Evolvability in the Fossil Record

Article Type: Review

Word Count

Main Text: 8,227

5 Figures

1 Box

Authors (*corresponding)

Alan C. Love* (aclove@umn.edu): Department of Philosophy and Minnesota Center for Philosophy of Science, University of Minnesota

Mark Grabowski (m.w.grabowski@ljmu.ac.uk): Research Centre in Evolutionary Anthropology and Palaeoecology, School of Biological and Environmental Sciences, Liverpool John Moores University

David Houle (dhoule@bio.fsu.edu): Department of Biological Science, Florida State University

Lee Hsiang Liow (l.h.liow@nhm.uio.no): Natural History Museum & Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo

Arthur Porto (aporto3@lsu.edu): Department of Biological Sciences and Center for Computation and Technology, Louisiana State University

Masahito Tsuboi (masahito.tsuboi@biol.lu.se): Department of Biology, Lund University; Department of Biosciences, University of Oslo

Kjetil L. Voje (k.l.voje@nhm.uio.no): Natural History Museum, University of Oslo, Oslo

Gene Hunt (hunte@si.edu): Department of Paleobiology, National Museum of Natural History, Smithsonian Institution

Acknowledgments: This paper emerged out of a collaboration initiated by our participation in the project “Evolvability: a new and unifying concept for evolutionary biology?” (2019-2020), which was funded by the Norwegian Academy of Science and Letters and hosted by the Centre for Advanced Study (Oslo) in 2019-2020. We are grateful to the project group leaders, Thomas Hansen and Christophe Pelabon, for their organizational efforts and all of the Fellows involved in the project for stimulating discussions on many of the issues treated in the manuscript.

Abstract

The concept of evolvability—the capacity of a population to produce and maintain evolutionarily relevant variation—has become increasingly prominent in evolutionary biology. Although paleontology has a long history of investigating questions of evolvability, often invoking different but allied terminology, the study of evolvability in the fossil record has seemed intrinsically problematic. How can we surmount difficulties in disentangling whether the causes of evolutionary patterns arise from variational properties of traits or lineages rather than due to selection and ecological success? Despite these challenges, the fossil record is unique in offering growing sources of data that span millions of years and therefore capture evolutionary patterns of sustained duration and significance otherwise inaccessible to evolutionary biologists. Additionally, there are a variety of strategic possibilities for combining prominent neontological approaches to evolvability with those from paleontology. We illustrate three of these possibilities with quantitative genetics, evolutionary developmental biology, and phylogenetic models of macroevolution. In conclusion, we provide a methodological schema that focuses on the conceptualization, measurement, and testing of hypotheses to motivate and provide guidance for future empirical and theoretical studies of evolvability in the fossil record.

Keywords: constraints, disparity, evo-devo, evolutionary potential, integration, macroevolution, modularity, quantitative genetics, variability, versatility

1. Introduction

The term “evolvability” refers to those characteristics that confer a disposition to evolve under a causal stimulus, such as natural selection. In most modern usages, evolvability is tied to the ability of a population to produce and maintain evolutionarily relevant genetic variation (Crother and Murray 2019). Evolvability has become increasingly prominent in evolutionary biology through studies that adopt approaches from quantitative genetics (Hansen and Houle 2008), evolutionary developmental biology (evo-devo) (Hendrikse et al. 2007, Tiozzo and Copley 2015), phylogenetic models of macroevolution (Hunt and Slater 2016), and experimental evolution (Colegrave and Collins 2008).

Although paleontology has a long history of investigating questions of evolvability (e.g., Eldredge and Gould 1972, Simpson 1944), often invoking different but allied terminology (e.g., constraint or versatility *sensu* Vermeij 1973a, b), the study of evolvability in the fossil record has seemed intrinsically problematic. This is largely because of difficulties in disentangling whether the causes of evolutionary patterns arise from variational properties of traits or lineages (‘evolvability’ features) or properties of the abiotic or biotic environment (sources of selection and ecological success), both of which might generate similar if not indistinguishable outcomes in deep time (Jablonski 2017a, b, 2020, Jackson 2020). These difficulties are often viewed as a rationale for investigating questions about evolvability using other, primarily neontological approaches because they harbor the promise of discriminating between variational contributions and selective components, especially through experimental manipulation of variables related to trait generation or selective regime under highly controlled conditions (Colegrave and Collins 2008, Payne and Wagner 2019).

Importantly, this rationale ignores the distinctive empirical resources that paleontological studies bring to questions about evolvability. Despite the challenges of disentangling the variational and environmental causal factors responsible for evolutionary trajectories in the history of life, the fossil record is unique in offering data that span millions of years and therefore capture evolutionary processes of sustained duration and significance that are otherwise inaccessible to evolutionary biologists (Bell 2014, Dilcher 2000, Jablonski and Shubin 2015; Jackson 2020). Studies of the differential rates of evolution in lineages and the propensity for specific traits to contribute to adaptive radiations can relate directly to evolvability and may be best seen empirically in the realm of the fossil record. At a minimum, it is critical to find an appropriate balance between what paleontology is uniquely positioned to offer to the study of evolvability, and what is impossible as a result of the absence or loss of pertinent information.

Paleontology not only offers an unparalleled and ever-increasing data resource, but its value is augmented when combined strategically with other approaches to evolvability. Consider first the intersection of quantitative genetics and paleontology. A pioneering study (Cheverud 1988) suggested that a standardized summary of the multivariate phenotypic relationships among a set of traits (i.e., the **P** matrix, or *phenotypic* variance-covariance matrix), which can be derived from fossil evidence, could serve as a reliable proxy for the standardized summary of multivariate genetic relationships among a set of traits (i.e., the **G** matrix, or additive *genetic* variance-covariance matrix), whose geometry shapes the direction of evolution in response to selection (Lande 1979). Subsequently, a number of studies have profitably analyzed morphological traits using fossil data that take advantage of this proxy inference (e.g., Brombacher et al. 2017a, Hunt 2007, Renaud et al. 2006). Another example is found in combining evo-devo approaches with paleontology. On the assumption that ontogenetic processes are conserved from fossil taxa to their modern relatives, sufficient developmental information can yield predictions about which traits and lineages may be more likely to produce abundant phenotypic variation relevant to evolutionary processes (Urdy et al. 2013), such as models of mammalian molar development predicting the evolvability of different dental traits (e.g., Jernvall 2000, Salazar-Ciudad and Jernvall 2010). A third example is how evolvability has been analyzed using phylogenetic models at the level of macroevolution, where different modes of evolutionary change for traits (e.g., directional selection or punctuated equilibrium) can be correlated with key variables such as speciation rates in a lineage (Rabosky 2012). Increased evolvability was implicated in dramatic morphological evolution via changes to the structure of developmental modules using phylogenetic models (Parins-Fukuchi 2020).

Our aim in the present paper is to make an explicit case for studying evolvability in the fossil record with special attention to the advances that can be derived from fruitful cross-disciplinary collaborations in evolutionary biology. We begin by recalling touchstones in the history of paleontology where questions about evolvability were under scrutiny, sometimes in the guise of alternative terminology, and accent the unique position of the fossil record for informing questions about evolvability. Next, we illustrate in detail how paleontology is working in combination with other approaches to yield new insights into evolvability, focusing on three primary partnerships: quantitative genetics, evo-devo, and phylogenetically-informed macroevolutionary modeling. In conclusion, we offer a methodological schema that focuses on the conceptualization, measurement, and testing of hypotheses for investigating evolvability that yields several potential avenues of research on outstanding questions that exploit both the distinctive contribution of paleontology and the interdisciplinary synergy available with other approaches in evolutionary biology. Overall, this generates a strong motivation for empirical and theoretical studies of evolvability in the fossil record.

2. Evolvability and Paleontology: Classical Studies and Controversies

The modern concept of evolvability focuses on the variational properties of traits, especially how the relationship between genotype and phenotype mediated by development establishes the potential or capacity of traits to respond to drift and selection (Houle 1992, Wagner and Altenberg 1996). Many paleontological studies examine phenotypic variation, including those pertaining to taxonomic richness, morphological disparity, functional diversity, and morphological change in single or multiple related lineages, but few directly invoke the term "evolvability" or attempt to connect directly with its modern meanings. However, some classics in paleontological literature stand out as pioneering ideas closely aligned with and sometimes pre-dating the modern evolvability concept. For example, in *Tempo and Mode in Evolution*, G.G. Simpson noted that the "capacity of … animals to differ" is distinct from the expression of differences among individuals (i.e., realized variation) and the inheritance of phenotypes (Simpson 1944, 30), a perspective he saw explicitly in earlier paleontological discussions (e.g., Rosa 1899). In a related vein, Vermeij (1973b) claimed that there was an "increase in potential versatility of form" through geological time, which he supported with an example where the number of parameters required to describe coiling in gastropods increased over the Phanerozoic. According to Vermeij, groups with a greater potential versatility of form replaced those that exhibited this capacity to a lesser degree (Vermeij 1973a).

Beyond these classic exemplars, there is a rich literature on the temporal dynamics of morphological disparity among taxa that points to changing patterns of new traits and trait combinations over long time spans (reviewed in Foote 1997, Hughes et al. 2013); some groups, once evolved, seem constrained in morphospace, whereas previously occupied morphospaces, once vacated, are sometimes not reoccupied. While these patterns are the combined outcome of both evolvability and ecological success or failure (i.e., selection), the potential for evolvability explanations has long been recognized, usually considered in terms of evolutionary constraints—the lack of evolvability in some guise (Allmon and Ross 1990, Blake 1980, Erwin 2007, Gould 1989, Jablonski 2020, Maynard Smith et al. 1985, Raup 1967, Wright 2017). This connection between morphospace exploration and evolvability was perhaps most explicit in discussions of the dramatic explosion of disparity in the Cambrian Period. Two classes of (non-mutually exclusive) hypotheses have been commonly considered: (i) those that emphasize ecological opportunities afforded by nearly unoccupied early Paleozoic ecosystems or environmental triggers such as changes in the amount of dissolved oxygen in seawater that facilitate the formation of biomineralized skeletons, and (ii) those that posit genetic or developmental processes facilitating elevated expression of morphological variation in the Cambrian (Erwin 1994, Erwin and Valentine 2013, Valentine 1995, Webster 2007, Webster 2019). The first class of explanation invokes selection, whereas the second relates to evolvability.

Similarly, stasis within fossil species can be explained by mechanisms that are either extrinsic, related to natural selection, or intrinsic, related to the variational potential of populations. Eldredge and Gould’s (1972) original suggestion for the cause for stasis was that variational constraints would be relaxed at speciation. This particular explanation did not fare well, as Gould himself later acknowledged (Gould 2002), but more modern versions suggested that stasis may result when traits lack variation or if most variation is bound up in correlations with other traits (Hansen and Houle 2004, 2008). In opposition to these explanations are those that view stasis as a consequence of stabilizing natural selection (Charlesworth et al. 1982, Estes and Arnold 2007, Hunt and Rabosky 2014).

Extinction as a failure of evolvability is implicit in Van Valen's Red Queen Hypothesis (Van Valen 1973), which envisions species at a constant risk of extinction because they must continually adapt in the face of changing environments and to other species that are continually improving. However, studies that explicitly test whether evolvability (as reflected in, for example, trait variation) protects against species extinction are still quite rare (Hopkins 2011, Kolbe et al. 2011, Liow 2007). Quantitative genetics suggests distinctive strategies for measuring trait variation in the fossil record that can provide further unique insights into evolvability on geological time scales.

3. Quantitative Genetics in the Rock Record

3.1 Estimating G-matrices from P-Matrices Drawn from Fossils

Evolutionary quantitative genetics is a theoretical framework linking selection and genetic variation to evolutionary change (Lynch and Walsh 1998; Walsh and Lynch 2018). Central to this framework is the "Lande equation" (Lande 1976, 1979, Lande and Arnold 1983), which permits the response to selection to be decomposed into the (a) pattern of genetic variation and covariation among traits (summarized in the *genetic* variance-covariance [**G**] matrix, see below), and (b) strength and direction of selection on individual traits (e.g., from environmental factors). This decomposition formally separates evolutionary change into evolvability-related and selection-related components. Since most traits do not exist as autonomous units and are unable to respond to selection independently of other traits (Cheverud 1982a, Hansen et al. 2003a, Hansen and Houle 2008, Lande 1979, Lynch and Walsh 1998, Walsh and Blows 2009), a multivariate theoretical formulation of natural selection and variation is necessary to gain a more satisfactory understanding of evolutionary change.

In the Lande equation, $\Delta z = \mathbf{G}\beta$, Δz is the response to selection (a vector with trait changes for multiple traits), **G** is the additive genetic variance-covariance matrix (or **G**-matrix) among those traits, which quantifies the role of the genetic system in evolution, and β is the selection gradient, which quantifies the amount and direction of selection on each trait independent of other traits (Fig. 1). Hansen and colleagues (Hansen et al. 2003a, Hansen and Houle 2008, Hansen et al. 2003b) later used the Lande equation to develop a theoretical framework connecting the **G**-matrix to short-term evolutionary potential through the concept of evolvability (Houle 1992), which describes a population's ability to evolve in the direction of selection when stabilizing selection is absent (Hansen and Houle 2008). One way of calculating a trait's evolvability is to divide its additive genetic variance by the trait mean squared (i.e., evolvability equals a mean-standardized additive genetic variance). This measure of evolvability predicts an expected proportional response to selection that is as strong as that on fitness. For example, an evolvability of 0.10 means the expected response in the trait mean per generation is 10% given selection as strong as selection on fitness itself. Typically, observed directional selection is on the order of 10% as strong as the selection on fitness (Hereford et al. 2004). This concept of evolvability also can be used to generate hypotheses about the direction and strength of selection (β) responsible for past evolutionary change (Δz).

[Figure 1 near here]

For many evolutionary biologists and paleontologists, quantitative genetics seems irrelevant for studies of macroevolution, including paleontology, because of empirical evidence and theoretical considerations that imply **G** can evolve rapidly (e.g., Pigliucci 2006). If **G** is likely to evolve within short time scales, then its power to predict evolution is severely limited. Although we know **G** can evolve (Steppan et al. 2002), directions of diversification among populations are often aligned with above-average genetic variation. Schluter (1996) was the first to show empirically that **G** can be sufficiently stable to have a detectable influence on the direction of evolution across macroevolutionary timescales and also suggested that genetic constraints would predict phenotypic divergence along "lines of least genetic resistance"—phenotypic divergence in directions aligned with above-average additive genetic variance (Fig. 1B). Several studies have found such a pattern (e.g., Bégin and Roff 2004, Blows and Higgie 2003, McGlothlin et al. 2018). More recently, the concept of evolvability has been used as a general framework to compare genetic variation measured within populations with rates of phenotypic divergence among populations across traits. There is growing evidence that evolvability can predict patterns of macroevolution at surprisingly long-time scales. For example, Bolstad et al. (2014) found that patterns of genetic variation in contemporary populations of *Dalechampia* predicted macroevolutionary divergence within the genus. Houle et al. (2017) showed that standing genetic variation within a population of *Drosophila melanogaster* was strongly correlated with phenotypic divergence across the Drosophilidae, which represents 40 million years of evolution. These studies suggest that evolutionary quantitative genetics may be applicable to much longer time scales than were considered previously. Paleontologists are well situated to contribute to the testing of the generality of these results.

Robustly estimating **G** normally requires measurements of large numbers of families of known pedigree. Properly estimating **G** is a major undertaking even in living populations

(Charmantier et al. 2014, Dochtermann 2011, Steppan et al. 2002); it is close to impossible for most taxa found in the fossil record. Fortunately, the phenotypic variance-covariance matrix, **P**, is both estimable in many paleontological settings and a possible proxy for **G**. **P** is the sum of **G** and other sources of phenotypic variation, notably the effects of the environment. Cheverud (1988) conjectured that **P** can be proportional to **G** based on three cogent arguments (Fig. 2). First, quantitative traits have reasonably high heritabilities (Hansen et al. 2011), meaning that **G** accounts for a large proportion of the variation measured by **P**. Second, the non-genetic variation in **P** arises through the same developmental and physiological pathways that structure **G**, and thus may have a similar pattern (Cheverud 1984). Finally, if the first two legs of the conjecture hold, estimates of **P** may better estimate the true **G** than direct estimates of **G**. The precision of a **G** matrix is a function of the number of families, while the precision of a **P** matrix is a function of the (much larger) number of individuals measured.

[Figure 2 near here]

Evidence for and against Cheverud’s conjecture has accumulated (de Oliveira et al. 2009, Kohn and Atchley 1988, Martínez-Abadías et al. 2012, Porto et al. 2009, Roff 1995, 1996, 1997, Simons and Roff 1996). The validity of the conjecture in non-morphological traits is controversial (Atchley et al. 1981, Hadfield et al. 2007, Lofsvold 1986), while a recent review confirms that **P** and **G** are generally similar for the morphological traits that paleontologists can measure (Sodini et al. 2018). Therefore, using **P** as an estimate of **G** enables paleontologists to strategically utilize evolutionary quantitative genetic tools.

An important cautionary note when estimating **P** from fossil data is that the population variance of fossil samples may be inflated due to changes in the population mean over the timescale captured by the sample. However, fossil samples have been found to show levels of trait variances and covariances similar in magnitude with estimates from extant, non-time-averaged populations (Hunt 2004). In addition, the richness of the fossil record varies substantially among taxa and not all fossil species have sufficient sample size to robustly estimate **P**. Cheverud (1988) suggested at least 40 individuals were needed for a reasonably accurate **P** for **G** substitution, but larger sample sizes are required as the number of traits increase and to accurately estimate some evolvability statistics (Grabowski and Porto 2017). One potential solution is to use **P** (or **G**) matrices from extant species as a substitute for unknown fossil **G** matrices (Ackermann and Cheverud 2004, Baab 2018, Grabowski et al. 2011, Grabowski and Roseman 2015, Hansen and Voje 2011, Young et al. 2010). This assumes that the estimated **P** or **G** from the extant population is similar enough to **G** in the ancestral extinct population. Based on similarities between closely related extant species, a wide array of work (both neontological and paleontological) assumed that the **P** (or **G**) from an extant population is representative of the ancestral **G**, which has allowed researchers to make evolutionary inferences from phenotypic data across macroevolutionary timescales in ways that would be impossible otherwise (Ackermann and Cheverud 2004, Baab 2018, de Oliveira et al. 2009, Grabowski et al. 2011, Marroig and Cheverud 2010, Rolian 2009, Young et al. 2010).

Paleontological studies that estimate **P** matrices exemplify how some paleontological model systems are well-suited for exploring a potential role of evolvability in macroevolution. Hunt's (2007) study on phenotypic divergence in the ostracode genus *Poseidonamicus* is one example. Analyzing morphological traits from 51 fossil samples spanning a time-interval of about 40 million years, Hunt showed a positive relationship between phenotypic variation within fossil samples and the directions of evolutionary change in different lineages—a relationship that weakened with elapsed time. The ability to study the duration of the effect of trait variation on evolutionary change exemplifies a unique advantage paleontological data bring to the study of evolvability. Another example is Brombacher et al. (2017b), who estimated the phenotypic variances from 75 samples of two fossil lineages of planktonic foraminifera (*Truncorotalia crassaformis* and *Globoconella puncticulata*) across 500,000 years, and concluded that the within-sample covariance generally predicted evolution from one sample to the next. However, major changes in climate caused this prediction to break down. A similar conclusion was reached by Renaud et al. (2006) in their investigation of how variational properties of fossil samples of two closely related rodent species affected their evolution.

Perhaps the best example of a paleontological study system that can connect the concept of evolvability from quantitative genetics to long-term phenotypic evolution is work on the bryozoan genus *Metrarabdotos* (Cheetham et al. 1994). This capitalized on the clonal nature of bryozoans, which permits estimates of the broad-sense **G** matrix (a quantification of the effects of entire genotypes on resemblance between individuals) from variation among genetically identical zooids within a colony. In contrast, the **G** matrix of the Lande equation, also known as the narrow-sense **G** matrix, measures just the part of inheritance that causes offspring to resemble their parents. Evidence suggests that broad-sense and narrow-sense **G** matrices may be similar to each other for morphological traits that can be measured from fossils. Although methodological issues obfuscate the original conclusions on evolutionary tempo and mode within the clade (Voje et al. 2020), this work—along with that of others (Alex et al. 2016, Brombacher et al. 2017b, Hunt 2007, Renaud et al. 2006)—exemplifies how evolutionary quantitative genetics and the concept of evolvability can be operationalized in the fossil record.

3.2 Allometry, Evolvability, and Fossils

Another theoretical framework that connects paleontological data to evolvability is the study of allometry (Gould 1966, Huxley 1932, Lande 1985). Allometry is commonly expressed as a power function in the form of $Y = aX^b$, where X is overall size and Y is the size of a part. Depending on the level of comparison, three conceptually distinct kinds of allometry can be defined: (1) *ontogenetic allometry* characterizes variation among individuals at different growth stages from embryo to adult, (2) *static allometry* characterizes variation among individuals of the same life stages (typically adults), and (3) *evolutionary allometry* characterizes variation across species (Cheverud 1982b). The past decade saw developments in the concepts and tools used to study allometry in the context of quantitative genetics and evolvability (Houle et al. 2011, Pélabon et al. 2013, Voje et al. 2014). Ontogenetic and static allometries are particularly relevant for evolvability because they are summary statistics of two-dimensional **P** when one of the two traits is overall size. Size is often a "line of least evolutionary resistance" (Marroig and Cheverud 2005) and traits are commonly more evolvable in the direction predicted by the allometric relationship compared to other directions, similar to the concept of "genetic lines of least resistance" (e.g., Schluter 1996; Fig. 1B). Furthermore, the direction of trait evolution predicted by the allometric relation is often found to be conserved among taxa (Voje et al. 2014; Fig. 3), suggesting that patterns of developmental and genetic constraints are at play in channeling the evolutionary response in the trait in relation to changes in overall size (Pélabon et al. 2014).

[Figure 3 near here]

The study of allometry has a rich history within paleontology. Gould's (1974) famous study of the antler size of the Irish elk *Megaloceros giganteus* shows two results relevant for evolvability. First, the Irish elk had the predicted antler size of a species of its size from the pattern of evolutionary allometry across 20 extant species of the subfamily Cervinae. Second, within-species static allometry was similar to the among-species evolutionary allometry of antler and body size. Gould concluded that the seemingly extravagant antlers of the Irish elk evolved through heterochronic extrapolation of the common allometry of Cervinae. Work on horse-skull morphology is another example of allometric relationships in paleontology (Radinsky 1984, Robb 1935a, b). The similarity between the slope of ontogenetic and static allometry of the modern horse and evolutionary allometry among fossil horses was interpreted as constraining the morphological divergence in the family Equidae in morphospace (Simpson 1944).

The fossil record provides critical data to examine the hypothesis that morphological evolution is constrained to follow the direction of ontogenetic and static allometry due to low evolvability of the allometric slope ('the allometric-constraint hypothesis', reviewed in Pélabon et al. 2014, Voje et al. 2014). The previously mentioned study by Brombacher et al. (2017a) looked at traits in relation to size and tested the allometric-constraint hypothesis in two lineages of planktonic foraminifera. They estimated the static allometric slope at time step t and evaluated whether this predicted the direction of evolution in the bivariate means at time step $t+1$. Within a constant climatic phase, the static allometric slope predicted the direction of among-population morphological divergence, whereas the static slope failed to do so for the divergence across different climatic phases. This pattern corroborates the idea that allometry serves as a constraint over relatively short time scales, but that allometric slopes evolve and facilitate evolution away from the common allometric trajectory on longer timescales (Houle et al. 2019, Tsuboi et al. 2016, Voje and Hansen 2012, Voje et al. 2014). Within the same conceptual framework, Firmat et al. (2014) revealed a pattern supporting the allometric constraints in dental morphology of rodents, but again over a relatively short time scale (~600,000 years). The traits studied by both Brombacher et al. (2017a) and Firmat et al. (2014) were weakly correlated with size, which means that evolvability is only modestly reduced in directions away from the allometric relationship. It would be interesting to investigate sets of traits in the fossil record that show a stronger association with overall body size in future assessments of the evolvability of allometric slopes and the predictability of allometry on trait evolution.

4. Evolutionary Developmental Paleontology

4.1 Fossil Evo-Devo

Evolvability is at the center of evo-devo (Hendrikse et al. 2007). One key intersection between evo-devo and paleontology concerns morphologies recorded by fossils but not present among extant taxa, including intermediate states in important evolutionary transitions. Proximate developmental processes that underpin major evolutionary transitions have been inferred for an increasing number of examples, such as the mammalian inner ear (Luo 2011, Luo et al. 2015, Urban et al. 2017; Wang et al. 2019; Wang et al. 2021), arthropod segmentation (Chipman and Edgecombe 2019), tetrapod limbs (Stewart et al. 2020), and turtle shells (Lyson and Bever 2020, Schoch and Sues 2020). Insights from these paleo-evo-devo studies provide a richer understanding of how evolutionary innovations arise and of their importance in the history of life (Erwin 2012, Jablonski 2020, Wagner 2014; Urdy et al. 2013). However, cases in which researchers use developmental information to make predictions about the generation of phenotypic variation are most relevant to the topic of evolvability. Sufficient knowledge of developmental processes, coupled with assumptions or evidence that they are conserved from fossil taxa to their modern relatives, can offer an alternative to the quantitative genetic approach for predicting which traits and lineages may be more likely to produce abundant variation for natural selection and other evolutionary processes (Jackson 2020). For example, the structure of some gene regulatory networks may greatly limit the realization of variation in certain body plan traits, leading to their profound stability over time (Davidson and Erwin 2006). Cell-reflecting structures in ostracod carapaces offer another example. These structures allow for cell divisions to be inferred from ontogenetic changes in reticulation (Liebau 1991, Okada 1981), and it has been shown that some divisions in these sequences can be much more variable than others, shaping the variation present in fossil and modern populations (Hunt and Yasuhara 2010). In the remainder of this section, we discuss two trait systems—vertebral counts in amniotes and tooth development in mammals—for which the intersection of evo-devo, evolvability, and paleontological data have been especially productive.

4.1.1 Vertebrae counts in amniotes

The regionalization of the axial skeleton in amniotes has been well studied in terms of variation and evolutionary divergence. Vertebrae are divided into presacral, sacral, and caudal series, with the presacral series further subdivided into cervical, thoracic and lumbar series. It has long been known that the counts of vertebrae in these different series tend to be conserved in mammals but are more variable in reptile groups. Müller et al. (2010) showed that this pattern of variability is ancient: mammals share their conserved variation with basal synapsids, whereas even basal reptile groups show high evolutionary lability in vertebrae counts in different axial regions. Cervical (neck) vertebrae counts, in particular, are nearly invariant among mammals; only manatees and three-toed sloths differ from the canonical mammalian complement of seven (Narita and Kuratani 2005). In contrast, many reptile groups are extremely variable in their vertebrae counts. Sauropterygians (pliosaurs, plesiosaurs, and their relatives) can have anywhere from 6 to 76 neck vertebrae (Soul and Benson 2017), and total vertebral counts in snakes can differ by several hundred across species (Lindell 1994).

Is the rarity of evolutionary changes in vertebral counts, especially in the neck region, caused by constraints (i.e., a low evolvability) of this suite of traits in mammals (Jones et al. 2018)? The simplest variational cause for a lack of evolutionary change is the lack of genetic variation. If mammalian development (almost) always produces axial skeletons with exactly seven neck vertebrae, then this trait would have (near) zero evolvability. Perhaps surprisingly, this seems not to be the case: studies of different mammal species have documented variation in cervical counts (Galis 1999, Galis et al. 2006, ten Broek et al. 2012, Varela-Lasheras et al. 2011). However, these studies also demonstrate that individuals bearing variant numbers of cervical vertebrae almost always bear other skeletal or soft-tissue anomalies, including lethal cancers. Therefore, evolutionary changes are limited not by the absence of variation, but instead by strong genetic correlations between vertebral counts and other traits that dramatically lower organismal fitness. The result is that very little of the variation in vertebral patterning is available for adaptive evolution.

The explanation that cervical vertebral counts are conserved in mammals because of low evolvability has been extended to consider differences in evolvability across other vertebral traits and between different lineages. Thoracic vertebrae variants are also associated with negative developmental anomalies, but the association is weaker than for cervical variants (Galis et al. 2006). Indeed, vertebral counts are less conserved in the thoracic region of mammals (Narita and Kuratani 2005). Some have argued that the two mammal lineages with evolutionary shifts in cervical vertebrae, manatees and sloths, have been able to do so because their relatively low metabolism can reduce the harmful side effects, especially those related to cancers (Varela-

Lasheras et al. 2011). Similarly, the lower incidence of cancer in birds and other reptiles may be related to the greater evolutionary lability of vertebral counts in these groups (Galis 1999), though additional factors can be identified that likely play a role (Varela-Lasheras et al. 2011).

4.1.2 Molar development in mammals

Developmental biologists have extensively explored the mouse as a model system for tooth development, with several decades of work elucidating the gene expression patterns and tissue interactions associated with tooth formation. Given that the fossil record of mammalian teeth is especially rich, there is great potential to marry this archive of tooth form with an accumulating understanding of tooth development.

Generative models of tooth formation have been crucial to making predictions about the evolvability of different dental traits (Jernvall 2000, Ortiz et al. 2018, Polly 1998, Salazar-Ciudad and Jernvall 2010). For example, the Inhibitory Cascade (IC) model makes predictions about the relative size of molars in the tooth row. Kavanagh et al. (2007) developed this model based on the experimental demonstration in mice that the first molar (M1) inhibits the formation of the second (M2), which, in turn, inhibits the third molar (M3). They then postulated a quantitative relationship that captured this behavior, with a parameter that represents the relationship between signal activation and inhibition in the developing tooth precursors. This model predicts that molars can develop on a continuum between equal sized (M1 = M2 = M3) and increasingly M1 dominated (M1 > M2 > M3), depending on the relative strength of activators to inhibitors. Moreover, the model predicts that M2 should always account for ⅓ of total molar size when three molars are present and that the slope between M2/M1 and M3/M1 should be exactly two. The IC model proposes that variation in relative tooth size should therefore be highly structured, with some configurations arising easily and others essentially forbidden (e.g., M2 > M1). Assuming the IC model is strictly true, evolvability should therefore be high in some dimensions and absent in others.

Initial data published with the IC model indicated that relative molar sizes in murine rodents followed its predictions (Kavanagh et al. 2007). Many subsequent studies have applied the IC model to other mammal groups, both extant and fossil. Halliday and Goswami (2013) assessed a large sample that included fossil mammals dating back to the Jurassic and found that molar ratios in most, but not all, taxa were similar to the IC predictions (Fig. 4). Other studies reported on different mammal clades, which yielded results that sometimes comported with IC predictions, but other times did not (Asahara 2013, Evans et al. 2016, Polly 2007, Renvoisé et al. 2009, Wilson et al. 2012). All the above studies looked at predicted (mean) tooth morphologies. As an independent prediction of the IC model, Roseman and Delezene (2019) derived the expected variances and covariances of tooth dimensions and found that these predictions were generally not matched closely by data from primates (see also Vitek et al. 2020).

[Figure 4 near here]

Like all models, the IC is a simplification of reality and therefore should not be expected to fully reproduce patterns in nature. Moreover, the experimental evidence that prompted the development of this model was drawn from a single mouse species. Many of the studies that found patterns at odds with IC predictions postulated that these deviations could result from evolutionary changes in tooth development processes that occurred between the focal clade and mice (reviewed in Roseman and Delezene 2019). This is to be expected because development evolves, and predictions based on an unchanged developmental program will thus decay in usefulness with increasing evolutionary time. It is noteworthy that the IC predictions were at least sometimes supported in taxa tens to hundreds of millions of years diverged from modern mice. Therefore, differences in evolvability due to features captured in developmental models can be quite persistent, perhaps much more so than those based on inferences from quantitative genetic parameters such as the **G** or **P** matrix.

4.2 Integration and modularity

Phenotypes are composed of parts recognized with anatomical names. For example, tetrapod bodies can be divided into forelimbs, hindlimbs, and axial regions, and these may be subdivided further. Morphological parts can be associated or integrated with others because they are specified by common genes, influenced by shared developmental pathways, or work together to achieve a particular function (e.g., locomotion for forelimbs and hindlimbs in many tetrapods) (Klingenberg 2008, Olson and Miller 1958). Empirical studies repeatedly suggest that such associated parts tend to be more correlated with each other than with unassociated parts. For example, individuals with larger than average forelimbs also will have larger than average hindlimbs. Modules refer to groups of traits that are integrated with each other but relatively independent of other sets of traits.

Modularity and integration clearly describe the apportionment of evolvability among traits, but the consequences for evolution depend on the relationship between modularity and selection. If the directions of selection are random over long time periods, phenotypic evolution will occur more rapidly in the directions of modular variation but less rapidly in other directions, leaving the overall rate of evolution unchanged from a non-modular architecture. Alternatively, modularity could be adaptive if mutually correlated traits are frequently selected in a direction consistent with their correlations. In this case, modular architecture will minimize the pleiotropic effects of adaptation on other modules and minimize the costs of adaptation. This enhances the overall rate of evolution (Cheverud 1996, Kirschner and Gerhart 1998, Riedl 1978, Wagner and Altenberg 1996). The claim that modularity is aligned with likely directions of selection is plausible, especially for modules related to function, though to date it is not well assessed by empirical evidence, perhaps due to the scarcity of estimates of natural selection on suites of traits (Melo et al. 2016). Modules are frequently identified on the basis of developmental, anatomical, or functional knowledge of the suite of traits under consideration or are inferred empirically from the patterns of covariation among traits (Goswami and Polly 2010, Klingenberg 2008), rather than from the nature of selective forces. The overall impact of modularity on evolution depends on whether modular architecture changes the overall evolvability. If modularity is achieved by reduction of variability in non-modular directions, it can readily reduce evolvability relative to a less-modular architecture (Hansen 2003).

Paleontologists have assessed patterns of integration and modularity within abundantly preserved species, explored differences in modularity between closely related species (Gerber and Hopkins 2011, Webster and Zelditch 2011a, b), and tracked changes in modularity and integration within lineages (Goswami et al. 2015, Maxwell and Dececchi 2013). In other cases, modularity has been assessed in extant populations and then applied to fossil taxa not normally preserved in high abundance (e.g., Young et al. 2010). All these studies provide important information about the stability of trait variational patterns, as well as about whether evolutionary divergence is shaped by the developmental organization of traits. Other studies address what is sometimes called “evolutionary modularity” (e.g., Felice et al. 2019, Larouche et al. 2018, Parins-Fukuchi 2020). This approach also looks at associations among sets of traits, but the variation examined is between species rather than within species. This body of work is interesting but less easily related to evolvability because, absent information about genetic variation or development within species, it is not possible to determine variational versus selective causes for these patterns.

5. Comparative Methods, Macroevolution, and Paleontology

Phylogenetically-informed macroevolutionary modeling is another area where increased integration with paleontology is possible and being realized. In particular, recent developments in phylogenetic comparative methods (PCMs) have led to an explosion of neontological interest in the study of macroevolutionary processes and patterns (reviewed in Garamszegi 2014, O'Meara 2012, Pennell and Harmon 2013). This is a direct consequence of the development of novel and powerful statistical models of trait and lineage evolution. As a result, there is a growing overlap in the type of research questions that paleontologists and neontologists can ask regarding long-term evolvability (Hunt and Slater 2016): What role does evolvability play in regulating lineage diversity and morphological disparity through time? How can we explain stasis over macroevolutionary timescales? To what extent can evolutionary novelties shape the patterns or rates of diversification? Can shifts in modularity induce changes in the rate of morphological diversification?

One of the benefits of this increased overlap in research interests is the emergence of interdisciplinary approaches, such as the addition of fossil data to molecular phylogenies (e.g., Slater et al. 2012) or the application of comparative methods to trees derived from fossil data (e.g., Mitchell et al. 2019). Another important benefit is the realization that paleontologists and neontologists working in a comparative framework can share a common set of mathematical models. This unification is essential for evolvability research because it provides the field with a robust statistical framework in which to test hypotheses regarding the impact of intrinsic organismal properties on long term evolutionary dynamics.

Quantitative genetic and developmental approaches to evolvability are clearly applicable at short time scales, but genetic and developmental systems evolve over longer time scales, which means that the evolvability of clades may diverge over time (but see Tsuboi et al. 2018). In addition, long term evolvability must encompass not only a lineage's ability to respond to selection, but also its capacity to survive repeated rounds of large-scale changes in its biotic and abiotic environment (Jablonski 2017a). Evolvability research on macroevolutionary time scales is, therefore, necessarily more complex than studies at microevolutionary time scales (Jablonski 2008). This makes it even more challenging to disentangle whether macroevolutionary patterns arise from variational properties of traits or lineages rather than selection or ecological opportunity.

At the macroevolutionary level, the confluence of phylogenetic methods with paleontological data allows researchers to focus on clade-level properties as a rich source of data. We highlight the study of two such properties: (a) morphological disparity and (b) lineage diversification through speciation and extinction rates.

5.1 Disparity

One of the clearest routes for combining neontological and paleontological data in the study of long term evolvability is through analyses of disparity or morphological diversity. Studies of disparity have traditionally been used to test the idea that the exploration of the morphospace is limited by the availability of ecological space (Harmon et al. 2003, Hughes et al. 2013, Yoder et al. 2010). Ecological opportunity would then be the major determinant of the rate of morphological diversification (Rainey and Travisano 1998) and the opening of adaptive zones would help to explain large radiations (Simpson 1944). Increasingly, however, biologists have come to recognize that intrinsic organismal factors might play a role in regulating the occupation of a multivariate morphospace (Wagner 2018). In particular, the pathways followed by a lineage are shaped not only by externally imposed evolutionary processes, but also by variational properties that steer evolution along paths with abundant variation and constrain it away from pathways that lack such variation.

A classic example is observed in mammals following the Cretaceous-Paleogene (K-Pg) extinction (Archibald and Deutschman 2001, Raia et al. 2013, Slater 2013). While several mammalian lineages survived the K-Pg extinction event, they have since followed remarkably different morphological diversification patterns. Placentals have diversified into a large array of forms, encompassing species that are aerial, arboreal, fossorial, aquatic or cursorial with body sizes that vary anywhere from 2 g to 1.5×10^8 g (Wilson and Reeder 2005). Marsupials, on the other hand, have remained far more conservative and display lower disparity than placentals for several skeletal elements, such as the mandible and dentition (Echarri and Prevosti 2015), skull (Bennett and Goswami 2013), shoulder girdle (Sears 2004), and limb bones (Cooper and Steppan 2010). Several authors have argued that the low disparity in skeletal forms among marsupials is a consequence of their altricial reproductive strategy, which requires juvenile marsupials to climb to one of their mothers' teats soon after birth and suckle earlier and for longer than placentals (Lillegraven 1975, Smith 2006). This strategy requires an early ossification of facial and limb skeletal elements (Bininda-Emonds et al. 2007, Sánchez-Villagra 2002) so that the neonate can both climb and suckle properly. Early ossification of skeletal structures, in turn, may limit the range of forms easily generated during marsupial ontogeny, which would allow for more derived skeletal morphologies and thereby limit ecomorphological diversification of the group (Fig. 5).

[Figure 5 near here]

Vermeij hypothesized that the number of dimensions in which the phenotype is capable of varying, which he termed versatility, is correlated with disparity (Vermeij 1973b). He pointed to the possibility that increasing versatility facilitated the evolution of more complex forms and enabled the evolution of key innovations that opened up new adaptive zones. More generally, increasing the dimensionality of possible phenotypes may allow the exploitation of a wider variety of ecological roles within each adaptive zone. Versatility is therefore a higher-level property not tied to any particular phenotype, enlarging the nature of evolvability characteristics that can influence disparity.

One of the main reasons why disparity provides fruitful grounds for integrating paleontological and neontological approaches to evolvability is that the mechanics of phylogenetically-informed disparity analysis is highly similar whether fossil taxa are incorporated or not. The main necessary components are essentially the same: (1) a tree with branches scaled to time units, (2) scores for each taxon in the morphological trait of interest, (3) a model for evolutionary change, and (4) the evolvability hypothesis being tested. However, several initial studies have demonstrated that even incorporating a little fossil information into a phylogenetic backbone can go a long way toward improving parameter estimates of the statistical models used for testing differential evolvability hypotheses (e.g., Pyron and Burbrink 2012) and the ability to differentiate between evolutionary models (e.g., Slater et al. 2012).

5.2 Lineage diversification

Another promising route for integration of neontological and paleontological data in the study of long term evolvability is through analyses of lineage diversification rates caused by extinction and speciation. There are several proposed routes by which organismal or population properties can affect diversification. At the trait level, morphological novelties are capable of reshaping lineage diversification rates, sometimes causing bursts of diversification (Rabosky et al. 2013). The study of such key innovations or trait-dependent diversification has enjoyed a significant revival in modern phylogenetics with the development of trait-dependent speciation and extinction models (FitzJohn 2010, 2012, Goldberg and Igić 2012, Goldberg et al. 2011, Magnuson-Ford and Otto 2012). In principle, such approaches can separately estimate effects of traits on speciation and extinction, using only observations from extant taxa and a phylogeny connecting them. However, these methods perform unreliably under many conditions when their strict assumptions are violated (Rabosky and Goldberg 2015), and more general arguments cast doubt on extant-only data being able to recover historical diversification dynamics (Louca and Pennell 2020). As a result, fossil data may be especially informative for testing whether traits—including those related to evolvability—influence speciation and extinction rates. Indeed, it is reasonable to hypothesize that high evolvability, through enhanced generation of potentially adaptive variants, can protect against extinction. In the fossil record, this survival could be disguised as pseudo-extinction. Evolvability also may plausibly influence the formation of new species as it can enhance responses to natural selection, which can be important under scenarios of ecological speciation (Schluter 2009) or for the survivorship of incipient species as they become established (Allmon and Sampson 2016).

However, tests for such associations between evolvability and speciation or extinction face a complication. The propensity to speciate or go extinct are properties of lineages, not traits (though traits, of course, can influence these probabilities). Differences in variation—and therefore, evolvability—are commonplace among traits. For example, the first principal component in multivariate trait datasets commonly accounts for a majority of the variance, whereas the trailing principal components harbor only trivial amounts of variation. If evolvability itself can evolve (Wagner and Draghi 2010), then this implies at least the possibility of differences in evolvability across lineages, but the frequency and strength of such differences are unclear. Taxonomic differences have been documented for genetic features related to evolvability, such as overall rates of mutation (Lynch 2010) and recombination (Stapley et al. 2017). Developmental or morphological features that have been associated with evolvability differences among clades include growth strategy in regular versus irregular echinoids (Hopkins and Smith 2015), the loosening of allometric relationships (Tsuboi et al. 2018), and the breaking of left-right symmetry in bivalves (Jablonski 2020).

Despite this obstacle, we noted earlier that several paleontological studies tested if extinction was predicted by phenotypic variation (Kolbe et al 2011, Hopkins 2011, Liow 2007). These studies captured variation levels using morphometric analysis of skeletal elements (Kolbe et al. 2011, Hopkins 2011) or through expert, but qualitative, opinions judging certain species to be unusually variable (Liow 2007). For the measured variation to be relevant for extinction, the focal traits must be construed as proxies for overall phenotypic variation, or they must be selectively important enough to influence adaptation and population survival. We do not know of other analyses of variation and extinction, or any studies that perform analogous analyses for origination, though the possibility is discussed in some detail by Jablonski (2020). Vermeij's (1973a) claim that taxa with higher versatility tend to preferentially replace those of lower versatility implies that this evolvability-related characteristic should increase speciation or decrease extinction, and possibly both together.

6. Conclusions and Prospects: Evolvability in the Fossil Record

Although evolvability is not commonly invoked by many paleontologists, the examples we have presented provide ample warrant for paleontologists to actively incorporate the concept of evolvability into investigations of the fossil record. Our enthusiasm should, however, be tempered with caution; disentangling the role of evolvability and the forces that cause evolution is always challenging. With these two points in mind, we offer a methodological schema for paleontologists and their interdisciplinary collaborators to initiate investigations of evolvability. To enable such a study, three distinct aspects must be addressed a priori: conceptualization (what counts as evolvability?); measurement (how is evolvability empirically measured, directly or indirectly?); and testing (what strategies are used to evaluate hypotheses about evolvability?).

The conceptualization of evolvability differs across the approaches canvassed herein. Classic studies focused on how variational properties could influence the propensity for evolutionary stability versus change, within the span of individual lineages, and more broadly over the history of life. Vermeij (1973a, b) used the potential versatility of form, whereas others have concentrated on morphological disparity (e.g., Hughes et al. 2013). Within quantitative genetics, evolvability is conceptualized specifically as the short-term evolutionary potential for a population to evolve in the direction of selection (Hansen and Houle 2008). Within evo-devo, evolvability is conceptualized in terms of how developmental processes can structure trait variation that is subject to selection and drift. Within phylogenetic macroevolutionary modeling, evolvability is analyzed retrospectively and is based on different measures of the evolutionary success of lineages (e.g., disparity or rates of speciation and extinction). Although interesting theoretical relationships might obtain between these different meanings, a clearly specified conceptualization is crucial for initiating an inquiry.

Once evolvability is conceptualized in a particular way, the next issue is to address methodologically is how evolvability will be measured. For instance, Vermeij operationalized "potential versatility of form" as the number of parameters required to describe coiling in gastropods. In quantitative genetics, short-term evolvability is defined as the mean-standardized additive genetic variance. In evo-devo studies, developmental considerations may form the basis for statements of relative evolvability (e.g., cervical vertebrae should be less evolvable than thoracic vertebrae in mammals), but these may not readily predict magnitudes of such effects. The specific type of measurement utilized will limit the kinds of evolutionary inference that might be drawn in distinctive ways that must be explicitly appreciated.

Finally, with evolvability operationalized, the final step in the methodological sequence is testing. Armed with measures or predictions of relative evolvability across traits or lineages, researchers can then measure evolutionary differences to assess the degree to which divergences match evolvability predictions. Are changes larger in more evolvable traits or lineages? Are low-evolvability traits or directions in morphospace especially conserved? Do lineages whose characteristics indicate high evolvability attain greater morphological disparity or experience greater diversification? To the extent that evolvability predictions are upheld, the argument is supported that variation—and not just selection—is important in shaping long-term evolutionary changes. When evolvability does not predict empirical divergences, it is likely that natural selection has been the dominant process determining the evolutionary outcomes. Compilation of numerous such tests can help to establish the relative importance of these two components of evolution for different kinds of traits, in different lineages, and over different time scales.

This three-step methodological schema offers a general template for approaching a variety of outstanding questions on evolvability in the fossil record (Box 1). These include major theoretical questions such as the relative importance of variation versus selection in the history of life, as well as more granular issues such as the timescales over which variational patterns, and thus evolvability predictions, are stable. Attention to the specifics of the schema provides detailed guidance for novel empirical and theoretical studies of evolvability in the fossil record. Paleontology, with its unique access to temporal data from the history of life, is positioned to make distinctive contributions to studies of evolvability and especially in interdisciplinary collaborations with other approaches in evolutionary biology.

LITERATURE CITED

Ackermann, R. R., and J. M. Cheverud. 2004. Detecting genetic drift versus selection in human evolution. Proceedings of the National Academy of Sciences USA 101:17946–17951.

Alex, H., M. Diogo, and M. Gabriel. 2016. A case study of extant and extinct Xenarthra cranium covariance structure: implications and applications to paleontology. Paleobiology 42:465–488.

Allmon, W. D., and R. M. Ross. 1990. Specifying causal factors in evolution: the paleontological contribution. Pp. 1–17. *In* R. M. Ross, and W. D. Allmon, eds. Causes of Evolution: A Paleontological Perspective. University of Chicago Press, Chicago.

Allmon, W. D., and S. D. Sampson. 2016. The stages of speciation: a stepwise framework for analysis of speciation in the fossil record. Pp. 121–167. *In* W. D. Allmon, and M. M. Yacobucci, eds. Species and Speciation in the Fossil Record. University of Chicago Press, Chicago.

Archibald, J. D., and D. H. Deutschman. 2001. Quantitative analysis of the timing of the origin and diversification of extant placental orders. Journal of Mammalian Evolution 8:107–124.

Asahara, M. 2013. Unique inhibitory cascade pattern of molars in canids contributing to their potential to evolutionary plasticity of diet. Ecology and Evolution 3:278–285.

Atchley, W. R., J. J. Rutledge, and D. E. Cowley. 1981. Genetic components of size and shape. II. Multivariate covariance patterns in the rat and mouse skull. Evolution 35:1037–1055.

Baab, K. L. 2018. Evolvability and craniofacial diversification in genus *Homo*. Evolution 72:2781–2791.

Bégin, M., and D. A. Roff. 2004. From micro- to macroevolution through quantitative genetic variation: positive evidence from field crickets. Evolution 58:2287–2304.

Bell, M. A. 2014. Patterns in palaeontology: trends of body-size evolution in the fossil record - a growing field. Palaeontology Online 4:1–9.

Bennett, C. V., and A. Goswami. 2013. Statistical support for the hypothesis of developmental constraint in marsupial skull evolution. BMC Biology 1:52.

Bininda-Emonds, O. R. P., J. E. Jeffery, M. R. Sánchez-Villagra, J. Hanken, M. Colbert, C. Pieau, L. Selwood, C. Ten Cate, A. Raynaud, C. Osabutey, and M. K. Richardson. 2007. Forelimb-hindlimb developmental timing across tetrapods. BMC Evolutionary Biology 7:182.

Blake, D. B. 1980. Homeomorphy in Paleozoic bryozoans: A search for explanations. Paleobiology 6:451–465.

Blows, M. W., and M. Higgie. 2003. Genetic constraints on the evolution of mate recognition under natural selection. American Naturalist 161:240–253.

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

Brombacher, A., P. A. Wilson, I. Bailey, and T. H. G. Ezard. 2017a. The breakdown of static and evolutionary allometries during climatic upheaval. The American Naturalist 190:350–362.

Brombacher, A., P. A. Wilson, and T. H. G. Ezard. 2017b. Calibration of the repeatability of foraminiferal test size and shape measures with recommendations for future use. Marine Micropaleontology 133:21–27.

Charlesworth, B., R. Lande, and M. Slatkin. 1982. A neo-Darwinian commentary on macroevolution. Evolution 36:474–498.

Charmantier, A., D. Garant, and L. E. B. Kruuk, eds. 2014. Quantitative Genetics in the Wild. Oxford University Press, Oxford

Cheetham, A. H., J. B. C. Jackson, and L.-A. C. Hayek. 1994. Quantitative genetics of bryozoan phenotypic evolution. II. Analysis of selection and random change in fossil species using reconstructed genetic parameters. Evolution 48:360–375.

Cheverud, J. M. 1982a. Phenotypic, genetic, and environmental morphological integration in the cranium. Evolution 36:499–516.

Cheverud, J. M. 1982b. Relationships among ontogenetic, static, and evolutionary allometry. American Journal of Physical Anthropology 59:139–149.

Cheverud, J. M. 1984. Quantitative genetics and developmental constraints on evolution by natural selection. Journal of Theoretical Biology 110:155–171.

Cheverud, J. M. 1988. A comparison of genetic and phenotypic correlations. Evolution 42:958–968.

Cheverud, J. M. 1996. Developmental integration and the evolution of pleiotropy. American Zoologist 36:44–50.

Chipman, A. D., and G. D. Edgecombe. 2019. Developing an integrated understanding of the evolution of arthropod segmentation using fossils and evo-devo. Proceedings of the Royal Society B: Biological Sciences 286:20191881.

Colegrave, N., and S. Collins. 2008. Experimental evolution: experimental evolution and evolvability. Heredity 100:464–470.

Cooper, W. J., and S. J. Steppan. 2010. Developmental constraint on the evolution of marsupial forelimb morphology. Australian Journal of Zoology 58:1–15.

Crother, B. I., and C. M. Murray. 2019. Early usage and meaning of evolvability. Ecology and Evolution 9:3784–3793.

Davidson, E. H., and D. H. Erwin. 2006. Gene regulatory networks and the evolution of animal body plans. Science 311:796–800.

de Oliveira, F. B., A. Porto, and G. Marroig. 2009. Covariance structure in the skull of Catarrhini: a case of pattern stasis and magnitude evolution. Journal of Human Evolution 56:417–430.

Dilcher, D. 2000. Toward a new synthesis: major evolutionary trends in the angiosperm fossil record. Proceedings of the National Academy of Sciences USA 97:7030–7036.

Dochtermann, N. A. 2011. Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. Evolution 65:1814–1820.

Echarri, S., and F. J. Prevosti. 2015. Differences in mandibular disparity between extant and extinct species of metatherian and placental carnivore clades. Lethaia 48:196–204.

Eldredge, N., and S. J. Gould. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Pp. 82-115. *In* T. J. M. Schopf, ed. Models in Paleobiology. Freeman, Cooper & Co., San Francisco.

Erwin, D. H. 1994. Major morphologic innovations. Acta Palaeontologica Polonica 38:281–294.

Erwin, D. H. 2007. Disparity: morphological pattern and developmental context. Palaeontology 50:57–73.

Erwin, D. H. 2012. Novelties that change carrying capacity. Journal of Experimental Zoology B (Mol Dev Evo) 318:460–465.

Erwin, D. H. 2019. Prospects for a general theory of evolutionary novelty. Journal of Computational Biology 26:735–744.

Erwin, D. H., and J. W. Valentine. 2013. The Cambrian Explosion: The Construction of Animal Biodiversity. Roberts, Greenwood, CO.

Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. The American Naturalist 169:227–244.

Evans, A. R., E. S. Daly, K. K. Catlett, K. S. Paul, S. J. King, M. M. Skinner, H. P. Nesse, J.-J. Hublin, G. C. Townsend, G. T. Schwartz, and J. Jernvall. 2016. A simple rule governs the evolution and development of hominin tooth size. Nature 530:477–480.

Felice, R. N., A. Watanabe, A. R. Cuff, E. Noirault, D. Pol, L. M. Witmer, M. A. Norell, P. M. O'Connor, and A. Goswami. 2019. Evolutionary integration and modularity in the archosaur cranium. Integrative and Comparative Biology 59:371–382.

Firmat, C., I. Lozano-Fernandez, J. Agusti, G. H. Bolstad, G. Cuenca-Bescos, T. F. Hansen, and C. Pélabon. 2014. Walk the line: 600000 years of molar evolution constrained by allometry in the fossil rodent *Mimomys savini*. Philosophical Transactions of the Royal Society B: Biological Sciences 369:20140057.

FitzJohn, R. G. 2010. Quantitative traits and diversification. Systematic Biology 59:619–633.

FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3:1084–1092.

Foote, M. 1997. The evolution of morphological diversity. Annual Review of Ecology and Systematics 28:129–152.

Galis, F. 1999. Why do almost all mammals have seven cervical vertebrae? Developmental constraints, Hox genes, and cancer. Journal of Experimental Zoology 285:19–26.

Galis, F., T. J. M. V. Dooren, J. D. Feuth, J. A. J. Metz, A. Witkam, S. Ruinard, M. J. Steigenga, and L. C. D. Wunaendts. 2006. Extreme selection in humans against homeotic transformations of cervical vertebrae. Evolution 60:2643–2654.

Garamszegi, L. Z. 2014. Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology: Concepts and Practice. Springer, Heidelberg.

Gerber, S., and M. J. Hopkins. 2011. Mosaic heterochrony and evolutionary modularity: the trilobite genus *Zacanthopsis* as a case study. Evolution 65:3241–52.

Goldberg, E. E., and B. Igić. 2012. Tempo and mode in plant breeding system evolution. Evolution 66:3701–3709.

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Systematic Biology 60:451–465.

Goswami, A., W. J. Binder, J. Meachen, and F. R. O’Keefe. 2015. The fossil record of phenotypic integration and modularity: A deep-time perspective on developmental and evolutionary dynamics. Proceedings of the National Academy of Sciences USA 112:4891–4896.

Goswami, A., and P. D. Polly. 2010. The influence of modularity on cranial morphological disparity in Carnivora and Primates (Mammalia). PLoS ONE 5:e9517.

Gould, S. J. 1966. Allometry and size in ontogeny and phylogeny. Biological Reviews of the Cambridge Philosophical Society 41:587–640.

Gould, S. J. 1974. The origin and function of ‘bizarre’ structures: antler size and skull size in the Irish Elk, *Megaloceros giganteus*. Evolution 28:191–220.

Gould, S. J. 1989. A developmental constraint in *Cerion*, with comments on the definition and interpretation of constraint in evolution. Evolution 43:516–539.

Gould, S. J. 2002. The Structure of Evolutionary Theory. Harvard University Press, Cambridge, MA.

Grabowski, M., J. D. Polk, and C. C. Roseman. 2011. Divergent patterns of integration and reduced constraint in the human hip and the origins of bipedalism. Evolution 65:1336–1356.

Grabowski, M., and A. Porto. 2017. How many more? Sample size determination in studies of morphological integration and evolvability. Methods in Ecology and Evolution 8:592–603.

Grabowski, M., and C. C. Roseman. 2015. Complex and changing patterns of natural selection explain the evolution of the human hip. Journal of Human Evolution 85:94–110.

Hadfield, J. D., A. Nutall, D. Osorio, and I. P. F. Owens. 2007. Testing the phenotypic gambit: phenotypic, genetic and environmental correlations of colour. Journal of Evolutionary Biology 20:549–557.

Halliday, T. J. D., and A. Goswami. 2013. Testing the inhibitory cascade model in Mesozoic and Cenozoic mammaliaforms. BMC Evolutionary Biology 13:79.

Hansen, T. F. 2003. Is modularity necessary for evolvability?: Remarks on the relationship between pleiotropy and evolvability. Biosystems 69:83–94.

Hansen, T. F., W. S. Armbruster, M. L. Carlson, and C. Pélabon. 2003a. Evolvability and genetic constraint in *Dalechampia* blossoms: genetic correlations and conditional evolvability. Journal of Experimental Zoology B (Mol Dev Evo) 296:23–39.

Hansen, T. F., and D. Houle. 2004. Evolvability, stabilizing selection, and the problem of stasis. Pp. 130–50. *In* M. Pigliucci, and K. Preston, eds. Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes. Oxford University Press, Oxford.

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

Hansen, T. F., C. Pélabon, W. S. Armbruster, and M. L. Carlson. 2003b. Evolvability and genetic constraint in *Dalechampia* blossoms: components of variance and measures of evolvability. Journal of Evolutionary Biology 16:754–766.

Hansen, T. F., and K. L. Voje. 2011. Deviation from the line of least resistance does not exclude genetic constraints: a comment on Berner et al. (2010). Evolution 65:1821–1822.

Harmon, L. J., J. A. Schulte, A. Larson, and J. B. Losos. 2003. Tempo and mode of evolutionary radiation in Iguanian lizards. Science 301:961–964.

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

Hopkins, M. J. 2011. How species longevity, intraspecific morphological variation, and geographic range size are related: a comparison using late Cambrian trilobites. Evolution 65:3252–3273.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. Genetics 130:195–204.

Houle, D., G. H. Bolstad, K. van der Linde, and T. F. Hansen. 2017. Mutation predicts 40 million years of fly wing evolution. Nature 548:447–450.

Houle, D., L. T. Jones, R. Fortune, and J. L. Sztepanacz. 2019. Why does allometry evolve so slowly? Integrative and Comparative Biology 59:1429–1440.

Houle, D., C. Pélabon, G. P. Wagner, and T. F. Hansen. 2011. Measurement and meaning in biology. The Quarterly Review of Biology 86:3–34.

Hughes, M., S. Gerber, and M. A. Wills. 2013. Clades reach highest morphological disparity early in their evolution. Proceedings of the National Academy of Sciences USA 110:13875–13879.

Hunt, G. 2004. Phenotypic variation in fossil samples: modeling the consequences of time-averaging. Paleobiology 30:426–443.

Hunt, G. 2007. Evolutionary divergence in directions of high phenotypic variance in the ostracode genus *Poseidonamicus*. Evolution 61:1560–1576.

Hunt, G., and D. L. Rabosky. 2014. Phenotypic evolution in fossil species: pattern and process. Annual Review of Earth and Planetary Sciences 42:421–441.

Hunt, G., and G. Slater. 2016. Integrating paleontological and phylogenetic approaches to macroevolution. Annual Review of Ecology, Evolution, and Systematics 47:189–213.

Hunt, G., and M. Yasuhara. 2010. A fossil record of developmental events: variation and evolution in epidermal cell divisions in ostracodes. Evolution & Development 12:635–646.

Huxley, J. S. 1932. Problems of Relative Growth. Methuen and Company, Limited, Essex.

Jablonski, D. 2008. Species selection: theory and data. Annual Review of Ecology, Evolution, and Systematics 39:501–524.

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. 2020. Developmental bias, macroevolution, and the fossil record. Evolution & Development 22:103–125.

Jablonski, D., and N. H. Shubin. 2015. The future of the fossil record: paleontology in the 21st century. Proceedings of the National Academy of Sciences USA 112:4852–4858.

Jackson, I. S. C. 2020. Developmental bias in the fossil record. Evolution & Development 22:88–102.

Jernvall, J. 2000. Linking development with generation of novelty in mammalian teeth. Proceedings of the National Academy of Sciences USA 97:2641–2645.

Jones, K. E., L. Benitez, K. D. Angielczyk, and S. E. Pierce. 2018. Adaptation and constraint in the evolution of the mammalian backbone. BMC Evolutionary Biology 18:172.

Kavanagh, K. D., A. R. Evans, and J. Jernvall. 2007. Predicting evolutionary patterns of mammalian teeth from development. Nature 449:427–432.

Kirschner, M., and J. Gerhart. 1998. Evolvability. Proceedings of the National Academy of Sciences USA 95:8420–8427.

Klingenberg, C. P. 2008. Morphological integration and developmental modularity. Annual Review of Ecology Evolution and Systematics 39:115–132.

Kohn, L. A. P., and W. R. Atchley. 1988. How similar are genetic correlation structures data from mice and rats. Evolution 42:467–481.

Kolbe, E. S., J. J. Zambito, E. C. Brett, J. Wise, and D. R. Wilson. 2011. Brachiopod shell discoloration as an indicator of taphonomic alteration in the deep-time fossil record. Palaios 26:682–692.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

Lande, R. 1979. Quantitative qenetic analysis of multivariate evolution applied to brain-body size allometry. Evolution 33:402–416.

Lande, R. 1985. Genetic and evolutionary aspects of allometry. Pp. 21-32. *In* W. L. Jungers, ed. Size and Scaling in Primate Biology. New York.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

Larouche, O., M. L. Zelditch, and R. Cloutier. 2018. Modularity promotes morphological divergence in ray-finned fishes. Scientific Reports 8:7278.

Liebau, A. 1991. Skulptur-Evolution bei Ostrakoden. Geologie und Paläontologie in Westfalen.

Lillegraven, J. A. 1975. Biological considerations of the marsupial-placental dichotomy. Evolution 29:707–722.

Lindell, L. E. 1994. The evolution of vertebral number and body size in snakes. Functional Ecology 8:708–719.

Liow, L. H. 2007. Does versatility as measured by geographic range, bathymetric range and morphological variability contribute to taxon longevity? Global Ecology and Biogeography 16:117–128.

Lofsvold, D. 1986. Quantitative genetics of morphological differentiation in *Peromyscus*. I. Tests of the homogeneity of genetic covariance structure among species and subspecies. Evolution 40:559–573.

Louca, S., and M. W. Pennell. 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature 580:502–505.

Luo, Z.-X. 2011. Developmental patterns in mesozoic evolution of mammal ears. Annual Review of Ecology, Evolution, and Systematics 42:355–380.

Luo, Z.-X., Q.-J. Meng, Q. Ji, D. Liu, Y.-G. Zhang, and A. I. Neander. 2015. Evolutionary development in basal mammaliaforms as revealed by a docodontan. Science 347:760–764.

Lynch, M. 2010. Evolution of the mutation rate. Trends in Genetics 26:345–352.

Lynch, M., and B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates, Inc., Sunderland, MA.

Lyson, T. R., and G. S. Bever. 2020. Origin and evolution of the turtle body plan. Annual Review of Ecology, Evolution, and Systematics 51:143–166.

Magnuson-Ford, K., and S. P. Otto. 2012. Linking the investigations of character evolution and species diversification. The American Naturalist 180:225–245.

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

Marroig, G., and J. M. Cheverud. 2010. Size as a line of least resistance II: direct selection on size or correlated response due to constraints? Evolution 64:1470–1488.

Martínez-Abadías, N., M. Esparza, T. Sjøvold, R. González-José, M. Santos, M. Hernández, and C. P. Klingenberg. 2012. Pervasive genetic integration directs the evolution of human skull shape. Evolution 66:1010–1023.

Maxwell, E. E., and T. A. Dececchi. 2013. Ontogenetic and stratigraphic influence on observed phenotypic integration in the limb skeleton of a fossil tetrapod. Paleobiology 39:123–134.

Maynard Smith, J., R. Burian, S. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, and L. Wolpert. 1985. Developmental constraints and evolution. Quarterly Review of Biology 60:265–287.

McGlothlin, J. W., M. E. Kobiela, H. V. Wright, D. L. Mahler, J. J. Kolbe, J. B. Losos, and E. D. Brodie III. 2018. Adaptive radiation along a deeply conserved genetic line of least resistance in Anolis lizards. Evolution Letters 2:310–322.

Melo, D., A. Porto, J. M. Cheverud, and G. Marroig. 2016. Modularity: genes, development, and evolution. Annual Review of Ecology, Evolution, and Systematics 47:463–486.

Meredith, R. W., J. E. Janečka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, E. Eizirik, T. L. L. Simão, T. Stadler, D. L. Rabosky, R. L. Honeycutt, J. J. Flynn, C. M. Ingram, C. Steiner, T. L. Williams, T. J. Robinson, A. Burk-Herrick, M. Westerman, N. A. Ayoub, M. S. Springer, and W. J. Murphy. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334:521–524.

Mitchell, J. S., R. S. Etienne and D. L. Rabosky. 2019. Inferring diversification rate variation from phylogenies with fossils. Systematic Biology 68:1–18.

Müller, J., T. M. Scheyer, J. J. Head, P. M. Barrett, I. Werneburg, P. G. P. Ericson, D. Pol, and M. R. Sánchez-Villagra. 2010. Homeotic effects, somitogenesis and the evolution of vertebral numbers in recent and fossil amniotes. Proceedings of the National Academy of Sciences USA 107:2118–2123.

Narita, Y., and S. Kuratani. 2005. Evolution of the vertebral formulae in mammals: a perspective on developmental constraints. Journal of Experimental Zoology B (Mol Dev Evo) 304:91–106.

O'Meara, B. C. 2012. Evolutionary inferences from phylogenies: a review of methods. Annual Review of Ecology, Evolution, and Systematics 43:267–285.

Okada, Y. 1981. Development of cell arrangement in ostracod carapaces. Paleobiology 7:276–280.

Olson, E. C., and R. L. Miller. 1958. Morphological Integration. University of Chicago Press, Chicago.

Ortiz, A., S. E. Bailey, G. T. Schwartz, J.-J. Hublin, and M. M. Skinner. 2018. Evo-devo models of tooth development and the origin of hominoid molar diversity. Science Advances 4:eaar2334.

Parins-Fukuchi, C. 2020. Mosaic evolution, preadaptation, and the evolution of evolvability in apes. Evolution 74:297–310.

Payne, J. L., and A. Wagner. 2019. The causes of evolvability and their evolution. Nature Reviews Genetics 20:24–38.

Pélabon, C., G. H. Bolstad, C. K. Egset, J. M. Cheverud, M. Pavlicev, and G. Rosenqvist. 2013. On the relationship between ontogenetic and static allometry. The American Naturalist 181:195–212.

Pélabon, C., C. Firmat, G. H. Bolstad, K. L. Voje, D. Houle, J. Cassara, A. Rouzic, and T. F. Hansen. 2014. Evolution of morphological allometry. Annals of the New York Academy of Sciences 1320:58–75.

Pennell, M. W. and L. J. Harmon. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. Annals of the New York Academy of Sciences 1289:90–105.

Pigliucci, M. 2006. Genetic variance–covariance matrices: a critique of the evolutionary quantitative genetics research program. Biology & Philosophy 21:1–23.

Polly, P. D. 1998. Variability in mammalian dentitions: size-related bias in the coefficient of variation. Biological Journal of the Linnean Society 64:83–99.

Polly, P. D. 2007. Evolutionary biology: development with a bite. Nature 449:413–415.

Porto, A., F. B. de Oliveira, L. T. Shirai, V. De Conto, and G. Marroig. 2009. The evolution of modularity in the mammalian skull I: morphological integration patterns and magnitudes. Evolutionary Biology 36:118–135.

Pyron, R. A., and F. T. Burbrink. 2012. Extinction, ecological opportunity, and the origins of global snake diversity. Evolution 66:163–178.

Rabosky, D. L. 2012. Positive correlation between diversification rates and phenotypic evolvability can mimic punctuated equilibrium on molecular phylogenies. Evolution 66:2622–2627.

Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Systematic Biology 64:340–355.

Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature Communications 4:1958.

Radinsky, L. 1984. Ontogeny and phylogeny in horse skull evolution. Evolution 38:1–15.

Raia, P., F. Carotenuto, F. Passaro, P. Piras, D. Fulgione, L. Werdelin, S. J., and F. M. 2013. Rapid action in the Palaeogene, the relationship between phenotypic and taxonomic diversification in Cenozoic mammals. Proceedings of the Royal Society of London B: Biological Sciences 280:20122244.

Rainey, P. B., and M. Travisano. 1998. Adaptive radiation in a heterogeneous environment. Nature 394:69–72.

Raup, D. M. 1967. Geometric analysis of shell coiling: coiling in ammonoids. Journal of Paleontology 41:43–65.

Renaud, S., J.-C. Auffray, and J. Michaux. 2006a. Conserved phenotypic variation patterns, evolution along lines of least resistance, and departure due to selection in fossil rodents. Evolution 60:1701–1717.

Renvoisé, E., A. R. Evans, A. Jebrane, C. Labruère, R. Laffont, and S. Montuire. 2009. Evolution of mammal tooth patterns: new insights from a developmental prediction model. Evolution 63:1327–1340.

Riedl, R. 1978. Order in Living Organisms: A Systems Analysis of Evolution. John Wiley & Sons, Ltd, New York.

Robb, R. C. 1935a. A study of mutations in evolution. I. Evolution in the equine skull. Journal of Genetics 31:39–46.

Robb, R. C. 1935b. A study of mutations in evolution. II. Ontogeny in the equine skull. Journal of Genetics 31:47–52.

Roff, D. A. 1995. The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture. Heredity 74:481–490.

Roff, D. A. 1996. The evolution of genetic correlations: an analysis of patterns. Evolution 50:1392–1403.

Roff, D. A. 1997. Evolutionary Quantiative Genetics. Chapman & Hall, New York.

Rolian, C. 2009. Integration and evolvability in primate hands and feet. Evolutionary Biology 36:100–117.

Rosa, D. 1899. La Riduzione progressiva della variabilità e i suoi rapporti coll'estinzione e coll'origine delle specie. Clausen, Turin.

Roseman, C. C., and L. K. Delezene. 2019. The inhibitory cascade model is not a good predictor of molar size covariation. Evolutionary Biology 46:229–238.

Salazar-Ciudad, I., and J. Jernvall. 2010. A computational model of teeth and the developmental origins of morphological variation. Nature 464:583–586.

Sánchez-Villagra, M. R. 2002. Comparative patterns of postcranial ontogeny in therian mammals: an analysis of relative timing of ossification events. Journal of Experimental Zoology (Mol Dev Evol) 294:264–273.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766–1774.

Schluter, D. 2009. Evidence for ecological speciation and Its alternative. Science 323:737–741.

Schoch, R. R., and H.-D. Sues. 2020. The origin of the turtle body plan: evidence from fossils and embryos. Palaeontology 63:375–393.

Sears, K. E. 2004. The role of constraints in the morphological evolution of marsupial shoulder girdles: evidence from comparative anatomy, paleontology, and embryology. Evolution 58:2353–2370.

Simons, A. M., and D. A. Roff. 1996. The effect of a variable environment on the genetic correlation structure in a field cricket. Evolution 50:267–275.

Simpson, G. G. 1944. Tempo and Mode in Evolution. Columbia University Press, New York.

Slater, G. 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. Methods in Ecology and Evolution 4:734–744.

Slater, G. J., L. J. Harmon, and M. E. Alfaro. 2012. Integrating fossils with molcular phylogenies improves inference of trait evolution. Evolution 66:3931–3944.

Smith, K. K. 2006. Craniofacial development in marsupial mammals: developmental origins of evolutionary change. Developmental Dynamics 235:1181–1193.

Sodini, S. M., K. E. Kemper, N. R. Wray, and M. Trzaskowski. 2018. Comparison of genotypic and phenotypic correlations: Cheverud's conjecture in humans. Genetics 209:941–948.

Soul, L. C., and R. B. J. Benson. 2017. Developmental mechanisms of macroevolutionary change in the tetrapod axis: a case study of Sauropterygia. Evolution 71:1164–1177.

Spiekman, S. N. F., and I. Werneburg. 2017. Patterns in the bony skull development of marsupials: high variation in onset of ossification and conserved regions of bone contact. Scientific Reports 7:43197.

Stapley, J., P. G. D. Feulner, S. E. Johnston, A. W. Santure, and C. M. Smadja. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philosophical Transactions of the Royal Society B: Biological Sciences 372:20160455.

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

Stewart, T. A., J. B. Lemberg, N. K. Taft, I. Yoo, E. B. Daeschler, and N. H. Shubin. 2020. Fin ray patterns at the fin-to-limb transition. Proceedings of the National Academy of Sciences USA 117:1612–1620.

ten Broek, C. M. A., A. J. Bakker, I. Varela-Lasheras, M. Bugiani, S. Van Dongen, and F. Galis. 2012. Evo-Devo of the human vertebral Column: on homeotic transformations, pathologies and prenatal selection. Evolutionary Biology 39:456–471.

Tiozzo, S., and R. Copley. 2015. Reconsidering regeneration in metazoans: an evo-devo approach. Frontiers in Ecology and Evolution 3:67.

Tsuboi, M., A. Kotrschal, A. Hayward, S. D. Buechel, J. Zidar, H. Lovlie, and N. Kolm. 2016. Evolution of brain-body allometry in Lake Tanganyika cichlids. Evolution 70:1559–1568.

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

Urban, D. J., N. Anthwal, Z.-X. Luo, J. A. Maier, A. Sadier, A. S. Tucker, and K. E. Sears. 2017. A new developmental mechanism for the separation of the mammalian middle ear ossicles from the jaw. Proceedings of the Royal Society B: Biological Sciences 284:20162416.

Urdy, S., L. A. B. Wilson, J. T. Haug and M. R. Sánchez-Villagra. 2013. On the unique perspective of paleontology in the study of developmental evolution and biases. Biological Theory 8:293–311.

Valentine, J. W. 1995. Why no new phyla after the Cambrian? Genome and ecospace hypotheses revisited. Palaios 10:190–194.

Van Valen, L. 1973. A new evolutionary law. Evolutionary Theory 1:1–30.

Varela-Lasheras, I., B. A.J., v. d. M. S.D., J. A. J. Metz, V. A. J., and F. Galis. 2011. Breaking evolutionary and pleiotropic constraints in mammals. On sloths, manatees and homeotic mutations. EvoDevo 2:11.

Vermeij, G. J. 1973a. Adaptation, versatility, and evolution. Systematic Zoology 22:466–477.

Vermeij, G. J. 1973b. Biological versatility and Earth history. Proceedings of the National Academy of Sciences USA 70:1936–1938.

Vermeij, G. J. 2015. Forbidden phenotypes and the limits of evolution. Interface Focus 5:20150028.

Vitek, N. S., C. C. Roseman, and J. I. Bloch. 2020. Mammal molar size ratios and the inhibitory cascade at the intraspecific scale. Integrative Organismal Biology 2(1).

Voje, K. L., and T. F. Hansen. 2012. Evolution of static allometries: adaptive change in allometric slopes of eye span in stalk-eyed flies. Evolution 67:453–467.

Voje, K. L., T. F. Hansen, C. K. Egset, G. H. Bolstad, and C. Pélabon. 2014. Allometric constraints and the evolution of allometry. Evolution 68:866–885.

Voje, K. L., E. D. Martino, and A. Porto. 2020. Revisiting a landmark study-system: no evidence for a punctuated mode of evolution in *Metrarabdotos*. The American Naturalist 195:899–917.

Wagner, G. P. 2014. Homology, Genes, and Evolutionary Innovation. Princeton University Press, Princeton, NJ.

Wagner, G. P., and L. Altenberg. 1996. Perspective: complex adaptations and the evolution of evolvability. Evolution 50:967–976.

Wagner, G. P., and J. Draghi. 2010. Evolution of evolvability. Pp. 379-400. *In* M. Pigliucci, and G. B. Müller, eds. Evolution — The Extended Synthesis. MIT Press, Cambridge.

Wagner, P. J. 2018. Early bursts of disparity and the reorganization of character integration. Proceedings of the Royal Society B: Biological Sciences 285:20181604.

Walsh, B., and M. W. Blows. 2009. Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. Annual Review of Ecology, Evolution, and Systematics 40:41–59.

Wang, H., J. Meng and Y. Wang. 2019. Cretaceous fossil reveals a new pattern in mammalian middle ear evolution. Nature 576:102–105.

Wang, J., J. R. Wible, B. Guo, S. L. Shelley, H. Hu, and S. Bi. 2021. A monotreme-like auditory apparatus in a Middle Jurassic haramiyidan. Nature 590:279–283.

Webster, M. 2007. A Cambrian peak in morphological variation within trilobite species. Science 317:499–502.

Webster, M. 2019. Morphological homeostasis in the fossil record. Seminars in Cell & Developmental Biology 88:91–104.

Webster, M., and M. L. Zelditch. 2011a. Evolutionary lability of integration in Cambrian ptychoparioid trilobites. Evolutionary Biology 38:144–162.

Webster, M., and M. L. Zelditch. 2011b. Modularity of a Cambrian ptychoparioid trilobite cranidium. Evolution & Development 13:96–109.

Wilson, D., and D. Reeder. 2005. Preface and introductory material. P. xxvi. *In* D. Wilson, and D. Reeder, eds. Mammal Species of the World: A Taxonomic and Geographic Reference. Johns Hopkins University Press, Baltimore.

Wilson, L. A. B., M. R. Sánchez-Villagra, R. H. Madden, and R. F. Kay. 2012. Testing a developmental model in the fossil record: molar proportions in South American ungulates. Paleobiology 38:308–321.

Wright, D. F. 2017. Phenotypic innovation and adaptive constraints in the evolutionary radiation of Palaeozoic crinoids. Scientific Reports 7:13745.

Yoder, J. B., E. Clancey, S. Des Roches, J. M. Eastman, L. Gentry, W. Godsoe, T. J. Hagey, D. Jochimsen, B. P. Oswald, J. Robertson, B. A. J. Sarver, J. J. Schenk, S. F. Spear, and L. J. Harmon. 2010. Ecological opportunity and the origin of adaptive radiations. Journal of Evolutionary Biology 23:1581–1596.

Young, N. M., G. P. Wagner, and B. Hallgrímsson. 2010. Development and the evolvability of human limbs. Proceedings of the National Academy of Sciences USA 107:3400–3405.

Figures

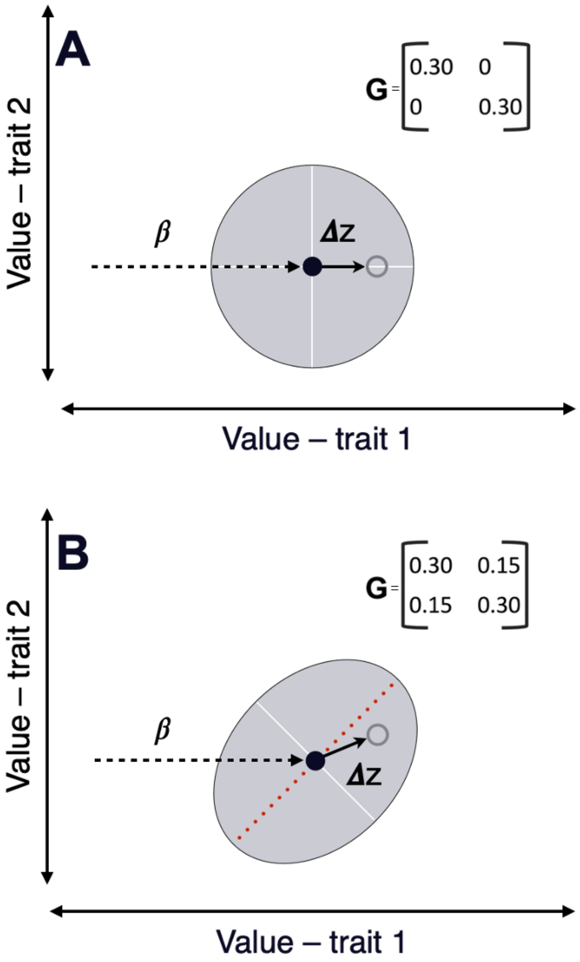

Figure 1. Geometry of the Lande equation. The three elements of the Lande equation are the direction and the magnitude of selection (*β*, dotted arrow), the amount of additive genetic variance in the direction of trait change (**G**, grey circle/ellipse) and the response to selection (Δ*z*, black arrow). The closed black and open grey points represent the trait mean before and after the selection event, respectively. (A) Δ*z* and *β* point in the same direction as there is no genetic covariance between the two traits. (B) The evolutionary response is deflected towards the direction with the largest amount of genetic variance (red dotted line) due to the genetic covariance between trait 1 and trait 2. The direction with the largest amount of genetic variance (i.e., the direction with highest evolvability) is what Schluter (1996) named the "line of least resistance." In the context of allometry (see Section 3.2), the direction of trait evolution predicted by the allometric relationship will be similar to the "genetic line of least resistance" if **P** closely resembles **G.**

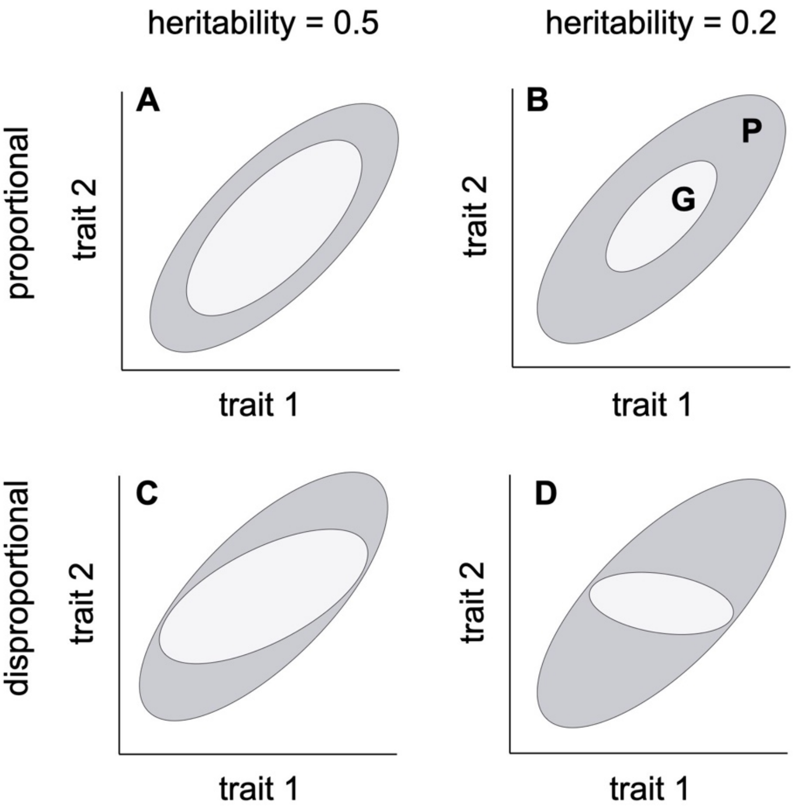

Figure 2. Heritability, proportionality of G and P, and Cheverud's conjecture. Panels schematically illustrate how **G** and **P** are related and how trait heritability affects the relationship. (A) and (B) show proportional **G** and **P**, (C) and (D) show disproportional **G** and **P**. (A) and (C) show traits with heritability of 0.5 (i.e., 50% of phenotypic variation is attributable to genetic variation), and (B) and (D) show traits with heritability of 0.2. In each panel, the dark grey ellipse represents the **P**-matrix and the light grey ellipse represents the **G**-matrix. In panels (C) and (D), **G** is rotated to be maximally dissimilar to **P**. Cheverud's conjecture holds in situations depicted in panel (A) and (B). In highly heritable traits, represented by (C), there is an upper limit on the dissimilarity between **G** and **P**.

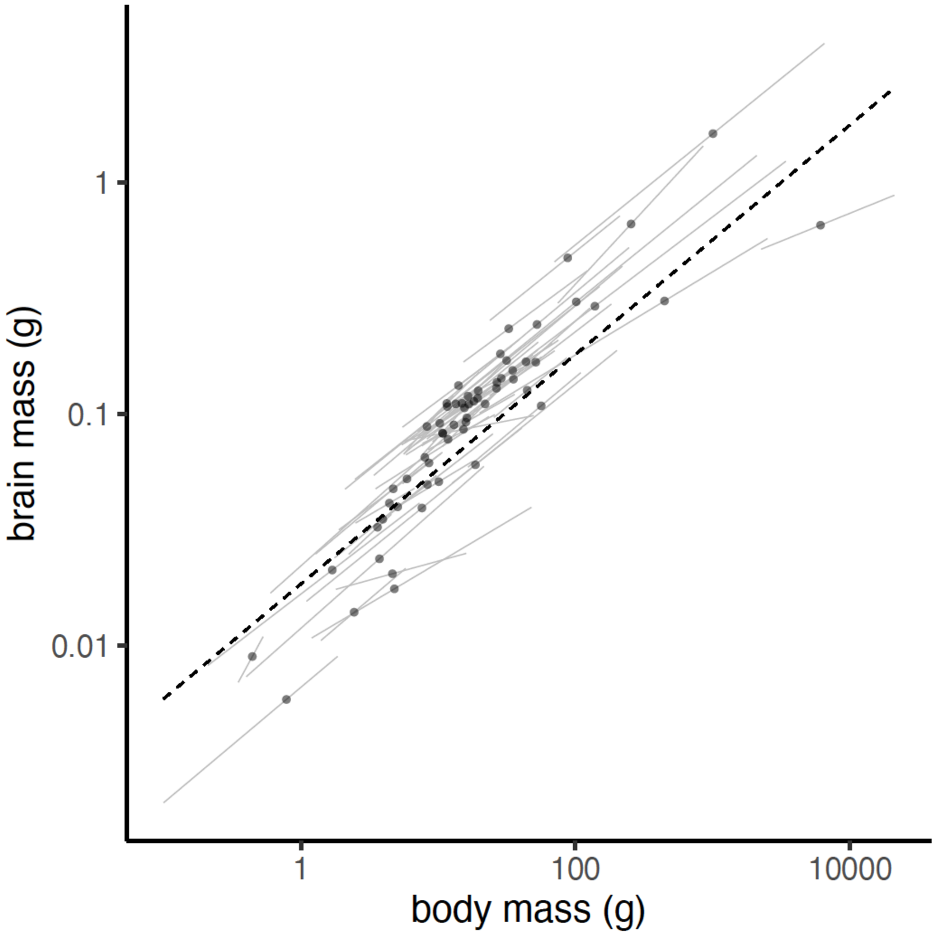

Figure 3. Brain mass - body mass allometry within and among species of the teleost order Perciformes. Grey lines represent the allometric relationship between brain mass and body mass among adult individuals within species (static allometry), whereas a dashed line represents the same relationship across species (evolutionary allometry). Static allometries are estimated using the Ordinary Least Squares. Evolutionary allometry is based on the phylogenetically-informed regression method reported in Tsuboi et al. (2018): log10(brain mass) = log10(body mass) x 0.496 - 1.73. Circles are species means (n = 94 species). The allometric slope of static (mean = 0.45, SD = 0.02) are similar to the slope of evolutionary allometry, which suggests that the static slopes are conserved over geological time scales and constrained in the direction of brain size evolution in Perciformes.

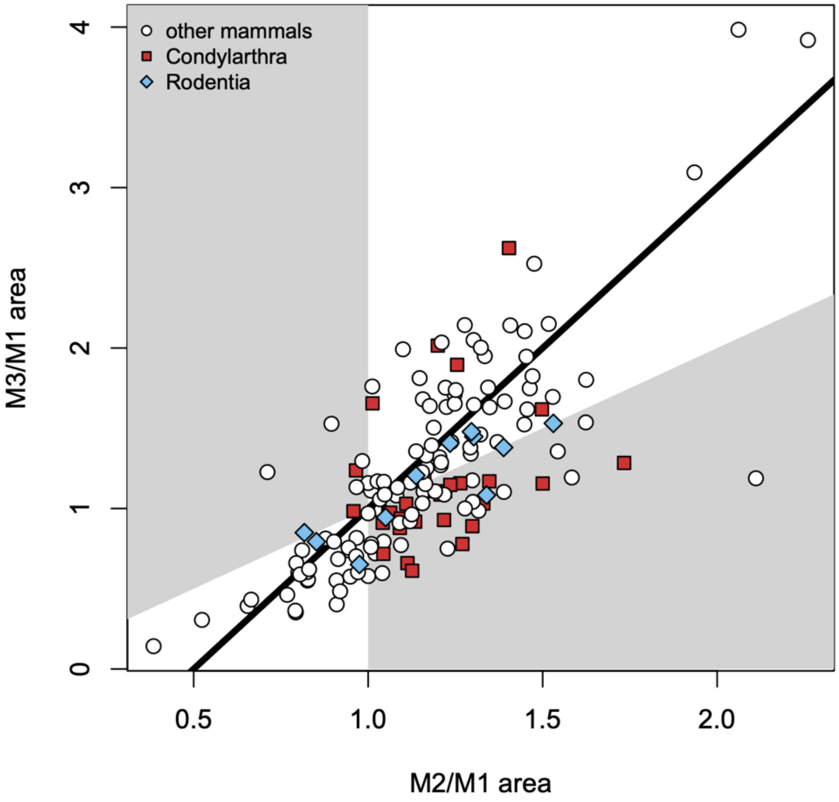

Figure 4. Area of third molar (M3) compared to the second molar (M2), each relative to the area of the first molar (M1). Each point represents tooth measurements from the tooth row of an individual mammal; many extinct and extant mammal groups are represented. Black line indicates the relationship predicted by the inhibitory cascade (IC) model of tooth development. Grey areas indicate tooth proportions that the strict IC model cannot produce. Most taxa have tooth dimensions consistent with the IC model, although some, such as rodents (blue diamonds), are more compatible than others, such as condylarths (red squares). Data are from Halliday and Goswami (2013).

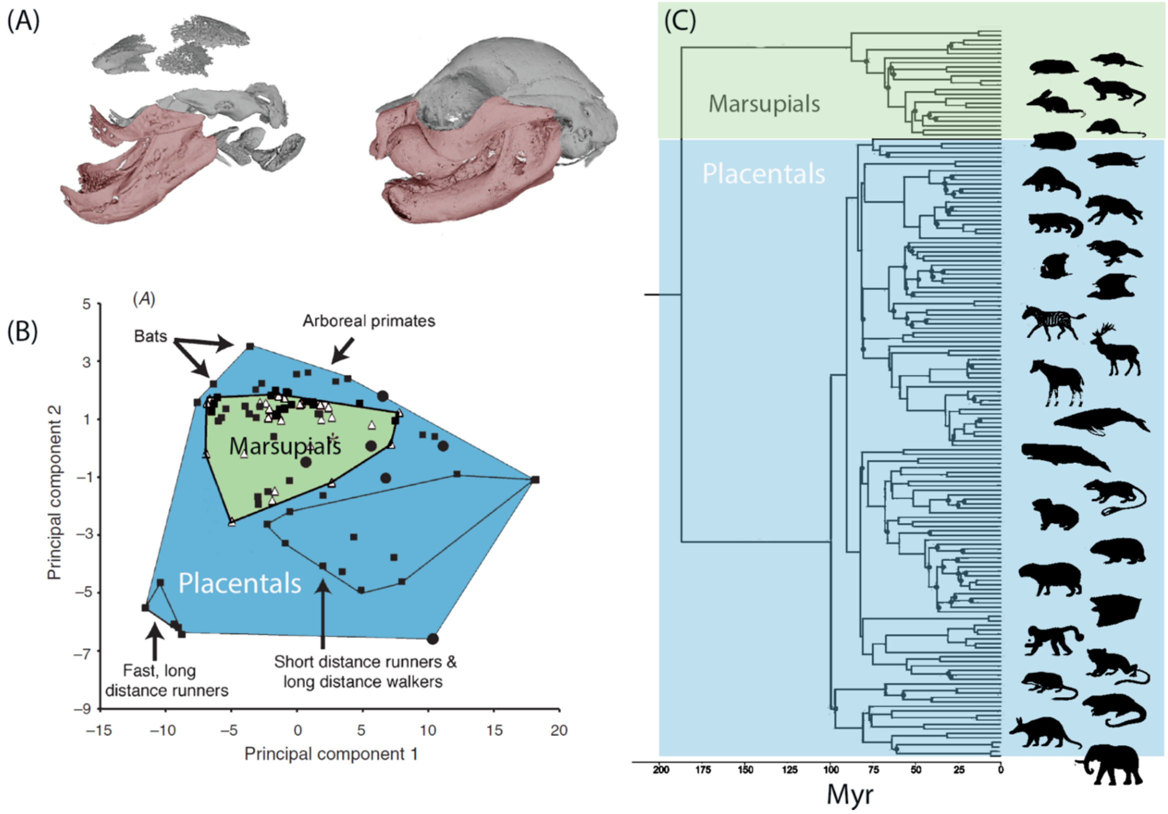

Figure 5. **Constraints on marsupial ecomorphological diversification.** (A) Example of an early ossifying skeletal element in marsupials. Here we illustrate two stages of cranial bone development in koalas (*Phascolarctos cinereus).* Note the relative early ossification of the dentary and maxilla (red) when compared to other cranial bones (gray). (Modified from Supplemental Fig. 5 in Spiekman and Werneburg 2017.) (B) Example of morphological disparity patterns observed when comparing placentals (blue) and marsupials (green). In this example, disparity patterns were obtained for forelimb traits. (Modified from Fig. 4 in Cooper and Steppan 2010.) (C) Phylogenetic time tree of mammalian families. (Modified from Fig. 1 in Meredith et al. 2011.)

BOX 1 - Outstanding Questions on Evolvability in the Fossil Record

- **General**

  - Is evolvability a major factor affecting differences in divergence among traits and lineages? Or are patterns of constraint and lability mostly determined by energetic and other trade-offs mediated by natural selection (Vermeij 2015)?
  - Are there secular trends in evolvability such that early appearing lineages may produce more phenotypic variation (Webster 2019) or have less potential versatility in form (Vermeij 1973a) compared to later appearing ones?
  - Does short-term evolvability, based on quantitative genetic theory, predict evolvability over longer time scales? Can other evolvability-related properties, such as developmental bias, be reliably measured in the fossil record (Jackson 2020)?

- **Paleontology and Quantitative Genetics**

  - Are phenotypic covariance (**P**) matrices usually good proxies for additive genetic (**G**) matrices?
  - How stable are **G** and **P** matrices over time?
  - Over what time scales are **G** and **P** predictive of evolutionary divergence?
  - Is the predictive power of **G** and **P** for evolutionary divergence due to genetic constraints or is selection shaping **G** and **P** to align with the direction of divergence?

- **Paleontology and Evo-devo**

  - Over what time scales are developmental processes conserved enough to make useful predictions about evolvability? Will this generally be longer than **G** and **P** matrices are stable?
  - How much insight can be derived about how modules originate and evolve from sampling fossil species? Do modules tend to form from the splitting of larger modules or from the integration of formerly independent units (Wagner and Altenberg 1996)?
  - There are many methods available for inferring modules from trait covariation, but they generally lack a strong theoretical basis. Can considerations of evolvability inform which methods are most appropriate or even suggest new approaches for recognizing modules?
  - What is the relationship between evolvability and innovation? Can the propensity for a lineage to produce novelties be operationalized, or are these processes too historically contingent to permit a general framework (Erwin 2019)?

- **Paleontology and Phylogenetic Macroevolution**

  - There are well documented differences in evolvability among traits. Are there also substantial differences in evolvability across lineages?
  - Is low evolvability associated with species extinction? Is high evolvability associated with radiation? How sensitive are existing methods for testing these associations with phylogenetic or paleontological data?
  - What is the relative importance of evolvability and selection in macroevolution?