Tsuboi, M, Sztepanacz, J, De Lisle, S, Voje, KL, Grabowski, M, Hopkins, MJ, Porto, A, Balk, M, Pontarp, M, Rossoni, D, Hildesheim, LS, Horta-Lacueva, QJ-B, Hohmann, N, Holstad, A, Lürig, M, Milocco, L, Nilén, S, Passarotto, A, Svensson, EI, Villegas, C, Winslott, E, Liow, LH, Hunt, G, Love, AC and Houle, D

The Paradox of Predictability Provides a Bridge Between Micro- and Macroevolution

The Paradox of Predictability Provides a Bridge Between Micro- and Macroevolution

Masahito Tsuboi[1,†], Jacqueline Sztepanacz[2], Stephen De Lisle[1,3], Kjetil L. Voje[4], Mark Grabowski[5], Melanie J. Hopkins[6], Arthur Porto[7], Meghan Balk[4], Mikael Pontarp[1], Daniela Rossoni[8], Laura S. Hildesheim[1], Quentin J.-B. Horta-Lacueva[1], Niklas Hohmann[9, 10], Agnes Holstad[11], Moritz Lürig[1], Lisandro Milocco[1], Sofie Nilén[1], Arianna Passarotto[1,12], Erik I. Svensson[1], Cristina Villegas[13], Erica Winslott[1], Lee Hsiang Liow[4, 14], Gene Hunt[15], Alan C. Love[16] and David Houle[8]

1. Department of Biology, Lund University, Lund, Sweden
2. Department of Ecology and Evolutionary Biology, University of Toronto, Canada
3. Department of Environmental and Life Sciences, Karlstad University, Karlstad, Sweden
4. Natural History Museum, University of Oslo, Oslo, Norway
5. Research Centre for Evolutionary Anthropology and Palaeoecology, School of Biological and Environmental Sciences, Liverpool John Moores University, UK
6. Division of Paleontology (Invertebrates), American Museum of Natural History, New York, NY, USA
7. Florida Museum of Natural History, University of Florida, Gainesville, USA
8. Department of Biological Science, Florida State University, Tallahassee, Florida, USA
9. Department of Earth Sciences, Utrecht University, Utrecht, The Netherlands
10. Institute of Evolutionary Biology, Biological and Chemical Research Centre, Faculty of Biology, University of Warsaw
11. Department of Biology, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway
12. Facultad de Biología, Universidad de Sevilla, Sevilla, Spain
13. Centro de Filosofia das Ciências da Universidade de Lisboa, 1749-016, Lisboa, Portugal.
14. Centre for Planetary Habitability, Department of Geosciences, University of Oslo, Oslo, Norway
15. Department of Paleobiology, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA
16. Department of Philosophy and Minnesota Center for Philosophy of Science, University of Minnesota, Minneapolis, MN, USA

[†]Corresponding Masahito Tsuboi (masa.tsuboi@gmail.com)

Acknowledgements

This article was crafted from discussions with participants of the Progress Meeting in Evolutionary Biology workshop "*Advancing 21st Century Evolutionary Biology through Evolvability in the Fossil Record*" that took place in Lund (Sweden), 23-26 April 2023. Funding was provided by the European Society for Evolutionary Biology in an award to MT, LHL and SL, by the Center for Advanced Studies in Norwegian Academy of Science and Letters through the project "*Evolvability: a unifying concept in evolutionary biology?*" awarded to Thomas F. Hansen and Christophe Pélabon, and by the Japanese Society for the Promotion of Science Overseas Research Fellowship awarded to MT (202270014). We thank Christina Rangefors for assisting with the organization of the workshop, the Department of Biology at Lund University for hosting the workshop, and the guest editor Thibault Latrille and three anonymous reviewers for providing constructive suggestions on an earlier version of the manuscript.

Data availability statement

Data and code to reproduce all results and figures presented in this study is deposited at https://www.doi.org/10.5281/zenodo.11059549.

Abstract

The relationship between the evolutionary dynamics observed in contemporary populations (microevolution) and evolution on timescales of millions of years (macroevolution) has been a topic of considerable debate. Historically, this debate centers on inconsistencies between microevolutionary processes and macroevolutionary patterns. Here, we characterize a striking exception: emerging evidence indicates that standing variation in contemporary populations and macroevolutionary rates of phenotypic divergence are often positively correlated. This apparent consistency between micro- and macroevolution is paradoxical because it contradicts our previous understanding of phenotypic evolution and is so far unexplained. Here, we explore the prospects for bridging evolutionary timescales through an examination of this "*paradox of predictability*." We begin by explaining why the divergence-variance correlation is a paradox, followed by data analysis to show that the correlation is a general phenomenon across a broad range of temporal scales, from a few generations to tens of millions of years. Then we review complementary approaches from quantitative-genetics, comparative morphology, evo-devo, and paleontology to argue that they can help to address the paradox from the shared vantage point of recent work on evolvability. In conclusion, we recommend a methodological orientation that combines different kinds of short-term and long-term data using multiple analytical frameworks in an interdisciplinary research program. Such a program will increase our general understanding about how evolution works within and across timescales.

Key words: evolutionary prediction, evolvability, genotype-phenotype map, macroevolution, quantitative genetics, phenotypic integration, phylogenetic comparative methods, time-series analysis

Introduction

"*How wonderful that we have met with a paradox. Now we have some hope of making progress.*"

Niels Bohr

The relationship between evolutionary change in contemporary populations (microevolution) and the pattern of evolution over millions of years (macroevolution) has been a topic of much controversy. Two important questions recur in these debates. First, are microevolutionary processes plus time adequate to *explain* macroevolutionary patterns? Many favor the view that the microevolutionary processes of mutation, selection, drift and gene flow operating within populations can, or at least should be able to, explain macroevolution (e.g., Charlesworth et al., 1982, Arnold, 2014). Others argue that additional processes occurring above the population level, such as species selection and biased extinctions (Erwin, 2000, Jablonski, 2017b), are also important. Second, can our knowledge of currently acting microevolutionary processes *predict* macroevolution across timescales (Estes and Arnold 2007)? It is widely accepted that consistent regimes of natural selection can result in convergence on similar adaptive solutions (Blount et al., 2018, Mas et al., 2020). However, in the absence of consistent selection over long periods of time, it was traditionally assumed that current microevolutionary changes would not be consistent enough over geological timescales to explain macroevolutionary patterns. Indeed, microevolutionary processes evaluated in contemporary populations are often inconsistent with macroevolutionary patterns (Hansen & Houle, 2004, Futuyma, 2010, Uyeda et al., 2011). This decoupling between micro- and macroevolutionary processes is often attributed to inferential bias due to increased noise in sparse fossil samples (Kidwell & Holland, 2002) and to factors that are unique to geological timescales, including dramatic ecological shifts (Simons, 2002), rare evolutionary transitions like the evolution of novel traits and key innovations (Jablonski, 2017a), or stochastic events, such as the asteroid strike that is considered a primary cause of the end-Cretaceous mass extinction (Chiarenza et al., 2020).

In recent years, a striking exception to this widely perceived micro-macro decoupling has gathered increasing evidential support. Many studies now show that the rate of phenotypic evolution across taxa, as reflected in divergence, is correlated with the standing variation estimated in contemporary populations. Schluter (1996) first showed that the trajectories of divergence between populations or closely related species were in directions closely aligned to those with the highest additive genetic variation: "genetic lines of least resistance." More recently, Houle et al. (2017) showed that the rate of evolution in 20 wing-shape traits measured in a sample of 110 drosophilid fly species that have diverged over at least 30 million years was tightly correlated with both mutational and additive genetic variation among those traits measured in *Drosophila melanogaster*. Among the 36 multivariate studies included in a recent review, 30 show positive correlations between variation and pattern of divergence among populations or species (Voje et al., 2023a). Two recent meta-analyses have revealed divergence-variance correlations among 33 plant species from 48 multivariate studies (Opedal et al., 2023) and in 409 traits in 123 contemporary species and 589 traits in 150 fossil lineages (Holstad et al., 2024). Such a general correlation between intra-population variation and rates of phenotypic divergence implies a predicable link between micro- and macro-evolution.

We refer to these findings as "*the paradox of predictability*" because this predictability runs contrary to expectations from the established theory of phenotypic evolution by mutation,

selection, and heritability (e.g., Simpson, 1944, Lande, 1979, Arnold, 2014). For example, a seemingly obvious explanation for the divergence-variance correlation is that the amount of variation limits the rate of evolution (Schluter, 1996). However, the amounts of genetic and mutational variation expected under a neutral model are sufficient to support rates of evolution orders of magnitude higher than we observe (Lynch, 1990, Hansen & Houle, 2004, Uyeda et al., 2011). Furthermore, as noted, there are a myriad of factors that can decouple micro- and macroevolution (e.g., Erwin, 2000, Simons, 2002, Jablonski, 2017b). Consequently, even as our ability to predict aspects of macroevolution from contemporary variation grows, we cannot yet explain why.

In this review, we explore the proposition—implied by the quote from Niels Bohr—that the paradox of predictability offers an exciting research opportunity to deepen our understanding of how microevolutionary processes are related to macroevolutionary patterns. We outline a research program that goes beyond the largely correlational evidence documenting the divergence-variance relationships to secure a mechanistic understanding of the processes that cause the correlations. Such an understanding should incorporate data, concepts, methods, and results from many disciplines that have explored different dimensions of the paradox from the standpoint of recent work on evolvability: the capacity of a population to evolve in response to selection (Wagner & Altenberg, 1996). Those disciplines include paleontology (Love et al., 2022, Jablonski, 2023), quantitative genetics (Hansen & Houle, 2008), evolutionary developmental biology (Hendrikse et al., 2007), experimental evolution (Colegrave & Collins, 2008), comparative morphology (Klingenberg, 2008), and theoretical biophysics (Kaneko & Furusawa, 2018). With this in mind, we first clarify why the divergence-variance correlation is a paradox. Second, we confirm the generality of the correlation between variation and the rate and direction of phenotypic divergence through analyses of newly assembled time-series data. Then we show how different evolvability-oriented disciplines, including quantitative genetics, the comparative study of morphology, evo-devo, and paleontology, offer powerful complementary approaches to resolving the paradox. We conclude with a list of outstanding questions that help to clarify how distinct disciplines and approaches might be integrated, thereby advancing research on the paradox of predictability.

Why the divergence-variance correlation is paradoxical

Evidence for the correlation of variation and evolutionary divergence comes primarily from the study of morphological traits like size and shape that show polygenic inheritance (e.g., Houle et al., 2017, McGlothlin et al., 2018, Rohner & Berger, 2023). Variation in evolutionary rates in this class of traits is determined by two factors: (i) evolvability, the capacity of a population to evolve (Wagner & Altenberg, 1996), and (ii) the factors that act on variation (including evolvability) to cause changes, such as natural selection or drift (Hansen, 2023). The roles of these two factors are neatly captured in the multivariate Lande equation from quantitative genetics (Lande, 1979):

$$\Delta\bar{\mathbf{z}} = \mathbf{G}\boldsymbol{\beta} \quad (1)$$

The change over one generation due to selection ($\Delta\bar{\mathbf{z}}$) is determined by the additive genetic co/variance matrix **G** (G-matrix) and the direction and strength of selection ($\boldsymbol{\beta}$). See Box S1 in the Supplementary Information [SI] for more explanation. Because the evolutionary

divergence represents the left-side of Equation 1 ($\Delta\bar{\mathbf{z}}$) integrated across multiple populations and/or taxa, the correlation between variation and divergence suggests that both **G** and the pattern of selection should play a role in shaping the rate and direction of evolution. Based on this theoretical framework, there are two paradoxical aspects of the divergence-variance correlation.

First, the prevalence of divergence-variance correlation means that variation within populations often remain similar over millions of years. Traditional models of the maintenance of genetic variation suggest that the stochasticity of mutation and drift plus the alteration of variation caused by natural selection will cause changes in genetic architecture (see "Quantitative Genetics and the Paradox of Predictability"). In addition, studies of the relationship between genotype and phenotype (the genotype-phenotype map, the GP-map) suggest that phenotypic evolution over macroevolutionary timescales is liable to change the underlying propensity for genetic changes to alter phenotypes (i.e., variability). As we detail in the section "Evo-Devo and the Paradox of Predictability", there are theoretical and empirical results suggesting that the GP-map can change the pattern of variation in short time scales. Despite these findings, there is ample evidence that the standing variation is remarkably consistent over long time periods (e.g., McGlothlin et al., 2018, Rossoni et al., 2019).

The second paradoxical aspect concerns why there should be a relationship between variation and long-term evolution (even given that variation is stable). There are two known classes of hypotheses that can explain divergence-variance correlations: "constraint" hypotheses, which share a premise that the amount of variation limits the rates of phenotypic divergence (Schluter, 1996), and "selection-shapes-variation" hypotheses that assume the same selective forces causing the pattern of long-term evolution also shapes the nature of variation to conform to that pattern (Cheverud, 1984, Riedl, 1978, Waddington, 1957). Both classes of hypotheses have serious difficulties in explaining the divergence-variance correlation in macroevolutionary timescales.

Over a short timescale of a few generations, we can readily explain a correlation between genetic variation and evolutionary rate because standing genetic variation is the fuel that enables the response to selection or drift in a single generation (Equation 1). If genetic drift is the cause of evolutionary change, or if selection varies randomly in direction, change will be proportional to the amount of variation in a population. We call this basic expectation a "constraint" hypothesis because it depends on genetic variation limiting the rate and direction of evolution. Constraint hypotheses, however, cannot readily be extrapolated to macroevolutionary timescales. First, many factors can alter the amounts of genetic variation that a population harbors (Pélabon et al., 2023), including selection, genotype-environment interactions, gene flow, and population size. Thus, genetic variation in one population at a particular time might not predict variation in descendant populations (e.g., Pujol et al., 2018), let alone variation in sister taxa. Second, macroevolutionary rates are generally too slow for levels of genetic variation to be constraining (Williams, 1992, Hansen & Houle, 2004). For example, Houle et al. (2017) calculated that the expected neutral evolutionary rate for *Drosophila* wing shape would be 10,000 times larger than observed, a pattern typical for a wide range of taxa (e.g., Lynch, 1990). Evolutionary rates due to selection would generally be even higher. A third challenge to constraint hypotheses is that empirical estimates of the scaling relationship between variation and divergence are near 1 (Houle et al., 2017, Holstad et al., 2024), so rates of divergence are proportional to variation. Genetic drift predicts a scaling relationship of 1, but—as noted above—it also predicts evolutionary rates much higher than the observed rates of evolution. If fluctuating directional

selection dictates the rate of evolution, we would expect a scaling relationship of 2, since both taxa would be independently subjected to directional selection. None of the simple models of phenotypic evolution, such as Brownian motion or Ornstein-Uhlenbeck processes can readily explain a scaling relationship near 1 (Bolstad et al., 2014, Houle et al., 2017, Holstad et al., 2024). A plausible model explaining the scaling relationship between 0 and 1 found in many datasets is a two-layered Ornstein-Uhlenbeck process where populations track fluctuating optima (Bolstad et al., 2014), a scenario that has recently been found to account for an effect of evolvability on evolutionary divergence in two large datasets of both extant and extinct populations (Holstad et al., 2024).

The second class of hypothesis, which Rupert Riedl termed the "imitative epigenotype" hypothesis (Riedl, 1978), is that the pattern of genetic variation matches the long-term pattern of selection. We label this a "selection-shapes-variation hypothesis" (see Olson & Miller, 1958, Riedl, 1978, Wagner et al., 1997, Houle & Rossoni, 2022 for related discussion in the literature). This class of hypotheses invert the cause and effect of the constraints hypothesis by positing that within-population variation is increased by past directional and disruptive selection and/or decreased by consistent stabilizing selection (e.g., Waddington, 1957, Riedl, 1978, Cheverud, 1984, Wagner, 1996). Models have found plausible conditions under which this matching of selective and variational patterns can evolve (Draghi & Wagner, 2008, Pavlicev et al., 2011, Draghi & Whitlock, 2012, Jones et al., 2014). The most likely route for past selection to shape current variation is through changes to variability—the tendency for genetic variation to cause changes in the phenotype (Wagner & Altenberg 1996). However, conditions where variability evolves in a counterintuitive manner are readily identified, such as an increase in variability when stabilizing selection is too strong (Wagner et al., 1997, Hermisson et al., 2003, Hansen et al., 2006). Over macroevolutionary timescales, dramatic and unpredictable changes in selective regimes are also likely (Simons, 2002, Jablonski, 2017a), calling into question whether selection provides an adequate basis for macroevolutionary predictions.

The lack of a clear explanation for the pattern becomes more troubling with every study that confirms the correlation (Opedal et al., 2023, Voje et al., 2023a, Holstad et al., 2024), particularly over long temporal scales (Hunt, 2007, Houle et al., 2017, Tsuboi et al., 2018, McGlothlin et al., 2018, Rohner & Berger, 2023). Our own analysis of evolutionary divergence and variation drawn from literature, which includes both paleontological and neontological data (Box 1; see SI for methodological details), aligns with the finding of Holstad et al. (2024) that a positive correlation exists between variation and divergence also in paleontological data and suggests that the relationship is a general feature of phenotypic evolution across a broad range of temporal scales.

[Box 1 is inserted here]

We acknowledge that our formulation of the paradox partly reflects our own view of what constitutes a paradox. The purpose of our review is to offer a point of departure for future studies on related issues and paradoxes. Also, there are two definitions of macroevolution in the literature, one concerns distinct processes that are exclusively present at higher levels (e.g., Jablonski, 2017a, Erwin, 2000) and the other concerns all processes at and above the species level (e.g., Charlesworth et al., 1982, Futuyma, 2010). Our review primarily adheres to the latter that uses the species level as the defining feature of macroevolution. With these cautionary notes

in mind, we turn to elaborating how quantitative genetics, comparative morphology, evo-devo, and paleontology can contribute to solving the paradox of predictability.

Quantitative genetics and the paradox of predictability

Quantitative genetics forms the theoretical foundation of the paradox of predictability and provides a basis to formulate the two classes of hypotheses introduced in the previous section and predictions for empirical tests. Here, we address the potential causes of G-matrix stability and what accounts for the predictive relationship between variation and long-term evolution.

Variation, variability and genetic architecture

Although the quantitative-genetic explanations come in different flavors, they share the core premise that the paradox rests on a trait-by-trait assessment of variation and selection. When each trait of an organism is considered independently, the standing genetic variation is typically high, indicating that a lack of genetic variation should not be a limiting factor for evolution. However, organisms do not comprise a collection of independent traits. Pleiotropy, where a single genetic variant affects two or more traits (Pavlicev & Cheverud, 2015), is widespread (Stearns, 2010) and leads to genetic covariances between traits. For example, the most variable aspect of morphology is often overall size, and the sizes of most body parts are typically positively correlated with each other, suggesting that genetic variants affect the sizes of multiple body parts in similar ways (e.g., allometry; Marroig & Cheverud, 2005, Voje et al., 2014). It is widely accepted that genetic covariances caused by pleiotropy determine the short-term pattern of evolutionary divergence (reviewed in Walsh & Blows, 2009), and emerging evidence indicates a role for genetic covariance in determining the pattern of long-term divergence (Sztepanacz & Houle, 2021). In addition, genetic variance is unevenly distributed in phenotype space. Some trait combinations have large amounts of genetic variation, such as the axis in multivariate morphospace that describe a general size of an organism (Marroig & Cheverud, 2005, Voje et al., 2014), whereas other combinations have far less.

In addition, it is not just the magnitude of genetic variance that is important for evolution. The same amount of genetic variance can be caused by different genetic architectures, and thus the specific genetic architecture may also be important. This points to explorations of the developmental, functional, and molecular mechanisms that translate genetic perturbations into differences in phenotype (e.g., the genotype-phenotype map or GP-map). Some GP-maps result in large changes in phenotype for a given change in the genome, which confers a steep and potentially rugged form to the GP-map. Conversely, a population in a flat part of the map will generate little genetic variation per mutation. Such knowledge on the form of the GP-map will improve our explanations of why variation and variability differ across traits and taxa.

Over long timescales, variability (i.e., the propensity to vary) should be the predominant determinant of evolvability (Lande, 1976, Lynch & Hill, 1986, Lynch, 1990). Like standing variation, variability—often evaluated as the mutational variance—is often concentrated in a small part of phenotypic space (Hine et al., 2018, Dugand et al., 2021), although mutational variance can affect all parts of the phenotype space (Lynch, 1990, Houle & Fierst, 2013). In two dipteran species, mutational variances are strongly correlated with both standing variation and rates of phenotypic divergence among populations or species, suggesting that mechanisms

underlying variability evolve slowly, resulting in predictability over macroevolutionary timescales (Houle et al., 2017, Rohner & Berger, 2023).

What intrinsic factors could make genetic architecture stable?

Evidence from quantitative trait locus (QTL) mapping and genome-wide association studies (GWAS) indicate that most quantitatively varying traits have a highly polygenic basis (Boyle et al., 2017, Sella & Barton, 2019). A useful model of variation in quantitative genetics is the infinitesimal model, where a trait is determined by an infinite number of alleles with infinitely small effects (Barton et al., 2017). Although the infinitesimal model cannot be literally true, infinitesimal-like properties arise quickly when traits have a polygenic genetic architecture and when effective population sizes are large (Barton et al., 2017). This hypothesis is consistent with the remarkable success of artificial selection experiments (Hill & Caballero, 1992), as well as predictions of trait evolution over ecological timescales of tens to hundreds of generations (Hill & Kirkpatrick, 2010), and suggests that an infinitesimal-like genetic basis is common for quantitative traits. Importantly, the more polygenic a trait is, and the more intermediate allele frequencies are, the slower genetic variances will change under selection or drift (Barton, 2022). This can help to explain long-term stability of genetic variance despite large evolutionary changes in trait means.

A handful of studies have hypothesized how genetic architecture may differ between traits that have high and low genetic variation. For example, in an artificial selection experiment, Hine et al. (2014) applied directional selection to many multivariate trait combinations of **G**. They found that trait combinations with high genetic variation responded consistently to directional selection in all replicate populations. However, trait combinations with low genetic variance showed considerable variation in their responses, although on average they did respond as predicted. The authors suggested that the sampling of low frequency deleterious alleles in low-variance trait combinations was responsible for the stochastic responses observed. Similarly, Sztepanacz and Blows (2017) applied disruptive artificial selection to multivariate trait combinations that harbor the highest and lowest amount of additive genetic variance (i.e., $\boldsymbol{g}_{max}$ and $\boldsymbol{g}_{min}$) and found that the response to selection in $\boldsymbol{g}_{min}$ was consistent with a genetic architecture of few deleterious alleles held at low frequency in the starting population. If this implies a difference in the number, effect size, and pleiotropic effects of alleles that underlie trait combinations with high versus low genetic variation, mechanisms underlying such differences could help to explain long-lasting evolutionary constraints.

An infinitesimal-like, polygenic architecture for standing variation will make the G-matrix resistant to change from changes in allele frequency. As noted above, the pattern of variability, captured in the mutational variance co/variance matrix (**M**), is likely to play an important role in dictating genetic architecture. If **M** is stable, short-term changes in **G** are likely to represent fluctuations around a long-term expectation set by the variability of the phenotype (Svensson & Berger, 2019, Pélabon et al., 2023). Therefore, to explain the stability of **G**, we also need to explain how the structure of the GP-map could keep **M** stable.

Following de Visser et al. (2003) and Hansen (2006), we distinguish three kinds of selective forces that potentially shape the GP-map: intrinsic, adaptive, and congruent. Under the "intrinsic" hypothesis, the GP-map is shaped by the evolution of the mean phenotype. Once a population is driven to a high fitness phenotype, the GP-map that allowed the population to achieve that phenotype is preserved by stabilizing selection for the phenotype. This intrinsically

generated GP-map might be shared by clades of organisms, such as tetrapod vertebrates or radial echinoderms. The other two hypotheses suggest that the GP-map is also shaped by the nature of the genetic or phenotypic variation they allow a population to generate. In the "adaptive" scenario, genetic architecture is shaped directly by selection on the nature of the genetic variation that mutation produces. This might include minimizing mutations likely to have particularly deleterious effects or generating mutations more likely to be favored by changing environmental conditions. In the "congruence" scenario, the genetic architecture is shaped by selection on the organismal response to environmental variation, which indirectly shapes genetic architecture and the **M** matrix. If adaptive or congruent reshaping of the GP-map is a primary determinant of variability, **G** matrix stability rests on the stability of the selective forces that shape that variability. One simulation study suggested that the adaptive reshaping of variational properties of the GP-map leads to a more stable **G** over a timescale of a few thousand generations (Jones et al., 2007). Although we have little empirical evidence about adaptive or congruent selection to explain a stable **M** matrix, either would have to be very consistent, or the evolutionary response in the GP-map would have to be quite slow (see further discussion below).

What extrinsic factors could make genetic architecture stable?

Factors extrinsic to the organism—selection—can shape variation through either "adaptive" or "congruence" mechanisms (see above). Selection affects allele frequencies and short-term genetic variation due to directional, non-linear (stabilizing or disruptive) and epistatic selection (Hansen & Wagner, 2023). Directional and non-linear effects on genetic variation are summarized as $\Delta\mathbf{G}_s = \mathbf{G}(\boldsymbol{\gamma} - \beta\beta^{\mathrm{T}})\mathbf{G}$, where $\boldsymbol{\gamma}$ is a matrix describing multivariate nonlinear selection. Positive and negative values of $\boldsymbol{\gamma}$ represent disruptive and stabilizing selection, respectively. Expanding the right-hand side of the equation, we can see that change in **G** due to directional selection can be described as $-\mathbf{G}\beta\beta^{\mathrm{T}}\mathbf{G} = -\Delta\bar{z}\Delta\bar{z}^{T}$, indicating a reduction in genetic variance in the direction of expected evolutionary change in trait means. Non-linear effects are driven by stabilizing and disruptive selection, captured in $\mathbf{G}\boldsymbol{\gamma}\mathbf{G}$, which describes the forces respectively reducing and increasing genetic variation in the associated dimensions of trait space. Selection on epistatic effects results in alteration of the effects of alleles at all epistatically interacting loci. It is important to realize that genetic effects may be subject to all three types of selection simultaneously. For example, stabilizing selection creates epistasis, because the fitness of a variant depends on both the form of selection and the genetic background the variant occurs in. Variants with epistatic effects also necessarily have direct effects on the phenotype (Hansen & Wagner, 2001).

Selection may reshape variation through an adaptive mechanism when **G** is aligned with the shape of multivariate stabilizing selection (Waddington, 1957, Riedl, 1978, Olson & Miller, 1958), such that **G** and $-\boldsymbol{\gamma}^{-1}$ share common leading eigenvectors (Arnold et al., 2001). For example, Pavlicev et al. (2011) demonstrated the existence of a class of genetic variants that change the relationship between traits and hypothesize that these allow the pattern of pleiotropy to evolve so as to generate coordinated covariation in functionally related traits, and suppress covariance between unrelated traits. A simulation study demonstrates the plausibility of this process (Jones et al., 2014). Three studies have found evidence for alignment between **G** or **P** (phenotypic co/variance matrix) and stabilizing selection (Blows et al., 2004, Hohenlohe & Arnold, 2008, Punzalan & Rowe, 2016), supporting a role for nonlinear selection in maintaining patterns of multivariate genetic variation (but see Henry & Stinchcombe, 2023 for an example of

poor alignment between nonlinear selection and **G**). Therefore, one explanation for the stability of **G** is that the pattern of multivariate selection may be stable over macroevolutionary timescales.

A related version of the selection-shapes-variation hypothesis is that the long-term pattern of directional selection shapes **G**, rather than a fixed form of stabilizing selection. Periods of directional selection can favor epistatic variants that increase variability (Draghi & Wagner, 2008). Empirical evidence supports the possibility that episodic directional selection is common in nature. Studies of contemporary selection often find directional selection (Hereford et al., 2004, Morrissey, 2016), despite the expectation that the predominant pattern of selection should be stabilizing. However, theoretical work showed that fluctuating selection can generate long-term net stabilizing selection (Haller & Hendry, 2014, Tufto, 2015), as can antagonistic directional selection acting on genetically correlated traits (Keightley & Hill, 1990). Empirically, the pattern of changes in body size over timescales less than a million years suggests that the optimal body size fluctuates within broad limits, which could be explained by fluctuating selection (Uyeda et al., 2011). Contemporary selection studies suggest that differences in precipitation partly predict variation in directional selection over both space and time (Siepielski et al., 2017). Consequently, traits for which directional selection is driven by precipitation can be expected to regularly experience switches in the direction of selection, setting the stage for selection to shape variability that facilitates adaptation. Finally, studies of the genetic variation underlying multivariate trait combinations are consistent with stabilizing selection on genetic variation despite directional phenotypic selection (McGuigan & Blows, 2009, Sztepanacz & Rundle, 2012, Delcourt et al., 2012). This kind of adaptive reshaping of variability and variation could directly resolve the paradox of predictability, as the frequency and magnitude of changes in mean phenotype generate the selection on variability.

The other mechanism for a selection-shapes-variation hypothesis is the congruence scenario: direct selection for appropriate responses to environmental variation shapes **G**. The pathways that shape environmental and genetic variation are likely to be shared within an organism (Cheverud, 1988). Those pathways are always under selection to have an adaptive norm of reaction across environments. In some cases, the high fitness solution is to minimize responsiveness, favoring a flat norm of reaction; in other cases, a strong plastic response and a steep norm of reaction will be selected for. The shape of these environmental norms of reaction then dictates whether genetic variants have large or small effects on the trait. That selection on environmental responsiveness can reshape variability and variation is supported by theory (Wagner et al., 1997, Kaneko & Furusawa, 2006, Draghi & Whitlock, 2012). Recently, correlations between environmental variation and the pattern of macroevolution were documented in several studies (Noble et al., 2019, Rohner & Berger, 2023), suggesting the relevance of a congruence scenario to helping explain the paradox of predictability. To do so, shifts in optimal phenotypes over evolutionary time must correspond to the optimal pattern of environmental responsiveness, similar to the requirements for stabilizing selection on genetic variation to help explain the paradox.

The need for process-based macroevolutionary models of evolution in variance

Above, we identified several mechanisms that can shape the evolution of multivariate genetic variance. However, it remains difficult to reconcile how these mechanisms might jointly influence the evolution of **G** (and thus **P**) over longer timescales. A key conceptual challenge is

the lack of tractable, phenomenological process models for the evolution of genetic variation in deep time. Some studies have applied Brownian motion (BM) models to understand variance evolution across a phylogeny (Haber, 2011, McGlothlin et al., 2022). Although this approach may provide a pragmatic solution, it is conceptually fraught for two reasons: (i) since BM is derived from a model of evolution in trait means (Felsenstein, 1988, Lande, 1976), there is no guarantee that the same model is consistent with variance evolution; and, (ii) variances are themselves bounded but BM is an unbounded process. We may instead expect variances to evolve under an Ornstein-Uhlenbeck (OU) process (Hansen, 1997), where long-term stabilizing selection around a phenotypic optimum and pleiotropic mutation provides a restraining bound in the face of drift and recombination. However, we know little about how phenotypic optima may change through time, making it unclear how to parameterize such a model (Hansen, 2012, Rolland et al., 2023).

Similarly, the dynamics of selection over macroevolutionary timescales remain underexplored. There are two promising points of departure. The first is an exploration of ecological and demographic factors that lead to long-term stability in the pattern of selection. These include periodically changing abiotic environmental factors such as precipitation and temperature (Siepielski et al., 2017), as well as self-regulating biotic interactions that can preserve phenotypic variation through negative frequency-dependent selection and other stabilizing evolutionary forces (Chevin et al., 2022). Several recent macro-ecological models are also pushed in this direction (Rangel et al., 2018, Pontarp et al., 2019, Hagen et al., 2021). Climate emulators of past climate are used as spatiotemporal environmental templates on which macro-evolutionary diversification processes are modeled (Holden et al., 2019, Rangel et al., 2018). Nevertheless, the connection between environmental variability and selection dynamics (e.g., Siepielski et al., 2017) needs further exploration.

The second is a refinement of our understanding of the macroevolutionary analogue of the adaptive landscape (i.e. the covariance between phenotype and relative fitness; see Arnold et al., 2001). Currently, there are several heuristics that describe the macroevolutionary adaptive landscape, including adaptive zones (Simpson, 1944), primary optima (Hansen, 1997), species selection surfaces (Rolland et al., 2023), and dynamic fitness landscapes (Pontarp et al., 2024). Elucidating how these heuristics are conceptually and empirically related to the adaptive landscape of contemporary populations (Gupta et al., 2023, Beausoleil et al., 2023) would be an important step towards explaining the paradox of predictability.

Resolution of the paradox needs insights beyond quantitative genetics

The traditional expectation is that **G** should vary over long timescales, leaving the distribution of selection as the principal cause of the pattern of macroevolution (Simpson, 1944, Hansen & Houle, 2004, Estes & Arnold, 2007, Uyeda et al., 2011, Arnold, 2023). This expectation is now challenged by mounting evidence of a correlation between variation and macroevolutionary divergence. Based on considerations reviewed in this section, it is necessary to shift our attention from variation to variability, the propensity of mutation to supply phenotypic variation (Lande, 1976, Lynch & Hill, 1986, Lynch, 1990). Most plausible models for the maintenance of polygenic variation rest on mutation, including mutation-drift, mutation-selection, and, indirectly, migration-selection balance. The exceptions are balancing selection models, but this type of model would not expect a correlation between variation and evolutionary rate. Furthermore, there are relatively few polymorphisms known to be maintained by balancing

selection (Johnson & Barton, 2005). This implies that the key to the divergence-variance correlation is the underlying pattern of mutation, which is a function of the GP-map. We therefore hold that a resolution of the paradox of predictability requires not only an understanding of quantitative genetics but also the investigation of processes that shape genetic variation and the evolutionary forces acting over macroevolutionary timescales.

Comparative morphology and the paradox of predictability

Although constraint and selection-shapes-variation hypotheses have the potential to resolve the paradox fully or partially, we need data to empirically test predictions of these different mechanisms. Here we focus on comparative morphological work that offers testable functional, developmental, and ecological hypotheses for why variation remains stable over macroevolutionary timescales (Klingenberg, 2008, Klingenberg, 2014). To connect comparative morphological studies to the mechanisms and processes underlying the paradox, it is crucial to recognize that patterns of variance and covariance can be evaluated at different levels (Table 1). The first is *variability*—the potential or propensity to vary (Wagner & Altenberg, 1996). The second is *genetic variation*—the realized variation among individuals within a population. The third is *divergence*—the consequence of evolutionary changes among populations or species (Felsenstein, 1988). The final level is *environmental variation*, including plastic responses to the environment and unexplained variation around the trait mean. There is a clear causal relationship between the first three: variability is the source of variation, and variation directly affects divergence (i.e., responses to selection, see Equation 1), at least in the short-term. Environmental variation provides a tool to evaluate the congruence mechanism introduced earlier (e.g., Rohner & Berger, 2023). Below, we first summarize measurements at each level and then illustrate how close attention to mechanisms, causes, and consequences of variational properties at different levels will help researchers to make progress towards explaining the paradox of predictability.

Measurements

Variability

Variability can be measured as mutational inputs to phenotypic variance, which is summarized in the mutational co/variance matrix, **M** (Lande, 1980). **M** can be estimated in a mutation accumulation experiment (Halligan & Keightley, 2009), which requires special protocols and involves potentially enormous logistical costs. Consequently, **M** is difficult to estimate in most organisms and there are only a handful of available estimates (Pavličev et al., 2023). Recently, Rohner and Berger (2023) showed that non-genetic variation is correlated with variability, variation, and divergence in the single taxon they studied. To do this, they measured the fluctuating asymmetry (FA) covariance matrix (**F**), which summarizes deviations from perfect symmetry between left and right homologues of paired bilateral traits (Van Valen, 1962). **F** is probably caused by somatic-genetic and environmental perturbations during the developmental processes of growing phenotypes and may reflect variability of phenotypes due to developmental noise. Importantly, **F** is readily measured in any population of symmetrical organisms, including fossil taxa (Smith, 1998, Webster & Zelditch, 2011a, Webster & Zelditch, 2011b). Our understanding of the generality of **F** as a measure of variability and its biological significance is still in its infancy (Rohner & Berger, 2023, Saito et al., 2024), but if this relationship is general,

it could considerably improve our ability to study variability and elevate the congruence mechanism to a leading explanation for the paradox of predictability.

Genetic variation

The additive genetic covariance matrix, **G**, is the most widely used descriptor of variation in the literature of quantitative genetics. **G** is the part of the phenotypic covariance matrix, **P**, that is attributable to genetic kinship. Precise estimates of **G** require data from a large number of families with a known pedigree (or genetic relatedness), making robust estimates of **G** challenging even in laboratory populations (Steppan et al., 2002). In comparison, **P** is far easier to estimate with high precision because the unit of measurement is the individual and not the family. There is an ongoing debate about whether **P**, the sum of **G** and other sources of phenotypic variation, is a reasonable proxy for **G** (reviewed in Love et al., 2022). However, there are conceptual and empirical grounds to support the substitution of **G** by **P** for morphological traits (Kohn & Atchley, 1988, Roff, 1995, Porto et al., 2009, Sodini et al., 2018, but see Delahaie et al., 2017). A recent meta-analysis showed a tight and proportional relationship between genetic and phenotypic variance across a broad range of traits and taxa (Holstad et al., 2024). Typically, fossil record studies need to rely on **P** as a proxy because **G** is unavailable. However, broad-sense **G** (i.e., the sum of all genetic variances) can be estimated from fossilizable clonal organisms (Cheetham et al., 1994).

Integration—the degree of covariation among traits—and modularity—the degree to which sets of integrated traits covary less than other such sets (Houle & Pélabon, 2023)—affects the variational properties of a group of traits described by **G** or **P**. Using estimates of **G** or **P**, we can study how integration affects evolution through *conditional evolvability* (Table 1), which quantifies the ability of a trait or trait combination to respond to directional selection when other traits are under strong stabilizing selection (Hansen et al., 2003, Hansen & Houle, 2008). The most widely used measurement of conditional evolvability assumes stabilizing selection in all traits except for the focal trait (global conditional evolvability in Table 1). An alternative conceptualization is to identify traits that are under stabilizing selection based on ecological and functional evidence, then condition only with respect to those traits (Nilsson et al., 2024). From conditional evolvability, autonomy (*a*) can be estimated as the fraction of evolvability that remains after covariances are considered. If traits are entirely independent (i.e., no integration), then phenotypic space has as many dimensions as traits, but trait correlations can reduce the number of effective dimensionalities (McGuigan & Aguirre, 2016). Another metric of integration is the eigenvalue dispersion (Pavlicev et al., 2009, Wagner, 1984), which quantifies the sphericity of a matrix. A simplified version of this metric is the effective number of dimensions (Kirkpatrick, 2009) that evaluates the contribution of the leading eigenvector (i.e. $\boldsymbol{g}_{max}$) to the total variance (see Watanabe, 2022 for further details about eigenvalue dispersion indices).

Divergence

The most intuitive way to measure divergence is to evaluate the variance of mean phenotypes across populations or species, which in a multivariate case yields the divergence co/variance matrix (**D**). **D** is determined by a shared evolutionary history, adaptation to a shared ecological niche, adaptation to a lineage-specific ecological niche, and environmental variation. The

component due to shared evolutionary history can be estimated using relatedness among populations or species. Assuming a multivariate Brownian motion (BM) model of evolution, one can use the branch length of a phylogeny, either measured in unit of expected number of substitution or millions of years, to estimate the rate of evolution matrix, **R**, in an analogous manner as to when genetic relatedness information is used to estimate **G** (Lynch, 1991, Housworth et al., 2004).

When estimates of variance (**P** and **G**), variability (**M** and **F**), or divergence (**D** and **R**) are available from multiple populations, species, or groups of species, the divergence of covariance matrices can be evaluated. Among a variety of metrics and approaches for comparing covariance matrices (reviewed in Aguirre et al., 2014), geodesic distance (Mitteroecker & Bookstein, 2009) and response difference (Hansen & Houle, 2008) provide useful summary statistics that can be combined with variability and variance measures. The former characterizes the shortest paths between a pair of matrices, taking into consideration all aspects of the matrix (e.g., size, shape, and orientation) (Mitteroecker & Bookstein, 2009, Houle & Fierst, 2013); the latter characterizes the difference in response to the same selection gradient when two populations are assumed to evolve from the same multivariate mean phenotype (Milocco & Salazar-Ciudad, 2022).

Bridging variational levels in evolutionary analysis

Equipped with variational properties measured at different levels, we can study the relationships among them. Fortunately, most measurements (see Table 1) can be extracted from co/variance matrices estimated at different scales of biological organization (e.g., among individuals within a population, among populations, or among species). An intuitive approach to analysis is therefore to (i) estimate a variational property at one level, (ii) estimate the same (or other) property using the co/variance matrix at another level, and then (iii) evaluate their relationships using regressions. Although methodologically straightforward, this practice includes an important conceptual challenge. To illustrate this, consider two models of phenotypic evolution that each capture the constraint and selection-shapes-variation hypotheses. Lynch and Hill (1986) considered a case where the divergence (**D** and **R**) is governed predominantly by drift and mutation in the absence of selection to yield:

$$\mathbf{D} \propto 2\mathbf{M} \tag{2}$$

where **M** is the mutational co/variance matrix. Focusing on selection, Felsenstein (1988) and Zeng (1988) presented a model where divergence is determined by the product of multivariate selection, $\boldsymbol{\gamma}$, and the G-matrix.

$$\mathbf{D} \propto \mathbf{G}\boldsymbol{\gamma}\mathbf{G} \tag{3}$$

In real biological systems, divergence is almost certainly influenced by both drift and selection, suggesting that both constraints (Equation 2) and selection (Equation 3) shape the pattern of divergence. Their contributions can then be tested by relating **M**, **G** and $\boldsymbol{\gamma}$ with **D** (or **R**). Hypothetical scenarios in Equation (2) and (3) are illustrated in Figure 2a and 2b, considering **P** as an estimate of the underlying **M** and **G**. In these ideal cases, we will be able to separate the effect of constraints and selection by comparing the effect of **P** and $\boldsymbol{\gamma}$ on **D** using standard

statistical tests. However, selection is likely to change the variational properties themselves, as has been suggested both theoretically (Jones et al., 2004, Jones et al., 2014) and empirically (Blows et al., 2004, Hohenlohe & Arnold, 2008, Punzalan & Rowe, 2016). This results in a situation where the correlations among variational properties are consistent with both hypotheses (Figure 2c).

Attention to the distinctive causes, consequences, and concepts of variation at different levels will facilitate more nuanced and deeper understandings of the causes of the relationship—or lack thereof—between these properties at lower and higher levels of organization. For example, integration at the population level is due primarily to pleiotropy, where shared genes and developmental pathways create covariation patterns (Pavlicev & Cheverud, 2015). Measurements at this level can be used to evaluate evolvability (variance) or its underlying mechanistic causes (variability), and their evolutionary consequences, by plugging these estimates into the Lande-Arnold equation. The pattern of co/variances among species (e.g., Guillerme et al., 2023) in turn quantifies both the patterns of internal constraints and their consequences on evolutionary changes. At this level, several different factors and evolutionary processes are at play, including pleiotropy, correlational selection, mutational input, and genetic drift (Felsenstein, 1988, Zeng, 1988, Armbruster & Schwaegerle, 1996, Svensson et al., 2021). All these factors and processes may in part be influenced by higher-level evolutionary processes, such as species selection (Erwin, 2000, Jablonski, 2017b, De Lisle et al., 2021) or species drift (Chevin, 2016).

These considerations illustrate the complementary nature of the data acquired at different levels of organization and timescales. Although quantification of variational properties at the population level and below (i.e., variation and variability) is paramount for disentangling the relative contribution of intrinsic, adaptive, and congruent causes for observed patterns (Armbruster & Schwaegerle, 1996, Zelditch & Goswami, 2021, Love et al., 2022), covariation patterns at higher levels are necessary to understand how the possibilities offered by evolvability are explored and realized in biological systems (Hopkins et al., 2016). In other words, analyses bridging variational properties across multiple levels require careful attention to the connection between biological meaning and measurement—measurement theory (Houle et al., 2011, Voje et al., 2023b). For example, Machado et al. (2023) recently showed that a model of phenotypic macroevolution inspired by developmental processes substantially outperforms simpler models in explaining the pattern of macroevolution in mammalian tooth morphology. This highlights the direct benefit of mechanism- and process-oriented approaches to evolutionary analyses for advancing our understanding of the paradox of predictability.

Evo-devo and the paradox of predictability

Among the most promising approaches to solving the paradox of predictability is evolutionary developmental biology (evo-devo) because it has the potential to span levels of organization and timescales. Evolvability is a core area of evo-devo research (Hendrikse et al., 2007). However, the problem agenda of evolvability in evo-devo is radically different from that of evolvability in quantitative genetics. Quantitative genetics puts the generative mechanisms underlying evolvability in a black box and takes the observed pattern of variation and variability as the point of departure. Conversely, evo-devo seeks to open the black box, asking how and why differences in evolvability may arise. An evo-devo approach to the paradox of predictability facilitates exploring the causes of variability, and the potential for variability itself to evolve.

The genotype-phenotype map

The mutational, phenotypic, genetic, and environmental variance and covariance matrices (Box S1, Table 1) are all population-level summaries of phenotypic variation. The change and stability of these variation through evolution is therefore affected by the developmental processes that generate variation in phenotypes from variation in genetic and environmental factors (Hendrikse et al., 2007, Hallgrímsson et al., 2023). A useful conceptual framework to study these effects of development on the evolution of phenotypic variation is the genotype-phenotype map (GP-map), which describes how genotypes are translated into phenotypic variation through cellular properties, development, and physiology (Lewontin, 1974).

Studies of the GP-map, either in specific developmental systems (Green et al., 2017) or in abstract formal frameworks (Carter et al., 2005, Le Rouzic et al., 2013), have demonstrated that the GP-map can be highly complex and nonlinear for many classes of traits. These nonlinearities can have profound effects on the structure of co/variance matrices and how they evolve, particularly for populations under directional selection. For example, additive genetic variance can increase or decrease under the same directional selection regime depending on whether gene interactions are, on average, positive or negative (Carter et al., 2005). Milocco and Salazar-Ciudad (2022) showed that additive genetic variance can evolve in multiple ways depending on the curvature of the GP-map, differing from the expectations for a linear map (Jones et al., 2012). The existence of non-linear GP-maps in nature is supported by the effects of variants of major effect genes, such as *Bmp4* in Darwin's finches and cichlids (Abzhanov et al., 2004, Parsons & Albertson, 2009) or *EDA* in populations of sticklebacks (Barrett et al., 2008), where transient effects during the process of fixation would be expected to decouple variation and evolution.

Because variance can evolve differently depending on the curvature of the GP-map when a population is under directional selection, we can interpret these changes in variance as a *collateral* effect of selection on trait means in a population. As the mean of the population changes, the region of the GP-map that the population experiences also changes. These different regions may have different local properties that will result in different mappings of genetic to phenotypic variation (Milocco & Salazar-Ciudad, 2022). It has been argued that this mechanism, known as the directional epistasis, may be the primary force for the evolution of variance (Hansen & Wagner, 2023). A recent empirical study has suggested that the directional epistasis may be common in nature (Bourg et al., 2024).

The strong non-linearities that emerge in these model GP-maps deepen the paradox of predictability. If small changes in trait means lead to large changes in variability, we do not expect that the variation in any populations will be stable enough to predict divergence across macroevolutionary timescales. However, as we have emphasized, empirical evidence suggests that divergence-variance correlations are surprisingly common (Opedal et al., 2023, Voje et al., 2023a, Holstad et al., 2024, Box 1).

The apparent prevalence of divergence-variance correlations despite epistatic interactions and a non-linear GP-map might be explained in several ways. First, our observations may represent a biased sample of GP-maps. The divergence-variance relationship has so far been shown mainly in slowly evolving phenotypes, such as wings of Diptera, which may evolve in a linear region of the GP-map. Although divergence-variance correlations have also been observed in systems where substantial changes in trait means are documented (Bolstad et al., 2014, McGlothlin et al., 2018), it is possible that we have not yet sampled a sufficient variety of GP-

maps in nature. Identification of systems in which variation and evolutionary rate are decoupled (Tsuboi et al., 2018), or comparing variability and variation observed from empirical data with those derived from developmental models (Milocco & Salazar-Ciudad, 2020, Milocco & Salazar-Ciudad, 2022), would be informative to address this possibility. Second, nonlinearities in the GP-map may not refer to timescale *per se* but to the movement of population means in genotype and phenotype space. Genetic variance could evolve rapidly for populations under directional selection when the GP-map is complex and where the means are expected to diverge. In contrast, such rapid shifts in the mean and variance may be rare when populations experience stabilizing selection. This explanation returns us to the open question of what is a common form of long-term selection in nature (e.g., Hansen, 2012, Rolland et al., 2023).

Gene expression

Gene expression is a key element of the GP-map, influencing all other aspects of organismal form and function. Gene expression can be treated as a quantitative trait at the interface of genotype and phenotype. This makes gene expression traits useful models for the study of both genetic and environmental effects. Since expression can be measured at a genome-wide scale, it gives us the ability to study variability and variation comprehensively. Consequently, the variation in expression offers a powerful tool to draw inferences about the GP-map.

Studies of gene expression illuminate the generative mechanism of phenotypic variance. For instance, dose-response curves of gene expression may be a central component of GP-map nonlinearity. Experiments in mice show that the expression level of genes *Fgf8* and *Wnt9b*, which affect skull development, underlie phenotypic variation (Green et al., 2019, Green et al., 2017). Interestingly, in the case of *Fgf8*, non-linearity enhanced *robustness* (Green et al., 2017)– a mechanism that buffers the impact of environmental and genetic perturbations on phenotypic expression (reviewed in Signor & Nuzhdin, 2018). Robustness may provide a plausible molecular mechanism for why rates of phenotypic evolution can be orders of magnitude slower than expected from genetic and mutational variance (Lynch, 1990, Houle et al., 2017).

Additionally, *cis* (polymorphisms of the gene itself) or *trans* (external products targeting the expressed gene) regulatory elements may contribute to the long-term stability of phenotypic variation. These elements can interact in a compensatory fashion and further enhance robustness at the phenotypic level. These co-adapted compensatory mechanisms could maintain phenotypic variation within limited bounds, contributing to the stability of **G**. Empirical studies show that phenotypic variation rapidly changes when these mechanisms are disrupted through migration or hybridization (Mack et al., 2016, McGirr & Martin, 2019). Expression stability, plasticity, and mutational effect also are determined by the position of genes within regulatory networks (Uller et al., 2018) and their redundancy in such networks (Hallgrimsson et al., 2019). Moreover, gene expression variability is influenced by microRNAs (Siciliano et al., 2013), a class of abundant, small non-coding RNAs in many developmental processes, including morphogenesis (Giraldez et al., 2005, Horta-Lacueva et al., 2023). Documenting such molecular mechanisms and asking how these mechanisms translate into variability and variation at the level of functional phenotypes will be a fruitful future path for resolving the paradox of predictability.

Paleontology and the paradox of predictability

Paleontology has a rich history of studying evolvability (reviewed in Love et al., 2022). The fossil record stands as the sole direct source of temporal data from the history of life, making it a valuable resource for addressing unanswered questions related to the paradox of predictability. Some of the most striking places to observe the direct pay-offs of fossil data are in reconstructing ancestral character states (Ksepka et al., 2020), establishing a trend (Puttick, 2018), securing more realistic rate data (Liow et al., 2023), and helping to sample phenotypic spaces that are not represented by extant organisms (Mitchell, 2015). Although utilizing and interpreting variational properties of fossil data is challenging, research has shown that estimates of variational properties from fossil samples closely approximate standing variation in contemporary populations (Hunt, 2004, also see discussion on time-averaging in Box 1). Studies examining the extent to which within-population variation predicts divergence beyond microevolutionary timescales consistently provide evidence for a positive divergence-variance relationship (Renaud et al., 2006, Hunt, 2007, Brombacher et al., 2017, Voje et al., 2020, Holstad et al., 2024). Notably, all measurements summarized in Table 1 can be secured from paleontological datasets except for **M**, provided the phenotypic co/variance matrix is reasonably representative of the underlying **G** (Love et al., 2022). Overall, the distinct merits of paleontology for making robust evolutionary inferences hold great promise for studying the paradox of predictability.

Recently, Holstad et al. (2024) analyzed two large datasets of contemporary and fossil samples and found a strong relationship between evolvability and divergence in both datasets. Our analysis (Box 1) confirms the finding of Holstad et al. (2024) that the divergence-variance relationship holds in the fossil record, and suggests two questions that may illuminate the generative mechanisms behind the paradox of predictability: (1) Why do some datasets show stronger and tighter divergence-variance relationships than others? (2) Can we use additional information to increase the predictability of macroevolutionary divergence from standing variation within a population? For instance, the two species of planktonic foraminifera from Brombacher et al. (2017) showed varying strength of divergence-variance relationships (*Globoconella puncticulata*: slope ± SE = 0.39 ± 0.22, r^2 = 0.04, *Truncorotalia crassaformis*: slope ± SE = 0.62 ± 0.22, r^2 = 0.10, Table S2, Fig. S2). Interestingly, the species that underwent morphological shifts in three temporal regimes that are radically different in their abiotic environments (*G. puncticulata*) exhibited a weaker relationship, supporting the idea that the strength and scaling exponent of divergence-variance relationship depends partly on the selective regime. This provides circumstantial evidence favoring selection-shapes-variation hypotheses over constraint hypotheses. Thus, datasets that include shifts in the environment (Hunt et al., 2015, Brombacher et al., 2017, Antell et al., 2021), ecology (Liow et al., 2017), or other variables that lead to shifts in selective regimes (Di Martino & Liow, 2021) will be particularly valuable to disentangle and test the two major explanations of the paradox. Adding more data to our compilation through a more rigorous and systematic search of the literature and community databases (e.g., the Mammal Diversity Database, the Paleobiology Database, Neotoma Database, FuTRES Database) should illuminate these questions more thoroughly.

The paradox of predictability encourages neontological researchers to utilize fossil record data. Some neontological biologists may be working on a group of organisms that has fossil data but are unaware of its potential to inform their evolutionary research. New researchers just starting out and deciding which system to adopt might make a different decision if they were aware of available fossil data. Established scientists may not appreciate how fossil data can

improve or enrich evolutionary inferences in their system or might be unsure how to use it. In all three cases, attention to the distinct and complementary roles of quantitative genetics (genetic data and theory development), comparative morphology (measurements and empirical tests), and evo-devo (verifications and extensions of key assumptions) clarify how paleontological data and approaches can be integrated fruitfully into neontological research. Considering the irreplaceable role of fossil record as the source of longitudinal observation of evolution over geological timescales, we argue that the ultimate test for all hypotheses proposed in this paper will require some form of paleontological data.

Conclusions and Outlook

The prevalence of positive correlations between standing variation in contemporary populations and macroevolutionary rates and direction of phenotypic divergence is paradoxical. However, it also offers novel prospects for bridging evolutionary timescales. Historically, the relationship between microevolution and macroevolution has been posed primarily as a substantive, single question with two options: (i) Macroevolution is the summation of microevolutionary processes (Charlesworth et al., 1982, Arnold, 2014), or (ii) Macroevolution involves distinct processes from microevolution (Erwin, 2000, Jablonski, 2017b). More than a decade ago, Uyeda et al. (2011) argued that "We lack a comprehensive understanding of evolutionary pattern and process because short-term and long-term data have rarely been combined into a single analytical framework" (p. 15908). Our discussion of the paradox of predictability points to a common analytical framework that could bridge timescales in evolutionary analysis. Instead of a single question to be answered, we are challenged to identify short-term and long-term data that can be related to each other. Building such relationships is an open-ended research program that can be pursued using multiple analytical frameworks. Our paper is intended as an illustration of how this might be done in the context of evolvability research framework.

Although the four areas of evolvability-oriented research (quantitative genetics, paleontology, comparative morphology and evo-devo) highlight the complementary nature of data, theories, methods, and empirical insights that can be acquired from each approach, implementation of cross-disciplinary research is challenging (Liow et al., 2023). To incentivize and facilitate future collaborations, the distinct roles and major limitations of each approach must be recognized.

Quantitative genetics offers a theoretical definition and clarifies what counts as evolvability to make evolutionary inferences from two estimable aspects of a population (**G** and a selection gradient). The application of this approach is limited by two assumptions. First, the distribution of the phenotypes should be continuous. Second, to compare quantitative-genetic parameters across populations and species, the traits should be amenable to mean-standardization (Hansen & Houle, 2008). These assumptions make it difficult for quantitative genetics to study discrete traits such as color polymorphism and alternative life histories (e.g., mating systems) or traits measured on scales that cannot be standardized in a way that makes them comparable among populations or species, such as ratios, count, or Julian date (i.e. breeding date in birds, Hadfield & Reed, 2022). Currently, the scope of the paradox of predictability is limited to the realm wherein these specific assumptions are fulfilled.

Comparative morphology provides empirical data to test theoretical predictions and conceptual ideas. Identification and evaluation of ecological, functional, and developmental hypotheses explaining the observed patterns of variance and covariance is the primary role of

this area (Klingenberg, 2008, Melo et al., 2016, Zelditch & Goswami, 2021). The ever-growing availability of large molecular phylogenies and open-source phenotypic databases (e.g., Munoz & Price, 2019) suggest that phylogenetic comparative studies of morphology represent a promising path to assessing the generality of the paradox. However, caution is warranted because the relationship between measurements and their biological meaning is frequently overlooked (Houle et al., 2011, Voje et al., 2023b). Our advocacy for a methodological orientation to bridging timescales in evolutionary analysis foregrounds these concerns. Attention to the distinctive causes, consequences, and concepts in measurements of variance and covariance patterns at different levels of biological organization, based in sound measurement theory, will be critical to facilitating a richer integration of comparative morphology with other research approaches (Machado et al., 2023).

To that end, evo-devo contributes causal explanations of variation that both address the paradox and deepen it. On the one hand, a GP-map motivated by empirical knowledge of developmental processes (Milocco & Salazar-Ciudad, 2020) and the identification of molecular mechanisms that produce phenotypic variation (Green et al., 2017) will inform us about when the evolvability framework from quantitative genetics does (and does not) work. Given the generality of the divergence-variance relationship in quantitative and polygenic traits, understanding of the developmental and molecular mechanisms that characterize the GP-map of this class of traits will be necessary to explain the paradox. On the other hand, the prevalence of non-linearity in model GP-maps, potentially causing rapid change in variational properties (Milocco & Salazar-Ciudad, 2022, Hansen & Wagner, 2023), and genes with major effects (Abzhanov et al., 2004, Parsons & Albertson, 2009) contradict the apparent generality of the divergence-variance relationship. This likely reflects the focus of current evo-devo research, which preferentially studies characters that are unlikely to satisfy the infinitesimal assumption of quantitative genetics. One possibility for facilitating a better integration of frameworks from evo-devo and quantitative genetics is for both fields to pay closer attention to genetic architecture that is intermediate between infinitesimal and single-locus models, such as “omnigenic” models (Boyle et al., 2017).

Paleontology provides temporal data on millennial or longer timescales and offers the ultimate empirical test for the paradox of predictability. It significantly broadens the temporal scope of the paradox and provides realistic estimates of divergence. Fossil time-series provide longitudinal data of phenotypic change beyond contemporary timescales (Box 1). In addition, the fossil record is the only source of data on intermediate temporal spans that are greater than 10^2 years but less than 10^5 years. Although data from this time span is currently rare (Fig. 1), targeting fossil systems that allow for fine-grained temporal resolution (Liow et al., 2023, Harmon et al., 2021) may address this deficit. The abundance of paleontological datasets with these resolutions (https://www.neotomadb.org) suggests rich yet underutilized resources for morphological studies. Applicability of the evolvability framework in the fossil record must grapple with a variety of issues, such as the sparseness and incompleteness of fossil samples. There are ways to address some of these issues through strategic collaborations (Love et al., 2022) and new tools. Notably, advances in the automated extraction of information from digital images (Lürig et al., 2021) hold great promise for improving the quality of fossil record data (Brombacher et al., 2017, Di Martino & Liow, 2021, Yu et al., 2024).

The paradox of predictability offers numerous promising avenues for linking micro- and macroevolution. However, our excitement should be tempered because this predictability is difficult to explain using existing evolutionary models of phenotypic evolution—prediction is

not explanation. The demonstration of evolutionary predictability without an investigation to understand its explanatory causes and mechanisms is "little more than alchemy" (Endler, 1986, p. 164). Yet there are a variety of questions that can be pursued that will help in resolving the paradox (Table 2). Further progress is contingent on interdisciplinary collaboration that integrates multiple research approaches, the deployment of new conceptual and methodological tools, and tests of explicit and explanatory hypotheses. Given that this progress is already on display (Voje et al., 2023a, Opedal et al., 2023, Rohner & Berger, 2023, Liow et al., 2023, Saito et al., 2024, Holstad et al., 2024), we anticipate that efforts to resolve the paradox of predictability could bridge micro- and macroevolution.

Conflict of interest

We have no conflict of interest to declare.

References

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R. & Tabin, C. J. 2004. Bmp4 and morphological variation of beaks in Darwin's finches. *Science* **305**: 1462-1465.

Aguirre, J. D., Hine, E., McGuigan, K. & Blows, M. W. 2014. Comparing G: multivariate analysis of genetic variation in multiple populations. *Heredity* **112**: 21-29.

Antell, G. S., Fenton, I. S., Valdes, P. J. & Saupe, E. E. 2021. Thermal niches of planktonic foraminifera are static throughout glacial?interglacial climate change. *Proceedings of the National Academy of Sciences of the United States of America* **118**.

Armbruster, W. S. & Schwaegerle, K. E. 1996. Causes of covariation of phenotypic traits among populations. *Journal of Evolutionary Biology* **9**: 261-276.

Arnold, S. J. 2014. Phenotypic Evolution: The Ongoing Synthesis. *American Naturalist* **183**: 729-746.

Arnold, S. J. 2023. *Evolutionary Quantitative Genetics*. Oxford University Press, Oxford, UK.

Arnold, S. J., Pfrender, M. E. & Jones, A. G. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* **112**: 9-32.

Barrett, R. D. H., Rogers, S. M. & Schluter, D. 2008. Natural selection on a major armor gene in threespine stickleback. *Science* **322**: 255-257.

Barton, N. H. 2022. The "new synthesis". *Proceedings of the National Academy of Sciences* **119**: e2122147119.

Barton, N. H., Etheridge, A. M. & Veber, A. 2017. The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* **118**: 50-73.

Beausoleil, M. O., Carrión, P. L., Podos, J., Camacho, C., Rabadán-González, J., Richard, R., Lalla, K., Raeymaekers, J. A. M., Knutie, S. A., De León, L. F., Chaves, J. A., Clayton, D. H., Koop, J. A. H., Sharpe, D. M. T., Gotanda, K. M., Huber, S. K., Barrett, R. D. H. & Hendry, A. P. 2023. The fitness landscape of a community of Darwin's finches. *Evolution* **77**: 2533-2546.

Bell, M. A., Sadagursky, M. S. & Baumgartner, J. V. 1987. Utility of lacustrine deposits for the study of variation within fossil samples. *Palaios* **2**: 455-466.

Blount, Z. D., Lenski, R. E. & Losos, J. B. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**: eaam5979.

Blows, M. W., Chenoweth, S. F. & Hine, E. 2004. Orientation of the genetic variance-covariance matrix and the fitness surface for multiple male sexually selected traits. *American Naturalist* **163**: 329-340.

Bolstad, G. H., Hansen, T. F., Pélabon, C., Falahati-Anbaran, M., Pérez-Barrales, R. & Armbruster, W. S. 2014. Genetic constraints predict evolutionary divergence in Dalechampia blossoms. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **369**: 20130255.

Bourg, S., Bolstad, G. H., Griffin, D., Pélabon, C. & Hansen, T. F. 2024. Directional epistasis is common in morphological divergence. *Evolution* **78**: 934-950.

Boyle, E. A., Li, Y. I. & Pritchard, J. K. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**: 1177-1186.

Brombacher, A., Wilson, P. A., Bailey, I. & Ezard, T. H. G. 2017. The Breakdown of Static and Evolutionary Allometries during Climatic Upheaval. *American Naturalist* **190**: 350-362.

Carter, A. J. R., Hermisson, J. & Hansen, T. F. 2005. The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theoretical Population Biology* **68**: 179-196.
Charlesworth, B., Lande, R. & Slatkin, M. 1982. A Neo-Darwinian Commentary on Macroevolution. *Evolution* **36**: 474-498.
Cheetham, A. H., Jackson, J. B. C. & Hayek, L. A. C. 1994. Quantitative Genetics of Bryozoan Phenotypic Evolution .2. Analysis of Selection and Random Change in Fossil Species Using Reconstructed Genetic-Parameters. *Evolution* **48**: 360-375.
Cheverud, J. M. 1984. Quantitative Genetics and Developmental Constraints on Evolution by Selection. *Journal of Theoretical Biology* **110**: 155-171.
Cheverud, J. M. 1988. A Comparison of Genetic and Phenotypic Correlations. *Evolution* **42**: 958-968.
Chevin, L. M. 2016. Species selection and random drift in macroevolution. *Evolution* **70**: 513-525.
Chevin, L. M., Gompert, Z. & Nosil, P. 2022. Frequency dependence and the predictability of evolution in a changing environment. *Evolution Letters* **6**: 21-33.
Chiarenza, A. A., Farnsworth, A., Mannion, P. D., Lunt, D. J., Valdes, P. J., Morgan, J. V. & Allison, P. A. 2020. Asteroid impact, not volcanism, caused the end-Cretaceous dinosaur extinction. *Proceedings of the National Academy of Sciences* **117**: 17084-17093.
Colegrave, N. & Collins, S. 2008. Experimental evolution: experimental evolution and evolvability. *Heredity* **100**: 464-470.
Cuvier, G. 1828. *Le règne animal distribué d'après son organisation : pour servir de base a l'histoire naturelle des animaux et d'introduction a l'anatomie comparée*. Chez Déterville, A Paris.
De Lisle, S. P., Punzalan, D., Rollinson, N. & Rowe, L. 2021. Extinction and the temporal distribution of macroevolutionary bursts. *Journal of Evolutionary Biology* **34**: 380-390.
de Visser, J. A. G. M., Hermisson, J., Wagner, G. P., Meyers, L. A., Bagheri, H. C., Blanchard, J. L., Chao, L., Cheverud, J. M., Elena, S. F., Fontana, W., Gibson, G., Hansen, T. F., Krakauer, D., Lewontin, R. C., Ofria, C., Rice, S. H., von Dassow, G., Wagner, A. & Whitlock, M. C. 2003. Perspective: Evolution and detection of genetic robustness. *Evolution* **57**: 1959-1972.
Delahaie, B., Charmantier, A., Chantepie, S., Garant, D., Porlier, M. & Teplitsky, C. 2017. Conserved G-matrices of morphological and life-history traits among continental and island blue tit populations. *Heredity* **119**: 76-87.
Delcourt, M., Blows, M. W., Aguirre, J. D. & Rundle, H. D. 2012. Evolutionary optimum for male sexual traits characterized using the multivariate Robertson-Price Identity. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 10414-10419.
Di Martino, E. & Liow, L. H. 2021. Trait–fitness associations do not predict within-species phenotypic evolution over 2 million years. *Proceedings of the Royal Society B* **288**: 20202047.
Draghi, J. & Wagner, G. R. 2008. Evolution of evolvability in a developmental model. *Evolution* **62**: 301-315.
Draghi, J. A. & Whitlock, M. C. 2012. Phenotypic Plasticity Facilitates Mutational Variance, Genetic Variance, and Evolvability Along the Major Axis of Environmental Variation. *Evolution* **66**: 2891-2902.

Dugand, R. J., Aguirre, J. D., Hine, E., Blows, M. W. & McGuigan, K. 2021. The contribution of mutation and selection to multivariate quantitative genetic variance in an outbred population of Drosophila serrata. *Proceedings of the National Academy of Sciences of the United States of America* **118**: e2026217118.

Endler, J. A. 1986. *Natural selection in the wild*. Princeton University Press.

Erwin, D. H. 2000. Macroevolution is more than repeated rounds of microevolution. *Evolution & Development* **2**: 78-84.

Estes, S. & Arnold, S. J. 2007. Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence on all timescales. *American Naturalist* **169**: 227-244.

Felsenstein, J. 1988. Phylogenies and Quantitative Characters. *Annual Review of Ecology and Systematics* **19**: 445-471.

Futuyma, D. J. 2010. Evolutionary Constraint and Ecological Consequences. *Evolution* **64**: 1865-1884.

Gingerich, P. D. 1983. Rates of Evolution - Effects of Time and Temporal Scaling. *Science* **222**: 159-161.

Giraldez, A. J., Cinalli, R. M., Glasner, M. E., Enright, A. J., Thomson, J. M., Baskerville, S., Hammond, S. M., Bartel, D. P. & Schier, A. F. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**: 833-838.

Gradstein, F. M., Ogg, J. G., Schmitz, M. D. & Ogg, G. M. 2020. *Geologic time scale 2020*. Elsevier.

Grant, P. R. & Grant, B. R. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**: 707-711.

Green, R. M., Fish, J. L., Young, N. M., Smith, F. J., Roberts, B., Dolan, K., Choi, I., Leach, C. L., Gordon, P., Cheverud, J. M., Roseman, C. C., Williams, T. J., Marcucio, R. S. & Hallgrimsson, B. 2017. Developmental nonlinearity drives phenotypic robustness. *Nature Communications* **8**: 1970.

Green, R. M., Leach, C. L., Diewert, V. M., Aponte, J. D., Schmidt, E. J., Cheverud, J. M., Roseman, C. C., Young, N. M., Marcucio, R. S. & Hallgrimsson, B. 2019. Nonlinear gene expression-phenotype relationships contribute to variation and clefting in the A/WySn mouse. *Developmental Dynamics* **248**: 1232-1242.

Guillerme, T., Bright, J. A., Cooney, C. R., Hughes, E. C., Varley, Z. K., Cooper, N., Beckerman, A. P. & Thomas, G. H. 2023. Innovation and elaboration on the avian tree of life. *Science Advances* **9**: eadg1641.

Gupta, A., Svensson, E. I., Frietsch, H. & Tsuboi, M. 2023. Stasis in the adaptive landscape of two ecologically similar and sympatric damselfly species. *bioRxiv*: 2022.10.19.512907.

Haber, A. 2011. A comparative analysis of integration indices. *Evolutionary Biology* **38**: 476-488.

Hadfield, J. D. & Reed, T. E. 2022. Directional selection and the evolution of breeding date in birds, revisited: Hard selection and the evolution of plasticity. *Evolution Letters* **6**: 178-188.

Hagen, O., Flück, B., Fopp, F., Cabral, J. S., Hartig, F., Pontarp, M., Rangel, T. F. & Pellissier, L. 2021. gen3sis: A general engine for eco-evolutionary simulations of the processes that shape Earth's biodiversity. *Plos Biology* **19**.

Haller, B. C. & Hendry, A. P. 2014. Solving the Paradox of Stasis: Squashed Stabilizing Selection and the Limits of Detection. *Evolution* **68**: 483-500.

Hallgrímsson, B., Aponte, J. D., Vidal-Garcia, M., Richbourg, H., Green, R., Young, N. M., Cheverud, J. M., Calof, A. L., Lander, A. D. & Marcucio, R. S. (2023) The Developmental Basis for Evolvability. In: *Evolvability: A Unifying Concept in Evolutionary Biology?*, (Hansen, T. F., Houle, D., Pavličev, M. & Pélabon, C., eds.). pp. 171-198. The MIT Press.

Hallgrimsson, B., Green, R. M., Katz, D. C., Fish, J. L., Bernier, F. P., Roseman, C. C., Young, N. M., Cheverud, J. M. & Marcucio, R. S. 2019. The developmental-genetics of canalization. *Seminars in Cell & Developmental Biology* **88**: 67-79.

Halligan, D. L. & Keightley, P. D. 2009. Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. *Annual Review of Ecology Evolution and Systematics* **40**: 151-172.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**: 1341-1351.

Hansen, T. F. 2006. The evolution of genetic architecture. *Annual Review of Ecology Evolution and Systematics* **37**: 123-157.

Hansen, T. F. (2012) Adaptive landscapes and macroevolutionary dynamics. In: *The adaptive landscape in evolutionary biology*, (Svensson, E. I. & Calsbeek, R., eds.). pp. 205-226. Oxford University Press, Oxford.

Hansen, T. F. (2023) Variation, Inheritance, and Evolution: A Primer on Evolutionary Quantitative Genetics. In: *Evolvability: A Unifying Concept in Evolutionary Biology?*, (Hansen, T. F., Houle, D., Pavličev, M. & Pélabon, C., eds.). pp. 73-100. The MIT Press.

Hansen, T. F., Alvarez-Castro, J. M., Carter, A. J. R., Hermisson, J. & Wagner, G. P. 2006. Evolution of genetic architecture under directional selection. *Evolution* **60**: 1523-1536.

Hansen, T. F., Carlson, M. L. & PÉlabon, C. 2003. Evolvability and genetic constraint in Dalechampia blossoms: genetic correlations and conditional evolvability. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **296**: 23-39.

Hansen, T. F. & Houle, D. 2004. Evolvability, stabilizing selection, and the problem of stasis. *Phenotypic integration: studying the ecology and evolution of complex phenotypes. Oxford University Press, Oxford*: 130-150.

Hansen, T. F. & Houle, D. 2008. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* **21**: 1201-1219.

Hansen, T. F. & Wagner, G. P. 2001. Modeling genetic architecture: A multilinear theory of gene interaction. *Theoretical Population Biology* **59**: 61-86.

Hansen, T. F. & Wagner, G. P. (2023) The Evolution of Evolvability. In: *Evolvability: A Unifying Concept in Evolutionary Biology?*, (Hansen, T. F., Houle, D., Pavličev, M. & Pélabon, C., eds.). pp. 121-145. The MIT Press.

Harmon, L. J., Pennell, M. W., Henao-Diaz, L. F., Rolland, J., Sipley, B. N. & Uyeda, J. C. 2021. Causes and Consequences of Apparent Timescaling Across All Estimated Evolutionary Rates. *Annual Review of Ecology, Evolution, and Systematics, Vol 52, 2021* **52**: 587-609.

Hendrikse, J. L., Parsons, T. E. & Hallgrimsson, B. 2007. Evolvability as the proper focus of evolutionary developmental biology. *Evolution & Development* **9**: 393-401.

Henry, G. A. & Stinchcombe, J. R. 2023. Strong selection is poorly aligned with genetic variation in Ipomoea hederacea: implications for divergence and constraint. *Evolution* **77**: 1712-1719.

Hereford, J., Hansen, T. F. & Houle, D. 2004. Comparing strengths of directional selection: How strong is strong? *Evolution* **58**: 2133-2143.
Hermisson, J., Hansen, T. F. & Wagner, G. P. 2003. Epistasis in polygenic traits and the evolution of genetic architecture under stabilizing selection. *American Naturalist* **161**: 708-734.
Hill, W. G. & Caballero, A. 1992. Artificial Selection Experiments. *Annual Review of Ecology and Systematics* **23**: 287-310.
Hill, W. G. & Kirkpatrick, M. 2010. What Animal Breeding Has Taught Us about Evolution. *Annual Review of Ecology, Evolution, and Systematics* **41**: 1-19.
Hine, E., McGuigan, K. & Blows, M. W. 2014. Evolutionary Constraints in High-Dimensional Trait Sets. *American Naturalist* **184**: 119-131.
Hine, E., Runcie, D. E., McGuigan, K. & Blows, M. W. 2018. Uneven Distribution of Mutational Variance Across the Transcriptome of Drosophila serrata Revealed by High-Dimensional Analysis of Gene Expression. *Genetics* **209**: 1319-1328.
Hohenlohe, P. A. & Arnold, S. J. 2008. MIPoD: A hypothesis-testing framework for microevolutionary inference from patterns of divergence. *American Naturalist* **171**: 366-385.
Holden, P. B., Edwards, N. R., Rangel, T. F., Pereira, E. B., Tran, G. T. & Wilkinson, R. D. 2019. PALEO-PGEM v1.0: a statistical emulator of Pliocene-Pleistocene climate. *Geoscientific Model Development* **12**: 5137-5155.
Holstad, A., Voje, K. L., Opedal, Ø. H., Bolstad, G. H., Bourg, S., Hansen, T. F. & Pélabon, C. 2024. Evolvability predicts macroevolution under fluctuating selection. *Science* **384**: 688-693.
Hopkins, M. J., Haber, A. & Thurman, C. L. 2016. Constraints on geographic variation in fiddler crabs (Ocypodidae: Uca) from the western Atlantic. *Journal of Evolutionary Biology* **29**: 1553-1568.
Horta-Lacueva, Q. J. B., Jonsson, Z. O., Thorholludottir, D. A. V., Hallgrimsson, B. & Kapralova, K. H. 2023. Rapid and biased evolution of canalization during adaptive divergence revealed by dominance in gene expression variability during Arctic charr early development. *Communications Biology* **6**: 897.
Houle, D., Bolstad, G. H., van der Linde, K. & Hansen, T. F. 2017. Mutation predicts 40 million years of fly wing evolution. *Nature* **548**: 447-450.
Houle, D. & Fierst, J. 2013. Properties of spontaneous mutational variance and covariance for wing size and shape in Drosophila melanogaster. *Evolution* **67**: 1116-1130.
Houle, D. & Pélabon, C. (2023) Measuring Evolvability. In: *Evolvability: A Unifying Concept in Evolutionary Biology?*, (Hansen, T. F., Houle, D., Pavličev, M. & Pélabon, C., eds.). pp. 101-119. The MIT Press.
Houle, D., Pelabon, C., Wagner, G. P. & Hansen, T. F. 2011. Measurement and Meaning in Biology. *The Quarterly Review of Biology* **86**: 3-34.
Houle, D. & Rossoni, D. M. 2022. Complexity, Evolvability, and the Process of Adaptation. *Annual Review of Ecology, Evolution, and Systematics* **53**: 137-159.
Housworth, E. A., Martins, E. P. & Lynch, M. 2004. The phylogenetic mixed model. *American Naturalist* **163**: 84-96.
Hunt, G. 2004. Phenotypic variance inflation in fossil samples: an empirical assessment. *Paleobiology* **30**: 487-506.

Hunt, G. 2007. Evolutionary divergence in directions of high phenotypic variance in the ostracode genus Poseidonamicus. *Evolution* **61**: 1560-1576.
Hunt, G., Hopkins, M. J. & Lidgard, S. 2015. Simple versus complex models of trait evolution and stasis as a response to environmental change. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 4885-4890.
Jablonski, D. 2017a. Approaches to Macroevolution: 1. General Concepts and Origin of Variation. *Evolutionary Biology* **44**: 427-450.
Jablonski, D. 2017b. Approaches to Macroevolution: 2. Sorting of Variation, Some Overarching Issues, and General Conclusions. *Evolutionary Biology* **44**: 451-475.
Jablonski, D. (2023) Evolvability and Macroevolution. In: *Evolvability: A Unifying Concept in Evolutionary Biology?*, (Hansen, T. F., Houle, D., Pavličev, M. & Pélabon, C., eds.). pp. 351-372. The MIT Press.
Johnson, T. & Barton, N. 2005. Theoretical models of selection and mutation on quantitative traits. *Philosophical Transactions of the Royal Society B-Biological Sciences* **360**: 1411-1425.
Jones, A. G., Arnold, S. J. & Burger, R. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution* **58**: 1639-1654.
Jones, A. G., Arnold, S. J. & Burger, R. 2007. The mutation matrix and the evolution of evolvability. *Evolution* **61**: 727-745.
Jones, A. G., Burger, R. & Arnold, S. J. 2014. Epistasis and natural selection shape the mutational architecture of complex traits. *Nature Communications* **5**: 3709.
Jones, A. G., Burger, R., Arnold, S. J., Hohenlohe, P. A. & Uyeda, J. C. 2012. The effects of stochastic and episodic movement of the optimum on the evolution of the G-matrix and the response of the trait mean to selection. *Journal of Evolutionary Biology* **25**: 2210-2231.
Kaneko, K. & Furusawa, C. 2006. An evolutionary relationship between genetic variation and phenotypic fluctuation. *Journal of Theoretical Biology* **240**: 78-86.
Kaneko, K. & Furusawa, C. 2018. Macroscopic Theory for Evolving Biological Systems Akin to Thermodynamics. *Annual Review of Biophysics* **47**: 273-290.
Keightley, P. D. & Hill, W. G. 1990. Variation Maintained in Quantitative Traits with Mutation Selection Balance - Pleiotropic Side-Effects on Fitness Traits. *Proceedings of the Royal Society B-Biological Sciences* **242**: 95-100.
Kidwell, S. M. & Holland, S. M. 2002. The quality of the fossil record: Implications for evolutionary analyses. *Annual Review of Ecology and Systematics* **33**: 561-588.
Kirkpatrick, M. 2009. Patterns of quantitative genetic variation in multiple dimensions. *Genetica* **136**: 271-284.
Klingenberg, C. P. 2008. Morphological Integration and Developmental Modularity. *Annual Review of Ecology Evolution and Systematics* **39**: 115-132.
Klingenberg, C. P. 2014. Studying morphological integration and modularity at multiple levels: concepts and analysis. *Philosophical Transactions of the Royal Society B-Biological Sciences* **369**: 20130249.
Kohn, L. A. P. & Atchley, W. R. 1988. How Similar Are Genetic Correlation Structures - Data from Mice and Rats. *Evolution* **42**: 467-481.
Kowalewski, M., Casebolt, S., Hua, Q., Whitacre, K. E., Kaufman, D. S. & Kosnik, M. A. 2018. One fossil record, multiple time resolutions: Disparate time-averaging of echinoids and mollusks on a Holocene carbonate platform. *Geology* **46**: 51-54.

Ksepka, D. T., Balanoff, A. M., Smith, N. A., Bever, G. S., Bhullar, B.-A. S., Bourdon, E., Braun, E. L., Burleigh, J. G., Clarke, J. A. & Colbert, M. W. 2020. Tempo and Pattern of Avian Brain Size Evolution. *Current Biology* **30**: 2026-2036.

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. 1980. The Genetic Covariance between Characters Maintained by Pleiotropic Mutations. *Genetics* **94**: 203-215.

Le Rouzic, A., Alvarez-Castro, J. M. & Hansen, T. F. 2013. The Evolution of Canalization and Evolvability in Stable and Fluctuating Environments. *Evolutionary Biology* **40**: 317-340.

Lewontin, R. C. 1974. *The genetic basis of evolutionary change*. Columbia University Press New York.

Liow, L. H., Di Martino, E., Krzeminska, M., Ramsfjell, M., Rust, S., Taylor, P. D. & Voje, K. L. 2017. Relative size predicts competitive outcome through 2 million years. *Ecology Letters* **20**: 981-988.

Liow, L. H., Uyeda, J. & Hunt, G. 2023. Cross-disciplinary information for understanding macroevolution. *Trends in Ecology & Evolution* **38**: 250-260.

Love, A., Grabowski, M., Houle, D., Liow, L. H., Porto, A., Tsuboi, M., Voje, K. L. & Hunt, G. 2022. Evolvability in the Fossil Record. *Paleobiology* **48**: 186-209.

Lürig, M., Donoughe, S., Svensson, E. I., Porto, A. & Tsuboi, M. 2021. Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Frontiers in Ecology and Evolution* **in press**.

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

Lynch, M. 1991. Methods for the Analysis of Comparative Data in Evolutionary Biology. *Evolution* **45**: 1065-1080.

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

MacFadden, B. (1989) Dental character variation in paleopopulations and morphospecies fossil horses and extant analogs. In: *The evolution of perissodactyls*, (D. R. Prothero & Schoch, R. M., eds.). pp. 128-141. Oxford University Press, New York.

Machado, F. A., Mongle, C. S., Slater, G., Penna, A., Wisniewski, A., Soffin, A., Dutra, V. & Uyeda, J. C. 2023. Rules of teeth development align microevolution with macroevolution in extant and extinct primates. *Nature Ecology & Evolution* **7**: 1729–1739.

Mack, K. L., Campbell, P. & Nachman, M. W. 2016. Gene regulation and speciation in house mice. *Genome Research* **26**: 451-461.

Marroig, G. & Cheverud, J. M. 2005. Size as a line of least evolutionary resistance: Diet and adaptive morphological radiation in new world monkeys. *Evolution* **59**: 1128-1142.

Mas, A., Lagadeuc, Y. & Vandenkoornhuyse, P. 2020. Reflections on the Predictability of Evolution: Toward a Conceptual Framework. *Iscience* **23**: 101736.

McGirr, J. A. & Martin, C. H. 2019. Hybrid gene misregulation in multiple developing tissues within a recent adaptive radiation of Cyprinodon pupfishes. *Plos One* **14**: e0218899.

McGlothlin, J. W., Kobiela, M. E., Wright, H. V., Kolbe, J. J., Losos, J. B. & Brodie, E. 2022. Conservation and Convergence of Genetic Architecture in the Adaptive Radiation of Anolis Lizards. *American Naturalist* **200**: E207-E220.

McGlothlin, J. W., Kobiela, M. E., Wright, H. V., Mahler, D. L., Kolbe, J. J., Losos, J. B. & Brodie, E. D. 2018. Adaptive radiation along a deeply conserved genetic line of least resistance in Anolis lizards. *Evolution Letters* **2**: 310-322.
McGuigan, K. & Aguirre, J. 2016. Quantitative genetic variation, comparing patterns of. *Encyclopedia of Evolutionary Biology*: 377-387.
McGuigan, K. & Blows, M. W. 2009. Asymmetry of genetic variation in fitness-related traits: Apparent stabilizing selection on gmax. *Evolution* **63**: 2838-2847.
Melo, D., Porto, A., Cheverud, J. M. & Marroig, G. 2016. Modularity: Genes, Development, and Evolution. *Annual Review of Ecology, Evolution, and Systematics* **47**: 463-486.
Milocco, L. & Salazar-Ciudad, I. 2020. Is evolution predictable? Quantitative genetics under complex genotype-phenotype maps. *Evolution* **74**: 230-244.
Milocco, L. & Salazar-Ciudad, I. 2022. Evolution of the G Matrix under Nonlinear Genotype-Phenotype Maps. *American Naturalist* **199**: 420-435.
Mitchell, J. S. 2015. Extant-only comparative methods fail to recover the disparity preserved in the bird fossil record. *Evolution* **69**: 2414-2424.
Mitteroecker, P. & Bookstein, F. 2009. The Ontogenetic Trajectory of the Phenotypic Covariance Matrix, with Examples from Craniofacial Shape in Rats and Humans. *Evolution* **63**: 727-737.
Morrissey, M. B. 2016. Meta-analysis of magnitudes, differences and variation in evolutionary parameters. *Journal of Evolutionary Biology* **29**: 1882-1904.
Munoz, M. M. & Price, S. A. 2019. The Future is Bright for Evolutionary Morphology and Biomechanics in the Era of Big Data. *Integrative and Comparative Biology* **59**: 599-603.
Nilsson, K. J., Tsuboi, M., Opedal, Ø. & Runemark, A. 2024. Colonization of a novel host plant reduces phenotypic variation. *Evolutionary Biology* **51**: 269-282.
Noble, D. W. A., Radersma, R. & Uller, T. 2019. Plastic responses to novel environments are biased towards phenotype dimensions with high additive genetic variation. *Proceedings of the National Academy of Sciences of the United States of America* **116**: 13452-13461.
Olson, E. C. & Miller, R. L. 1958. *Morphological integration*. University of Chicago Press, Chicago.
Opedal, Ø. H., Armbruster, W. S., Hansen, T. F., Holstad, A., Pélabon, C., Andersson, S., Campbell, D. R., Caruso, C. M., Delph, L. F. & Eckert, C. G. 2023. Evolvability and trait function predict phenotypic divergence of plant populations. *Proceedings of the National Academy of Sciences* **120**: e2203228120.
Parsons, K. J. & Albertson, R. C. 2009. Roles for Bmp4 and CaM1 in Shaping the Jaw: Evo-Devo and Beyond. *Annual Review of Genetics* **43**: 369-388.
Pavličev, M., Bourg, S. & Le Rouzic, A. (2023) The Genotype-Phenotype Map Structure and Its Role in Evolvability. In: *Evolvability: A Unifying Concept in Evolutionary Biology?*, (Hansen, T. F., Houle, D., Pavličev, M. & Pélabon, C., eds.). pp. 147-169. The MIT Press.
Pavlicev, M. & Cheverud, J. M. 2015. Constraints Evolve: Context Dependency of Gene Effects Allows Evolution of Pleiotropy. *Annual Review of Ecology, Evolution, and Systematics* **46**: 413-434.
Pavlicev, M., Cheverud, J. M. & Wagner, G. P. 2009. Measuring Morphological Integration Using Eigenvalue Variance. *Evolutionary Biology* **36**: 157-170.
Pavlicev, M., Norgard, E. A., Fawcett, G. L. & Cheverud, J. M. 2011. Evolution of Pleiotropy: Epistatic Interaction Pattern Supports a Mechanistic Model Underlying Variation in

Genotype-Phenotype Map. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* **316b**: 371-385.
Pélabon, C., Morrissey, M. B., Reid, J. M. & Sztepanacz, J. L. (2023) Can We Explain Variation in Evolvability on Ecological Timescales? In: *Evolvability: A unifying concept in evolutionary biology?*, (Hansen, T. F., Houle, D., Pavlicev, M. & Pelabon, C., eds.). pp. 267-287. MIT Press.
Pontarp, M., Brännström, Å. & Petchey, O. L. 2019. Inferring community assembly processes from macroscopic patterns using dynamic eco-evolutionary models and Approximate Bayesian Computation (ABC). *Methods in Ecology and Evolution* **10**: 450-460.
Pontarp, M., Lundberg, P. & Ripa, J. 2024. The succession of ecological divergence and reproductive isolation in adaptive radiations. *Journal of Theoretical Biology* **587**: 111819.
Porto, A., de Oliveira, F. B., Shirai, L. T., De Conto, V. & Marroig, G. 2009. The Evolution of Modularity in the Mammalian Skull I: Morphological Integration Patterns and Magnitudes. *Evolutionary Biology* **36**: 118-135.
Pujol, B., Blanchet, S., Charmantier, A., Danchin, E., Facon, B., Marrot, P., Roux, F., Scotti, I., Teplitsky, C., Thomson, C. E. & Winney, I. 2018. The Missing Response to Selection in the Wild. *Trends in Ecology & Evolution* **33**: 337-346.
Punzalan, D. & Rowe, L. 2016. Concordance between stabilizing sexual selection, intraspecific variation, and interspecific divergence in Phymata. *Ecology and Evolution* **6**: 7997-8009.
Puttick, M. N. 2018. Mixed evidence for early bursts of morphological evolution in extant clades. *Journal of Evolutionary Biology* **31**: 502-515.
Rangel, T. F., Edwards, N. R., Holden, P. B., Diniz, J. A. F., Gosling, W. D., Coelho, M. T. P., Cassemiro, F. A. S., Rahbek, C. & Colwell, R. K. 2018. Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves. *Science* **361**.
Renaud, S., Auffray, J. C. & Michaux, J. 2006. Conserved phenotypic variation patterns, evolution along lines of least resistance, and departure due to selection in fossil rodents. *Evolution* **60**: 1701-1717.
Riedl, R. 1978. *Order in living organisms: a systems analysis of evolution*. John Wiley & Sons.
Roff, D. A. 1995. The Estimation of Genetic Correlations from Phenotypic Correlations - a Test of Cheveruds Conjecture. *Heredity* **74**: 481-490.
Rohner, P. T. & Berger, D. 2023. Developmental bias predicts 60 million years of wing shape evolution. *Proceedings of the National Academy of Sciences* **120**: e2211210120.
Rolland, J., Henao-Diaz, L. F., Doebeli, M., Germain, R., Harmon, L. J., Knowles, L. L., Liow, L. H., Mank, J. E., Machac, A. & Otto, S. P. 2023. Conceptual and empirical bridges between micro-and macroevolution. *Nature Ecology & Evolution* **7**: 1181-1193.
Rossoni, D. M., Costa, B. M. A., Giannini, N. P. & Marroig, G. 2019. A multiple peak adaptive landscape based on feeding strategies and roosting ecology shaped the evolution of cranial covariance structure and morphological differentiation in phyllostomid bats. *Evolution* **73**: 961-981.
Saito, K., Tsuboi, M. & Takahashi, Y. 2024. Developmental noise and phenotypic plasticity are correlated in Drosophila simulans. *Evolution Letters* **8**: 397-405.
Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution* **50**: 1766-1774.
Sella, G. & Barton, N. H. 2019. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics* **20**: 461-493.

Siciliano, V., Garzilli, I., Fracassi, C., Criscuolo, S., Ventre, S. & di Bernardo, D. 2013. miRNAs confer phenotypic robustness to gene networks by suppressing biological noise. *Nature Communications* **4**: 2364.

Siepielski, A. M., Morrissey, M. B., Buoro, M., Carlson, S. M., Caruso, C. M., Clegg, S. M., Coulson, T., DiBattista, J., Gotanda, K. M., Francis, C. D., Hereford, J., Kingsolver, J. G., Augustine, K. E., Kruuk, L. E. B., Martin, R. A., Sheldon, B. C., Sletvold, N., Svensson, E. I., Wade, M. J. & MacColl, A. D. C. 2017. Precipitation drives global variation in natural selection. *Science* **355**: 959-962.

Signor, S. A. & Nuzhdin, S. V. 2018. The Evolution of Gene Expression in cis and trans. *Trends in Genetics* **34**: 532-544.

Simons, A. M. 2002. The continuity of microevolution and macroevolution. *Journal of Evolutionary Biology* **15**: 688-701.

Simpson, G. G. 1944. *Tempo and mode in evolution*. Columbia University Press.

Smith, L. H. 1998. Asymmetry of Early Paleozoic trilobites. *Lethaia* **31**: 99-112.

Sodini, S. M., Kemper, K. E., Wray, N. R. & Trzaskowski, M. 2018. Comparison of Genotypic and Phenotypic Correlations: Cheverud's Conjecture in Humans. *Genetics* **209**: 941-948.

Stearns, F. W. 2010. One Hundred Years of Pleiotropy: A Retrospective. *Genetics* **186**: 767-773.

Steppan, S. J., Phillips, P. C. & Houle, D. 2002. Comparative quantitative genetics: evolution of the G matrix. *Trends in Ecology & Evolution* **17**: 320-327.

Svensson, E. I., Arnold, S. J., Burger, R., Csillery, K., Draghi, J., Henshaw, J. M., Jones, A. G., De Lisle, S., Marques, D. A., McGuigan, K., Simon, M. N. & Runemark, A. 2021. Correlational selection in the age of genomics. *Nature Ecology & Evolution* **5**: 562-573.

Svensson, E. I. & Berger, D. 2019. The Role of Mutation Bias in Adaptive Evolution. *Trends in Ecology & Evolution* **34**: 422-434.

Sztepanacz, J. L. & Blows, M. W. 2017. Artificial Selection to Increase the Phenotypic Variance in Gmax Fails. *American Naturalist* **190**: 707-723.

Sztepanacz, J. L. & Houle, D. 2021. Allometry constrains the evolution of sexual dimorphism in Drosophila across 33 million years of divergence. *Evolution* **75**: 1117-1131.

Sztepanacz, J. L. & Rundle, H. D. 2012. Reduced Genetic Variance among High Fitness Individuals: Inferring Stabilizing Selection on Male Sexual Displays in Drosophila Serrata. *Evolution* **66**: 3101-3110.

Tomasovych, A., Gallmetzer, I., Haselmair, A. & Zuschin, M. 2022. Inferring time averaging and hiatus durations in the stratigraphic record of high-frequency depositional sequences. *Sedimentology* **69**: 1083-1118.

Tsuboi, M., van der Bijl, W., Kopperud, B. T., Erritzoe, J., Voje, K. L., Kotrschal, A., Yopak, K. E., Collin, S. P., Iwaniuk, A. N. & Kolm, N. 2018. Breakdown of brain-body allometry and the encephalization of birds and mammals. *Nature Ecology & Evolution* **2**: 1492-1500.

Tufto, J. 2015. Genetic evolution, plasticity, and bet-hedging as adaptive responses to temporally autocorrelated fluctuating selection: A quantitative genetic model. *Evolution* **69**: 2034-2049.

Uller, T., Moczek, A. P., Watson, R. A., Brakefield, P. M. & Laland, K. N. 2018. Developmental Bias and Evolution: A Regulatory Network Perspective. *Genetics* **209**: 949-966.

Uyeda, J. C., Hansen, T. F., Arnold, S. J. & Pienaar, J. 2011. The million-year wait for macroevolutionary bursts. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 15908-15913.

Van Valen, L. 1962. A study of fluctuating asymmetry. *Evolution* **16**: 125-142.
Voje, K. L., Di Martino, E. & Porto, A. 2020. Revisiting a Landmark Study System: No Evidence for a Punctuated Mode of Evolution in Metrarabdotos. *The American Naturalist* **195**: 899-917.
Voje, K. L., Grabowski, M., Holstad, A., Porto, A., Tsuboi, M. & Bolstad, G. H. (2023a) Does Lack of Evolvability Constrain Adaptation? If So, on What Timescales? In: *Evolvability: A Unifying Concept in Evolutionary Biology?*, (Thomas F. Hansen, D. H., Mihaela Pavličev, and Christophe Pélabon, ed.). pp. 289-306. The MIT Press, Massachusetts, USA.
Voje, K. L., Hansen, T. F., Egset, C. K., Bolstad, G. H. & Pelabon, C. 2014. Allometric Constraints and the Evolution of Allometry. *Evolution* **68**: 866-885.
Voje, K. L., Saulsbury, J. G., Starrfelt, J., Latorre, D. V., Rojas, A., Kinneberg, V. B., Liow, L. H., Wilson, C. J., Saupe, E. E. & Grabowski, M. 2023b. Measurement theory and paleobiology. *Trends in Ecology & Evolution* **38**: 1165-1176.
Waddington, C. H. 1957. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. Routledge.
Wagner, G. P. 1984. On the Eigenvalue Distribution of Genetic and Phenotypic Dispersion Matrices - Evidence for a Nonrandom Organization of Quantitative Character Variation. *Journal of Mathematical Biology* **21**: 77-95.
Wagner, G. P. 1996. Apparent stabilizing selection and the maintenance of neutral genetic variation. *Genetics* **143**: 617-619.
Wagner, G. P. & Altenberg, L. 1996. Perspective: Complex adaptations and the evolution of evolvability. *Evolution* **50**: 967-976.
Wagner, G. P., Booth, G. & Bagheri, H. C. 1997. A population genetic theory of canalization. *Evolution* **51**: 329-347.
Walsh, B. & Blows, M. W. 2009. Abundant Genetic Variation plus Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology Evolution and Systematics* **40**: 41-59.
Watanabe, J. 2022. Statistics of eigenvalue dispersion indices: Quantifying the magnitude of phenotypic integration. *Evolution* **76**: 4-28.
Webster, M. & Zelditch, M. L. 2011a. Evolutionary Lability of Integration in Cambrian Ptychoparioid Trilobites. *Evolutionary Biology* **38**: 144-162.
Webster, M. & Zelditch, M. L. 2011b. Modularity of a Cambrian ptychoparioid trilobite cranidium. *Evolution & Development* **13**: 96-109.
Yu, C. Y., Qin, F. B., Watanabe, A., Yao, W. Q., Li, Y., Qin, Z. C., Liu, Y. M., Wang, H. B., Qigao, J. Z., Hsiang, A. Y., Ma, C., Rayfield, E., Benton, M. J. & Xu, X. 2024. Artificial intelligence in paleontology. *Earth-Science Reviews* **252**: 104765.
Zelditch, M. L. & Goswami, A. 2021. What does modularity mean? *Evolution & Development* **23**: 377-403.
Zeng, Z. B. 1988. Long-Term Correlated Response, Interpopulation Covariation, and Interspecific Allometry. *Evolution* **42**: 363-374.

Table 1. Summary of measurements describing different aspects of variational properties in multivariate traits.

Level	Metric	Equation	Source	Measure
Variability	mutational co/variance matrix (**M**)	-	Lande 1976	Mutational rate of input of genetic variance
	fluctuating asymmetry co/variance matrix (**F**) *	cov($\mathbf{z_l}$-$\mathbf{z_r}$)	Van Valen 1962	Developmental variability caused by local genetic/environmental perturbations
Variation†	evolvability (e)	$\beta'\mathbf{G}\beta$	Houle 1992	The expected response to a unit directional selection gradient
	global conditional evolvability (c_{global})	$(\beta'\mathbf{G^{-1}}\beta)^{-1}$	Hansen et al. 2003	The realized evolvability along a vector when all other directions are under stabilizing selection
	evolvability conditioned on a selected trait (c_x)	$G_z - (\mathbf{G_{zx}G^{-1}G_{zx}^t})$	Hansen et al. 2003	The realized evolvability of a focal trait (z) when a specified trait (x) is under stabilizing selection
	autonomy (a) and integration	$c^{\ddagger}/e$	Hansen and Houle 2008	The fraction of e independent of trait covariances. Integration is $1-a$
	eigenvector dispersion	var($\boldsymbol{\lambda}$)	Wagner 1984, Pavlicev et al. 2009	The degree to which variances are concentrated on few dimensions
Divergence	divergence matrix (**D**)	cov($\bar{\mathbf{z}}$)	-	Divergence among population/species
	rate of evolution matrix (**R**)	-	Lynch 1991	Rates of evolution according to a multivariate Brownian motion model
	geodesic distance	$\sqrt{\Sigma_{i=1}^{n}\log(\lambda_{c\cdot i})^2}$	Mitteroecker and Bookstein 2009	The shortest path between a pair of matrices
	response difference	$\sqrt{\beta'(\mathbf{G_1}-\mathbf{G_2})^2\beta}$	Hansen and Houle 2008	The magnitude of the difference between populations in response to a common β if they evolved from the same multivariate mean

Abbreviations: **G**: mean-scaled additive genetic co/variance matrix, z_l and z_r: phenotypes measured from left- and right-side of a bilaterally symmetric trait, β: mean-standardized selection gradient, $\boldsymbol{\lambda}$: eigenvalues of co/variance matrix, $\bar{z}$: mean phenotype value of a population or a species, $\lambda_{c\cdot i}$: eigenvalues of $\mathbf{C} = \mathbf{G_1^{-1}G_2}$

Notes: *: also a measurement of environmental variation. †: equations to evaluate genetic variation are presented. To measure environmental variation, replace **G** with environmental co/variance matrix (**E**). ‡: can either be c_{global} or c_x

Table 2. Outstanding questions relevant to addressing the paradox of predictability.

Level	Outstanding questions
Variability	• Is fluctuating asymmetry (FA) correlated with variability? If so, can we disentangle genetic and microenvironmental causes of FA? • Can FA be reliably measured in the fossil record (Smith, 1998)? • Is FA co/variance (**F**) correlated with mutational co/variance (**M**) (Rohner & Berger, 2023)?
Between variability and variation	• How can the prevalence of large-effect non-linearity in the GP-map (Milocco & Salazar-Ciudad, 2022) be reconciled with the paradox of predictability? • What are the molecular mechanisms of phenotypic robustness (Green et al., 2017)? How common are they? • Is the multivariate selection surface (γ) aligned with **M** and **F** (Svensson & Berger, 2019)?
Variation	• To what extent can phenotypic co/variance (**P)** be used as a proxy for additive genetic co/variance (**G**), both in morphological and other traits? • What are the genetic and developmental mechanisms that generate a resemblance among **P**, **G**, and environmental co/variance (**E**)? • Can the pattern of integration, autonomy, and modularity explain the residual variance of the divergence-variance relationship?
Between variation and divergence	• How can spatiotemporal variation in contemporary selection (Siepielski et al., 2017) give rise to a stable macroevolutionary adaptive zone (Simpson, 1944) and primary optimum (Hansen, 1997)? • Under what circumstances do divergence-variance relationships break down? Is the breakdown related to taxonomic levels (Guillerme et al., 2023), ecological shifts (Brombacher et al., 2017), or novel developmental pathways (Tsuboi et al., 2018)?
Divergence	• What is the tempo and mode of evolution in co/variance matrices? • How can process-based macroevolutionary models of co/variance be constructed?

Box 1: A case study: the paradox of predictability on generational timescales and beyond.

To illustrate the generality of the paradox of predictability, we examined divergence-variance relationships using both extant and fossil data based on phenotypic time-series—arrays of phenotypes measured from sequences of populations inferred to have ancestor-to-descendant relationships. Data from the long term sampling of contemporary populations have been central to the study of predictability in microevolution (Grant & Grant, 2002). When integrated with similar data from fossil lineages, the continuity for morphological changes from generational to geological timescales may be illuminated (Gingerich, 1983, Uyeda et al., 2011).

Fossil time-series differ from contemporary ones in that they are subject to much greater temporal uncertainty. The ages of contemporary populations are essentially known without error, whereas dating

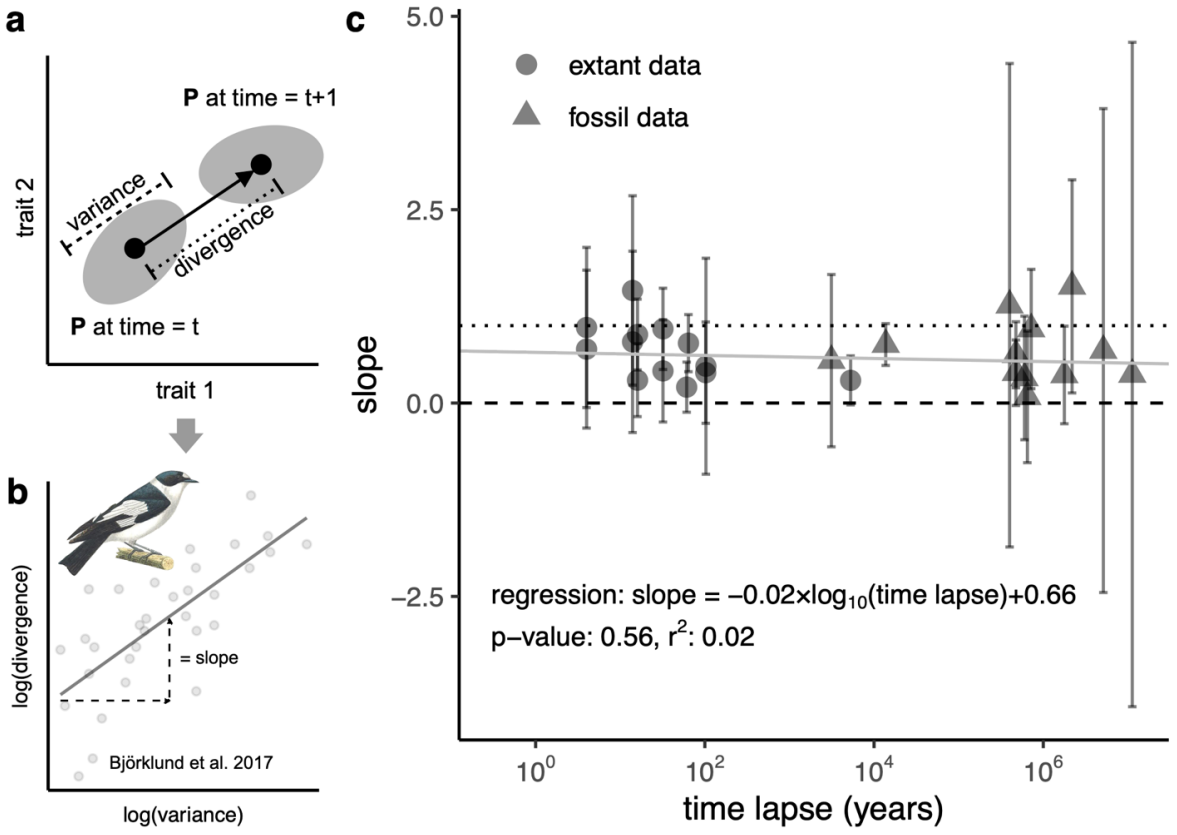

Figure 1. The paradox of predictability tested with time-series data. (a) Schematic illustration of our analysis. If variance and divergence are correlated, the log-log slope of the regression between these variables will be positive. The slope of 1 (= isometry) indicates that the divergence is exactly proportional to the variance. (b) An example data representing four morphological traits measured in the Collard Flycatcher *Ficedula albicollis*. Image taken from Cuvier (1828). (c) The relationship between divergence-variance slope and time lapse of each time-series dataset (N datasets = 25). No relationship (slope = 0) and isometry (slope = 1) are indicated with dashed and dotted lines, respectively. Estimates from extant and fossil time-series are shown in different symbols with error bars representing ± 1.96 standard errors (SEs). A grey line shows a weighted GLS (by SEs) regression between slope and time lapse. Regression estimates and statistical tests are also shown.

sediments is complex and challenging (Gradstein et al., 2020). Moreover, fossil samples are generally time-averaged, as bioturbation and other sediment-mixing processes combine individual organisms that lived over a range of ages into a single sampled layer. Durations of this time-averaging can be quite short, as in annually resolved sediment layers, but they also can be substantial: up to millennia for robustly skeletonized taxa (e.g. Kowalewski et al., 2018, Tomasovych et al., 2022). This time-averaging is a particular concern for the estimation of variational properties (Table 1), because any evolutionary changes occurring within the time-averaged interval will be collapsed into within-sample variation. This contaminates estimates of variation with divergence. Fortunately, empirical studies across a wide range of taxa, traits, and environments have found this variation inflation effect to be quite small (Bell et al., 1987, MacFadden, 1989, Hunt, 2004, Di Martino & Liow, 2021). As a result, it seems that variational properties measured from fossil samples generally reflect standing phenotypic variation, despite time-averaging.

We compiled published data from both contemporary and fossil systems (see SI for the data and details of analyses). To be included, studies had to: (1) provide data from two or more morphological variables, (2) be measured on a ratio scale, and (3) collect from three or more temporal populations of the same species in the same area for which estimates of elapsed time also were available. Our initial compilations found 15 studies containing 25 time series that fulfilled our criteria. For each ancestor-descendant pair of sequential populations, we computed the phenotypic variance in the direction of evolutionary divergence, along with the magnitude of this divergence (Fig. 1A). Ordinary least-squares regression was used to test if variance predicted divergence magnitudes within each time-series. We summarized these relationships by their slopes (Fig. 1B).

All 25 datasets showed a positive slope, indicating that evolutionary changes were generally larger in morphological directions of high evolvability (Fig. 1C). The relationship is noisy and only five of these slopes are statistically different from zero individually. Many of these time-series are short, which introduces substantial noise in slope estimates. However, we see no mechanism by which this or other artifacts can plausibly account for the extreme preponderance of positive relationships between variance and divergence.

Hence, contemporary and fossil datasets both show the paradoxical divergence-variance correlations, despite their vast differences in temporal resolution. There is no relationship between elapsed time and slope (Fig. 1C). This is inconsistent with previous studies finding a decay in predictability over longer, macroevolutionary timescales (Schluter, 1996, Hunt, 2007). This pattern is expected if the G-matrix itself evolves over time, but the rapidity and specific pattern of the decline are not easily predicted. From the viewpoint of selection-shapes-variation hypotheses, this pattern might be explained if selection intensity changed over time. One explanation for the discrepancy is that divergence and evolvability may be subject to greater error in the fossil data because of variable fossil preservation, time-averaging, and lower sample sizes. This would inflate the noise in estimates of variance in the fossil studies. We explored some of these possibilities (SI results), but we were unable to find evidence that error associated with fossil data could have artifactually generated these results (Fig. S5). Our analyses are consistent with Holstad et al. (2024) that the divergence-variance correlation is a common feature of phenotypic evolution in both neontological and paleontological samples.

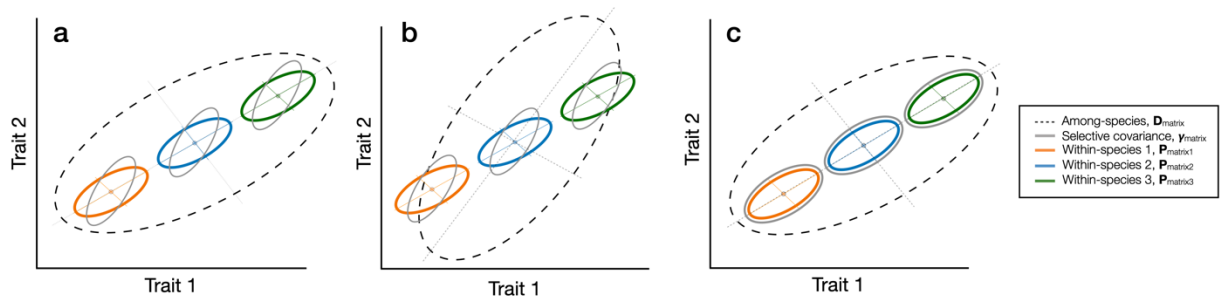

Figure 2. Schematic illustration of how divergence ($\mathbf{D}_{matrix}$, dashed ellipse), within-species matrices ($\mathbf{P}_{matrix1,2,3}$, orange, blue and green ellipses), and non-linear selection ($\boldsymbol{\gamma}_{matrix}$, gray solid ellipse) would be related. The bivariate mean is in the center of each matrix (dots) and the two major axes of variation (eigenvectors) are also depicted. (a) **D** is proportional to **P**, suggesting that the pattern of divergence is shaped by the pattern of variation, supporting the constraint hypothesis. (b) **D** is proportional to $\boldsymbol{\gamma}$, suggesting that the pattern of divergence is shaped by multivariate stabilizing selection, supporting the selection-shapes-variation hypothesis. (c) **D** is proportional to both **P** and $\boldsymbol{\gamma}$. The relationship is consistent with both hypotheses, and one cannot disentangle the two from these relationships.