Introduction

Whether quantitative genetic processes can explain macroevolutionary processes has been an important question in evolutionary biology since paleontologist George G. Simpson applied the concept of an adaptive landscape to explain the origin of grazing horses by natural selection and environment-driven changes in fitness (Simpson, 1944). While Simpson did not literally apply quantitative genetic models to his fossil data, he did adopt quantitative genetic concepts like adaptive peaks.
Figure 1  Modes of selection used by Simpson (1944) to explain macroevolutionary patterns, with today's equivalent terms. The top four examples result in no net change in the group (red dot shows the group mode), but affect the pattern of within-group variation (cream colored area). Centripetal selection, which is the same as stabilizing selection, pushes variation toward the mode; centrifugal selection, which is the same as disruptive selection, pushes variation away from the mode equally in all directions. Simpson envisioned many combinations of selective patterns, two of which are shown in the bottom panels. Centrifugal selection that is unequal results in directional change in the group mode. A combination of centrifugal selection on individual groups but centripetal selection between groups, perhaps as a result of intergroup competition, results in fractionating selection (diversifying selection in current terminology).

from Sewall Wright's work on mutation, allele frequencies, selection, and population divergence (Wright, 1932). Simpson argued that the long-term evolution documented by the fossil record – macroevolution, including not only the origin and evolution of species, but also their phylogenetic and phenotypic diversification – could be explained by the same microevolutionary population processes that are studied by quantitative geneticists (Figure 1). Simpson's goal was to demonstrate that the varied rates of evolution in phenotypic traits measured from sequences of fossils could be explained by varied regimes of natural selection. His integration of genetics, paleontology, and macroevolution made him one of the architects of the Modern Synthesis theory of evolution, which in the 1930s and 1940s replaced the orthogenetic, saltationist, and neo-Lamarckian theories that were current at the beginning of the twentieth century (Bowler, 1992).

Controversies

The Modern Synthesis view that macroevolution is the result of population-level microevolutionary processes quickly dominated the field but not without controversy. Simply put, the Synthesis view is that evolution is a population-level process, that phenotypic variation in populations arises from mutation and recombination, that population means change from one generation to the next by natural selection that differentially preserves those phenotypes that increase individual fitness or by the chance sampling process of drift. Coupled with factors that isolate one population from another, these processes are arguably sufficient to account for macroevolutionary patterns observed at a geological time scale such as the origin and diversification of major clades, changes in taxonomic diversity through time, or phenotypic changes within a long-lived species through time.

Challenges to this Synthesis view have come primarily from scientists who have argued that additional, non-population processes are required to explain some macroevolutionary patterns. Examples that are relevant to this discussion are punctuated equilibrium, the hypothesis that most species do not change much during their history except for rapid bursts of change at speciation (Eldredge and Gould, 1972); developmental constraints, the hypothesis that processes of organismal development channel and constraint variation and thus strongly influence macroevolutionary outcomes (Alberch, 1982); neutral evolution, the hypothesis that most evolutionary change occurs by random fixation of mutations or variants rather than by selection based on improved fitness (Kimura, 1983); and species or clade selection, the hypothesis that selective processes not only at the level of individual fitness but also on emergent traits of species and clades (Jablonski, 2008). Scientific debate about the impact of these three processes on macroevolution has encouraged the development of fully quantitative models for phenotypic evolution and their application to comparative and paleontological data to test whether quantitative genetic patterns observed at the population level are commensurate with long-term macroevolutionary patterns observed in comparative phylogenetic and paleontological data.

Nonmetric Traits

Early applications of quantitative genetics to macroevolution fell into two categories: nonmetric trait analysis, which was modeled on the analysis of allele frequencies, and continuous trait analysis, the theory for which was slower to develop. Nonmetric trait analysis focused on the frequencies of minor variations of the phenotype, such as numbers of foramina (small openings in bones for vessels and nerves), numbers of cusps on teeth, numbers of bristles on fruit flies, and numbers of spots on moth wings. The frequencies of
these traits, which were either known or assumed to be heritable, were compared within and between populations and species to determine the amount of divergence and whether drift alone, including founder effects, could explain the divergence or whether selection needed to be invoked. The nonmetric trait approach was grounded in the early genetic work of Sewell Wright (e.g., Wright, 1934) and extensively elaborated by Ford (1945), whose seminal evolutionary work on industrial melanism in moths used it; Berry (1978), who used it to study founder effect, selection, and phylogeography in mainland and island populations; and Sjøvold (1977), who enlarged the statistical toolkit for analyzing nonmetric trait divergence at macroevolutionary scales. Theory and process of the evolution nonmetric phenotypic traits were developed at length in the book *Phenetics* by Yablokov (1986), a student of the prominent geneticist Timofeev-Ressovsky. Despite the potential of nonmetric traits for studying the relationship of quantitative genetic processes and macroevolution, this approach has received little attention since the 1980s.

**Continuous Quantitative Phenotypic Traits**

**Selection Regimes and Rates of Evolution**

While nonmetric traits were important in early research on evolution of the phenotype, most studies of genetic processes and macroevolution have been based on quantitative traits. Advances in multivariate morphometrics and phylogenetic theory of quantitative traits have allowed complex new questions to be asked, as described in more detail below. Simpson’s early work, and those who followed his lead, involved relatively simple phenotypes, usually sizes of teeth and bones that could be measured in both living and fossil animals, and focused on questions that could be answered by comparing rates of evolution, direction of selection, or intensity of selection.

Simpson, Haldane, and Kurten made early contributions to our understanding of how selection changes as environments change by measuring rates of evolution in living taxa from different environmental regimes and measuring sequences of fossils through major environmental transitions (Simpson, 1944; Haldane, 1949; Kurten, 1959). Simpson (1944) classified evolutionary rates into three classes based on the distribution of their magnitudes within a taxonomic group or clade: horotelic distributions are the norm, with many bursts of rapid evolution interspersed with lower rates; bradytelic distributions are characterized by slow rates with most near zero, arising in situations where evolution is constrained by stabilizing selection, diminished genetic variation, or other factors; tachytelic distributions are unusually rapid and include exceptionally fast bursts of ‘quantum evolution’ as lineages cross from one adaptive zone to another. Simpson explained all three of these modes of evolution – stabilizing, random, and directional modes of selection are their parallels – in terms of population genetic processes, especially variation in the environmental and evolutionary context controlling the intensity of selection (Figure 1). Wright’s (1932) shifting balance explanation for how a species could shift from one adaptive peak to another was an important model for Simpson’s explanation for the origin of higher taxa, or adaptive radiations, by colonization of new adaptive zones, an idea that has been tested by comparing rates of evolution (Kurten, 1959), comparing modes of selection in a quantitative phylogenetic framework (Schluter, 2000), and by measuring phylogenetic evolution of multivariate phenotypic traits in morphospaces (Polly, 2008a). The study of rates of evolution was formally placed in a quantitative genetic framework by Gingerich, who proposed that rates be measured in units of haldanes, which is the average change in standard deviations per generation, that they be scaled to take into account the probability of evolutionary reversals that occur over long time periods, and that measurements of macroevolutionary divergence be scaled in a comparable way (Gingerich, 1993, 2001). This work made it possible to convert phenotypic rates of evolution measured over thousands or millions of generations to quantitative genetic parameters such as heritability and selection intensity. Selection has also been measured directly from sequences of well preserved fossils from which demographic age structures and, hence, individual survivorship can be estimated (Van Valen, 1965; Bell, 1988).

**Adaptive Peaks**

Although the adaptive landscape was used as a metaphor for the processes governing variation in rates of phenotypic evolution and the origin of taxa, it was not placed in a fully quantitative framework until Lande’s work in the late 1970s. Lande presented equations that describe the intensity and direction of selection on the population mean of a quantitative phenotypic trait at any point on an adaptive peak (Figure 2), derived estimates of the

![Figure 2](https://example.com/figure2.png)
rate and direction of evolutionary response to that selection, and described the trade-offs between drift and selection on the evolution of the phenotype (Lande, 1976). Lande’s equations were based firmly in population quantitative genetic principles of population variance, heritability, fitness, and multilocus genetic underpinnings of phenotypic traits. One of the goals of the quantitative framework was to determine whether challenges to the Modern Synthesis, especially the hypotheses that most evolution was neutral or random, that developmental and other constraints channel the direction of long-term evolution, and that most phenotypic change occurs at speciation, were consistent with quantitative genetic principles (Charlesworth et al., 1982), so it was elaborated to include adaptive peak shifts as model for long-term changes in the environmental context that defines fitness (Lande, 1986) and for correlated phenotypic traits (Lande and Arnold, 1983). Lande’s own conclusions were that drift, selection, and a phenotypic version of Wright’s shifting balance theory were adequate to explain, perhaps even to predict, patterns of punctuated equilibrium and that developmental and other constraints on evolution were consistent with quantitative genetic models for multivariate phenotypic traits (Lande, 1985, 1986). Importantly, expansions of Lande’s model now form the basis for most of the quantitative analysis of phenotypic evolution (Arnold et al., 2001; Gingerich, 2009; Hansen, 2013; Doebli, 2013; Bell, 2013).

Phylogenetic Comparative Methods

In parallel to adaptive peak models, methods for analyzing traits in a phylogenetic context were also developed. Felsenstein (1973) proposed a maximum-likelihood algorithm for reconstructing phylogeny from continuous phenotypic traits when they have evolved by a Brownian motion process such as drift or randomly changing selection. Unlike previously existing morphometric tree-building algorithms such as the unweighted pair-group averaging (UPGMA) method, or then-emerging discrete-trait algorithms such as parsimony, Felsenstein’s method was explicitly grounded in probabilistic evolutionary quantitative genetic theory and balanced the probability of character reversals with character change using the statistical properties of Brownian motion. In the simplest situations, the phenotypic divergence between taxa evolving by Brownian motion is a function of their rate of evolution and shared phylogenetic history. A single phenotypic trait evolving in an ancestor-descendant lineage by Brownian motion will have a possible distribution of descendant phenotypes whose mean is equal to the ancestor’s phenotype and whose variance is equal to the squared rate of change per generation times the number of generations elapsed (Felsenstein, 1988). Because the statistics of Brownian motion are non-directional, the same distribution describes the range of possible ancestor phenotypes given an observed descendant phenotype. The likelihood of phenotypes of the ancestor of two observed descendants is described by the joint probabilities of the descendants. By taking into account the ancestral likelihood distributions of many taxa, the overall likelihoods of competing phylogenetic hypotheses can be estimated and the most likely one identified. Using the same logic, the likelihood that traits evolved by Brownian motion can be evaluated if the phylogenetic tree is already known, their rate of evolution can be estimated, the likely ancestral phenotypes can be estimated, and the covariances that arise from shared common ancestry can be removed in order to study non-phylogenetic factors that cause trait covariance, such as genetic correlations or developmental constraints (Felsenstein, 1988; Harvey and Pagel, 1991). Furthermore, the statistical framework can be parameterized so that non-Brownian motion modes of evolution can also be evaluated (Martins and Hansen, 1997; O’Meara et al., 2006; Revell and Collar, 2009; Hunt and Rabosky, 2014). These other modes include Ornstein–Uhlenbeck processes that arise from centrifugal (stabilizing) selection associated with an adaptive landscape with a broad peak, and diversifying (directional) selection. Collectively, these tools for analyzing phenotypic traits among taxa whose phylogeny is known are called phylogenetic comparative methods.

Hypothesis Testing

Because they are underpinned by a common quantitative genetic framework, the tools derived from Lande’s quantitative theory for the evolution of phylogenetic traits and Felsenstein’s phylogenetic methods can be combined to test quantitative genetic hypotheses about macroevolution, or to generate hypotheses about population-level genetic processes from comparative phenotypic data. Population-level parameters such as phenotypic variance, heritability, population size, and selection intensity can be extrapolated over hundreds, thousands, or millions of generations using any one of the several existing models of evolutionary pattern (Brownian motion, stasis or stabilizing selection, directional or diversifying selection, shifting adaptive peak, etc.) and the resulting predictions can be compared to data collected from species with comparable phenotypic divergence times. The expected rate of phenotypic divergence in the phenotype will be a function of population-level phenotypic variance, population size, and heritability (Lande, 1976). If these parameters are known for all the species being compared, then the average selection intensity and the long-term pattern of selection can be inferred. Conversely, the population parameters can be inferred if the long-term patterns are known from paleontological or phylogenetic data and tested at the population level.

An example of macroevolutionary hypothesis testing based on quantitative genetic predictions is Gingerich’s (2001) study rates of evolution in lineages of fossil species tracked through millions of years. He found that average per-generation selection intensity was about 0.2 phenotypic standard deviations per generation, similar to the rapid rates of change measured in lab selection experiments for large and small body size. Interestingly, long-term diversification (maximum range of phenotypes in a clade) was less than expected if evolution continued directionally at this rate over millions of generations, suggesting that, while evolution is not constrained over short timescales, there are large-scale constraints on phenotypic disparity, a finding that has been replicated in other macroevolutionary studies covering many taxonomic groups at a several phylogenetic timescales (e.g., Hunt, 2007; Evans et al., 2012).
Units of Analysis

An important requirement for comparing data drawn from different evolutionary scales is that the units of analysis must be comparable (Gingerich, 2001; Houle et al., 2011). Evolutionary divergence times, for example, can be measured in units of millions of years (meganna), in molecular sequence differences, or in phenotypic character counts, but the time scale of the evolutionary process is the generation. Similarly, phenotypes can be measured in millimeters, kilograms, or Procrustes units, but it is the amount of change proportional to the genetic and phenotypic variance that is relevant to processes such as selection and drift. In order to compare population-level and macroevolutionary processes, the units of analysis must be standardized across the scales on which data are observed. Rates of phenotypic evolution are most easily compared across scales when they are measured as standard deviations per generation, a unit known as the ‘halfane’ (Gingerich, 1993), which is the same unit as quantitative genetic parameters such as selection intensity.

An example of how scaling matters is determining whether an observed Brownian motion pattern of species divergence arises from neutral genetic drift or selective drift. Neutral genetic drift is a process due to chance sampling from parent to offspring generation and is dependent on population size. Divergence between species due to drift is therefore a function of the average population size over time, the number of generations since common ancestry, and the genetic variance in the trait and will have a Brownian motion pattern (Lande, 1976). Selective drift is a process in which a trait is evolving by directional natural selection, but the direction and magnitude of selection are changed randomly over time (Kimura, 1954).

Polly (2004) used this scaling relationship to show that divergence in tooth morphology in a species flock of shrews whose last common ancestor lived about 40 million years ago had diverged randomly (Brownian motion) but at a rate much faster than expected from neutral genetic drift, even if heritability in tooth structure was exceptionally high and population sizes were consistently small. Only selection could explain the observed divergences, yet there was no evidence that selection followed a consistent, directional trend, despite the clear functional role that teeth have.

Morphometrics and Multivariate Phenotypes

This comparative framework for hypothesis testing is especially powerful when it can be applied to complex traits that are commonly preserved in the fossil record. Quantitative genetic studies have traditionally focused on traits that are easy to measure in lab or field settings, especially univariate traits such as body size, color, bristle number, and fecundity, but these traits are rarely preserved in fossils or in the museum research collections that are used for large-scale comparative phenotypic studies. Instead, complex structures such as teeth, bones, shells, and leaves are commonly studied in macroevolution, even though we know comparatively little about their developmental and genetic underpinnings. Developments in geometric morphometrics now allow extraordinarily complicated structures to be measured efficiently, allowing multivariate structures such as these to be studied quantitatively in large numbers (Bookstein, 1991; Dryden and Mardia, 1998). Studies of the genetics, development, geography, phylogeny, and phenomics of complex traits have abounded in the last decade and the opportunity for studying phenotypic evolution across these scales has never been greater.

Complex traits present new challenges, both in terms of analytical methods and evolutionary theory. Morphometric analysis of the shapes of two-dimensional structures, even extraordinarily complex ones, has become trivially easy, merely requiring a source of digital photos and software for capturing Cartesian coordinates that represent the structure of interest (Bookstein, 1991). The mathematical dimensionality of the resulting morphometric traits is often very large, but nevertheless easily tractable in mathematical terms. The analysis of three-dimensional shape, which includes most real biological structures, is far more challenging. Capturing three-dimensional data requires specialized equipment such as CT-scanning or laser scanning, and the methods available for representing the full three-dimensional structure of a trait as Cartesian coordinates, whose orientation on the object must be homologous, are difficult to apply to varied structures and fraught with concerns about biological homology, algorithmic optimization, and unintentional weighting of one part of a structure over another (Guzn et al., 2005; Polly, 2005b; Shen et al., 2009).

The high dimensionality of morphometric traits itself is an issue, because the number of mathematical variables represented by the landmark coordinates used to represent the structure may be much greater than the trait’s biological dimensionality. Three-dimensional morphometric representations of mammal teeth, for example, can easily have several hundred mathematical dimensions, each of which can be thought of as a potentially independent sub-trait, but quantitative trait locus (QTL) analysis and molecular developmental studies suggest that a much smaller number of genes may be involved in variation in tooth phenotypes (Workman et al., 2002; Harjunmaa et al., 2014). Because rates of evolution, standard deviation units, trait covariances, and other fundamental parameters are themselves dimensional, a new consensus about the concept of trait dimensionality is urgently needed. Multivariate traits might be thought of as coherent individual phenotypes with no sub-trait (Adams, 2014), they might be thought of in terms of the number of different genes underpinning them (Workman et al., 2002), they might be thought of in terms of the number of developmental factors that control them (Salazar-Ciudad and Jernvall, 2010), or they might be thought of in terms of their number of functional units (Stinchcombe et al., 2012).

An important challenge for quantitative phenotypic evolution is the study of evolutionary novelties – the gain and loss of limbs and digits, the multiplication of body segments, and the origin of eyes. The origin of novel structures is one of the most interesting aspects of macroevolution, and a better understanding of which is the primary goal of evolutionary developmental biology (Raff, 1996). However, the origin of new structures is almost impossible to quantify in morphometrics because each homologous point on the structure must be present in all structures in the analysis. Some morphometric approaches, such as outlines and surfaces, are ‘homology free’ in that they can accommodate the gain and loss of features on
a larger homologous structure, but even these cannot be used to represent an evolutionary transformation such as the evolution of limbs (Polly, 2008b). The most promising approach for studying evolution that involves the origin and loss of traits is the nonmetric approach described above, but no new developments have been made in this area.

An even more important challenge is that the evolutionary transformations that are reconstructed as shortest paths through a multivariate mathematical morphospace may not be the shortest biological paths from one phenotype to another (Polly, 2008b). The quantitative genetic and morphometric frameworks that were introduced above are based on trajectories through mathematical spaces, but as evolutionary distances increase to the macroevolutionary scale, the correspondence between mathematical and biological transformations becomes less certain. Dynamic relationships between tissues and gene signaling during development are known to produce nonlinear transformation in phenotypes that jump across gaps in morphometric spaces (Salazar-Ciudad and Jernvall, 2010). Thus, a continuous linear change in the level of expression of a developmental signaling gene may result in a discontinuous jump in the phenotype, one that may involve the origin of novel features. This phenomenon has important implications for quantitative theory for phenotypic evolution and is the object of ongoing study (Wolf et al., 2001; Hansen, 2008; Rice, 2008). Morphometric algorithms may require elaboration to provide a nonlinear mapping between developmental genetic spaces and morphospaces (Figure 3).

**Conclusion**

The challenges raised about the ability of the Modern Synthesis to reconcile microevolutionary processes with macroevolutionary patterns have as yet only been partially answered, in large part for want of data and analytical methods. The rapid advances in both now offer opportunities for comprehensively addressing these questions. For instance, the preponderance of data suggest that phenotypic evolution is usually more rapid than expected by truly neutral processes like drift over short timescales, but over broad timescales there appear to be constraints operating that prevent lineages from becoming as disparate as they might, given their per-generation rates of evolution (Gingerich, 2001; Hunt, 2007; Evans et al., 2012). However, despite the huge variety of data that underpins that observation, most of the traits that have been analyzed are essentially proxies for body size. Body size has fairly obvious constraints imposed on the small end by the requirements of cellular function and on the large end by the force of gravity (McNeill Alexander, 1998). Data for other kinds of traits are incomplete, but those that are available suggest that the pattern is different.

Rates of evolution in complex traits, such as the form of teeth and bones, for example, is still usually faster than expected from drift and long-term divergence often appears to be constrained; however, the constraints appear to change to allow lineages to move in new directions that their ancestors and relatives could not. The lifting of constraints is associated with large, seemingly correlated changes in form and function. For example, ankle structure in mammalian carnivores has evolved for about 50 million years constrained to a relatively small area of morphospace within which lineages have evolved and revolved similar structures; only the pinniped lineage (seals and sea lions) have broken out of those constraints, and out of the terrestrial locomotor system of their relatives (Polly, 2008a), consistent with Simpson’s concept of adaptive zones (Simpson, 1944; Figure 4). The evolutionary dynamics of developmental systems appear to generate similar patterns in which phenotypes evolve within constrained areas of morphospace then jump to new areas at points where the system passes a threshold (Salazar-Ciudad and Jernvall, 2010). Added to the complexity of these processes is the complexity of selection surfaces and adaptive landscapes for...
highly multivariate traits, which may have ‘holes’ that cannot be crossed, isometric lines of equal fitness in which phenotypes can wander randomly, and complex peaks and valleys that defy the simple metaphor of a landscape (Gavrilets, 1999). Some processes, such as species selection, are only starting to be formulated in a quantitative framework that will allow their effects to be compared in the same equations. Notably, Simpson (2011) has used a modified version of Price’s (1972) evolutionary equation to partition phenotypic change in clades into change arising from evolution within lineages and change that arises from turnover in traits due to differential survival of lineages.

Opportunities exist not only for empirical research on how evolution works, but on methodological developments to deal with the evolution of complex traits. In addition to the morphometric challenges raised above, new ways of conceiving evolutionary modes are needed for multivariate traits. For univariate traits, the only alternatives to Brownian motion are directional evolution and stasis. Multivariate traits evolving on multivariate selection surfaces behave in ways that hardly fit the latter definitions. The models used to estimate evolutionary modes are therefore difficult to apply to multivariate evolution in a meaningful way, even though considerable progress has been made to develop multivariate equations (Arnold et al., 2001; Gavrilets, 1999). Furthermore, most of the quantitative apparatus that is available for estimating rates and modes of phenotypic evolution are algorithms that find a single optimal parameter. The complex patterns and processes of multivariate evolution require methods that allow the relative support for competing hypotheses to be compared, ones that use a maximum-likelihood or Bayesian statistical framework. Some such methods are already available (Schluter et al., 1997; O’Meara et al., 2006; Goldberg and Igić, 2008; Slater et al., 2012), but more are needed, as are development of new models for the evolution of nonmetric traits and, ideally, synthesis with continuous multivariate traits.

Supplementary Material

[Note: This multimedia item is essentially an animated version of Figure 4.] Multimedia Animation 1 related to this article can be found online at doi:10.1016/B978-0-12-800049-6.000555.X.

See also: Adaptive Landscapes. Divergence and Diversification, Quantitative Genetics of. Systems in Evolutionary Systems Biology.
References


Further Reading
