

Article

Are our phylomorphospace plots so terribly tangled? An investigation of disorder in data simulated under adaptive and nonadaptive models

C. Tristan STAYTON*

Department of Biology, Bucknell University, 337 Biology Building, Lewisburg, PA, 17837, USA

*Address correspondence to C. Tristan Stayton. E-mail: tstayton@bucknell.edu

Handling editor: Martha Muñoz

Received on 4 May 2020; accepted on 17 August 2020

Abstract

Contemporary methods for visualizing phenotypic evolution, such as phylomorphospaces, often reveal patterns which depart strongly from a naïve expectation of consistently divergent branching and expansion. Instead, branches regularly crisscross as convergence, reversals, or other forms of homoplasy occur, forming patterns described as “birds’ nests”, “flies in vials”, or less elegantly, “a mess”. In other words, the phenotypic tree of life often appears highly tangled. Various explanations are given for this, such as differential degrees of developmental constraint, adaptation, or lack of adaptation. However, null expectations for the magnitude of disorder or “tangling” have never been established, so it is unclear which or even whether various evolutionary factors are required to explain messy patterns of evolution. I simulated evolution along phylogenies under a number of varying parameters (number of taxa and number of traits) and models (Brownian motion, Ornstein–Uhlenbeck (OU)-based, early burst, and character displacement (CD)) and quantified disorder using 2 measures. All models produce substantial amounts of disorder. Disorder increases with tree size and the number of phenotypic traits. OU models produced the largest amounts of disorder—adaptive peaks influence lineages to evolve within restricted areas, with concomitant increases in crossing of branches and density of evolution. Large early changes in trait values can be important in minimizing disorder. CD consistently produced trees with low (but not absent) disorder. Overall, neither constraints nor a lack of adaptation is required to explain messy phylomorphospaces—both stochastic and deterministic processes can act to produce the tantalizingly tangled phenotypic tree of life.

Key words: adaptation, character displacement, early burst, optima, OU model

Ever since the development of phylomorphospace visualization techniques (Bookstein et al. 1985 ; Klingenberg and Ekau 1996), researchers have explored patterns of phenotypic evolution by projecting phylogenies onto multivariate trait spaces. In many instances, both formal (Figueirido et al. 2013; Goswami et al. 2014; Martín-Serra et al. 2014; Felice et al. 2018) and informal researchers have noted that the patterns of evolution thus visualized appear less

orderly than expected. In some cases, researchers have proposed mechanisms to explain the perceived “messiness” of empirical phylomorphospaces. Underlying all of these explanations is the idea that the patterns observed with phylomorphospace techniques, absent some intervening factor, would be much more orderly than what is actually seen in nature. Thus ultimately, although phylomorphospace plots are merely visualization tools, this mismatch is

believed to contain some fundamental information about phenotypic evolution for the group in question.

However, null expectations for phylomorphospace patterns are usually not provided in such studies for even a qualitative comparison with the actual data. Researchers can and do of course test many aspects of their data sets against null models (for rates of evolution, differences in disparity among groups, and so on), but the features that make phylomorphospace patterns seem “messy”—the presence of many overlapping or crossing branches within a limited space—are not usually among them. To the best of my knowledge, researchers have not quantified or measured such features, nor studied the null expectations for such features under various scenarios. Even areas that might seem to lend themselves to such investigations, such as the study of convergent evolution (Ingram and Mahler 2013; Stayton 2008, 2015) do not provide such data, and although the seminal phylomorphospace analysis paper (Sidlauskas 2008) also developed the related topic of “lineage density”, null expectations for that measure remain understudied.

Thus it is somewhat surprising that researchers should be so consistent in noting that the phylomorphospace patterns seen in their data do not match their expectations. What are these expectations, and from what are they derived? I suspect that the expectations for what a phylomorphospace “should” look like come from informal, nonempirical diagrams which either demonstrate phylomorphospace concepts (e.g., Sidlauskas 2008, Figures 1 and 4), or which can be interpreted as doing so. Perhaps one of the most popular and certainly one of the oldest such diagrams is Darwin’s famous “I

think” diagram from his First Notebook on the Transmutation of Species (Figure 1A). Darwin’s sketch is clearly not a phylomorphospace plot in the contemporary sense, and though it may or may not reflect some implicit information about phenotypic differences among taxa or lineages, the patterns in this or similar diagrams could be interpreted by a casual reader as reflecting phenotypic evolution. Similarly, schematic diagrams of multivariate evolution over time may inform scholars’ expectations of how empirical phylomorphospaces will appear.

And what are the patterns shown in such diagrams? Continual divergence and radiation, perhaps with a trend, on many scales, and little to no crossing of branches (even if the diagrams are intended to illustrate convergent evolution!; Stayton 2008, 2015). Obviously, this style makes for cleaner and easier-to-interpret diagrams, but its suitability as an expectation for actual evolution, adaptive, or otherwise, is uncertain. The expectations produced by such sketches may be reinforced by the fact that some famous examples of evolution used to illustrate phylomorphospace concepts (e.g., the morphology of *Anolis* lizard ecomorphs, Figure 1B) also appear relatively orderly, though not completely without entanglement. Other empirical phylomorphospace spots are occasionally well-ordered as well (Figure 1C). However, it is still unknown whether these examples are representative of the patterns expected under various evolutionary conditions.

The goals of this article are, first, to define ways to quantify the qualitative impression of tangled disorder within phylogenies, and second, to explore the effects that various evolutionary parameters

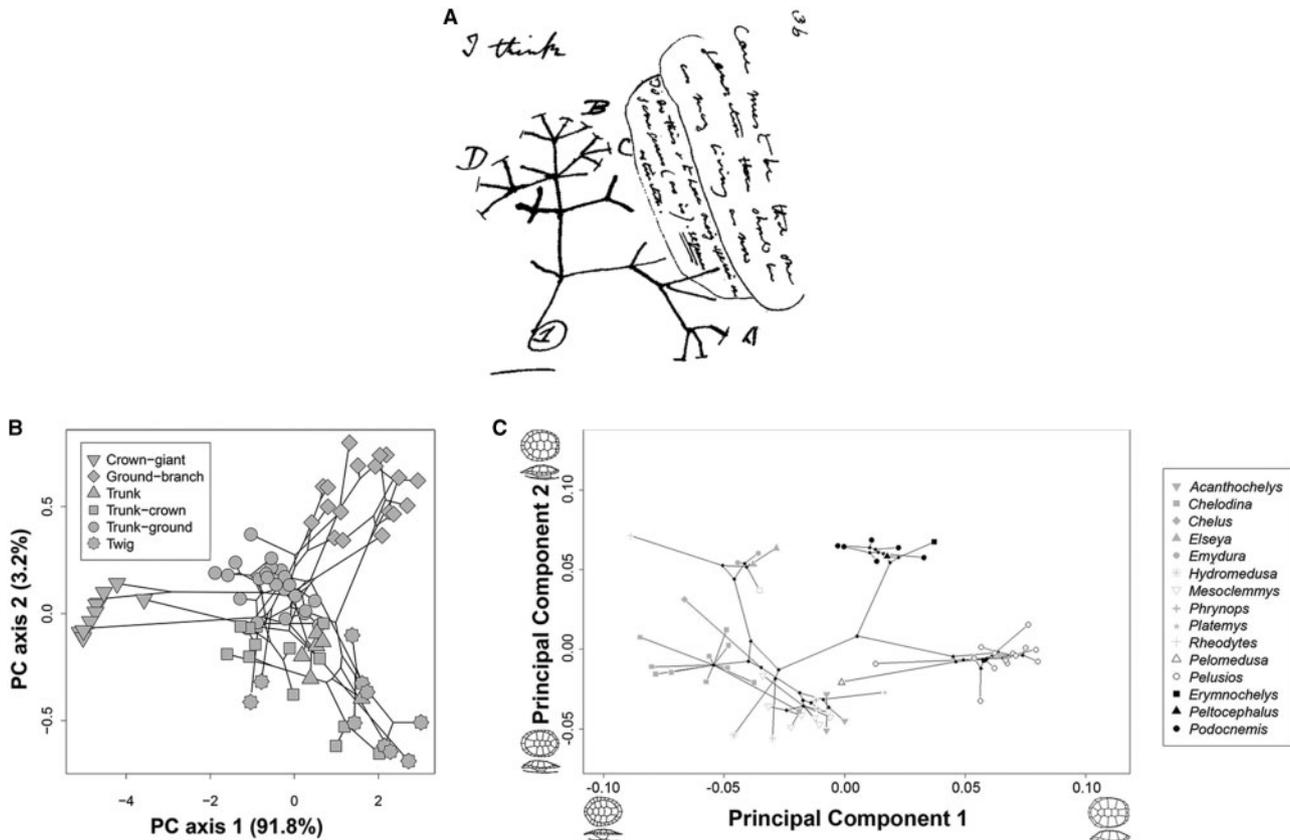


Figure 1. (A) Darwin’s “I think” diagram from his First Notebook on the Transmutation of Species (1837). (B) A relatively orderly and well-known phylomorphospace visualization of *Anolis* lizard ecomorph evolution; data from Revell (2012). (C) A relatively orderly phylomorphospace plot of pleurodire turtle shell shape; from Wise and Stayton (2017).

have on those measures. Besides exploring the effects of the number of taxa or number of dimensions on disorder (the term used in this article in lieu of more informal words such as “tangledness” or “messiness”) in phylogenies, I also explore the effects of various evolutionary models: the familiar Brownian motion (BM), Ornstein–Uhlenbeck (OU), and early-burst (EB) models, as well as a model which incorporates character displacement (CD) and a new model that uses information on functional performance to simulate evolution on complex adaptive landscapes.

Why is this important? After all, a mismatch between scientists’ perceptions of the pattern of evolution and the actual pattern of evolution might be of some sociological interest, but what is the salience to an actual understanding of the history of life? First, perceptions of patterns guide areas of inquiry, influence the questions which researchers consider, determine the topics chosen for study, and even influence the direction of grant funding. If, for example, high levels of disorder are assumed to reflect strong limitations on evolution (Felice et al. 2018), then research may be directed toward studies of developmental constraint in systems that may or may not manifest such constraints (at least not very strongly). Similarly, if disorder is believed to reflect certain features of adaptive evolution (Figueirido et al. 2013; Martín-Serra et al. 2014), then researchers may disregard searches for constraints or other influences on seemingly “messy” data.

Second, and more substantively, although phylomorphospace plots are simply convenient abstractions, they do reflect underlying patterns in data. Thus by studying these plots, and the often-surprising amounts of disorder which they reveal, researchers also study broader-scale patterns of evolutionary diversification. Such patterns then inform perennial debates in evolutionary biology, such as those on the relative amounts of divergence versus convergence among taxa (see Stayton 2015), the relative roles of adaptation and constraint in shaping diversification (Losos 2011), or the relative importance of adaptive radiation in the history of life (Simpson 1953; Harmon et al. 2010). This information, in turn, can then be used to guide studies of all of the processes involved in biological diversification. Darwin’s (en)tangled bank is an image of diversity and interaction, all produced by a common set of processes (Darwin 1859), but which processes most tangle the tree of life?

Materials and Methods

Quantifying disorder

I used 2 measures to quantify the degree to which phenotypic evolution produces a “tangled” pattern when visualized using phylomorphospace techniques. The first method could only be used for 2-dimensional data: each lineage on a phylogeny was compared with all other lineages, and each time 2 lineages “cross” one another (that is, each time there was an intersection between lineages) was counted, providing a measure here called “crosses”. Code to calculate this measure is provided as a Supplementary File. Second, lineage density (as defined by Sidlauskas 2008) was quantified. The total amount of phenotypic evolution that occurred on a phylogeny (that is, the sum of all branch lengths in phenotypic space) was calculated. This was then divided by the p th root, where p is the number of traits, of the total amount of shape space occupied by all of the taxa on a phylogeny. I quantified morphospace occupation volume using a Gaussian kernel density estimate, which avoids problems of outliers or non-normally distributed data, using the *hypervolume_gaussian* function in the R package “hypervolume” (Blonder et al. 2014). Large amounts of evolution restricted to very

small areas necessarily produce large lineage densities as well as a tangled appearance to phylomorphospace plots.

Preliminary study of the effect of tree size and number of traits

I performed all simulations in R (R Core Team 2020). First, I conducted a number of simulations of undirected evolution (BM model) to assess the effect of tree size and dimensionality on phylomorphospace disorder. Trees with $n = 32, 64, 128, 256,$ and 512 tips were used. Data were simulated on $p = 2, 3, 5,$ and 10 traits. Computation time for hypervolumes made the use of larger numbers of tips or traits impractical. A total of 1,000 random trees were generated for each combination of tip and trait numbers using the *rtree* command in the R package “ape” (Paradis and Schliep 2018). I simulated evolution under a BM model along each tree, for each trait, using the *fastBM* command in the R package “phytools” (Revell 2012), with σ^2 (the rate of evolution) set to 1, the ancestral state set to 0 for all traits, and no correlation between evolution of any traits. Uncorrelated traits were specifically chosen to determine whether substantial amounts of disorder are suggestive of correlations between traits which may result, among other things, from developmental constraints (Goswami et al. 2014; Felice et al. 2018).

After each simulation, ancestral states for each trait were reconstructed using the *fastAnc* command in “phytools”. I then calculated the number of crosses (for $p = 2$ only) and lineage density for each tree. Note that, as these analyses are conducted on simulations, the actual ancestral states were available. However, this would not typically be the case in empirical studies of evolution. Since the goal of this study is to provide expectations for phylomorphospace messiness against which empirical findings can be compared, the number of crosses and lineage density were calculated from reconstructed ancestors. This will necessarily change some measures away from the “true” values—for the ancestral state methods used here, all ancestors are reconstructed as weighted averages of descendants, such that ancestors can never be reconstructed outside of the range of their descendants even though some of the “real” ancestors may have occurred outside of that range, with potential effects on “crosses” and definite effects on lineage density—but it makes for more useful null models against which real data can be assessed. Note also that no dimensionality-reduction techniques (such as principal components analysis) were used on any data sets—analyses were conducted using all trait dimensions for all simulations.

Quantifying the effect of different modes of evolution

Next I conducted a series of simulations to determine the role that different modes of evolution might play in generating disorder within evolutionary phenotypic data. Here, trees with $n = 32$ or 256 tips were used (to compare effects between relatively small and relatively large trees), and $p = 2, 5,$ or 10 traits (to compare effects between relatively small and relatively large data sets).

I explored 4 different models of evolution (beyond BM). First, to investigate the potential effect of adaptation on diversification and disorder, evolution was simulated according to an OU process (Butler and King 2004). In all cases, σ^2 was set to 1, α (the strength of the restraining force in the OU model) was set to 1, the ancestral state was set to 0, and there were no evolutionary correlations between traits. The number of peaks was set at one greater than the dimensionality of the data (e.g., 6 peaks were simulated for the 5-dimensional data set) to ensure that the peaks did not constrain evolution to a lower-dimensional “slice” of phenotypic space—these

numbers match reasonably with empirical studies which have found strong support for anywhere from 2 to 15 peaks in multidimensional data (Ingram and Mahler 2013; Mahler et al. 2013; Sheftel et al. 2013; Arbour and López-Fernández 2014; Grundler and Rabosky 2014; Ingram and Kai 2014; Shoval et al. 2012; Alméjida et al. 2015; Tandler et al. 2015; Blom et al. 2016; Esquerré and Keogh 2016; Khabbazian et al. 2016; Moen et al. 2016; Davis and Betancur-R 2017; Arbour et al. 2019; Slater and Friscia 2019).

Peaks were located at equal distances from one another, slightly inside the average range of tip values seen in corresponding BM simulations, in order to generate approximately equal variance between the OU and BM simulations. Lineages were modeled as experiencing transitions between adaptive regimes (i.e., between “peaks” to which they are responding) according to a stochastic model. The transition matrix for each tree size was chosen to provide approximately 5 transitions within a 32-taxon tree, and 40 transitions within a 256-taxon tree, matching observed numbers of transitions from empirical studies (Blom et al. 2016; Alméjida et al. 2015; Arbour and López-Fernández 2014; Arbour et al. 2019; Davis and Betancur-R 2017; Slater and Friscia 2019; Khabbazian et al. 2016; Esquerré and Keogh 2016; Ingram and Kai 2014; Evans et al. 2017; Zelditch et al. 2017; Mahler et al. 2013; Grundler and Rabosky 2014; Moen et al. 2016).

I modeled transitions between regimes separately for each tree using the *sim.history* command in “phytools”. Once all parameters were set, OU evolution was simulated using the *mvSIM* command in the R package “mvMORPH” (Clavel et al. 2015). After each simulation, the number of crosses (for $p = 2$) and lineage density were calculated for each tree. Unfortunately, a realistic number of transitions per tree meant that smaller trees ($n = 32$) usually did *not* produce situations where all adaptive peaks were visited, and thus where the dimensionality explored by evolving clades was less than that of the full data set, skewing calculations of occupied hypervolume. For this reason, simulations with 32 tips and 5 or 10 traits were not used in subsequent analyses.

In addition I developed a new method to simulate evolution on a more complex and potentially more realistic adaptive landscape. This method also utilizes OU concepts and mathematics, but applies them to a much wider set of empirically-derived, continuous peaks rather than the few discrete peaks usually employed in OU modeling.

This method thus represents an extension of a set of techniques (Tseng 2013; Polly et al. 2016; Stayton 2019a, 2019b) inspired by Pareto front concepts (Shoval et al. 2012; Sheftel et al. 2013; see also references in Niklas 1999; McGhee 2006 for earlier applications). Specifically, the techniques start with a set of validated performance landscapes for multiple functions of a given structure (Tandler et al. 2015; Stayton 2019a, 2019b). Performance landscapes are multivariate representations of the relationship between phenotype and functional performance (see Arnold 1983, 2003; Arnold et al. 2001), inspired by the adaptive landscape concepts frequently evoked in evolutionary biology (Wright 1931; Gavrilets 2004; Kaplan 2008; Ghokale et al. 2009), and are often used to generate predictions regarding diversification (Alfaro et al. 2004, 2005). In the newer techniques, these performance landscapes are assigned relative weights (ranging from 0 to 1 for each function, and always summing to 1 for all functions), reflecting their relative selective importance for a given species, and then combined to determine the optimal shape for that combination of weights. This is repeated for every combination of relative weights to derive a set of “optima”—that is, the set of predicted phenotypes that should

evolve under *any* set of selective regimes on the functions under investigation. Here, the optima derived for turtle shells under 3 functions (Stayton 2019a) were used as a representative set of peaks.

For the new evolutionary modeling method developed here, these optima are used as the basis for simulations. Once the optima are available and a tree has been generated, the simulations proceeded as follows: first, the *relative weights* associated with each of the 3 functions are evolved according to a BM model, using the *fastBM* command in the R package “phytools” (Revell 2012), with σ^2 (the rate of evolution) set to 1 and the ancestral states set to 0.33 (i.e., starting with equal importance for all functions). Ancestral states are then calculated from the tip values using the *fastAnc* command in “phytools”. Values were rescaled to keep the relative weights of all functions, but so that all weights summed to 1 for each node.

After the evolution of relative weights had been reconstructed, each ancestral node was assigned an adaptive regime according to an OU model, with the “peak” being the optimum which corresponds to the set of weights for each ancestor, and σ^2 and α set to 1. Essentially this models evolution on a landscape with many peaks and with smooth transitions between adjacent peaks. Evolution of the 2 phenotypic traits was then simulated on the tree using the *OUwie.sim* command in the R package “OUwie” (Beaulieu et al. 2012). As with more conventional OU models, the locations of the optima were scaled so that the simulations produced data with approximately the same variance and range as did BM simulations on trees with the same numbers of tips. These procedures were conducted 1,000 times, and crosses and lineage densities were calculated for all iterations.

Next I used an EB model (Blomberg et al. 2003; Harmon et al. 2010) to explore the ways in which adaptive radiation of taxa might produce disparity and disorder. A total of 1,000 trees were generated for each combination of tip and trait numbers, and branch lengths were then transformed using the *rescale.phylo* command in the R package “geiger” (Pennell et al. 2014), with $a = -0.75$. Evolution was then simulated according to a BM model with $\sigma^2 = 1$ and all ancestral states = 0. This transformation effectively moves most phenotypic change to early branches, potentially generating multiple clusters of species close to one another but far from other species in different clusters. The number of crosses (for $P = 2$) and lineage densities were calculated for each iteration.

Finally, I studied a model incorporating CD using a method developed by Drury et al. (2016, 2018) that imposes a tendency on lineages to evolve greater phenotypic distance from one another. This mode of evolution was implemented using the *sim.divergence-geo* command in the R package RPANDA (Morlon et al. 2016). To maximize the potential impact of this mode of evolution, all taxa were modeled as interacting with all others. σ was set to 1×10^{-6} and m (the strength of the tendency for taxa to evolve greater phenotypic distance) was set at 0.02. These parameters were chosen because they approximate values found in empirical data and because they ultimately produce the same variance of tip data seen in equally-sized trees under a BM model. Again, 1,000 simulations were run for each combination of tree size and trait number and the number of crosses ($p = 2$) and lineage densities were calculated for each simulation.

Analysis of results

I first explored the relationships between the measures of disorder, tree size, and number of dimensions using simple linear regression. However, since larger trees have more branches and thus more

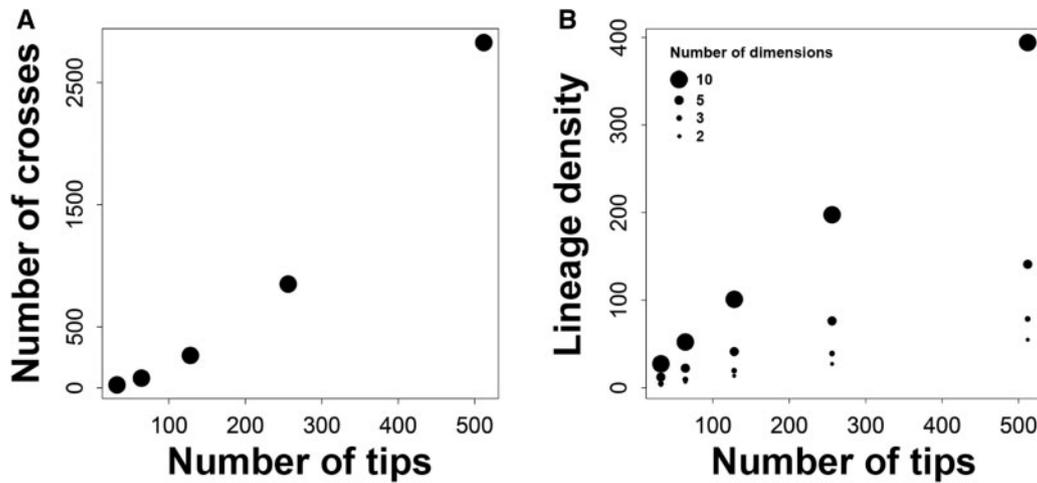


Figure 2. (A) Average number of crosses among branches versus tree size, for data simulated according to a BM model (1,000 simulations per point). (B) Average lineage density versus tree size, for different numbers of traits, for data simulated according to a BM model (1,000 simulations per point).

opportunities for crossing, it was necessary to scale the “crosses” measure before regression. It was not possible to derive an analytical maximum to the number of crosses which could occur in a tree for a given size. Therefore, 2 separate scalings were used: one with the “crosses” scaled to the number of tips on the tree, and with the “crosses” scaled to the maximum observed across all simulations. No rescaling was necessary for lineage density. Variables were log-transformed before analysis.

I also used analysis of variance (ANOVA) to assess the joint effects of tree size and number of traits on disorder. However, this was only possible with lineage density (because only this measure can be applied to data sets with $P < 2$). Lineage density was the response variable and tree size and the number of traits were factors.

Next I explored the relationships between the measures of disorder and different modes of evolution using ANOVA. First, I performed a set of “naïve” ANOVA with either number of crosses or lineage density as the response variable and mode of evolution as a factor for each combination of tree size and number of traits. I also conducted ANOVA with either number of crosses or lineage density as the response variable and with mode of evolution, tree size, and (for lineage density) number of traits as factors.

Next, a new set of ANOVA was conducted which attempted to control for the different amounts of variation observed from different modes of evolution. Here, “crosses” and lineage density were still the response variables and mode of evolution and the average variance across all variables were factors. I conducted ANOVA separately for all combinations of tree sizes and number of traits. All ANOVA were conducted using the *aov* command in R. *Post hoc* tests were used to investigate pairwise differences between individual modes of evolution.

Results

Tree size and number of traits

Larger trees produce greater amounts of disorder under a BM model of evolution (Figure 2 and Tables 1, 2). This holds with disorder measured as the number of “crosses” even if the numbers are rescaled (Table 1), and with lineage density (Table 2). ANOVA indicated that increases in both tree size and number of traits result in

Table 1. Linear regression of “crosses”, as well as “crosses” scaled using 2 different methods, against tree size for 2-dimensional data

Measure	r^2	$P <$
“Crosses”	0.968	2.46×10^{-2}
“Crosses”/number of taxa	0.992	2.83×10^{-4}
“Crosses”/maximum observed “Crosses”	0.918	1.02×10^{-2}

Table 2. Linear regression of lineage density against tree size for various numbers of traits

Number of traits	r^2	$P <$
2	0.847	5.06×10^{-11}
3	0.875	2.48×10^{-10}
5	0.964	9.74×10^{-6}
10	0.996	1.83×10^{-8}

increased lineage density ($P < 2.00 \times 10^{-16}$ for both factors; Figure 2B). There was a significant interaction as well—the relationship between lineage density and tree size is steeper for higher numbers of traits ($P < 8.54 \times 10^{-14}$, Figure 2B).

Different modes of evolution

The different modes of evolution are associated with significantly different levels of disorder. This holds across different tree sizes and numbers of dimensions, whether disorder is measured by the number of “crosses” of branches (ANOVA, 32-taxon trees: $F = 1668$, $P < 2 \times 10^{-16}$, 256-taxon trees: $F = 4315$, $P < 2 \times 10^{-16}$; Figure 3) or by lineage density (Table 3 and Figure 4). Evolution according to the CD model produces the lowest amount of disorder, then the EB mode, then BM. The OU-based models produced the highest amounts of disorder (Figure 4). *Post hoc* tests indicated that all modes of evolution produced significantly different amounts of disorder (all pairwise tests produced $P < 0.005$ for all tree sizes, numbers of dimensions, or measures of disorder). These relationships also hold if variation in the data (measured as the average variance of the tips across all axes) is accounted for (Table 4; all *post hoc* tests of pairwise differences produced $P < 0.001$).

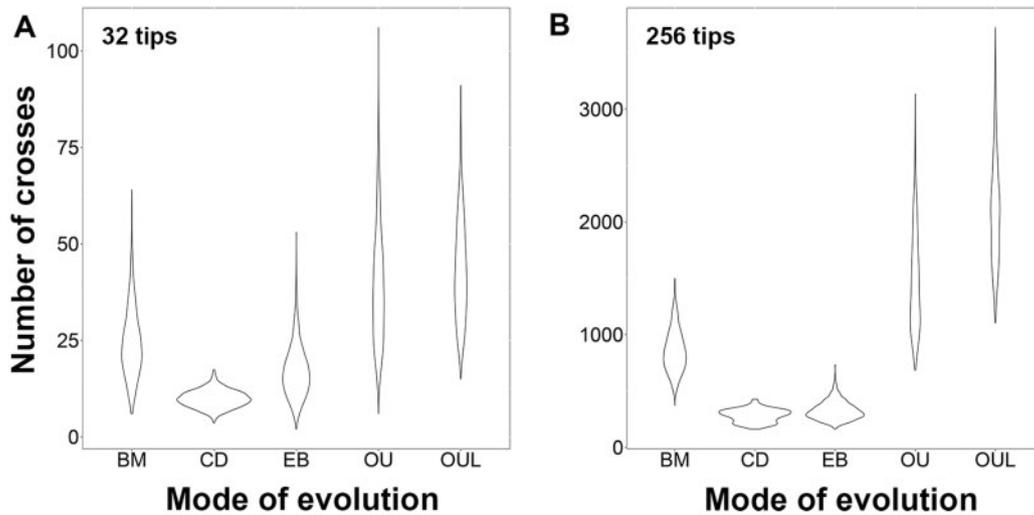


Figure 3. Differences in the disorder of phylomorphospace plots, measured as the number of crossing branches generated by BM, EB, CD, OU, and landscape-based OU (OUL) models, shown in violin plots. Each plot shows the results of 1,000 simulations for each model. **(A)** 32-tip phylogeny. **(B)** 256-tip phylogeny.

Table 3. ANOVA results for differences in the amounts of disorder, measured by lineage density, produced by different modes of evolution, for different tree sizes and different numbers of dimensions

Number of tips	Number of dimensions	F	P <
32	2	31,508	2×10^{-16}
32	5	23,115	2×10^{-16}
32	10	25,391	2×10^{-16}
256	2	50,519	2×10^{-16}
256	5	35,827	2×10^{-16}
256	10	90,999	2×10^{-16}

Table 4. ANOVA results for differences in the amounts of disorder, measured by lineage density, produced by different modes of evolution, for different tree sizes, and different numbers of dimensions, and controlled for tip disparity (measured as the mean variance across all dimensions)

Number of tips	Number of dimensions	F	P <
32	2	62,154	2×10^{-16}
32	5	41,466	2×10^{-16}
32	10	46,957	2×10^{-16}
256	2	1,33,790	2×10^{-16}
256	5	67,557	2×10^{-16}
256	10	3,54,324	2×10^{-16}

Discussion

Substantial amounts of disorder, as visualized by phylomorphospace plots and quantified by crossing branches or lineage density, can be generated in comparative data even in the absence of adaptation, constraint, or other influences on evolution beyond random drift (Figure 2). For an easily-conceptualized example: for a tree of 256 taxa and with 2-dimensional data, each branch in the

phylomorphospace plot is expected to cross 3 other branches *on average* even if evolution proceeds according to a BM model with no constraints.

Disorder in comparative data consistently increases as tree size and the number of traits increase when evolution is simulated according to a BM model (Figure 2; Tables 1 and 2). The results for crossing branches seem to follow an exponential trend, which raised concerns that the counts are simply the result of larger opportunities for crossing branches in larger trees, but the pattern holds even if the higher number of branches on larger trees is accounted for (Table 1). This by itself is an interesting result—all else being equal, researchers should expect phylomorphospace plots of larger data sets to be more disordered than smaller ones. The reason for this pattern seems to be that larger trees simply present *proportionally* more extensive opportunities for the phenomena which produce disorder (such as convergence, reversal, and so on) to manifest—evolution on a 4-taxon tree will by necessity almost always appear divergent, for example, although a single early reversion to an area of phenotypic space close to the ancestral states will result in large amounts of disorder in a larger tree even if subsequent evolution is mostly divergent. Similar arguments apply for the increasing disorder seen with increasing numbers of traits. The lineage density equation accounts for the number of dimensions, so stochastically perhaps there is a greater probability that convergence or reversal will occur along at least some dimensions in a high-dimensionality data set, with subsequent effects on lineage density as a whole.

Of greater importance are the effects of different modes of evolution on disorder. In some cases, these results are intuitive: the model involving CD consistently showed much lower levels of disorder than all other modes of evolution, for example. By promoting the evolution of greater distance among taxa, divergence is facilitated and reversal, parallelism, and convergence are inhibited. As such it is unsurprising that this model produces low lineage densities and few crossing branches (and predictable evolution, if the low levels of variance of results are any indication). This model could perhaps be taken as a baseline for the minimal amount of disorder that can be expected in an evolutionary data set. However, it should be noted that the parameters used here, which include interaction and displacement among all species, may not be realistic for many clades (it

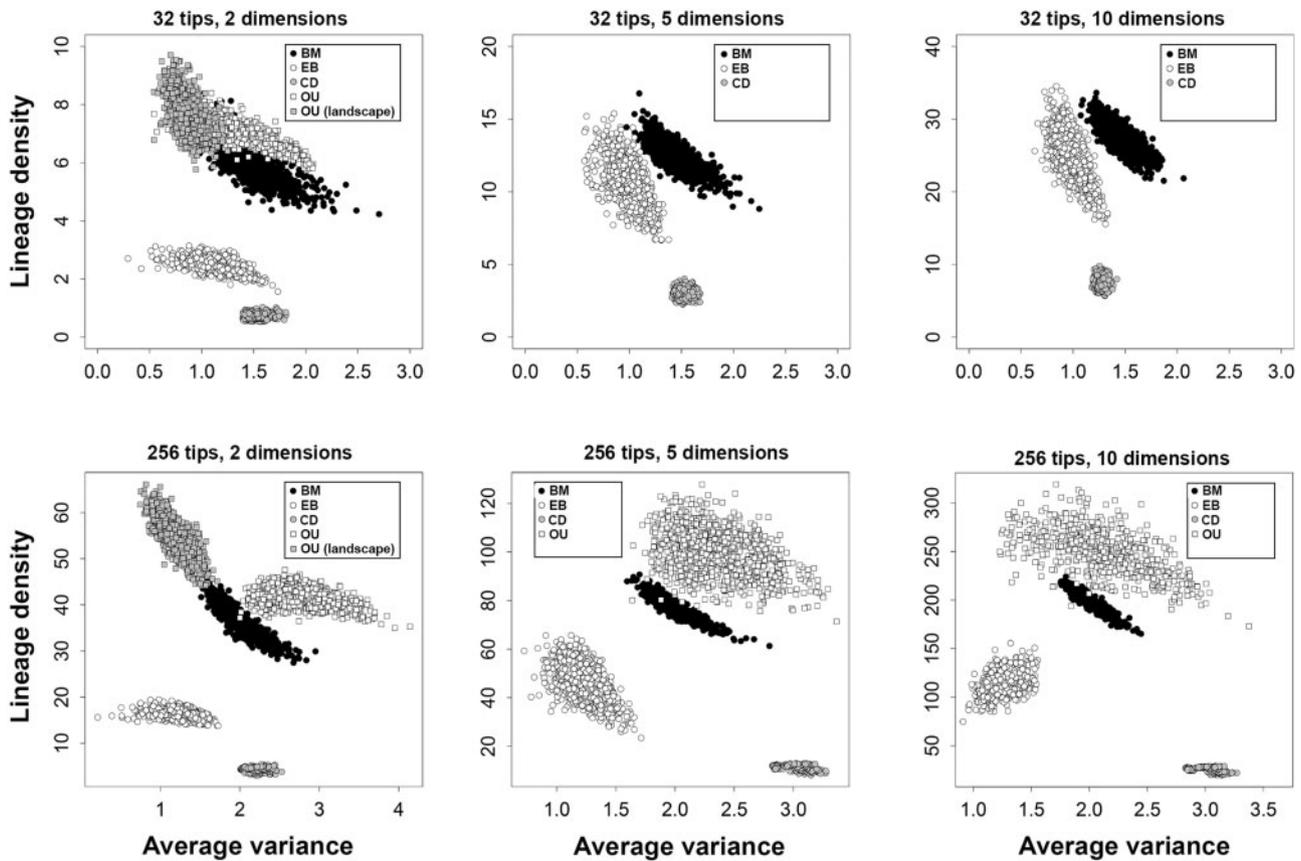


Figure 4. Differences in the disorder of phylomorphospace plots, measured as lineage density generated by BM, EB, CD, OU, and landscape-based OU models. Individual plots show results on trees with different numbers of tips and with data simulated for different numbers of traits. Each plot shows the results of 1,000 simulations for each model. Lineage density is plotted against the mean variance of tip data for standardized comparisons, because different modes of evolution produce different amounts of variation across tips in the phylogeny, and because lineage density clearly varies with mean variance for at least some models.

is probably unlikely that all taxa would co-occur and interact). In addition, it is important that some disorder does still occur even in this “best-case” scenario for orderly phylomorphospaces.

The EB model also produced relatively low levels of disorder. Here, rapid rates of evolution early in a clade’s history make a “starburst” pattern (or a nested series of “starbursts”) of disparity more likely—the major clades in a phylogeny rapidly evolve toward usually-separate areas of phenotypic space, and then diversify. Although this does not have a great effect on the number of crossing branches (after all, there are still plenty of opportunities for branches to cross *within* clusters; Figure 3), the effect on lineage density is substantial (Figure 4) because this model produces a disproportionate number of taxa in clusters close to the periphery of the occupied area of phenotypic space (Figure 5B). Empirical examples of such distributions are rare, perhaps because they are difficult to see unless the centroid of the distribution remains empty, but they do occur (Wise and Stayton 2017; Burns and Sidlauskas 2019) and even if the pattern itself is visually unobserved, a surprisingly low lineage density for a clade might serve as a motivation to investigate the possibility of an adaptive radiation within.

Notably, it was the EB and CD models, and not the OU models (see below), whose phylomorphospace plots corresponded most closely to the intuitive exemplar images of orderly, divergent evolution (compare Figure 5B,C with Figure 1). Could it be that biologists implicitly assume a greater degree of early divergence or competitive

exclusion than is usually found in evolving clades (EB evolution at least may be rare—see Harmon et al. 2010—although more multi-variate studies are needed; e.g., Arbour and López-Fernández 2013)?

Along those lines, intuitions that adaptive evolution should produce well-ordered data seem to be incorrect, as both OU-based models produced significantly higher levels of disorder than did evolution according to any other model. Why? An investigation of representative phylomorphospace plots (Figure 5) and a comparison with the EB results provides some hints. Both OU and EB tend to produce “clumpy” or clustered distributions of taxa, with evolution approximating a BM model within clusters. In OU models, these clusters occur around peaks and the restraining force of selection may tend to slightly increase homoplasy and hence crossing of branches within those regions. More importantly, however, lineages can move between clusters in the OU model. Not only does this promote more crossing of branches when lineages switch peaks (especially as there are a limited number of paths between peaks), but it also greatly increases the amount of evolution that occurs within the space defined by the peaks, increasing lineage density. Unless the number of transitions is very low (and again, the parameters used in this study produced biologically-reasonable frequencies of transitions—on average 4.7 for the 32-taxon trees and 44.3 for the 256-taxon trees) evolution among a set of discrete peaks will produce much the same effect as that proposed for constraints

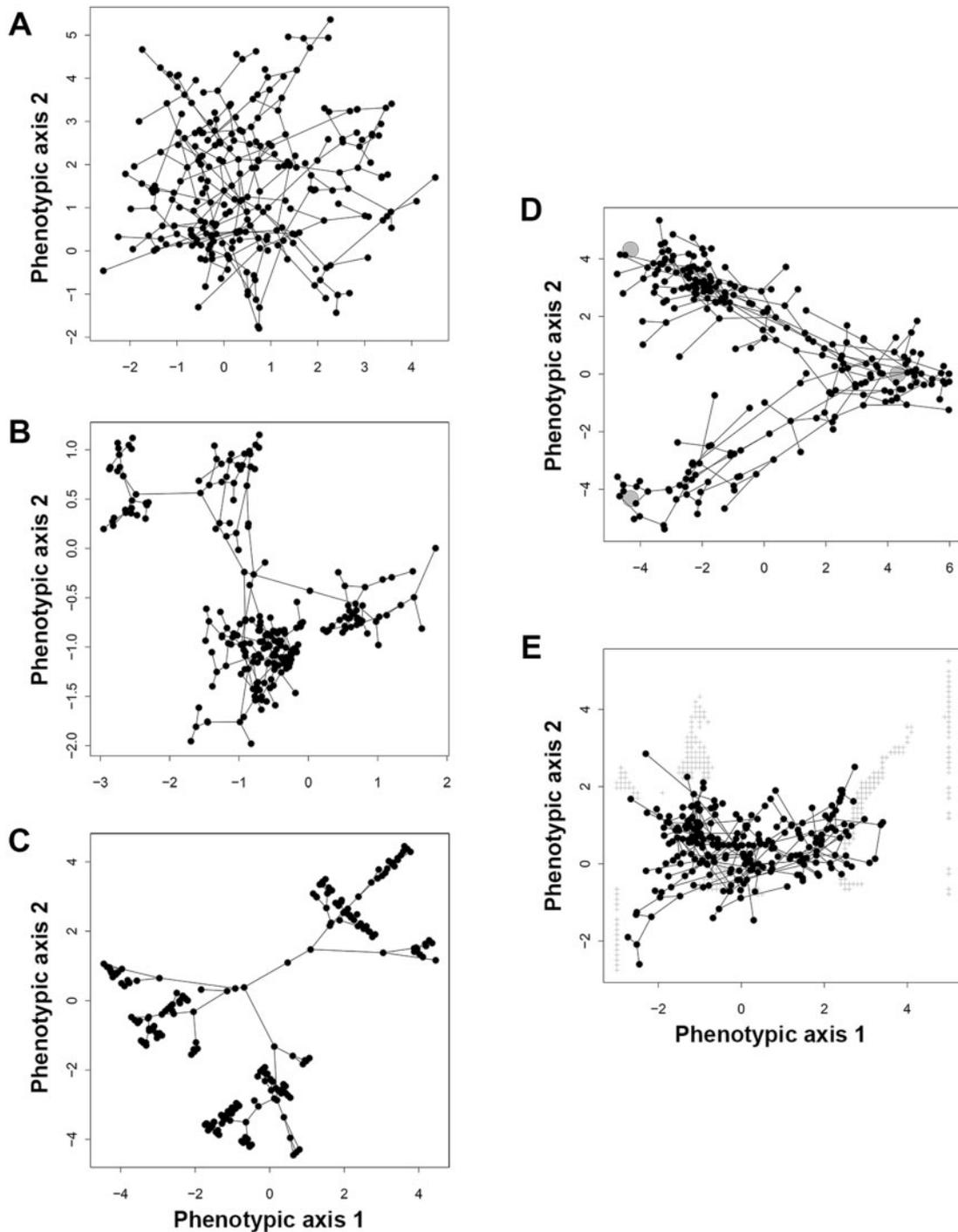


Figure 5. Representative phylomorphospace plots of data generated on a 128-tip tree by: (A) A BM model. (B) An EB model. (C) A CD model (following Drury et al. 2016, 2018). (D) A 3-peak OU model with the locations of peaks indicated by large grey circles (following Ingram and Mahler 2013). (E) A landscape-based OU model with the locations of optima indicated by grey crosses.

(Goswami et al. 2014; Felice et al. 2018)—plenty of evolution in a small space with concomitant increases in disorder. These results strongly indicate that researchers should not take the presence of high levels of disorder as indicative of a lack of selection, nor necessarily of constraints on evolution.

All of the preceding arguments apply to the landscape-based OU model as well, with some modifications. Here, evolution takes place among a greater number of less widely-spaced peaks, and with

many more transitions between peaks. Numerous crosses still occur as lineages evolve toward nearby peaks, although the absence of peaks in certain regions of phenotypic space still ensures that evolution will occur within a limited range. Moreover, the greater number of peaks, as well as the *far* greater number of paths between peaks, probably facilitate increased crossing of branches relative to the traditional OU model. The greater number of closely-packed peaks probably produces a less clustered distribution of tips than do the

traditional OU models. However, this also means that a lower percentage of available peaks are likely to be “discovered”, and thus that a smaller proportion of adaptive phenotypic space will be explored, producing higher lineage densities (see Figure 5E). Indeed, only rarely did lineages end up evolving in response to optima located throughout phenotypic space, even in runs using trees with 256 tips. These landscape-based models were intended to provide a more realistic and grounded view of evolution, and have been shown to predict species distributions quite well (and better than more traditional OU models; Stayton 2019a). However, they have seen only limited use so far—future studies will be needed to establish their more general applicability and to determine if empirical data and disorder fit these models better than more traditional OU models.

From a biological standpoint, then, this study supports a number of general conclusions. First, substantial amounts of disorder, sufficient to produce phylomorphospace plots with “messy” appearances, can be produced through many different modes of evolution (Figure 5). Second, selection *per se* is neither necessary nor sufficient to produce particularly high or low levels of disorder. For example, both the CD and OU models involve selection (explicitly for the OU models and implicitly for the CD model) but produce widely different amount of disorder. The nature of selection seems important here—adaptation in response to a static adaptive landscape may promote disorder, although selection in response to (potentially closely-related) competitors might inhibit it. Additionally, the EB model, which includes no selection at all, also produced well-order phylomorphospaces and low levels of disorder. Third, neither correlation among characters nor limits to evolution (as might be produced by developmental constraints) were modeled here, indicating that neither of these factors is necessary to produce highly disordered phylomorphospace plots, although they might certainly exacerbate disorder when present. Future studies could benefit from additional exploration of various evolutionary parameters (correlations between or limits to character evolution, changes in evolutionary rates, different modes of speciation, and different sources of selective factors) on disorder—clearly our intuitive expectations of what a phylomorphospace “should” look like can obtain, but only in specific and sometimes nonintuitive circumstances.

Practically, the disorder measures described here may be used to generate expected distributions for various models against which to compare empirical data and thus investigate potential modes of evolution. For example, a data set of turtle shell shape (Stayton et al. 2018) with 280 taxa produced 3117 crosses and a lineage density of 65.9 when the first 2 PCs were examined. These are high values, and consistent only with an OU model, likely one with many peaks (which is satisfying, because this is the data set that inspired the landscape-based OU model), suggesting strong selection but a lack of competitive CD among the lineages. Studies with similarly messy phylomorphospaces might also find evidence for this mode of selection among their data.

More broadly, if authors of future studies wish to draw conclusions from the appearance of phylomorphospace plots (or, more importantly, from signal in the underlying data), they should first compare their observations to null models to ensure that their conclusions match expectations for their favored mode of evolution, or even whether any explanation for the pattern is warranted at all. Evolution, it seems, can be complicated enough without the influence of processes like adaptation and constraint. Relatively ordered data which produce clean and intuitive phylomorphospace plots (Figure 1B) may be the exception rather than the rule among life. When Darwin wrote of a “entangled bank” (specifically, his second,

more famous, and more metaphorical reference to such a bank from *The Origin of Species* (1859); this is the one appearing in the final paragraph of that book), he was referring to a vast array of highly diverse and interacting species; this study suggests that individual clades, even absent interaction with other taxa, can produce trees that are tangled enough on their own.

Acknowledgments

Dr Martha Muñoz and the editorial staff at *Current Zoology* were exceptionally accommodating when the original deadline for submission was complicated by COVID-19; I thank them for their patience and understanding through the submission process. I also thank the executive and handling editors, and D. Adams, B. Sidlauskas, and an anonymous reviewer for helpful comments and extensive efforts on behalf of this manuscript.

Supplementary Material

Supplementary material can be found at <https://academic.oup.com/cz>.

References

- Alfaro ME, Bolnick DI, Wainwright PC, 2004. Evolutionary dynamics of complex biomechanical systems: an example using the four-bar mechanism. *Evolution* 58:495–503.
- Alfaro ME, Bolnick DI, Wainwright PC, 2005. Evolutionary consequences of many-to-one mapping of jaw morphology to mechanics in labrid fishes. *Am Nat* 165:E140–E154.
- Almécija S, Smaers JB, Jungers WL, 2015. The evolution of human and ape hand proportions. *Nat Commun* 6:7717.
- Arbour JH, López-Fernández H, 2013. Ecological variation in South American geophagine cichlids arose during an early burst of adaptive morphological and functional evolution. *Proc R Soc B* 280:20130849.
- Arbour JH, López-Fernández H, 2014. Adaptive landscape and functional diversity of Neotropical cichlids: implications for the ecology of evolution of Cichlinae (Cichlidae; Cichliformes). *J Evol Biol* 27:2431–2442.
- Arbour JH, Curtis AA, Santana SE, 2019. Signatures of echolocation and dietary ecology in the adaptive evolution of skull shape in bats. *Nat Commun* 10:2036.
- Arnold SJ, 1983. Morphology, performance, and fitness. *Am Zool* 23:347–361.
- Arnold SJ, 2003. Performance surfaces and adaptive landscapes. *Integr Comp Biol* 43:367–375.
- Arnold SJ, Pfenner ME, Jones AG, 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* 112/113:9–32.
- Beaulieu JM, Jhwueng D-C, Boettiger C, O’Meaura BC. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* 66:2369–2383.
- Blom MPK, Horner P, Moritz C, 2016. Convergence across a continent: adaptive diversification in a recent radiation of Australian lizards. *Proc R Soc B* 283:20160181.
- Blomberg SP, Garland T, Jr., Ives AR, 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745.
- Blonder B, Lamanna C, Violle C, Enquist BJ, 2014. The *n*-dimensional hypervolume. *Global Ecol Biogeogr* 23:595–609.
- Bookstein FL, Chernoff B, Elder RL, Humphries JM, Smith GR, et al., 1985. *Morphometrics in Evolutionary Biology; the Geometry of Size and Shape Change*. Philadelphia (PA): Academy of Natural Sciences.
- Burns MD, Sidlauskas BL, 2019. Ancient and contingent body shape diversification in a hyperdiverse continental fish radiation. *Evolution* 73:569–587.
- Butler MA, King AA, 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am Nat* 164:683–695.
- Clavel J, Escarguel G, Merceron G, 2015. mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. *Meth Ecol Evol* 6:1311–1319.

- Darwin C, 1859. *On the Origin of Species by Mean of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.
- Davis AM, Betancur-R R, 2017. Widespread ecomorphological convergence in multiple fish families spanning the marine-freshwater interface. *Proc R Soc B* 284:20170565.
- Drury J, Clavel J, Manceau M, Morlon H, 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. *Syst Biol* 65:700–710.
- Drury JP, Grether GF, Garland T Jr, Morlon H, 2018. An assessment of phylogenetic tools for analyzing the interplay between interspecific interactions and phenotypic evolution. *Syst Biol* 67:413–427.
- Esquerré D, Keogh JS. 2016. Parallel selective pressures drive convergent diversification of phenotypes in pythons and boas. *Ecol Lett* 19:800–809.
- Evans KM, Waltz B, Tagliacollo V, Chakrabarty P, Albert JS, 2017. Why the short face? Developmental disintegration of the neurocranium drives convergent evolution in neotropical electric fishes. *Ecol Evol* 7:1783–1801.
- Felice RN, Randau M, Goswami A, 2018. A fly in a tube: macroevolutionary expectations for integrated phenotypes. *Evolution* 72:2580–2594.
- Figueirido B, Tseng ZJ, Martín-Serra A, 2013. Skull shape evolution in durophagous carnivorans. *Evolution* 67:1975–1993.
- Gavrilets S, 2004. *Fitness Landscapes and the Origin of Species*. Princeton (NJ): Princeton University Press.
- Ghokale C, Iwasa Y, Nowak MA, Traulsen A, 2009. The pace of evolution across fitness valleys. *J Theor Biol* 259:613–620.
- Goswami A, Smaers JB, Soligo C, Polly PD, 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. *Philos T R Soc B* 369:20130254.
- Grundler MC, Rabosky DL, 2014. Trophic divergence despite morphological convergence in a continental radiation of snakes. *Proc R Soc B* 281:20140413.
- Harmon LJ, Losos JB, Davies J, Gillespie RG, Gittleman JL et al. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396.
- Ingram T, Kai Y, 2014. The geography of morphological convergence in the radiations of pacific *Sebastes* rockfishes. *Am Nat* 184:E115–E131.
- Ingram T, Mahler DL, 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise AIC. *Meth Ecol Evol* 4:416–425.
- Kaplan J, 2008. The end of the adaptive landscape metaphor? *Biol Philos* 23:625–638.
- Khabbazian M, Kriebel R, Rohe K, Ané C, 2016. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Meth Ecol Evol* 7:811–824.
- Klingenberg CP, Ekau W, 1996. A combined morphometric and phylogenetic analysis of an ecomorphological trend: pelagization in Antarctic fishes (Perciformes: nototheniidae). *Biol J Linn Soc* 59:143–177.
- Losos JB, 2011. Convergence, adaptation, and constraint. *Evolution* 65:1827–1840.
- Mahler DL, Ingram T, Revell LJ, Losos JB, 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science* 341:292–295.
- Martín-Serra A, Figueirido B, Palmqvist P, 2014. A three-dimensional analysis of morphological evolution and locomotor performance of the carnivoran forelimb. *PLoS One* 9:e85574.
- McGhee GR, 2006. *The Geometry of Evolution: Adaptive Landscapes and Theoretical Morphospaces*. Cambridge: Cambridge University Press.
- Moen DS, Morlon H, Wiens JJ, 2016. Testing convergence versus history: convergence dominates phenotypic evolution over 150 million years in frogs. *Syst Biol* 65:146–160.
- Morlon H, Lewitus E, Condamine FL, Manceau M, Clavel J, et al., 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Meth Ecol Evol* 7:589–597.
- Niklas K, 1999. Evolutionary walks through a land plant morphospace. *J Exp Bot* 50:39–52.
- Paradis E, Schliep K, 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528.
- Pennell M, Eastman J, Slater G, Brown J, Uyeda J et al., 2014. Geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30:2216–2218.
- Polly PD, Stayton CT, Dumont ER, Pierce SE, Rayfield EJ et al., 2016. Combining geometric morphometrics and finite element analysis with evolutionary modeling: towards a synthesis. *J Vert Paleont* 36:e111225.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Revell LJ, 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Meth Ecol Evol* 3:217–223.
- Sheftel H, Shoval O, Mayo A, Alon U, 2013. The geometry of the Pareto front in biological phenotype space. *Ecol Evol* 3:1471–1483.
- Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O et al., 2012. Evolutionary trade-offs, Pareto optimality and the geometry of phenotype space. *Science* 336:1157–1160.
- Simpson GG, 1953. *The Major Features of Evolution*. New York, NY: Columbia University Press.
- Slater GJ, Friscia AR, 2019. Hierarchy in adaptive radiation: a case study using the Carnivora (Mammalia). *Evolution* 73:524–539.
- Sidlauskas B, 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. *Evolution* 62:3135–3156.
- Stayton CT, 2008. Is convergence surprising? An examination of the frequency of convergence in simulated datasets. *J Theor Biol* 252:1–14.
- Stayton CT, 2015. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution* 69:2140–2153.
- Stayton CT, 2019a. Performance in three shell functions predicts the phenotypic distribution of hard-shelled turtles. *Evolution* 73:720–734.
- Stayton CT, 2019b. Performance surface analysis identifies consistent functional patterns across 10 morphologically diverse terrestrial turtle lineages. *Integr Comp Biol* 59:346–357.
- Stayton CT, O'Connor LF, Nisivoccia NM, 2018. The influence of multiple functional demands on morphological diversification: a test on turtle shells. *Evolution* 72:1933–1949.
- Tendler A, Mayo A, Alon U, 2015. Evolutionary tradeoffs, Pareto optimality and the morphology of ammonite shells. *BMC Syst Biol* 9:12.
- Tseng ZJ, 2013. Testing adaptive hypotheses of convergence with functional landscapes: a case study of bone-cracking hypercarnivores. *PLoS One* 8:e65305.
- Wise TB, Stayton CT, 2017. Side-necked versus hidden-necked: a comparison of shell morphology between pleurodiran and cryptodiran turtles. *Herpetologica* 73:18–29.
- Wright S, 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.
- Zelditch ML, Ye J, Mitchell JS, Swiderski DL, 2017. Rare ecomorphological convergence on a complex adaptive landscape: body size and diet mediate evolution of jaw shape in squirrels (Sciuridae). *Evolution* 71:633–649.