneous mutations, assuming effects of null mutations are roughly proportional to average effect of mutations in the corresponding genes. Note that the original data contained 37 natural strains (Yvert et al. 2013), of which 21 belong to the "mosaic" group (Liti et al. 2009; Peter et al. 2018)—their phylogenetic relationships with other *S. cereviase* strains vary among genomic regions—so cannot be included in our analysis that requires considering phylogenetic relationships. Most of the yeast strains (all except BY4743 and Y12) belong to the "Wine/European" group (Liti et al. 2009; Peter et al. 2018) and form a clade distinct from other strains (**Fig. D-1**). As a result, divergence between this clade and other strains represents a large proportion of evolutionary divergence among the yeast strains has a substantial contribution to evolutionary (co)variances and correlations.

We found that the frequency distribution of COR_E across all trait pairs differs significantly from that of COR_M (**Fig. 5-1b**), suggesting the action of selection. For each pair of traits, we transformed the COR_M and COR_E to *Z*-scores using Fisher's *r*-to-*Z* transformation and conducted *Z*-test to determine whether the two correlations are significantly different. Of the 24,090 trait pairs examined, 6743 pairs (or 28.0%) have a COR_E that deviates significantly from COR_M at the false discovery rate (FDR) of 5% (**Table 1**, **Data S1**), suggesting that natural selection has shaped the coevolution of many trait pairs. To investigate whether the above result is biased because of the use of each trait in many trait pairs, we randomly arranged the 220 traits into 110 non-overlapping pairs and counted the number of pairs with COR_E significantly different from COR_M. This was repeated 1,000 times to yield 1,000 estimates of the proportion of trait pairs with significantly different COR_E and COR_M. The middle 95% of these estimates ranged from 14.5% to 40.1%, with the median estimate being 28.2%, almost identical to the result (28.0%) from all pairwise comparisons. Hence, there is no indication that using overlapping trait pairs has biased the estimate of the fraction of trait pairs with significantly different COR_E and COR_M.

To further test selection, we simulated neutral evolution along the yeast tree 1000 times under a Brownian motion model with the observed mutational covariance matrix *M* used as the mutational input, generating 1,000 simulated datasets. Before the simulation, we confirmed that the sampling error of our estimated *M* is negligible, likely because of the large number of mutants used in *M* estimation (**Table D-1**; see Materials and Methods). From each simulated dataset, we calculated the number of trait pairs with COR_E significantly different from COR_M. Only in 0.7% of the simulated data did we find this number equal to or greater than that from the actual data (**Table 5-1**), indicating that the observed evolutionary correlations between traits

cannot be explained by the neutral Brownian motion model. The distribution of mutational effects can be asymmetric and skewed (Hodgins-Davis et al. 2019) while it is assumed normal in the Brownian motion model. Nevertheless, simulations showed that mutational bias will not render COR_E deviate from COR_M in the absence of selection and will not enlarge the variance of COR_E (**Table D-2**; see Materials and Methods).

We divided the 6743 cases of significantly different COR_E and COR_M into three categories. In the first category, the trait correlation generated by mutation is strengthened by natural selection during evolution. A total of 2,727 trait pairs are considered to belong to this "strengthened" category (**Table 5-1**) because they satisfy the following criteria: COR_E and COR_M have the same sign and $|COR_E| > |COR_M|$, or COR_E and COR_M have different signs but only COR_E is significantly different from 0 (at the nominal *P*-value of 0.05) (**Fig. 5-2a**). In the second category, the trait correlation generated by mutation is weakened by natural selection during evolution. A total of 1,221 trait pairs satisfying the following criteria are classified into this "weakened" category (**Table 5-1**): COR_E and COR_M have the same sign and $|COR_E| < |COR_M|$, or COR_E and COR_M have different signs but only COR_M is significantly different from 0 (**Fig. 5-2b**). In the last category, the trait correlation generated by mutation is reversed in sign by natural selection during evolution. A total of 2,795 trait pairs satisfying the following criteria are in this "reversed" category (**Table 5-1**): COR_E and COR_M have different signs and are both significantly different from 0 (**Fig. 5-2c**).

To assess the robustness of the selection signals detected, we repeated the above analysis using COR_M estimated from 89 mutation accumulation (MA) lines (Geiler-Samerotte et al. 2016) (**Fig. D-2a**). Again, the overall frequency distribution across all trait pairs differs significantly between COR_E and COR_M (**Fig. D-2b**). We found that 5,146 trait pairs exhibit a significantly

different COR_E from the corresponding COR_M (**Table 5-1**), supporting a role of selection in the coevolution of many trait pairs. When comparing the analysis using COR_M from gene deletion lines and that using COR_M from MA lines, we found 990 trait pairs to exhibit selection signals and fall into the same category in both analyses, including 275 pairs in the "strengthened" category, 223 pairs in the "weakened" category, and 574 pairs in the "reversed" category. All of these numbers substantially exceed the corresponding expected random overlaps ($P < 0.001$ based on 1,000 random draws in each case; the medians across the 1,000 draws are 271, 68 and 163, respectively), suggesting the reliability of both analyses. Although mutations in MA lines are more natural than those in gene deletion lines, the number of MA lines is much smaller than the number of gene deletion lines and only 187 of the original 220 traits were measured in the MA lines. For these reasons, we focused on the COR_M estimated from the gene deletion lines in subsequent analyses.

To examine the generality of the above yeast-based findings, we analyzed the 24 wing morphology traits of Drosophilid flies. The COR_M and COR_E have been previously estimated from 150 MA lines (Houle and Fierst 2013) and 110 Drosophilid species (Houle et al. 2017), respectively (**Fig. D-3a**). The overall frequency distribution across all trait pairs differs significantly between COR_E and COR_M (**Fig. D-3b**). Of the 276 pairs of traits, 144 (52.2%) showed a significantly difference between COR_E and COR_M (**Table 5-1**), indicating widespread actions of selection in the coevolution of fly wing morphology traits.

Together, these results demonstrate that, for many trait pairs, mutational and evolutionary correlations between morphological traits are more different than expected under neutrality. This observation suggests an important role of selection in shaping the strength and/or direction of trait correlation in evolution.

5.3.2 Effects of different selection regimes on trait-trait coevolution

The strengthened, weakened, and reversed trait correlations in evolution may have resulted from different selection regimes. Below we consider various selection regimes that could potentially explain these types of difference between COR_M and COR_E (**Fig. 5-3**). First, when a specific allometric relationship between two traits is selectively favored, the population mean trait values are expected to be concentrated near the fitness ridge or the optimal allometric line, resulting in a strong evolutionary correlation between the traits (i.e., a high $|COR_E|$) (**Fig. 5-3a**). Unless COR_M is already similar to COR_E, we expect to see strengthened or reversed COR_E depending on COR_M. Second, if there is a single fitness peak for an optimal combination of trait values and if there is sufficiently strong stabilizing selection on the optimal phenotype, the population mean phenotype should be restricted within a small range of the optimal phenotype in all directions in the phenotypic space regardless of the mutational variance. Consequently, COR_E is expected to be close to 0, which could account for a weakened evolutionary correlation relative to the mutational correlation (**Fig. 5-3b**). Finally, if the fitness optimum varies across lineages in a random fashion, the steady-state COR_E will be close to zero, potentially leading to the weakening of the evolutionary correlation relative to the mutational correlation (**Fig. 5-3c**); in a special case, if there are only two peaks, a strong correlation could occur as the population means end up centered about the two peaks, potentially leading to strengthening or reversal of the correlation.

To verify these predictions, we simulated the evolution of two traits. Under each parameter set, we simulated 50 independent replicate lineages and computed the correlation coefficient, or COR_E, between the traits across the replicate lineages at the end of the simulated

evolution. This was repeated 200 times to obtain an empirical distribution of COR_E. To evaluate the difference between COR_M and COR_E, we examined the location of COR_M in the distribution of COR_E; a significant ($P < 0.05$) difference is inferred if COR_M is in the left or right 2.5% tail of the COR_E distribution.

As expected, in the absence of selection, the distribution of COR_E is centered around COR_M (first block in **Table 5-2**). When a specific allometric relationship is selectively favored, a high $|COR_E|$ always emerges regardless of the COR_M used, resulting in either strengthened or reversed evolutionary correlations ($P < 0.005$ for all parameter sets examined; the second to fifth blocks in **Table 5-2**). By contrast, stabilizing selection of an optimal phenotype leads to weakened correlation across replicate lineages when $|COR_M|$ is not small (sixth block in **Table 5-2**). Finally, when different lineages have different phenotypic optima that are randomly picked from the standard bivariate normal distribution, weakened evolutionary correlations are generally observed except when COR_M is close to zero (bottom block in **Table 5-2**). These results suggest that the strengthened and reversed evolutionary correlations of yeast and fly morphological traits can result from selections of allometric relationships, while the weakened correlations are likely caused by selections of individual traits either when there is a single optimal phenotype or when the optimal phenotype randomly varies among lineages.

5.3.3 Selection for enhanced modularity of yeast morphological traits

While all of the above analyses focused on individual trait pairs, here we ask whether the overall trait correlation across divergent lineages is stronger or weaker than that created by mutation. As a measure of the overall level of trait correlation (i.e., overall integration), we calculated the variance of eigenvalues (V_{eigen}) of the correlation matrix from divergent lineages

and mutants, respectively. A greater V_{eigen} corresponds to a stronger overall correlation between traits because the eigenvalues become less evenly distributed as the absolute values of the correlation coefficients become larger (Pavlicev et al. 2009). However, the sample size (i.e., the number of strains) in the estimation of the correlation matrix also has an effect on V_{eigen}; a matrix estimated from a smaller sample naturally tends to have fewer positive eigenvalues and greater V_{eigen}. To exclude the influence of this factor, we randomly sampled the mutant strains to obtain 5000 control datasets. Because the rank number of the evolutionary correlation matrix is 15 for the yeast data (i.e., 15 positive eigenvalues), each control dataset also consists of 15 randomly drawn strains such that the corresponding mutational correlation matrix also has 15 positive eigenvalues. We examined the location of the observed V_{eigen} in this distribution and computed a *P*-value based on this location (see Materials and Methods). For the yeast traits, V_{eigen} of the observed evolutionary correlation matrix exceeds that in 96% of control datasets ($P = 0.08$ in a two-tailed test; **Table 5-3**). Furthermore, only two of the 5000 control datasets have V_{eigen} significantly different from that of the observed evolutionary correlation matrix (Fligner-Kileen test). Hence, there is little evidence for a difference between the overall evolutionary correlation and the overall mutational correlation in yeast. For the fly data, the number of positive eigenvalues is unlimited by the sample size for both the evolutionary and mutational correlation matrices, hence we directly compared V_{eigen} between the two matrices, but found them to be similar ($P = 0.459$, Figner-Kileen test; **Table 5-3**). We also compared the overall integration between yeast and flies using $V_{eigen}/(n\text{-}1)$, where n is the number of traits examined. $V_{eigen}/(n\text{-}1)$ equals 0.157 and 0.268 for the yeast mutational and evolutionary matrices, respectively, whereas the corresponding values in flies are 0.153 and 0.190, respectively.

In addition to the overall level of trait correlation, we asked whether the correlational structure of traits exhibits different levels of modularity among divergent lineages when compared with that among mutants. To this end, we used a covariance ratio (*CR*) test (Adams 2016) that compares covariance within and between pre-defined modules (see Materials and Methods). Specifically, we calculated *CR* for the evolutionary covariance matrix and compared it to the *CR* distribution based on 5000 mutational covariance matrices estimated from the randomly drawn subsets of mutants aforementioned. We treated the three non-overlapping categories of the yeast morphological traits—actin traits, nucleus traits, and cell wall traits (Ohya et al. 2005)—as three modules (**Data S1**). We found that the *CR* of the evolutionary covariance matrix exceeded that of every control dataset ($P < 0.001$; **Table 5-3**), suggesting natural selection for increased modularity in evolution.

5.4 Discussion

By comparing the trait-trait correlation across mutants (COR_M) with that across divergent lineages (COR_E) for 24,090 pairs of yeast cell morphology traits and 276 pairs of fly wing morphology traits, we detected the action of natural selection in trait-trait coevolution. The fraction of trait pairs showing evidence for selection is substantially higher in the fly (52%) than yeast (28%) data ($P < 10^{-4}$, chi-squared test). This is at least in part caused by a difference in statistical power, because the number of strains/species used for estimating COR_E is much greater for the fly (110) than yeast (16) data. It is likely that a higher fraction than 28% of the yeast trait pairs are subject to selection in their coevolution. Furthermore, our comparison between COR_E and COR_M intends to test selection on trait correlations common among the evolutionary lineages considered. If different evolutionary lineages have different trait

correlations, the COR_E estimated from all lineages may not be significantly different from COR_M even when selection occurs in some or all of the lineages. In other words, our test is expected to underestimate the proportion of trait pairs subject to selection.

We demonstrated by population genetic simulations that various selection regimes can explain differences between COR_M and COR_E. In particular, strengthened or reversed COR_E relative to COR_M can occur when a specific allometric relationship is preferred, while weakened COR_E can occur under directional or stabilizing selection of individual traits. A notable difference between the simulation results and empirical observations is that the simulations tend to end up with extreme values of $|COR_E|$ (i.e., close to either 1 or 0) except in the case of neutrality, whereas the empirically observed $|COR_E|$ is usually less extreme even when COR_M and COR_E are significantly different. This is due to the fact that the simulation results usually represent steady-state correlations across lineages. That is, the mean phenotype of each lineage is at or near the corresponding optimum (if any); consequently, $|COR_E|$ is close to 1 when the optimum is a line and close to 0 when the optimum is a single combination of two trait values. However, the population mean phenotypes may not be close to their optima in some strains because of recent changes of the optima or the sparsity of mutations toward the optima, the latter of which is well known as a potential hindrance to adaptation (Schluter 1996; Hansen and Houle 2008; Agrawal and Stinchcombe 2009; Blows and Mcguigan 2015). Another possibility is the existence of a wide range of preferred allometry such that there is no strong selection for extreme $|COR_E|$. Finally, selection may not result in the preferred allometry between two traits because of the constraints from unconsidered traits (Houle et al. 2019).

It is worth noting that the yeast natural strains had been cultured in synthetic media before phenotyping (Yvert et al. 2013) while the mutant strains were all grown in the rich

medium YPD (Ohya et al. 2005; Geiler-Samerotte et al. 2016). Hence, it remains a possibility that the difference between COR_E and COR_M reported here contains a component caused by the environmental difference in phenotyping. Notwithstanding, our analysis suggests that this component is small (see Materials and Methods), which is expected because both media are meant to provide an ideal, stress-free environment for yeast growth. This said, future phenotyping in the same medium will be needed to validate our findings.

While selection was detected for many trait pairs, a large fraction of trait pairs, especially in the yeast data, do not show a significant difference between COR_E and COR_M. These trait pairs may be divided into two groups. In the first group, COR_E and COR_M are actually different, but the difference is not found significant due to the limited statistical power. As mentioned, we believe that a substantial fraction of yeast trait pairs belong to this category due to the relatively low statistical power in detecting the difference between COR_E and COR_M in the yeast data. In the second group, COR_E truly equals COR_M, which could result from one of the following three scenarios. First, the specific trait-trait correlation does not impact fitness so evolves neutrally. Second, the two traits have an intrinsic, immutable relationship (such as the hypothetical traits of body size and twice the body size), so will yield equal COR_E and COR_M; this possibility can be tested by examining the correlation of the two traits across isogenic individuals that show non-heritable phenotypic variations (Geiler-Samerotte et al. 2020). The last and perhaps the most interesting scenario is that the trait-trait correlation impacts fitness and hence has driven the optimization of COR_M via a second-order selection (Wagner 2005; Hansen and Houle 2008; Ho and Zhang 2014) such that the first-order selection of mutations that affect the two traits is no longer needed. However, the relative frequencies of these three scenarios are unknown.

In addition to pairwise trait correlations, we tested hypotheses regarding the evolution of overall phenotypic integration and modularity. In the yeast data, we observed a higher modularity across natural strains than across mutants but did not find evidence for a change of overall phenotypic integration in evolution. These results support the view of increasing modularity during evolution (Wagner and Altenberg 1996; Wagner 1999; Wagner et al. 2007; Clune et al. 2013; Goswami et al. 2014) but also suggest that modularity is enhanced by both strengthening trait-trait correlations within modules and weakening trait-trait correlations across modules. We found the overall integration lower for the fly than yeast traits, but whether this observation indicates a difference between different types of traits (i.e., cellular traits and multicellular organisms' morphological traits) or between multicellular and unicellular organisms requires analyzing more species and traits.

Our analysis compared COR_M estimated from one yeast strain (BY) with COR_E estimated from 16 different strains, under the assumption of a constant COR_M across different strains. While it is a common practice to assume that the mutational architecture is more or less constant during evolution and to study phenotypic evolution by comparing mutational or genetic (co)variances in one species with those among different species (Lynch 1990; Ackermann and Cheverud 2004; Houle et al. 2017), genetic variations affecting the genetic (co)variances of phenotypic traits have been reported (Pavlicev et al. 2008; Jones et al. 2014; Jerison et al. 2017). As discussed earlier, such genetic variations may allow second-order selection of COR_M. For instance, it has been hypothesized that the optimization of mutational (co)variances driven by selection for mutational robustness and/or adaptability can lead to modularity (Wagner and Altenberg 1996; Wagner et al. 2007), but this process presumably takes a much longer time than is relevant to the present study. Even without second-order selection, COR_M could still vary

across strains because the pleiotropic effects of a mutation can vary by the environment and genetic background (Pavlicev and Cheverud 2015; Wei and Zhang 2019; Svensson et al. 2021). Regardless, in the future, it would be desirable to measure mutant phenotypes from multiple lineages to investigate whether COR_M evolves, how rapidly it evolves, and whether its evolution is largely neutral or adaptive.

In summary, we detected the action of natural selection in shaping trait-trait coevolution. Because the traits analyzed here, especially the yeast traits, were chosen almost exclusively due to their measurability, our results likely reflect a general picture of trait-trait coevolution. Measuring these yeast traits in additional divergent natural strains with clear phylogenetic positions could improve the statistical power and clarify whether the fraction of trait pairs whose coevolution is shaped by selection is much greater than detected here. Finally, the detection of selection for enhanced modularity of the yeast traits analyzed supports the hypothesis that modularity is beneficial (Wagner and Altenberg 1996; Goswami et al. 2014). The detection of selection in trait-trait coevolution and selection for enhanced modularity suggests that the current pleiotropic structure of mutation is not optimal. This nonoptimality could be due to the weakness of the second-order selection on mutational structure and/or a high dependence of the optimal mutational structure on the environment, which presumably changes frequently. Future studies on how the mutational structure evolves will likely further enlighten the mechanism of trait-trait coevolution.

5.5 Materials and Methods

5.5.1 Phenotypic data

The *S. cerevisiae* cell morphology traits were previously measured by analyzing fluorescent microscopic images. Three phenotypic datasets were compiled and analyzed in this study, including (i) 220 traits measured in 4,718 gene deletion lines that each lack an nonessential gene (Ohya et al. 2005), (ii) the same 220 traits measured in 37 natural strains (Yvert et al. 2013), and (iii) 187 of the 220 traits measured in 89 mutation accumulation (MA) lines (Geiler-Samerotte et al. 2016). When comparing patterns of trait correlation between two datasets, we used traits available in both datasets.

Before the analyses, we first standardized all trait values by converting each trait value to the natural log of the ratio of the original trait value to a reference. The standardized value of the *i*th trait in the *j*th strain is $\tilde{X}_{i,j} = \ln \frac{X_{i,j}}{X_{i,r}}$, where $X_{i,j}$ is the original trait value and $X_{i,r}$ is the trait value of the reference. For the gene deletion lines, the reference is the wild-type BY strain. For the MA lines, the reference is the progenitor strain used in MA. For natural strains, the reference is the same as the reference of the mutant strains to be compared with (i.e., wild-type BY or progenitor of the MA lines).

The locations of 12 vein intersections on the fly wing were previously measured in 150 MA lines of *Drosophila melanogaster* and a mutational covariance matrix was estimated (Houle and Fierst 2013). These traits were also measured in 110 Drosophilid species and an evolutionary covariance matrix was estimated with species phylogeny taken into account (Houle et al. 2017). Both matrices are based on log-scale trait values.

5.5.2 Influence of the sampling error on the *M* matrix

To evaluate the influence of sampling error on the estimated *M* matrix of yeast or fly, we took samples (vectors of phenotypes) from the multivariate distribution of *M* (4,817 samples for

yeast gene deletion data and 150 samples for fly MA data), estimated a covariance matrix ($\widetilde{M}$) from these samples, and calculated Pearson's correlation coefficient between the eigenvalues of M and $\widetilde{M}$. This was repeated 1,000 times and the distribution of the correlation coefficient was used to evaluate the potential impact of sampling error on M.

5.5.3 Impact of the environmental difference on the correlational structure of the yeast traits

Because the natural strains of yeast had been grown in synthetic media before phenotyping (Yvert et al. 2013) while the mutant strains were all grown in the rich medium YPD (Ohya et al. 2005; Geiler-Samerotte et al. 2016), we tested whether this environmental difference affected the correlational structure of the yeast morphological traits under consideration. Specifically, we examined whether the phenotype of the BY strain grown in synthetic media (referred to as "synthetic phenotype" for short) falls in the distribution of 123 biological replicates of BY grown in YPD (referred to as "YPD phenotypes" for short). The phenotypes were normalized in the way described earlier with the mean phenotype of the YPD replicates used as the reference. We decomposed YPD phenotypes into principal components (PCs) and focused on the first three PCs, which together explained 67.5% of the variance among the 123 YPD phenotypes. We then calculated the values of the three PC traits of the synthetic phenotype. The synthetic phenotype is in the central 95% of the distribution of the YPD phenotypes for each of the three PC traits, indicating a lack of major effect of the difference between synthetic and YPD media on the correlational structure of the yeast traits concerned.

5.5.4 Comparison between mutational and evolutionary correlations

To take into account the phylogenetic relationships among yeast strains in estimating COR_E, we utilized a distance-based tree previously inferred (Peter et al. 2018) (**Fig. D-1**). Strains with mosaic origins inferred in the same study (Peter et al. 2018) were removed before analysis, resulting in 16 remaining natural strains. Because the BY strain was not included in the data file in that study (Peter et al. 2018), W303, a laboratory strain closely related to BY, was chosen to represent BY. We obtained the evolutionary covariance matrix using the *ratematrix* function from the R package *geiger* (Revell et al. 2007; Pennell et al. 2014), which calculates evolutionary covariances using the independent contrast method (Felsenstein 1985). The evolutionary covariance matrix was then converted to the corresponding correlation matrix.

To test whether the observed pairwise trait correlation at the level of evolutionary divergence is significantly different from that expected by mutation alone for each pair of traits, we first converted both correlations to *Z*-scores by $Z = \frac{1}{2}[\ln(1+r) - \ln(1-r)]$, where r is the correlation coefficient. The testing statistic was then computed by $Z = \frac{Z_E - Z_M}{\sqrt{\frac{1}{n_E - 3} + \frac{1}{n_M - 3}}}$, where Z_E and Z_M are *Z*-scores converted from COR_E and COR_M, respectively, n_E is the number of independent contrasts, which equals the number of natural strains minus one, and n_M is the number of mutant strains. Two-sided *P*-value was calculated from each *Z* and converted to adjusted *P*-value following the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995). An adjusted *P*-value below 0.05 indicates selection.

To see how many trait pairs would show a significant difference between COR_E and COR_M under neutrality, we simulated neutral evolution along the phylogenetic tree that had been used in estimating COR_E. A Brownian motion model was used to simulate neutral phenotypic evolution such that the amount of evolution in branch i is $M_i l$, where M_i is a vector sampled from the multivariate normal distribution of the mutational covariance matrix M and l is the

branch length. Sampling was performed using the *rmvnorm* function in the R package *mvtnorm* (Genz 2020). The starting value of each trait is 0 in all simulations. The phenotypic value of each strain was obtained by adding up the amount of evolution on all branches ancestral to the strain. This was repeated 1,000 times to generate 1,000 datasets.

To account for the difference in V_{eigen} caused by different sample sizes in estimating the correlation matrices, we randomly sampled subsets of the gene deletion strains. Because the evolutionary correlation matrix has a rank number of 15 and has 15 positive eigenvalues, each subset consists of 15 strains randomly drawn from the 4718 gene deletion strains such that the mutational correlation matrix computed from each subset of mutants also has 15 positive eigenvalues. From each subset of strains, we computed V_{eigen}, leading to a null distribution of V_{eigen}. The observed V_{eigen} from the evolutionary correlation matrix is then compared to the null distribution; a significant difference is inferred if the observed value falls in either the left or right 2.5% tail.

To test whether there exists a significant modular structure among traits, we performed the covariance ratio (*CR*) test. For each pair of predefined modules, traits were first re-ordered such that traits belonging to each module were located in the upper-left and lower-right corners of the covariance matrix, respectively, and $CR = \sqrt{\frac{trace(M_{12}M_{21})}{\sqrt{trace(M_{11}^{*}M_{11}^{*})+trace(M_{22}^{*}M_{22}^{*})}}}$, where M_{12} and M_{21} are the upper-right and lower-left sections of the original covariance matrix, respectively, containing all between-module covariances, M_{11}^{*} is the upper-left section with diagonal elements replaced by zeros, M_{22}^{*} is the lower-right section with diagonal elements replaced by zeros, and $trace(M)$ denotes the trace, or the sum of diagonal elements, of matrix *M* (Adams 2016). Because three modules were defined in the yeast data, the average of all pairwise *CR* values was

used to represent the overall modularity. A test for selection on *CR* was performed following the test of selection on V_{eigen}.

5.5.5 Computer simulation of trait-trait coevolution under selection

In each simulation, we considered a pair of traits with equal amounts of mutational variance V_{M}, which is set to be 0.01. The mutational covariance matrix is thus $M = \begin{bmatrix} V_M & COV_M \\ COV_M & V_M \end{bmatrix} = \begin{bmatrix} V_M & V_M COR_M \\ V_M COR_M & V_M \end{bmatrix}$, where COV_{M} is the mutational covariance. The number of mutations is a random Poisson variable with the mean equal to 1. The phenotypic effect of a mutation is drawn from the multivariate normal distribution of *M* using the *rmvnorm* function in the R package *mvtnorm* (Genz 2020). The starting phenotype is (0, 0) in all simulations.

We considered a Gaussian fitness function of $f = \exp\left(-\frac{D^2}{2}\right)$, where *f* is the fitness and *D* is the distance between the current phenotype and the optimal phenotype. When there is a single fitness peak (i.e., the fitness optimum is a single point), *D* is the Euclidean distance defined by $\sqrt{{d_1}^2 + {d_2}^2}$, where d_1 and d_2 are the distances between the current phenotypic values of the two traits and their corresponding optima, respectively. When there is a fitness ridge (i.e., the fitness optimum is a line), *D* is the shortest distance from the current phenotype to the fitness ridge. The selection coefficient *s* equals $\frac{f}{f_{\text{WT}}} - 1$, where *f* and f_{WT} are the fitness values of the mutant and wild-type, respectively. The fixation probability of a newly arisen mutant is $P_f = \frac{1-\exp(-2s)}{1-\exp(-2N_e s)}$ in a haploid population (Kimura 1962), where the effective population size N_{e} was set at 10^4. After each unit time, the phenotypic effect of each mutation is added to the population

mean at a probability of $N_e P_f$; this probability is treated as 1 when $N_e P_f > 1$ or when there is no selection. Combinations of parameters used in the simulations are listed in Table 5-2.

In simulations where different lineages are assigned different optima, each lineage's optimum was obtained by independently drawing the optimal values of the two traits from the standard normal distribution. Before conducting simulations, we confirmed that the optima of the two traits are not correlated (correlation coefficient = 0.0882, $P = 0.54$, *t*-test).

5.5.6 Computer simulation of trait-trait coevolution under mutational bias

To investigate the effect of mutational bias on trait correlation, we introduced the bias coefficient B. Each mutation, after being sampled from a multivariate normal distribution described above, was rescaled using B. Let the mutational effect be $m = (m_1, m_2)$, where m_1 and m_2 are the effects on trait 1 and trait 2, respectively. The rescaled mutational effect, $\widetilde{m}$, is obtained by

$$\widetilde{m} = \begin{cases} mB & (m_1 > 0) \\ m/B & (m_1 < 0) \end{cases}.$$

Because mutational effects are first drawn from a pre-set multivariate normal distribution and then rescaled, we examined if COR_{M} estimated from the rescaled effects ($\widetilde{COR_{\mathrm{M}}}$) is different from the pre-set value of COR_{M}. For each pre-set value of COR_{M}, we obtained $\widetilde{COR_{\mathrm{M}}}$ from 5,000 rescaled mutations. This was repeated 200 times with different random mutations, yielding 200 $\widetilde{COR_{\mathrm{M}}}$ estimates. A series of different B values were used in the simulation (**Table D-2**). For comparison, we also estimated B from yeast gene deletion lines and found the maximal B of any trait to be 1.503. To estimate B for a trait from the yeast gene deletion lines, we calculated the mean trait value of all deletion lines with positive trait values (i.e., greater than the wildtype before log transformation) and mean trait value of all deletion

lines with negative values (i.e., smaller than the wildtype before log transformation). We then computed the ratio of their absolute values with the greater absolute value used as the numerator. The square root of the ratio is *B*. We found that COR_{M} is always near the center of the distribution of these 200 $\widetilde{COR_{\mathrm{M}}}$ estimates (**Table D-2**). Hence, mutational bias will not bias our test.

All analyses in this study were conducted in R (R Core Development Team 2010).

5.6 References

Ackermann, R. R. and J. M. Cheverud. 2004. Detecting genetic drift versus selection in human evolution. Proc Natl Acad Sci U S A 101:17946-17951.

Adams, D. C. 2016. Evaluating modularity in morphometric data: challenges with the RV coefficient and a new test measure. Methods Ecol Evol 7:565-572.

Agrawal, A. F. and J. R. Stinchcombe. 2009. How much do genetic covariances alter the rate of adaptation? P Roy Soc B-Biol Sci 276:1183-1191.

Arnold, S. J., R. Burger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the Evolution and Stability of the G-Matrix. Evolution 62:2451-2461.

Benjamini, Y. and Y. Hochberg. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J R Stat Soc B 57:289-300.

Blows, M. W. and K. Mcguigan. 2015. The distribution of genetic variance across phenotypic space and the response to selection. Mol Ecol 24:2056-2072.

Bolstad, G. H., J. A. Cassara, E. Marquez, T. F. Hansen, K. van der Linde, D. Houle, and C. Pelabon. 2015. Complex constraints on allometry revealed by artificial selection on the wing of Drosophila melanogaster. P Natl Acad Sci USA 112:13284-13289.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a metabolic theory of ecology. Ecology 85:1771-1789.

Clune, J., J. B. Mouret, and H. Lipson. 2013. The evolutionary origins of modularity. Proc Biol Sci 280:20122863.

Dochtermann, N. A. and N. J. Dingemanse. 2013. Behavioral syndromes as evolutionary constraints. Behav Ecol 24:806-811.

Fabre, A. C., C. Bardua, M. Bon, J. Clavel, R. N. Felice, J. W. Streicher, J. Bonnel, E. L. Stanley, D. C. Blackburn, and A. Goswami. 2020. Metamorphosis shapes cranial diversity and rate of evolution in salamanders. Nat Ecol Evol 4:1129-1140.

Felsenstein, J. 1985. Phylogenies and the Comparative Method. American Naturalist 125:1-15.

Gardner, K. M. and R. G. Latta. 2007. Shared quantitative trait loci underlying the genetic correlation between continuous traits. Mol Ecol 16:4195-4209.

Geiler-Samerotte, K. A., S. Li, C. Lazaris, A. Taylor, N. Ziv, C. Ramjeawan, A. B. Paaby, and M. L. Siegal. 2020. Extent and context dependence of pleiotropy revealed by high-throughput single-cell phenotyping. Plos Biol 18.

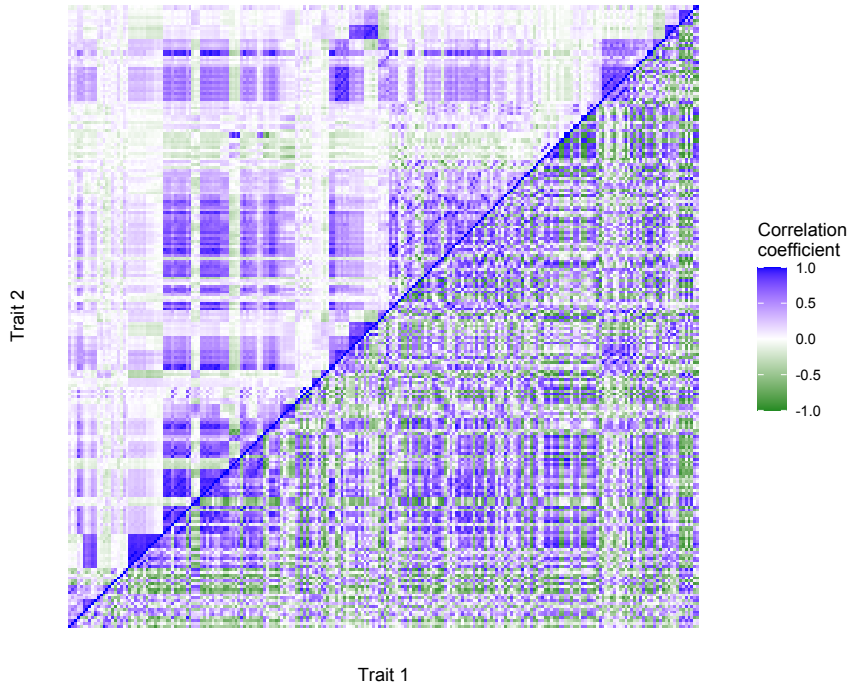
Geiler-Samerotte, K. A., Y. O. Zhu, B. E. Goulet, D. W. Hall, and M. L. Siegal. 2016. Selection Transforms the Landscape of Genetic Variation Interacting with Hsp90. Plos Biol 14.
Genz, A. B., F.; Miwa, T.; Mi, X.; Leisch, F.; Scheipl, F.; Hothorn, T. 2020. mvtnorm: Multivariate Normal and t Distributions. R package version 1.1-0, https://CRAN.R-project.org/package=mvtnorm.
Glazier, D. S. 2010. A unifying explanation for diverse metabolic scaling in animals and plants. Biol Rev 85:111-138.
Goswami, A., J. B. Smaers, C. Soligo, and P. D. Polly. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. Philos Trans R Soc Lond B Biol Sci 369:20130254.
Gould, S. J. 1966. Allometry and Size in Ontogeny and Phylogeny. Biol Rev 41:587-640.
Hansen, T. F. and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J Evolution Biol 21:1201-1219.
Ho, W. C., Y. Ohya, and J. Z. Zhang. 2017. Testing the neutral hypothesis of phenotypic evolution. P Natl Acad Sci USA 114:12219-12224.
Ho, W. C. and J. Zhang. 2014. The genotype-phenotype map of yeast complex traits: basic parameters and the role of natural selection. Mol Biol Evol 31:1568-1580.
Hodgins-Davis, A., F. Duveau, E. A. Walker, and P. J. Wittkopp. 2019. Empirical measures of mutational effects define neutral models of regulatory evolution in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 116:21085-21093.
Houle, D., G. H. Bolstad, K. van der Linde, and T. F. Hansen. 2017. Mutation predicts 40 million years of fly wing evolution. Nature 548:447-450.
Houle, D. and J. Fierst. 2013. Properties of Spontaneous Mutational Variance and Covariance for Wing Size and Shape in Drosophila Melanogaster. Evolution 67:1116-1130.
Houle, D., L. T. Jones, R. Fortune, and J. L. Sztepanacz. 2019. Why does allometry evolve so slowly? Integr Comp Biol 59:1429-1440.
Huxley, J. 1972. Problems of relative growth. Pp. xxvii, 312 p., New York, Dover.
Jerison, E. R., S. Kryazhimskiy, J. K. Mitchell, J. S. Bloom, L. Kruglyak, and M. M. Desai. 2017. Genetic variation in adaptability and pleiotropy in budding yeast. Elife 6.
Jones, A. G., R. Burger, and S. J. Arnold. 2014. Epistasis and natural selection shape the mutational architecture of complex traits. Nat Commun 5:3709.
Kimura, M. 1962. On the probability of fixation of mutant genes in a population. Genetics 47:713-719.
Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. Am Nat 157:245-261.
Lande, R. 1979. Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain - Body Size Allometry. Evolution 33:402-416.
Lande, R. 1980. The Genetic Covariance between Characters Maintained by Pleiotropic Mutations. Genetics 94:203-215.
Lande, R. 2007. The maintenance of genetic variability by mutation in a polygenic character with linked loci (Reprinted). Genetics Research 89:373-387.
Liti, G., D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. A. James, R. P. Davey, I. N. Roberts, A. Burt, V. Koufopanou, I. J. Tsai, C. M. Bergman, D. Bensasson, M. J. O'Kelly, A. van Oudenaarden, D. B. Barton, E. Bailes, A. N. Nguyen, M. Jones, M. A.

Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin, and E. J. Louis. 2009. Population genomics of domestic and wild yeasts. Nature 458:337-341.
Lynch, M. 1990. The Rate of Morphological Evolution in Mammals from the Standpoint of the Neutral Expectation. Am Nat 136:727-741.
Martin, R. D. 1981. Relative Brain Size and Basal Metabolic-Rate in Terrestrial Vertebrates. Nature 293:57-60.
Navalon, G., J. Marugan-Lobon, J. A. Bright, C. R. Cooney, and E. J. Rayfield. 2020. The consequences of craniofacial integration for the adaptive radiations of Darwin's finches and Hawaiian honeycreepers. Nat Ecol Evol 4:270-278.
Ohya, Y., J. Sese, M. Yukawa, F. Sano, Y. Nakatani, T. L. Saito, A. Saka, T. Fukuda, S. Ishihara, S. Oka, G. Suzuki, M. Watanabe, A. Hirata, M. Ohtani, H. Sawai, N. Fraysse, J. P. Latge, J. M. Francois, M. Aebi, S. Tanaka, S. Muramatsu, H. Araki, K. Sonoike, S. Nogami, and S. Morishita. 2005. High-dimensional and large-scale phenotyping of yeast mutants. P Natl Acad Sci USA 102:19015-19020.
Olson, E. C. and R. L. Miller. 1999. Morphological integration. University of Chicago Press, Chicago.
Pavlicev, M. and J. M. Cheverud. 2015. Constraints Evolve: Context Dependency of Gene Effects Allows Evolution of Pleiotropy. Annual Review of Ecology, Evolution, and Systematics, Vol 46 46:413-434.
Pavlicev, M., J. M. Cheverud, and G. P. Wagner. 2009. Measuring Morphological Integration Using Eigenvalue Variance. Evol Biol 36:157-170.
Pavlicev, M., J. P. Kenney-Hunt, E. A. Norgard, C. C. Roseman, J. B. Wolf, and J. M. Cheverud. 2008. Genetic variation in pleiotropy: Differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. Evolution 62:199-213.
Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216-2218.
Peter, J., M. De Chiara, A. Friedrich, J. X. Yue, D. Pflieger, A. Bergstrom, A. Sigwalt, B. Barre, K. Freel, A. Llored, C. Cruaud, K. Labadie, J. M. Aury, B. Istace, K. Lebrigand, P. Barbry, S. Engelen, A. Lemainque, P. Wincker, G. Liti, and J. Schacherer. 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556:339-344.
Pettersen, A. K., C. R. White, and D. J. Marshall. 2016. Metabolic rate covaries with fitness and the pace of the life history in the field. P Roy Soc B-Biol Sci 283.
Pigliucci, M. 2003. Phenotypic integration: studying the ecology and evolution of complex phenotypes. Ecol Lett 6:265-272.
Porto, A., H. Sebastiao, S. E. Pavan, J. L. VandeBerg, G. Marroig, and J. M. Cheverud. 2015. Rate of evolutionary change in cranial morphology of the marsupial genus Monodelphis is constrained by the availability of additive genetic variation. J Evol Biol 28:973-985.
R Core Development Team. 2010. R: A language and environment for statistical computing.
Revell, L. J., L. J. Harmon, R. B. Langerhans, and J. J. Kolbe. 2007. A phylogenetic approach to determining the importance of constraint on phenotypic evolution in the neotropical lizard Anolis cristatellus. Evol Ecol Res 9:261-282.
Roff, D. A., S. Mostowy, and D. J. Fairbairn. 2002. The evolution of trade-offs: Testing predictions on response to selection and environmental variation. Evolution 56:84-95.

Saltz, J. B., F. C. Hessel, and M. W. Kelly. 2017. Trait Correlations in the Genomics Era. Trends Ecol Evol 32:279-290.
Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766-1774.
Shoval, O., H. Sheftel, G. Shinar, Y. Hart, O. Ramote, A. Mayo, E. Dekel, K. Kavanagh, and U. Alon. 2012. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. Science 336:1157-1160.
Sih, A., A. Bell, and J. C. Johnson. 2004. Behavioral syndromes: an ecological and evolutionary overview. Trends Ecol Evol 19:372-378.
Simon, M. N., F. A. Machado, and G. Marroig. 2016. High evolutionary constraints limited adaptive responses to past climate changes in toad skulls. Proc Biol Sci 283.
Sinervo, B. and E. Svensson. 2002. Correlational selection and the evolution of genomic architecture. Heredity (Edinb) 89:329-338.
Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: evolution of the G matrix. Trends Ecol Evol 17:320-327.
Svensson, E. I., S. J. Arnold, R. Burger, K. Csillery, J. Draghi, J. M. Henshaw, A. G. Jones, S. De Lisle, D. A. Marques, K. McGuigan, M. N. Simon, and A. Runemark. 2021. Correlational selection in the age of genomics. Nat Ecol Evol.
Wagner, A. 2005. Robustness and evolvability in living systems. Princeton University Press, Princeton, NJ.
Wagner, G. P. 1989. Multivariate mutation-selection balance with constrained pleiotropic effects. Genetics 122:223-234.
Wagner, G. P. 1999. A research programme for testing the biological homology concept. Novartis Found Symp 222:125-134; discussion 134-140.
Wagner, G. P. and L. Altenberg. 1996. Perspective: Complex Adaptations and the Evolution of Evolvability. Evolution 50:967-976.
Wagner, G. P., M. Pavlicev, and J. M. Cheverud. 2007. The road to modularity. Nature Reviews Genetics 8:921-931.
Wagner, G. P. and J. Z. Zhang. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nature Reviews Genetics 12:204-213.
Walsh, B. and M. W. Blows. 2009. Abundant Genetic Variation plus Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. Annu Rev Ecol Evol S 40:41-59.
Walter, G. M., J. D. Aguirre, M. W. Blows, and D. Ortiz-Barrientos. 2018. Evolution of Genetic Variance during Adaptive Radiation. Am Nat 191:E108-E128.
Wang, Z., B. Y. Liao, and J. Zhang. 2010. Genomic patterns of pleiotropy and the evolution of complexity. Proc Natl Acad Sci U S A 107:18034-18039.
Watanabe, A., A. C. Fabre, R. N. Felice, J. A. Maisano, J. Muller, A. Herrel, and A. Goswami. 2019. Ecomorphological diversification in squamates from conserved pattern of cranial integration. Proc Natl Acad Sci U S A 116:14688-14697.
Wei, X. and J. Zhang. 2019. Environment-dependent pleiotropic effects of mutations on the maximum growth rate r and carrying capacity K of population growth. PLoS Biol 17:e3000121.
White, C. R., D. J. Marshall, L. A. Alton, P. A. Arnold, J. E. Beaman, C. L. Bywater, C. Condon, T. S. Crispin, A. Janetzki, E. Pirtle, H. S. Winwood-Smith, M. J. Angilletta, Jr., S. F. Chenoweth, C. E. Franklin, L. G. Halsey, M. R. Kearney, S. J. Portugal, and D. Ortiz-

Barrientos. 2019. The origin and maintenance of metabolic allometry in animals. Nat Ecol Evol 3:598-603.
Yvert, G., S. Ohnuki, S. Nogami, Y. Imanaga, S. Fehrmann, J. Schacherer, and Y. Ohya. 2013. Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. BMC Syst Biol 7:54.

a


Trait 2
Trait 1
Correlation coefficient
1.0
0.5
0.0
-0.5
-1.0

b

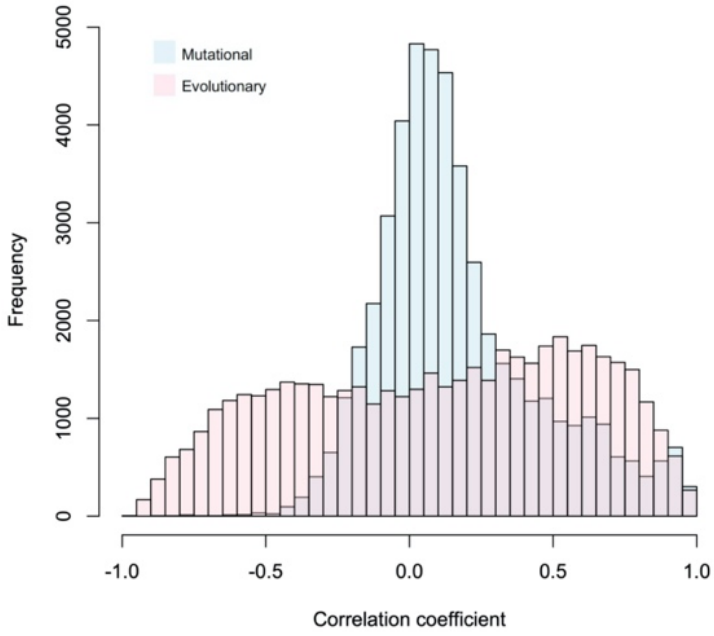
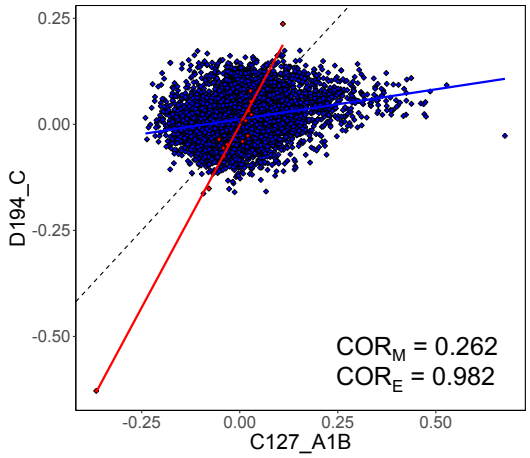
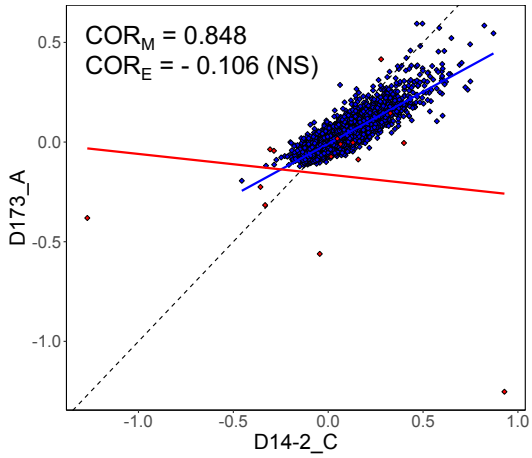

Mutational
Evolutionary
Frequency
0
1000
2000
3000
4000
5000
-1.0
-0.5
0.0
0.5
1.0
Correlation coefficient

Figure. 5-1. Mutational (COR_M) and evolutionary (COR_E) correlations for all pairs of the 220 yeast morphological traits. COR_M is based on yeast gene deletion lines. (a) COR_M (upper triangle) and COR_E (lower triangle) for all pairs of traits ordered according to their IDs. (b) Frequency distributions of COR_M and COR_E across all trait pairs. The two distributions are significantly different ($P < 10^{-10}$, Kolmogorov–Smirnov test).

a


0.25
0.00
-0.25
-0.50
D194_C
-0.25
0.00
0.25
0.50
C127_A1B
COR_M = 0.262
COR_E = 0.982

b


0.5
0.0
-0.5
-1.0
D173_A
-1.0
-0.5
0.0
0.5
1.0
D14-2_C
COR_M = 0.848
COR_E = - 0.106 (NS)

c

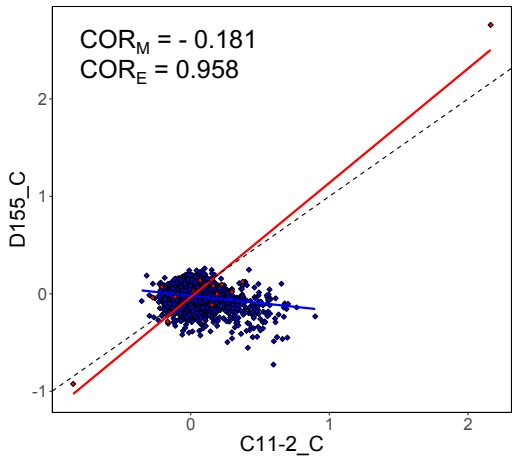
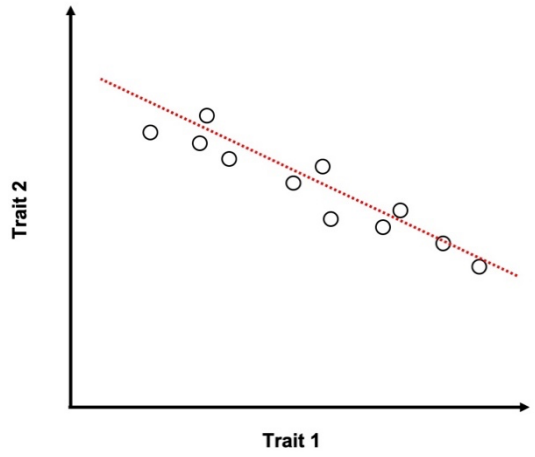
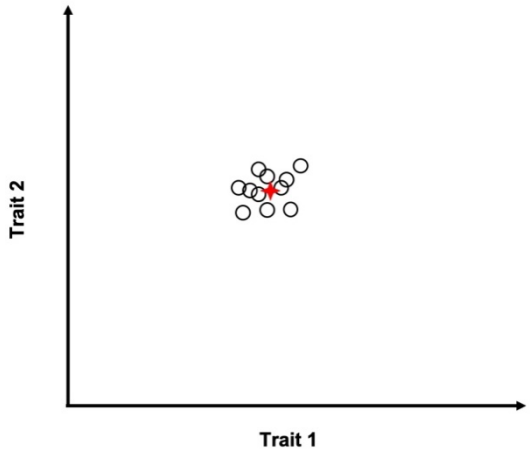

2
1
0
-1
D155_C
0
1
2
C11-2_C
COR_M = - 0.181
COR_E = 0.958

Figure. 5-2. Examples of yeast trait pairs with COR_E significantly different from the corresponding COR_M. (a) An example of evolutionarily strengthened correlation. (b) An example of evolutionarily weakened correlation. (c) An example of evolutionarily reversed correlation. Each blue dot represents a gene deletion line (a.k.a. mutant) while each red dot represents an independent contrast derived from natural strains. Blue and red lines are linear regressions between the standardized values of the two traits in mutants and independent contrasts, respectively, while the dotted blackline shows the diagonal ($y = x$). Trait IDs are shown along the axes. All COR_M and COR_E values shown are significantly different from 0 except when indicated by "NS" in the parentheses.

a



b



c

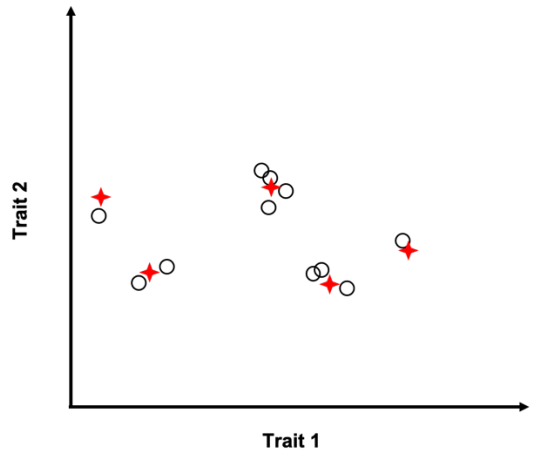


Figure. 5-3. Schematic illustration of predictions made by models of trait-trait coevolution. Each circle represents the equilibrium mean phenotype of two hypothetical traits (trait 1 and trait 2) of a diverging lineage. (a) When a specific allometric relationship is selectively favored, the population mean phenotypes are distributed along the fitness ridge (i.e., the optimal allometric line shown in red), resulting in a strong trait correlation across lineages. (b) When a specific value is selectively favored for each trait, the population mean phenotypes are concentrated near the optimal phenotype (marked by the red cross) and the trait correlation across lineages is weak. (c) When different lineages have different optimal phenotypes (marked by red crosses) that are randomly distributed, the trait correlation across lineages is weak.

Table. 5-1. Numbers of trait pairs with significantly different COR_E and COR_M in the yeast and fly data.

	Yeast						Fly (276 trait pairs)		
	COR_M from gene deletion lines (24,090 trait pairs)			COR_M from MA lines (17,391 trait pairs)					
	Observed[a]	Expected[b]	Fraction[c]	Observed	Expected	Fraction	Observed	Expected	Fraction
Strengthened	2727	145	3%	2395	17	3.5%	68	0	0%
Weakened	1221	78	1.4%	1348	6	0.1%	48	0	0%
Reversed	2795	51	0%	1403	2	0%	28	0	0%
Total	6743	302.5	0.7%	5146	31	0.7%	144	0	0%

[a]Number of trait pairs with significantly different COR_M and COR_E inferred by a Z-test.

[b]Number of trait pairs that show a significant difference between COR_M and COR_E derived from neutral Brownian motion simulations. The median from 1000 simulations is shown.

[c]Fraction of the 1000 Brownian motion simulations where the number of trait pairs with significantly different COR_M and COR_E exceeds the number under the "observed" column.

Table. 5-2. Parameters and results of simulations of trait-trait coevolution.

Optimum	COR_M	Median COR_E at the end of simulation	Fraction of simulations with $COR_E > COR_M$	COR_E compared with COR_M
No optimum	0.9	0.900	49.5%	No difference
	0.5	0.495	47.5%	No difference
	0.1	0.113	55.5%	No difference
$y = x$ *	0.9	1.000	100%	Strengthened
	0.5	1.000	100%	Strengthened
	0.1	1.000	100%	Strengthened
$y = 0.5x$	0.9	1.000	100%	Strengthened
	0.5	1.000	100%	Strengthened
	0.1	1.000	100%	Strengthened
$y = -0.5x$	0.9	-0.995	0%	Reversed
	0.5	-0.999	0%	Reversed
	0.1	-1.000	0%	Reversed
$y = - x$	0.9	-0.997	0%	Reversed
	0.5	-0.999	0%	Reversed
	0.1	-1.000	0%	Reversed
(0, 0)	0.9	0.0213	0%	Weakened
	0.5	0.00142	0.5%	Weakened
	0.1	-0.0109	24%	No difference
	0.9	0.0895	0%	Weakened

Drawn from	0.5	0.0874	0%	Weakened
$\boldsymbol{\mathcal{N}}(\mathbf{0}, \mathbf{1})$	0.1	0.0866	6%	No difference

*x and y respectively represent the values of the two traits considered.

Table. 5-3. Overall phenotypic integration (V_{eigen}) and modularity (CR) at the levels of mutation and evolutionary divergence. Values at the level of mutation for yeast are medians of 1,000 control datasets. *P*-values for yeast are computed from locations of the observed values in the corresponding distributions of 5000 control datasets, while the *P*-value for fly is from a Fligner-Killeen test.

Statistic	Taxon	Mutation	Divergence	*P*-value
V_{eigen}	Yeast	34.414	58.656	0.0788
	Fly	3.530	4.359	0.459
CR	Yeast	0.649	0.997	< 0.001

Chapter 6

Discussions and Future Directions

6.1 Introduction

Each of the main chapters has a discussion section where questions conceptually or technically related to works presented in the chapter are discussed. In this chapter, I discuss some additional questions in phenotypic evolution and my thoughts about them. Thoughts and views of mine presented in this chapter can be rather speculative and not yet supported by evidence. Nevertheless, thinking about these questions have prompted me to explore what used to be new to me and provided me with more opportunities to improve my works.

Some questions discussed here are what I consider as potential topics for future studies. I discuss here some thoughts of mine regarding how these questions might be addressed and possible difficulties. These questions do not necessarily have a direct connection to my previous works.

I also discuss some methodological questions that occurred to me while I was working on my previous projects. Models and methods broadly used in studies of phenotypic evolution do have their limitations, which can become challenges in future studies. In this chapter, I discuss some of such limitations and my thoughts on how these problems might be solved.

6.2 More models for evolution of molecular traits?

As some molecular traits have rather unique genetic architectures that differ from classically studied organismal traits, theoretical models incorporating details of their genetic architecture have been developed to help understand these traits' evolution. While some molecular traits have been dealt with in previous studies (Lynch and Hagner 2015; Lynch et al. 2016; Lynch 2018, 2020), dedicated theoretical frameworks for other traits' evolution are yet to be established. Although preexisting models were not developed for these traits in the first place, conceptual and mathematical approaches they used may be applied in modeling of new traits as well. In this section, I briefly discuss possibility of having such theoretical frameworks established for two special cases.

6.2.1 Evolution of post-transcriptional modifications

Post-transcriptional modifications of RNA molecules, such as A-to-I editing, happens upon recognition of sequence motifs by the responsible enzyme (Eggington et al. 2011; Grice and Degnan 2015; Nishikura 2016). Evolution of modification frequency at each site, therefore, is mediated by evolution of binding affinity between its flanking sequences and the enzyme, which is not unsimilar to other nucleic acid-protein interactions studied previously (Lynch and Hagner 2015): given expression level of the enzyme, probability that the enzyme binds to the RNA molecule at the site of interest is a function of the flanking sequences; therefore, there exists a finite number of states for binding affinity. The evolution of frequency at which modification takes place at a site is essentially transitions between these discrete states in face of mutation, drift, and selection.

Despite the similarity in genetic architectures, evolution of RNA modification at a site can be subject to additional constraints if the focal site is in the coding region. With a large

fraction flanking site mediating RNA modification at the focal site being nonsynonymous, the level of modification can be evolutionarily constrained: deleterious effects of flanking nonsynonymous mutations would likely exceed their beneficial effect on RNA modification, if any, since non-synonymous RNA modification events are essentially "partial" mutations. Such constraint may render certain modification events barely evolvable, preventing beneficial modifications from being enhanced and deleterious modifications from purged by selection, ultimately creating a phylogenetic distribution that is drastically different from what would be expected for other traits.

However, a theoretical model for RNA modification could be difficult to parameterize, as the distribution of mutations' phenotypic effects is largely unclear. Despite some sequence motifs associated with certain types of RNA modification have been identified, they are mostly weak and not sufficient to explain the observations (Gilbert et al. 2016; Liscovitch-Brauer et al. 2017). However, A-to-I editing in ascomycetes filamentous fungi, which is not catalyzed by homologs of known editing enzymes in other taxa, is an exception. Editing in these species is strong associated with an uracil (U) right upstream the editing site, with a vast majority of editing sites have an upstream U and sites with this U are much more likely to be edited (Liu et al. 2016; Liu et al. 2017; Teichert et al. 2017). Nonetheless, this motif cannot be the only determinant of editing level of a site, as there remains to be unexplained variations in editing levels.

6.2.2 Coevolution of transcription and translation

The rate at which protein molecules are produced is essentially a product of transcription rate and per-mRNA translational efficiency (i.e., amount of protein molecules translated from

each mRNA molecule). Mutations affecting transcription rate is expected to affect abundance of the protein had other variables unchanged, while mutations affecting translational efficiency only can affect protein abundance without altering mRNA abundance. Therefore, abundance of protein supposedly has more mutational input than abundance of the corresponding mRNA and is expected to diverge faster under neutrality.

It is often observed, however, that the abundance of a protein tends to be more evolutionarily conserved than the corresponding mRNA's abundance, which is commonly perceived to be a result of selection on abundance of the protein and mediated by compensatory evolution (Artieri and Fraser 2014; McManus et al. 2014; Wang et al. 2018; Wang et al. 2020): some slightly deleterious changes at the level of transcription can be compensated by another change at the translation level, and vice versa. However empirical estimations of translational efficiency are often not independent of the mRNA level and the observed correlation between mRNA level and translational efficiency can be, at least partly, a statistical artifact. Furthermore, coevolution of transcription and translation evolution has not yet received a formal theoretical investigation.

To this end, it would be helpful to have a theoretical framework for coevolution of transcription and translation. In this framework, there are "intermediate" traits (i.e., transcription and translation) that are not under selection per se and another trait (i.e., protein abundance) that is directly under selection. Value of the third trait is a function of the first two and, in this case, is by default their product. Mutations only directly affect the intermediate traits, yet each mutation's fitness effect is determined by its indirect effect on the trait under selection. Patterns of evolution for three traits would depend on mutational input for both intermediate traits, the effective population size and mode of selection.

Within this framework, expected results of different types of analysis can be derived given combination of parameters. One can explore what is required to generate a certain observation or if the observation is robust to certain variables. Specific questions that may be addressed include whether compensatory evolution is sufficient to explain the observed patterns of evolutionary conservation, whether other mechanisms (e.g., mutational covariance between transcription and translation, or regulation of degradation processes) need to be invoked, and whether observed patterns persist upon introduction of those extra mechanisms. Systematically examining different settings and different parameter value combinations could help interprete results of empirical studies.

6.3 Mechanistic basis of the $\boldsymbol{M}$ matrix

In models of phenotypic evolution, availability of mutational input is characterized by mutational variance V_M, or, when two or more traits are under consideration, the M matrix. The M matrix reflects combined effect of many mutations on phenotypic variance within the population, yet the same (co)variance can arise from different combinations: mutational variance of a trait is determined jointly by rate at which mutations affecting this trait occur and distribution of these mutations' effects; similarly, covariance between two traits is determined jointly by rate at which pleiotropic mutations affecting both traits occur and distribution of these mutations' effects on two traits. For instance, there would be zero covariance between two traits if there are no pleiotropic mutations; on the other hand, the covariance still also be zero when some or even all mutations are pleiotropic, as long as each mutation's effects on two traits are uncorrelated.

Under neutrality, pattern of long-term evolution depends only on structure of the M matrix and the exact underpinning of the M matrix does not make a difference. Selection, on the other hand, acts on individual mutations. As a result, the same mutational covariance can have different long-term consequences. For example, if trait A and trait B have zero covariance between them, the proportion of pleiotropic mutations out of all mutations affecting them could be anywhere between zero and one. In the presence of pleiotropic mutations, adaptive evolution can be constrained: adaptive evolution of trait A can be slowed down because some mutations affecting trait A might have deleterious pleiotropic effects on trait B, and vice versa. Similarly, the G matrix, which summarizes phenotypic (co)variance resulting from standing genetic variations, is in theory subject to the same problem: the same genetic (co)variance can have different genetic underpinnings.

Hence, to better understand how mutations and standing genetic variations contribute to long-term phenotypic evolution, it is necessary to know how level of pleiotropy of individual mutations is distributed; measuring only phenotypic (co)variance among mutants, in contrast, would be insufficient. Analyses of gene deletion lines in several model organisms have shown the number of traits affected by a gene deletion follows a rather skewed distribution, indicating most mutations are probably not strongly pleiotropic (Wang et al. 2010; Ho and Zhang 2014). With such data, contributions of pleiotropic and non-pleiotropic mutations to covariance between two traits could be teased apart, allowing better prediction about patterns in long-term evolution.

6.4 Comparing different levels of biological organization

Does evolution follow the same rules at different levels of biological organization? Studies of molecular traits have repeatedly uncovered nonadaptive evolution, yet organismal

traits like morphological traits at time show signal of adaptive evolution. A sharp contrast between molecular and morphological traits is seen in budding yeast *S. cerevisiae*: gene expression variation among strains and species is found to be largely neutral, while rate of divergence in cell morphological traits is positively correlated with trait importance (i.e., strength of correlation between phenotype and fitness) (Ho et al. 2017; Yang et al. 2017).

It is hypothesized that such a difference between levels of biological organization may arise from difference in proportion of substitutions with phenotypic effect at each level. Presumably, lower-level (i.e., more microscopic) traits give rise to higher-level (i.e., macroscopic) traits, such that substitutions affecting a higher-level trait must have effects at lower levels (Zhang 2018). Substitutions affecting a lower-level trait, on the other hand, may not have any effect at higher levels. As a result, evolution of lower-level traits tends to be more neutral, while evolution of higher-level traits is more likely to be selection-driven. For example, change in transcription level of a gene can alter concentration of the encoded protein, influence rate of biochemical interactions in which the protein involves, and, at last, affect organismal phenotypes this protein underlies. However, it is also possible that the change in transcription level is buffered at the level of translation or by degradation of gene products and therefore has no further phenotypic and fitness effects.

To test generality of this model, it is necessary to compare different phenotypic datasets that represent different levels of biological organization. Ideally, one should examine different phenotypic traits measured in the same set of species to understand the relationship between pattern of evolution and level of biological organization in a specific clade, and test generality of the finding by repeating the same analyses in more different taxa. However, it is rare that a phenotypic trait is measured in multiple closely related species, rendering aforementioned

comparisons unfeasible. If one would compare different traits of different species, comparability becomes a problem. For example, it is unknown if cell morphological traits of unicellular organisms are more comparable to cell morphological traits or morphological traits of multicellular organisms. At present, it remains hard to rigorously test hypotheses regarding relationship between level of biological organization and mechanisms of selection, and more comprehensive data would be helpful for solving the problem.

6.5 Evolution of traits with evolving genetic architectures

The M matrix is often assumed to have stable size and structure during evolution, in both theoretical and empirical studies (Lande 1976, 1979; Houle et al. 2017). Validity of various practices in studies of phenotypic evolution also rely on validity of this assumption. For example, hypotheses testing based on comparison of M and R matrices would only be valid if the M matrix estimated from extant species successfully recapitulate mutational input for past evolution.

Paradoxically, it is also widely acknowledged that genetic architectures of phenotypic traits are evolvable. On the one hand, there exist mutations that can modify structure the genotype-phenotype map (also referred to as modifier mutations for short): for instance, gene duplication followed by sub-functionalization may allow functions of the same gene reassigned to different genes, allowing different phenotypic traits to be genetically decoupled and reducing their genetic covariance (Force et al. 1999; Stoltzfus 1999; Conrad and Antonarakis 2007). On the other hand, selection for robustness and/or adaptability could lead to fixation of some modifier mutations, which, as introduced in earlier sections, could have been responsible for modularity of observed genotype-phenotype maps (Wagner and Altenberg 1996; Wagner et al.

2007; Wagner et al. 2008) and observed distributions of gene-level pleiotropy (Wagner et al. 2008; Wang et al. 2010).

Nonetheless, it remains challenging to have evolution of traits' genetic architectures incorporated into theoretical models, as rate and effect size distribution of modifier mutations remain largely unclear, making parameterization difficult. Mutation accumulation (MA) experiments and QTL mapping have previous been devised to identify mutations or genetic variations that affect the overall mutation rate (Gou et al. 2019; Liu and Zhang 2021). In theory, the same approaches could be used to identify other types of modifiers, such as those affecting mutational variances of a particular trait or those affecting mutational covariance between two traits (Pavlicev et al. 2008; Geiler-Samerotte et al. 2020). That said, it may require a large number of MA lines or segregants to perform the said screen since such modifiers are supposedly rarer than mutations directly affecting the phenotypic traits.

It is also desirable to have the M matrix measured in multiple species: high variability would indicate evolutionary changes are frequent, while lack of detection means it is safe to assume M is constant as long as only evolution after these species split is concerned. Unfortunately, it is very often that only M estimated from one species is available and it is necessary to have M measured in more sets of related species.

6.6 References

Artieri, C. G. and H. B. Fraser. 2014. Evolution at two levels of gene expression in yeast. Genome Res 24:411-421.

Conrad, B. and S. E. Antonarakis. 2007. Gene duplication: a drive for phenotypic diversity and cause of human disease. Annu Rev Genomics Hum Genet 8:17-35.

Eggington, J. M., T. Greene, and B. L. Bass. 2011. Predicting sites of ADAR editing in double-stranded RNA. Nat Commun 2:319.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531-1545.
Geiler-Samerotte, K. A., S. Li, C. Lazaris, A. Taylor, N. Ziv, C. Ramjeawan, A. B. Paaby, and M. L. Siegal. 2020. Extent and context dependence of pleiotropy revealed by high-throughput single-cell phenotyping. Plos Biol 18.
Gilbert, W. V., T. A. Bell, and C. Schaening. 2016. Messenger RNA modifications: Form, distribution, and function. Science 352:1408-1412.
Gou, L., J. S. Bloom, and L. Kruglyak. 2019. The Genetic Basis of Mutation Rate Variation in Yeast. Genetics 211:731-740.
Grice, L. F. and B. M. Degnan. 2015. The origin of the ADAR gene family and animal RNA editing. BMC Evol Biol 15:4.
Ho, W. C., Y. Ohya, and J. Z. Zhang. 2017. Testing the neutral hypothesis of phenotypic evolution. P Natl Acad Sci USA 114:12219-12224.
Ho, W. C. and J. Zhang. 2014. The genotype-phenotype map of yeast complex traits: basic parameters and the role of natural selection. Mol Biol Evol 31:1568-1580.
Houle, D., G. H. Bolstad, K. van der Linde, and T. F. Hansen. 2017. Mutation predicts 40 million years of fly wing evolution. Nature 548:447-+.
Lande, R. 1976. Natural-Selection and Random Genetic Drift in Phenotypic Evolution. Evolution 30:314-334.
Lande, R. 1979. Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain - Body Size Allometry. Evolution 33:402-416.
Liscovitch-Brauer, N., S. Alon, H. T. Porath, B. Elstein, R. Unger, T. Ziv, A. Admon, E. Y. Levanon, J. J. C. Rosenthal, and E. Eisenberg. 2017. Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. Cell 169:191-202.
Liu, H. and J. Zhang. 2021. The rate and molecular spectrum of mutation are selectively maintained in yeast. Nat Commun 12:4044.
Liu, H. Q., Y. Li, D. P. Chen, Z. M. Qi, Q. H. Wang, J. H. Wang, C. Jiang, and J. R. Xu. 2017. A-to-I RNA editing is developmentally regulated and generally adaptive for sexual reproduction in Neurospora crassa. P Natl Acad Sci USA 114:E7756-E7765.
Liu, H. Q., Q. H. Wang, Y. He, L. F. Chen, C. F. Hao, C. Jiang, Y. Li, Y. F. Dai, Z. S. Kang, and J. R. Xu. 2016. Genome-wide A-to-I RNA editing in fungi independent of ADAR enzymes. Genome Res 26:499-509.
Lynch, M. 2018. Phylogenetic divergence of cell biological features. Elife 7.
Lynch, M. 2020. The evolutionary scaling of cellular traits imposed by the drift barrier. P Natl Acad Sci USA 117:10435-10444.
Lynch, M., M. S. Ackerman, J. F. Gout, H. Long, W. Sung, W. K. Thomas, and P. L. Foster. 2016. Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet 17:704-714.
Lynch, M. and K. Hagner. 2015. Evolutionary meandering of intermolecular interactions along the drift barrier. P Natl Acad Sci USA 112:E30-E38.
McManus, C. J., G. E. May, P. Spealman, and A. Shteyman. 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res 24:422-430.
Nishikura, K. 2016. A-to-I editing of coding and non-coding RNAs by ADARs. Nat Rev Mol Cell Bio 17:83-96.

Pavlicev, M., J. P. Kenney-Hunt, E. A. Norgard, C. C. Roseman, J. B. Wolf, and J. M. Cheverud. 2008. Genetic variation in pleiotropy: Differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. Evolution 62:199-213.

Stoltzfus, A. 1999. On the possibility of constructive neutral evolution. Journal of Molecular Evolution 49:169-181.

Teichert, I., T. A. Dahlmann, U. Kuck, and M. Nowrousian. 2017. RNA Editing During Sexual Development Occurs in Distantly Related Filamentous Ascomycetes. Genome Biology and Evolution 9:855-868.

Wagner, G. P. and L. Altenberg. 1996. Perspective: Complex Adaptations and the Evolution of Evolvability. Evolution 50:967-976.

Wagner, G. P., J. P. Kenney-Hunt, M. Pavlicev, J. R. Peck, D. Waxman, and J. M. Cheverud. 2008. Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature 452:470-U479.

Wagner, G. P., M. Pavlicev, and J. M. Cheverud. 2007. The road to modularity. Nature Reviews Genetics 8:921-931.

Wang, S. H., C. J. Hsiao, Z. Khan, and J. K. Pritchard. 2018. Post-translational buffering leads to convergent protein expression levels between primates. Genome Biol 19:83.

Wang, Z., B. Y. Liao, and J. Z. Zhang. 2010. Genomic patterns of pleiotropy and the evolution of complexity. P Natl Acad Sci USA 107:18034-18039.

Wang, Z. Y., E. Leushkin, A. Liechti, S. Ovchinnikova, K. Mossinger, T. Bruning, C. Rummel, F. Grutzner, M. Cardoso-Moreira, P. Janich, D. Gatfield, B. Diagouraga, B. de Massy, M. E. Gill, A. Peters, S. Anders, and H. Kaessmann. 2020. Transcriptome and translatome co-evolution in mammals. Nature 588:642-647.

Yang, J. R., C. J. Maclean, C. Park, H. B. Zhao, and J. Z. Zhang. 2017. Intra and Interspecific Variations of Gene Expression Levels in Yeast Are Largely Neutral: (Nei Lecture, SMBE 2016, Gold Coast). Mol Biol Evol 34:2125-2139.

Zhang, J. Z. 2018. Neutral Theory and Phenotypic Evolution. Mol Biol Evol 35:1327-1331.

Appendix A

Parallel Transcriptomic Changes in the Origins of Divergent Monogamous Vertebrates?

A.1 Results and discussion

Comparing the neural transcriptomes of five phylogenetically independent pairs of monogamous and non-monogamous vertebrates, Young *et al.* claimed to have found evidence for "a universal transcriptomic mechanism underlying the evolution of monogamy in vertebrates" (Young et al. 2019). They stated that "while evolutionary divergence time between species or clades did not explain gene expression similarity, characteristics of the mating system correlated with neural gene expression patterns". Given the deep divergences among these clades, parallel evolution at the transcriptomic scale is highly unlikely, prompting us to reexamine their results.

We followed a previous study (Yang et al. 2017) to compute the distance between each pair of transcriptomes on the basis of expression level differences of orthologous genes and build a neighboring-joining tree of the 10 transcriptomes. Young *et al*.'s conclusion means that the transcriptomes from the five monogamous species should be clustered in the tree, in exclusion of

---

Work of this appendix has been published:
Jiang, D. H. and J. Z. Zhang. 2019. Parallel transcriptomic changes in the origins of divergent monogamous vertebrates? P Natl Acad Sci USA 116:17627-17628.

those from the non-monogamous species. By contrast, we found that the transcriptome of each monogamous species is clustered with that of its non-monogamous relative and that the neural transcriptome divergences followed the species phylogeny (**Fig. A-1**).

Despite the above finding, it is still of interest to identify individual genes that underwent concordant expression level changes in the five monogamous lineages and test whether such genes are more numerous than the chance expectation. Young *et al*. compared the expression levels of orthologous genes using DESeq2—a tool designed for testing differential gene expression between an experimental group and a control group (Love et al. 2014)—by treating the five monogamous species as replicates in the experimental group and the five non-monogamous species as replicates in the control group. This treatment is inappropriate because different species do not have the same expected expression levels and because the species are not independent from one another. Additionally, the DESeq2 analysis does not inform whether the number of significant cases exceeds the chance expectation, because even comparing two random groups of five species may yield some significant DESeq2 results.

To tackle the above problems, we directly compared the expression level of a gene between a monogamous species and its non-monogamous relative. A gene is considered to have concordant expression changes in monogamous lineages if the expression ratio between the two species within each clade is ≥ 2—commonly considered a substantial expression change—and the direction of the expression difference is the same among all clades. This analysis identified 15 concordant genes. To examine whether this number exceeds the chance expectation, we switched the monogamous status between the two within-clade species for one or more clades, yielding a total of 15 distinct negative controls. The mean number of concordant genes in these

controls is 7.2, while the maximum is 13. Thus, the observed number of concordant genes exceeds the random expectation ($P < 1/15 = 0.067$), but the excess is minute.

In conclusion, our reanalysis shows no transcriptome-wide parallel evolution in the repeated origins of vertebrate monogamy, although the neural expression levels of a very small number of genes may be generally associated with monogamy.

A.2 Methods

All analyses were performed in an *R* environment (R Core Development Team 2010). Transcriptomic tree is constructed using MEGA (Kumar et al. 2018).

A.3 References

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547-1549.

Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

R Core Development Team. 2010. R: A language and environment for statistical computing.

Yang, J. R., C. J. Maclean, C. Park, H. B. Zhao, and J. Z. Zhang. 2017. Intra and Interspecific Variations of Gene Expression Levels in Yeast Are Largely Neutral: (Nei Lecture, SMBE 2016, Gold Coast). Mol Biol Evol 34:2125-2139.

Young, R. L., M. H. Ferkin, N. F. Ockendon-Powell, V. N. Orr, S. M. Phelps, A. Pogany, C. L. Richards-Zawacki, K. Summers, T. Szekely, B. C. Trainor, A. O. Urrutia, G. Zachar, L. A. O'Connell, and H. A. Hofmann. 2019. Conserved transcriptomic profiles underpin monogamy across vertebrates. Proc Natl Acad Sci U S A 116:1331-1336.

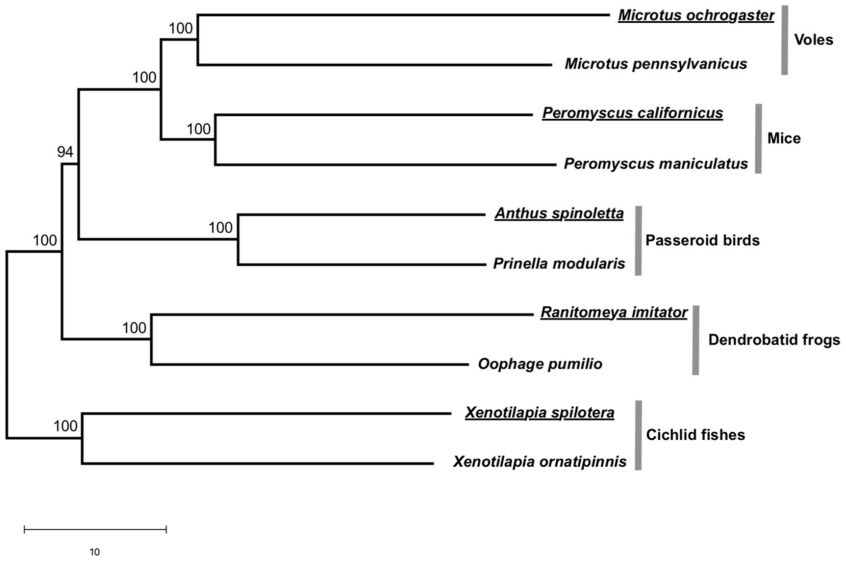


Figure. A-1. Neighbor-joining tree based on brain transcriptomes of species considered in this study. Number near each node is the bootstrap support. Branch lengths represent Euclidean distances between transcriptomes.

Appendix B

Supplementary Figures and Tables for Chapter 3

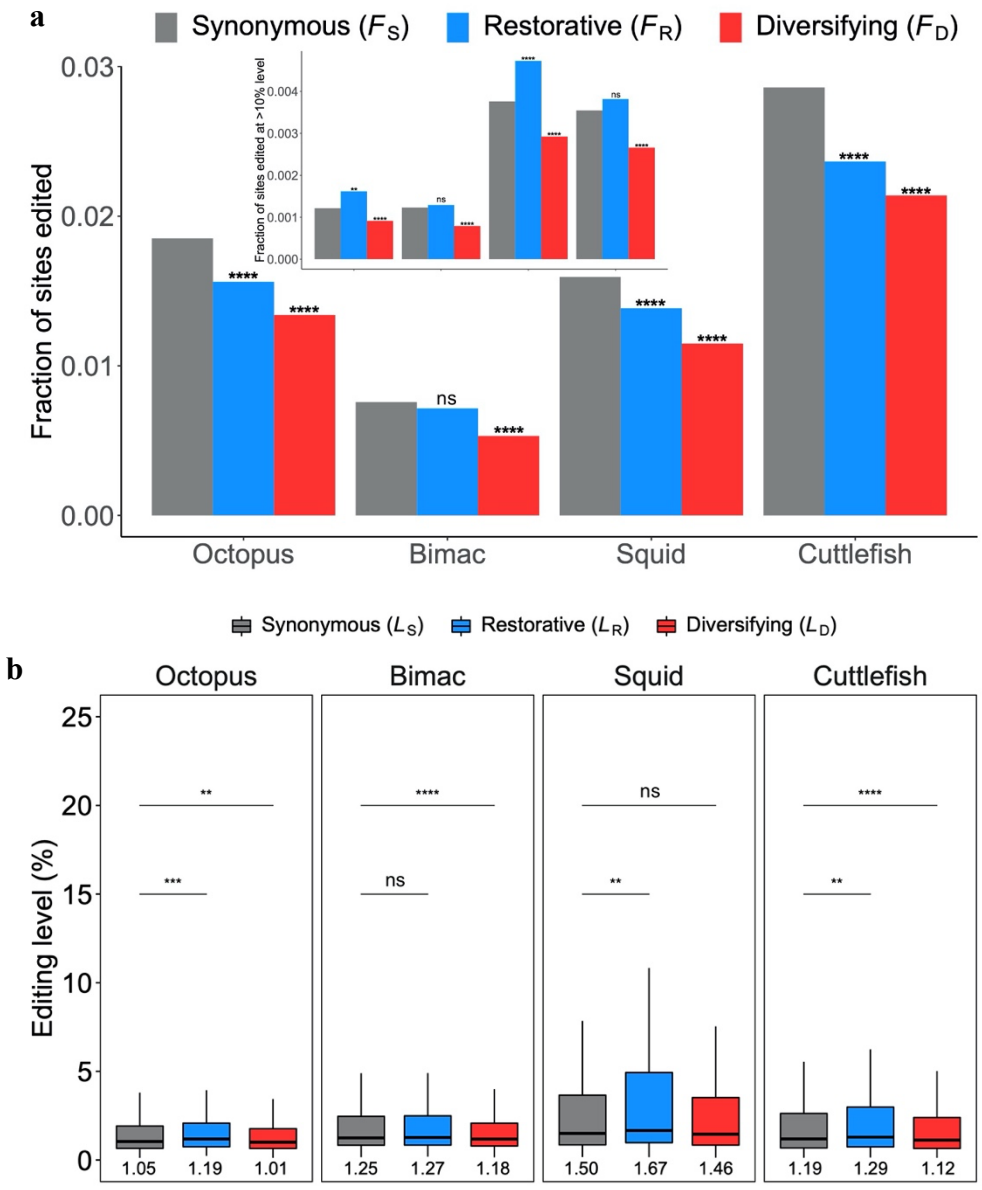


Figure. B-1. Respectively comparing restorative and diversifying editing with synonymous editing for species-specific editing supports the nonadaptive hypothesis of the preponderance of nonsynonymous editing in coleoids. (**a**) Frequencies of sites with synonymous (F_S), restorative (F_R), and diversifying (F_D) editing, respectively, in each of the four coleoids, for species-specific editing. Significant difference between F_S and F_R (or F_D) are indicated by stars above the bin of

F_R (or F_D) (*, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001; ns, not significant; chi-squared test). The inset shows fractions of sites with editing levels >10%. (**b**) Synonymous, restorative, and diversifying editing levels in each of the four coleoids for species-specific editing. The lower and upper edges of a box represent the first (qu_1) and third quartiles (qu_3), respectively, the horizontal line inside the box indicates the median (md), and the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(qu_3-qu_1)$. The median editing levels are also given below the corresponding boxes. Significant differences between L_S and L_R (or L_D) are indicated by stars (*, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001; ns, not significant; Mann-Whitney U test). Note that some identified species-specific editing may in fact be shared among species because of the low detectability of editing caused by low sequencing coverage and/or low editing level. While the group of identified species-specific editing may be contaminated with some shared editing, the group of identified shared editing is not contaminated with species-specific editing. Hence, true signals from species-specific editing can be inferred by comparing between the results here and those in Fig. 3-4. Specifically, the true signals from species-specific editing after the removal of the contamination from shared editing are inferred as follows. First, because shared editing between two species shows $L_D > L_S$ (Fig. 3-4b), removal of contamination from shared editing would strengthen the signal of $L_D < L_S$ observed in species-specific editing. In other words, the true signal of $L_D < L_S$ should be even stronger than what is shown here. Second, the true signal of $F_D < F_S$ and $F_R < F_S$ should be similar to that seen here, because the relationship among F_R, F_D, and F_S observed here is similar to that in Fig. 4a. Third, the true signal of $L_R > L_S$ should be similar to that seen here, because the relationship between L_R and L_S observed here is similar to that in Fig. 3-4b.

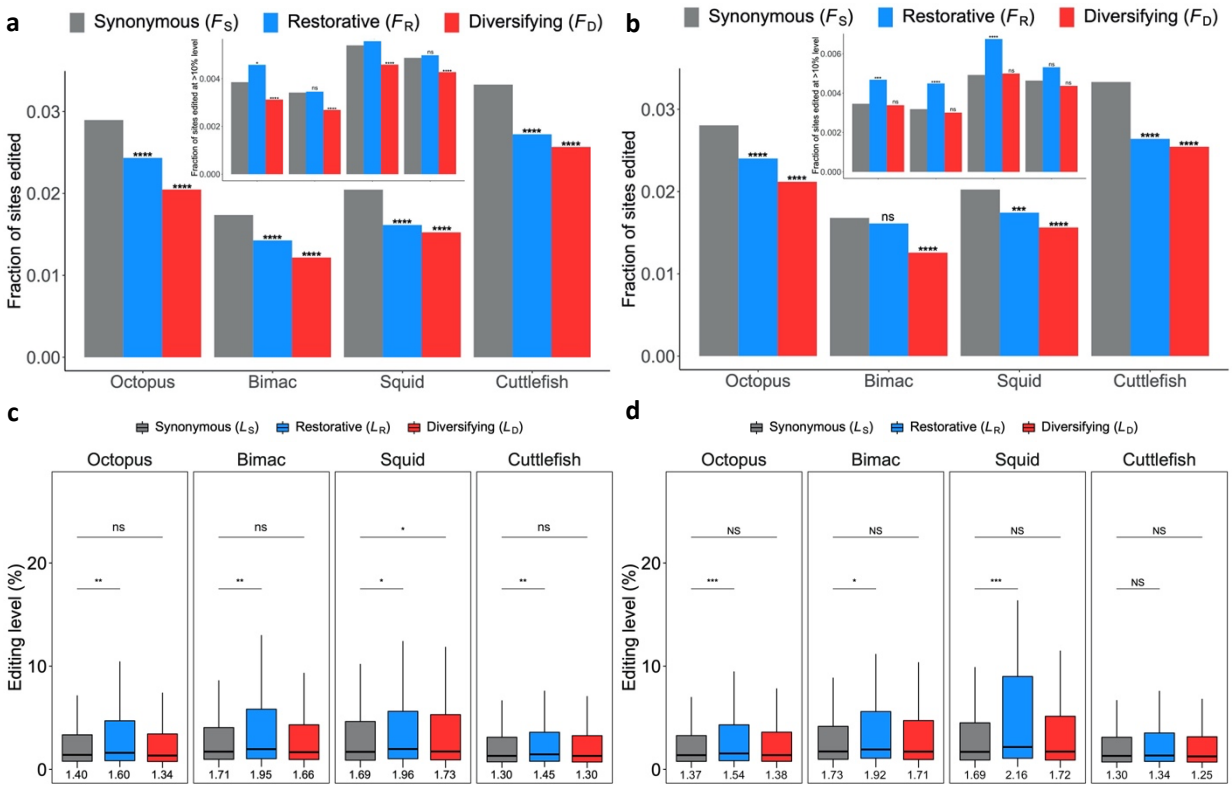


Figure. B-2. Comparison of editing levels between synonymous and nonsynonymous editing sites of different genes. Genes are ranked by d_N/d_S and are divided into two bins, with bin 1 comprising genes with odd ranks and bin 2 comprising genes with even ranks (see main text for details). (**a**) Comparison between editing frequencies of synonymous editing from bin 1 and those of nonsynonymous editing from bin 2. The inset shows fractions of sites with editing levels >10%. (**b**) Comparison between editing frequencies of synonymous editing from bin 2 and those of nonsynonymous editing from bin 1. The inset shows fractions of sites with editing levels >10%. (**c**) Comparison between editing levels of synonymous editing from bin 1 and those of nonsynonymous editing from bin 2. (**d**) Comparison between editing levels of synonymous editing from bin 2 and those of nonsynonymous editing from bin 1. All symbols have the same meanings as in Fig. 3-2.

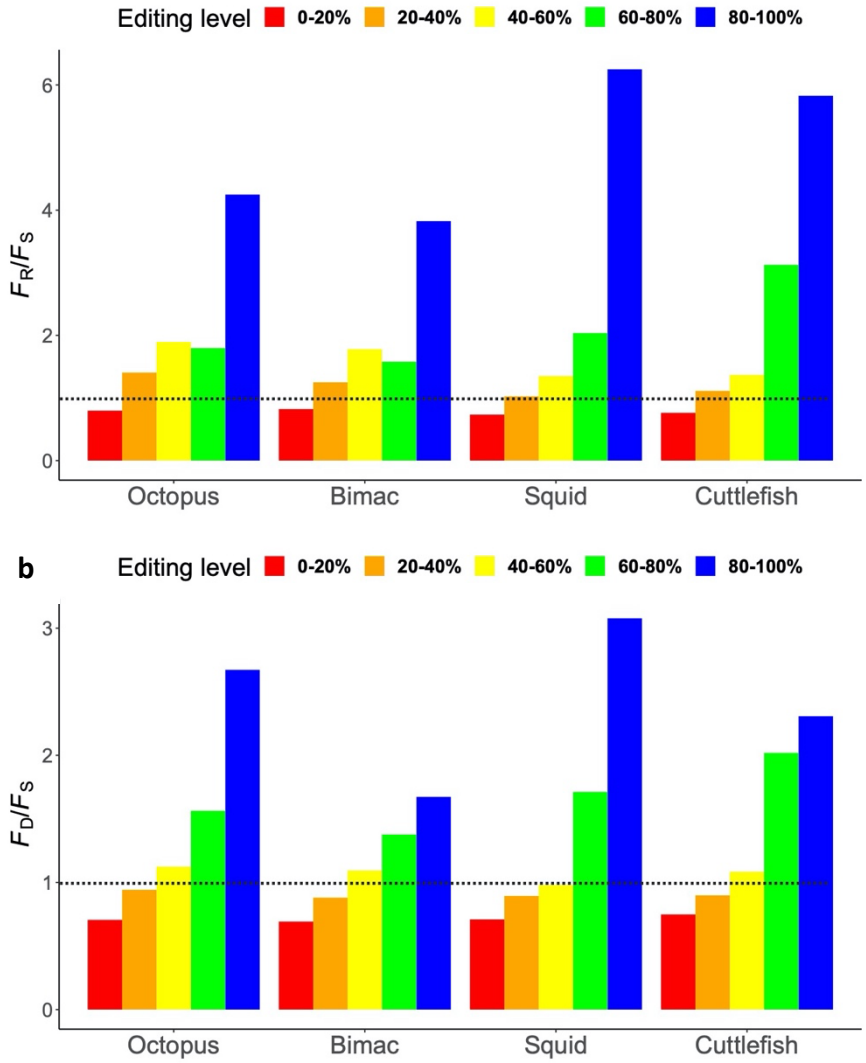


Figure. B-3. Ratios of editing frequencies of synonymous (F_S), diversifying (F_D), and radical (F_R) editing in various ranges of editing levels. (**a**) F_R/F_S. (**b**) F_D/F_S. Note the different Y-axis scales of the two panels. The location of one on Y-axis is marked by the dashed line in each panel.

Table B-1. Numbers of editing sites of each category. In the parentheses are numbers of species-specific editing.

		Nonsynonymous	
	Synonymous	Restorative	Diversifying
Octopus	21,104 (13,537)	2,099 (1,342)	35,449 (22,590)
Bimac	13,457 (5,890)	1,421 (664)	22,364 (9,505)
Squid	14,771 (11,474)	1,301 (1,046)	25,799 (18,849)
Cuttlefish	23,167 (19,870)	2,040 (1,785)	41,506 (34,556)
Shared between octopus and bimac	7,332	742	12,121
Shared between squid and cuttlefish	3,062	240	6,212
Shared among all four coleoids	235	15	738

Table B-2. Patterns of restorative and diversifying editing in 12 different tissues of the bimac. Stars indicate significant differences from synonymous editing (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; chi-squared test and Mann-Whitney U test are used for comparisons of editing frequencies and editing levels, respectively; #, not significant after the Bonferroni correction for multiple testing). As in the main analysis, editing levels are compared for sites covered by at least 400 RNA-seq reads. No restorative editing site satisfying this condition is found in one tissues; its L_R value is shown as "NA".

Tissue	**Editing frequencies**			**Median editing levels (%)**		
	Synonymous (F_S)	Restorative (F_R)	Diversifying (F_D)	Synonymous (L_S)	Restorative (L_R)	Diversifying (L_D)
Axial nerve cord (ANC)	0.0089	0.0083	0.0065****	1.63	2.58	1.78
Optical lobe (OL)	0.0036	0.0035	0.0027****	2.20	2.43	1.94
Subesophageal brain (sub)	0.0043	0.0039*#	0.0031****	2.38	3.46**	2.60
Supraesophageal brain (supra)	0.0031	0.0029*#	0.0022****	1.39	1.73***	1.51***
Posterior salivary gland (PSG)	0.000092	0.00010	0.000075**	0.20	0.28	0.17
Skin	0.00011	0.00013	0.000078****	0.18	0.36*#	0.19
Sucker	0.00017	0.00020	0.00014****	0.25	0.49*#	0.25
Retina	0.00024	0.00030	0.00018****	0.43	0.49	0.44
Ovary	0.000038	0.000047	0.000024***	0.46	NA	0.07
Testes	0.000067	0.000069	0.000048****	0.19	0.26	0.18
Viscera	0.000060	0.000072	0.000044****	0.19	0.16	0.23
Stage 15 embryo (ST15)	0.00028	0.00032	0.00022****	0.40	0.45	0.34

Table B-3. Proportion of editing sites belonging to each editing level range.

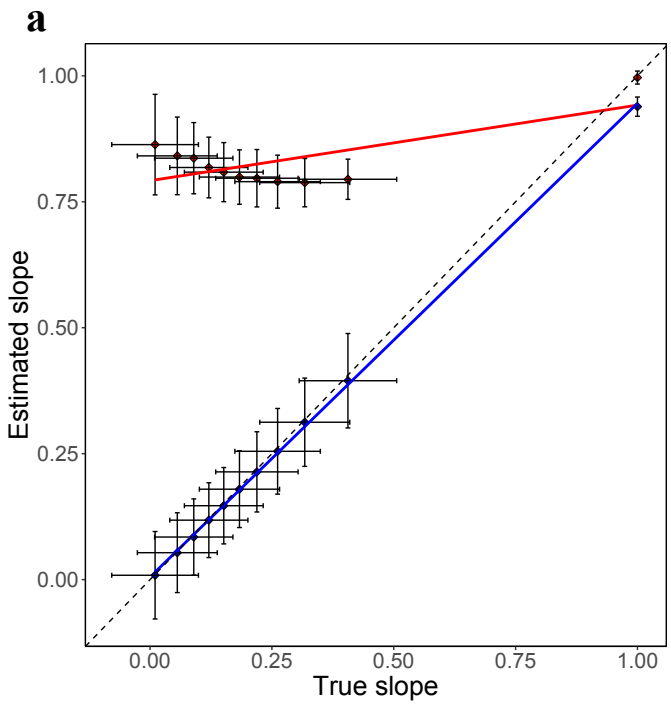
Species	Type	0-20%	20-40%	40-60%	60-80%	80-100%
Octopus	Synonymous	94.16%	3.72%	1.30%	0.67%	0.14%
	Restorative	88.76%	6.19%	2.91%	1.43%	0.71%
	Diversifying	91.21%	4.82%	2.00%	1.44%	0.52%
Bimac	Synonymous	90.64%	5.92%	2.27%	0.91%	0.26%
	Restorative	84.31%	8.37%	4.57%	1.62%	1.13%
	Diversifying	86.99%	7.23%	3.45%	1.73%	0.60%
Squid	Synonymous	86.78%	7.47%	3.30%	1.74%	0.71%
	Restorative	76.02%	9.15%	5.30%	4.23%	5.30%
	Diversifying	80.33%	8.71%	4.21%	3.89%	2.86%
Cuttlefish	Synonymous	93.50%	4.21%	1.48%	0.63%	0.18%
	Restorative	87.99%	5.78%	2.50%	2.45%	1.27%
	Diversifying	90.82%	4.91%	2.08%	1.66%	0.53%

Table B-4. Numbers of sites that are respectively edited, unedited, and substituted in the fourth coleoid among those sites that are edited in the other three coleoids examined.

	Edited A	Unedited A	A-to-G substitution	A-to-C/T substitution
Synonymous editing	235	616	98	31
Nonsynonymous editing	753	1,304	56	37
Diversifying editing	738	1,275	55	35

Appendix C

Supplementary Figures and Tables for Chapter 4



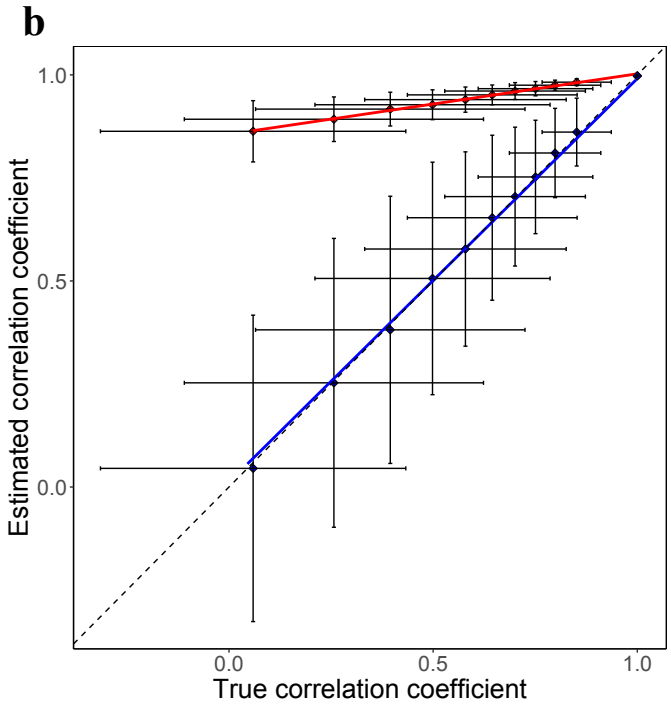


Figure C-1. Performances of Houle *et al.*'s method and the new method across a range of similarity between M and R, with M and R matrices subject to estimation error. (**a**) Linear regression slope of $\log_{10}(R_{obs}$ variance) on $\log_{10}(M_{obs}$ variance) estimated by Houle *et al.*'s method (red) and the new method (blue) plotted against that of $\log_{10}(R_{w}$ variance) on $\log_{10}(M_{original}$ variance) along eigenvectors of $M_{original}$. (**b**) Pearson's correlation coefficient between $\log_{10}(M_{obs}$ variance) and $\log_{10}(R_{obs}$ variance) estimated by Houle *et al.*'s method (red) and the new method (blue) plotted against that between $\log_{10}(M_{original}$ variance) and $\log_{10}(R_{w}$ variance) along eigenvectors of $M_{original}$. Each data point represents the mean of 100 simulations with the same weight parameter w (see Materials and Methods), and the error bars represent standard deviations.

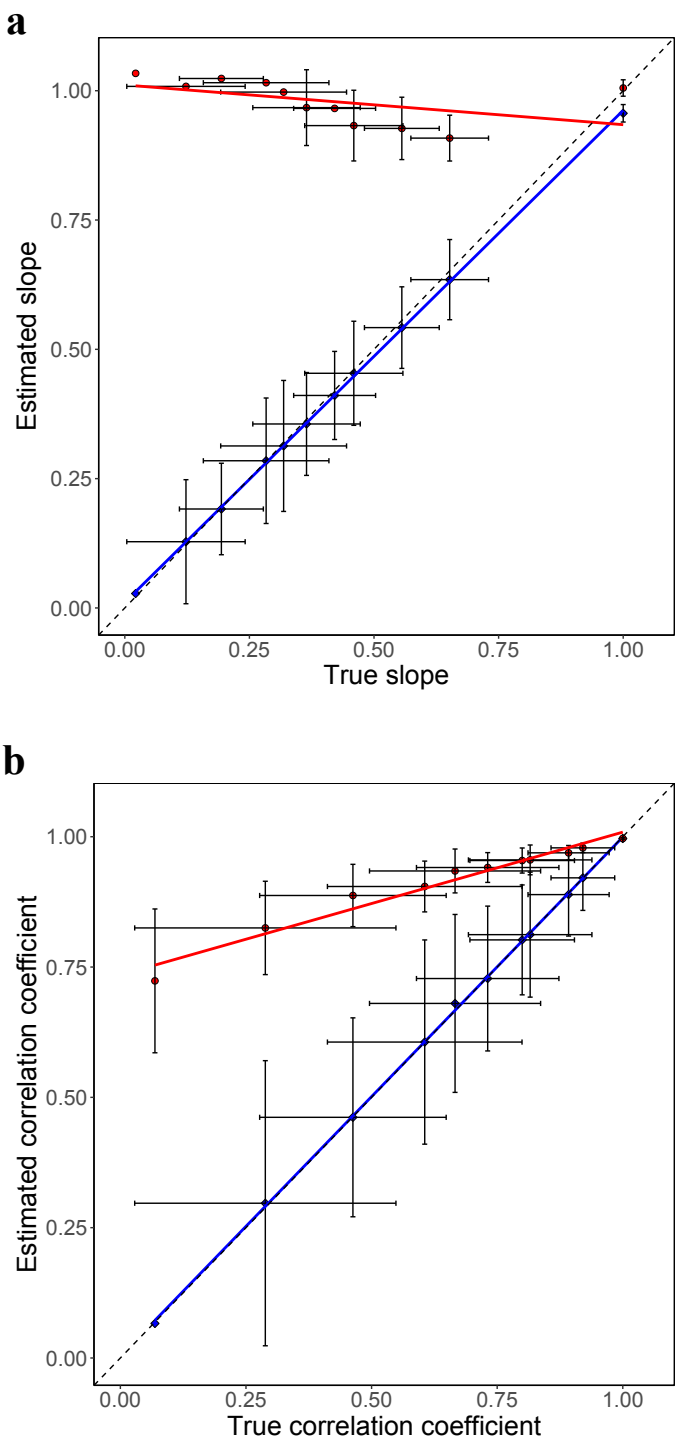


Figure C-2. Performances of Houle *et al.*'s method and the new method when the empirical M matrix is set to be $M_{original}$ and the Drosophilid tree is used to simulate trait evolution (see Materials and Methods). (**a**) Linear regression slope of $\log_{10}(R_{obs}$ variance) on $\log_{10}(M_{obs}$ variance) estimated by Houle *et al.*'s method (red) and the new method (blue) plotted against that

of $\log_{10}(R_w$ variance) on $\log_{10}(M_{original}$ variance) along eigenvectors of $M_{original}$. (**b**) Pearson's correlation coefficient between $\log_{10}(M_{obs}$ variance) and $\log_{10}(R_{obs}$ variance) estimated by Houle *et al*.'s method (red) and the new method (blue) plotted against that between $\log_{10}(M_{original}$ variance) and $\log_{10}(R_w$ variance) along eigenvectors of $M_{original}$. Each data point represents the mean of 50 simulations with the same weight parameter w (see Materials and Methods), and the error bars represent standard deviations.

Table C-1. Models examined by Houle *et al*. and their predictions regarding evolutionary rate, scaling exponent of evolutionary rate with mutational variance, and phylogenetic signal. Model predictions that are consistent with the corrected observations are marked by asterisks.

Model		**Predictions**		
		Evolutionary rate	**Scaling exponent**	**Phylogenetic signal**
Neutral evolution		High	1	High*
Fluctuating directional selection		High	2	High*
Divergent selection		High	2	Intermediate
Moving optimum	BM slow	Low*	~0	High*
	BM fast	High	~0	High*
	White noise	Low*	0	0
	OU slow	Low*	~0	Intermediate
	OU intermediate	Low*	0-1*	Low
	OU fast	Low*	~0	~0
Observations (Houle *et al*.)		Low	~1	High
Observations (corrected)		Low	0.54	High

Table C-2. Patterns of phenotypic evolution observed from computer simulation of non-neutral, orthogonal, focal traits with mutational pleiotropy.

Determinant of *M* variance	$\boldsymbol{\alpha}$	**Scaling exponent**	**Median relative divergence***	**Median phylogenetic signal**
Mutation rate	10^{-4}	0.53	4.37×10^{-5}	0.979
Mutation rate	10^{-3}	0.59	3.97×10^{-5}	0.98
Mutation rate	10^{-2}	0.51	3.77×10^{-5}	0.978
Mutation effect size	10^{-4}	0.51	4.82×10^{-6}	0.968
Mutation effect size	10^{-3}	0.52	4.74×10^{-6}	0.968
Mutation effect size	10^{-2}	0.42	4.74×10^{-6}	0.971

* Ratio of observed divergence (i.e., phenotypic variance across replicate lineages) and expected divergence (i.e., product of mutational variance and time).

Appendix D

Supplementary Figures and Tables for Chapter 5

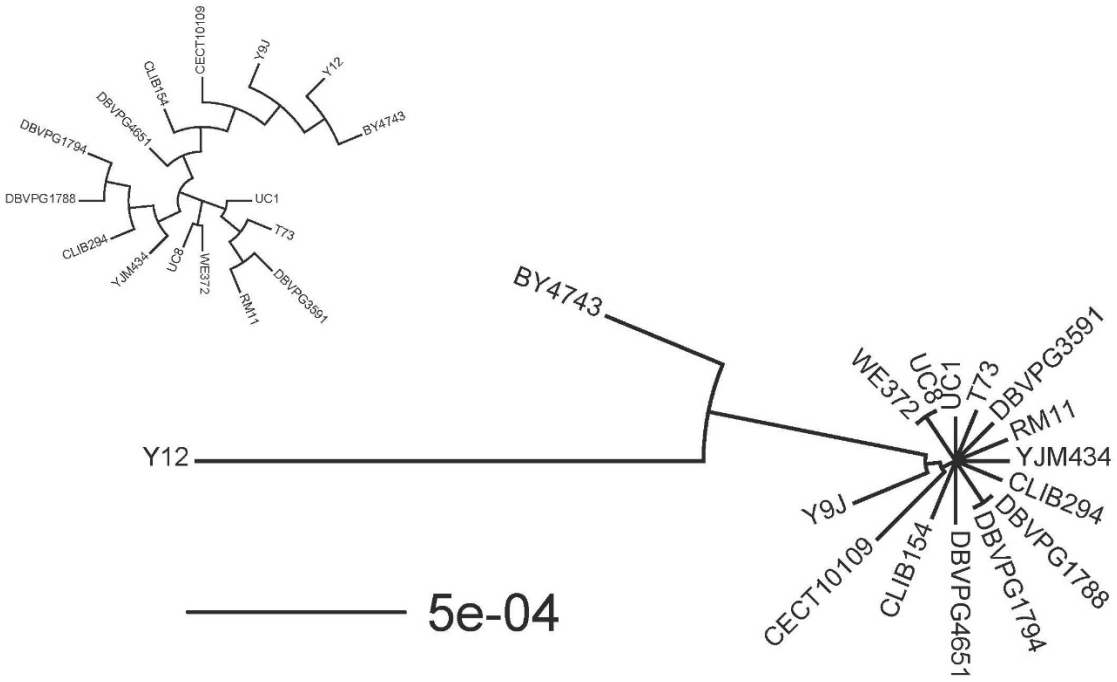
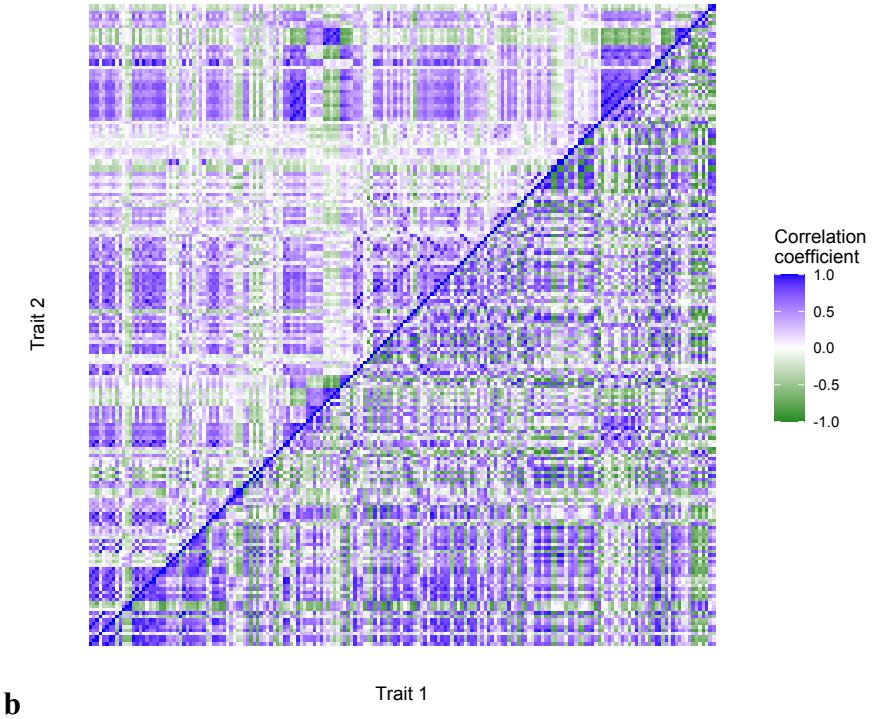


Figure. D-1. Neighbor-joining tree of the 16 natural yeast strains used in this study, based on 1,544,489 biallelic single nucleotide polymorphism (SNP) sites. Scale bar indicates genomic divergence level. The tree was based on the distance matrix downloaded from http://1002genomes.u-strasbg.fr/files/1011DistanceMatrixBasedOnSNPs.tab.gz. The inset at the top left coner shows the tree topology but the branch lengths are not drawn to scale.

a



b

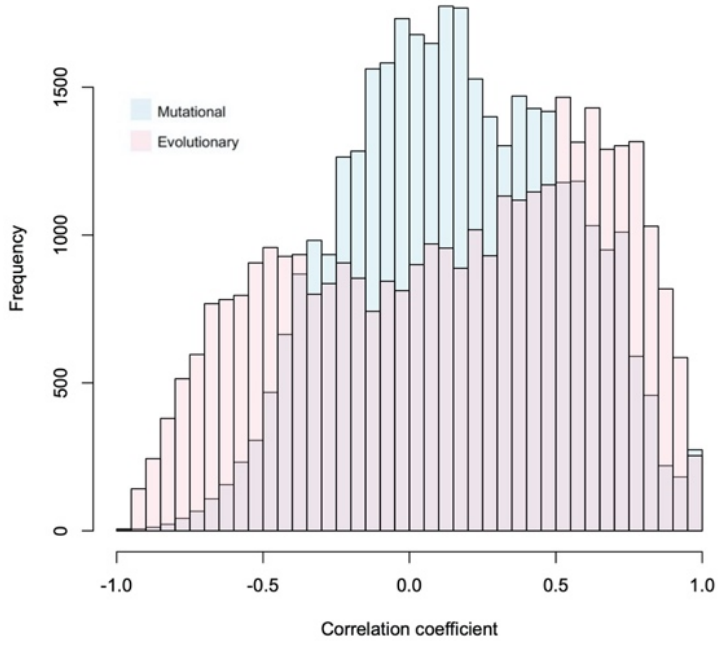
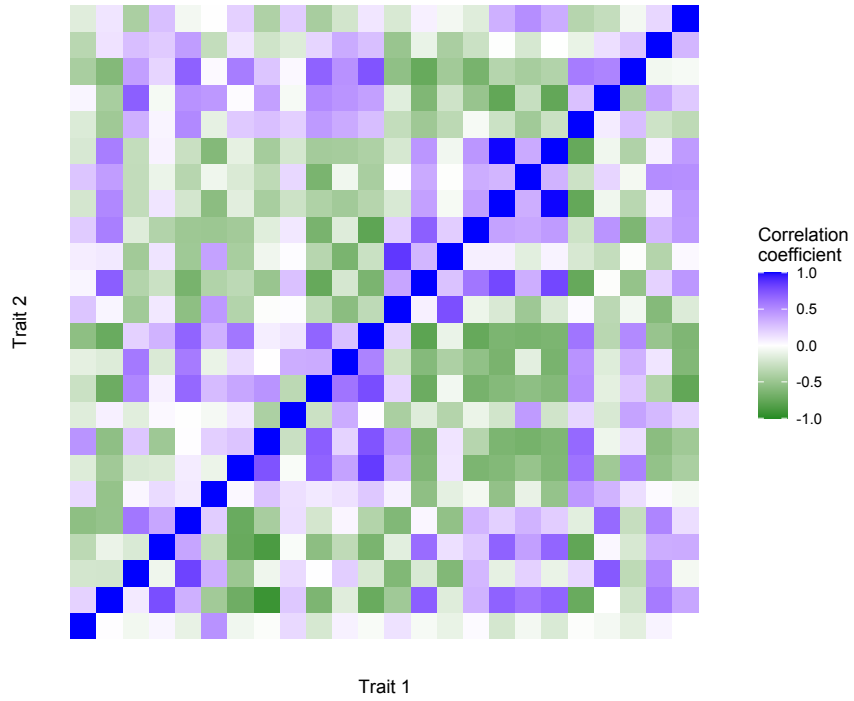


Figure. D-2. Mutational (COR_M) and evolutionary (COR_E) correlations for all pairs of the 187 yeast morphological traits. COR_M is based on yeast mutation accumulation lines. (a) COR_M (upper triangle) and COR_E (lower triangle) for all pairs of traits ordered according to their IDs.

(b) Frequency distributions of COR_M (blue) and COR_E (red) across all trait pairs. The two distributions are significantly different ($P < 10^{-10}$, Kolmogorov–Smirnov test).

a



b

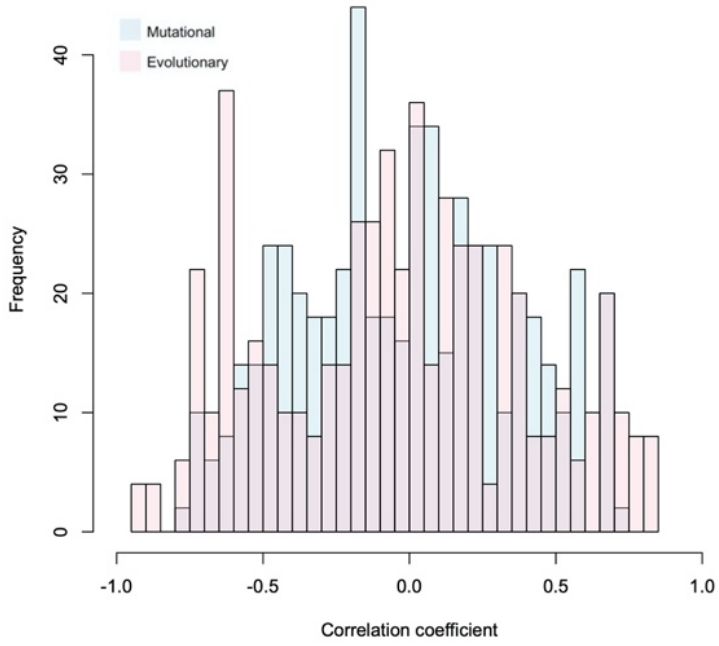


Figure. D-3. Mutational (COR_M) and evolutionary (COR_E) correlations for all pairs of the 24 fly wing morphological traits. (a) COR_M (upper triangle) and COR_E (lower triangle) for all pairs of

traits ordered in the same way as in the original dataset. (b) Frequency distributions of COR_M (blue) and COR_E (red) across all trait pairs. The two distributions are significantly different (P = 0.0015, Kolmogorov–Smirnov test).

Table D-1. Pearson's correlation in eigenvalues between M and the covariance matrix estimated from a sample from M ($\widetilde{M}$). Results from 1000 replications are shown.

Taxon	Sample size	Minimum correlation coefficient	Median correlation coefficient
Yeast	4,817	0.9993745	0.9999503
Fly	150	0.9698937	0.9967839

Table D-2. Parameters and results of simulations of trait-trait coevolution in the presence of mutational bias.

Mutational bias B	COR_M	Median $\widehat{COR_M}$	Fraction of $\widehat{COR_M} > COR_M$	Median COR_E at the end of simulation	Fraction of simulations with $COR_E > COR_M$	COR_E variance
2	0.9	0.900	45.5%	0.885	31.5%	0.0185
	0.5	0.500	49.5%	0.468	39.5%	0.0175
	0.1	0.102	54.5%	0.0999	50%	0.0210
10	0.9	0.900	50.5%	0.863	13%	0.0198
	0.5	0.499	46.5%	0.426	24.5%	0.0261
	0.1	0.101	54.5%	0.0886	46.5%	0.0168
100	0.9	0.900	52%	0.861	12%	0.0194
	0.5	0.500	49%	0.420	27%	0.0261
	0.1	0.101	52%	0.0856	46.5%	0.0211
1000	0.9	0.900	49.5%	0.863	14%	0.0218
	0.5	0.500	49%	0.443	30.5%	0.0228
	0.1	0.101	52.5%	0.0849	44.5%	0.0237