Abstract: Evolution is a fundamentally population level process in which variation, drift and selection produce both temporal and spatial patterns of change. Statistical model fitting is now commonly used to estimate which kind of evolutionary process best explains patterns of change through time using models like Brownian motion, stabilizing selection (Ornstein–Uhlenbeck) and directional selection on traits measured from stratigraphic sequences or on phylogenetic trees. But these models assume that the traits possessed by a species are homogeneous. Spatial processes such as dispersal, gene flow and geographical range changes can produce patterns of trait evolution that do not fit the expectations of standard models, even when evolution at the local-population level is governed by drift or a typical OU model of selection. The basic properties of population level processes (variation, drift, selection and population size) are reviewed and the relationship between their spatial and temporal dynamics is discussed. Typical evolutionary models used in palaeontology incorporate the temporal component of these dynamics, but not the spatial. Range expansions and contractions introduce rate variability into drift processes, range expansion under a drift model can drive directional change in trait evolution, and spatial selection gradients can create spatial variation in traits that can produce long-term directional trends and punctuation events depending on the balance between selection strength, gene flow, extirpation probability and model of speciation. Using computational modelling that spatial processes can create evolutionary outcomes that depart from basic population-level notions from these standard macroevolutionary models.

Key words: Brownian motion, evolutionary model fitting, evolutionary rate, metapopulation, trait evolution.

Evolution is not only a temporal process, but a spatial one. Just as species change through time, they vary in space. Their geographical ranges encompass assorted environmental conditions and harbour considerable genetic and phenotypic variation (Fig. 1). But spatial variation is more than just simple population variation, it is a gradient of differences between local populations or demes that arises from metapopulation population processes related to distance, barriers, gene flow, drift and selection. Spatial variation interacts with evolutionary processes in ways that produce patterns of trait change that differ markedly from the expectations of standard statistical evolutionary models like Brownian motion, directional selection and adaptive peak (Ornstein–Uhlenbeck) models. In this paper, I critically review the ways in which spatial processes cause departures from these standard evolutionary models. My aim is to better inform interpretations drawn from evolutionary model fitting. No existing statistical models fully capture the ways in which spatial processes modify standard evolutionary models, so no remedies are offered herein; however, the signatures of spatial processes can be recognized in some cases. This paper aims to inform the reader what those signatures might look like in palaeontological data.

Ever since Darwin (1859), variation has been understood to be a central component of evolution. Evolution in its quantitative formulation consists of change in a trait’s mean value as a result of the intergenerational sorting effects of drift and selection on the heritable variation within a population (Wright 1929; Fisher 1930; Lande 1976). The Modern Synthesis recognized that rates and directions of evolution are not simply consequences of selection, but are constrained by the size of a population, the amount of variation, and the rate at which variation is replenished by mutation (Fisher 1930; Haldane 1932; Dobzhansky 1937; Huxley 1942). Simpson’s (1944) adaptive landscapes with their centrifugal (stabilizing), centripetal (diversifying) and linear (directional) selection

© The Palaeontological Association
doi: 10.1111/pala.12410
regimes and their horotelic (normal), bradytelic (slow combined with depauperate diversity) and quantum (rapid adaptive peak jumping) rates were models based on principles of within-population variation to explain patterns of evolution in the fossil record. Evolutionary biologists consequently have made important strides in documenting within-species variation (Darwin 1868; Bateson 1894; Kinsey 1929; Mayr 1963; Berry 1977; Yablokov 1977, 1986).

Two very different concepts of within-species variation exist, however: unstructured variation and spatially structured variation. Unstructured variation is the classic model of Fisher (1930) in which a species consists of a single large, panmictic population with a spatially uniform gene pool. In this model, the average values and range of variation in traits are the same at all geographical locations. Selection and drift act on this species-wide population from one generation to the next to change the average trait value, a change that is again the same for the entire species. As described below, most of today’s statistical models of trait evolution are based on this ‘Fisherian’ view.

Spatially structured variation, the alternative, is a concept in which a species is a metapopulation composed of interacting local populations (or demes), each with their own trait means and local pool of variation (Levins 1968; Hanski 1999). In the metapopulation concept local populations may vary in population size and may occupy different habitats. Drift and habitat-specific selection tend to cause local populations to differentiate from one another thus producing spatially structured variation, processes that are counteracted by dispersal and gene flow that tend to make the local populations more alike (Levins 1968; Endler 1977). Geographical connectivity, dispersal ability, differences in local population fitness, and differences in interactions with other species govern the dynamics of metapopulations (Holt 1985; Hanski 1999; Thompson 2005). Waxing and waning of the geographic range of a metapopulation species due to environmental change, changes in resource availability, competition, or other factors, will modify the evolutionary dynamics of the metapopulation by altering the spatial distribution of its local populations, population sizes, selection gradients,
and patterns of connectivity (Brown 1996; Hewitt 1996; Gaston 2003; Stuart et al. 2004). Population subdivision thus allows drift and local selection to ‘experiment’ with new varieties in local populations while maintaining the overall viability of the species through dispersal and gene flow. The process of speciation itself also affects evolutionary processes in geographically heterogeneous metapopulation because daughter species are likely to differ in their trait means and ranges of variation because each is composed of only a subset of the total geographical range of variation in the parent species (consider the morphological difference if, for example, the British populations of shrew in Fig. 1B speciate) whereas daughter species from a large, unstructured panmictic population will be essentially identical at the time of speciation (Mayr 1942, 1963; Bush 1975).

In his ‘shifting balance theory’, Wright (1931, 1969) argued that without this kind of metapopulation experimentation a species cannot move from one adaptive peak to another. These factors all make evolution in a spatially variable metapopulation much more complex than in the panmictic population concept. The evolutionary outcomes of panmictic and metapopulation views are so great that they can be distinguished as ‘Fisharian’ and ‘Wrightian’ respectively (Wade & Goodnight 1998; Wade 2016; Wagner 2017).

The statistical models that are used today to infer evolutionary processes from palaeontological data are based on fundamentally ‘Fishierian’ principles that are ultimately derived from Simpson’s adaptive peak metaphor. Regardless of whether the models are applied to phylogenetic trees (Felsenstein 1985, 1988; McPeek 1995; Martins & Hansen 1997; Blomberg et al. 2003; Butler & King 2004; O’Meara et al. 2006; Harmon et al. 2008; Revell 2008; Revell et al. 2008, 2012; Harmon et al. 2010; Slater et al. 2012; Slater et al. 2015) or stratigraphic sequences of palaeontological samples (Bookstein 1987; Gingerich 1993; Roopnarine 2001; Polly 2004; Hunt 2006, 2007, 2012; Hannisdal 2007) they produce estimates of rates, directions and modes of evolution based on equations that treat each species as a single population whose size remains stochastically constant. While individual models vary, they are broadly derived from quantitative genetic equations that are used to predict the magnitude and direction of trait evolution.

The model of Lande (1976), among the first to evaluate the contributions of selection and drift to trait evolution in the fossil record, illustrates the basic principles that underpin most of today’s statistical models of evolution (Fig. 2). Lande developed probability models for predicting the patterns of evolutionary change that would be observed depending on whether selection or drift dominated, including the different expectations for directional vs stabilizing selection. He drew on Wright’s and Simpson’s adaptive landscape concept to model selection, which Lande implemented as a normally distributed function (\(W\) in Fig. 2A) for the fitness of different trait values (\(\ln|W|\)) relative to a selective optimum (\(\theta\)) represented by the peak of the normal curve (Lande 1976; Felsenstein 1988; Arnold et al. 2001). The contribution of selection to change in the average value of a trait (\(\Delta Z\)) in Lande’s model is the product of the selection coefficient \(b\) and the trait’s genetic variance, or \(\beta G\) (Fig. 2A). The selection coefficient is given by the slope of the fitness curve (\(\partial \ln|W|/\partial Z\)), which has a sign that always carries \(Z\) towards the peak where \(b\) becomes zero (Fig. 2A). Selection is stronger the farther the population mean is from the optimum, so narrow peaks keep the trait close to the optimum and broad ones allow the trait to vary widely. \(G\) is the trait’s additive genetic variance (heritable variance), which is the subset phenotypic variance (\(P\)) that comes from offspring resembling their parents. The proportion of \(P\) that is genetic is referred to as the trait’s heritability (\(h^2\)). Note that \(P\) can be estimated directly from measurements taken from a fossil sample and \(h^2\) is a proportion that ranges from 0 to 1, which allows one to estimate a range of credible values for \(G\). Drift also contributes to \(\Delta Z\) (Fig. 2A). Drift is a process that arises from stochastic sampling error in the trait’s mean from one generation to the next. Specifically, the expectation of the magnitude of change due to drift is \(\sqrt{G/N}\) where \(N\) is population size. Drift is equally likely to be in a negative or a positive direction, which makes it a type of Brownian motion or random walk process, the long term outcomes of which are well understood from a statistical perspective (Felsenstein 1988; Berg 1993).

Most of today’s statistical evolutionary models, especially those based on Brownian motion (random walk) or adaptive peaks (Ornstein–Uhlenbeck or OU processes), are grounded in the same principles as Lande’s (Felsenstein 1988; Polly 2004; O’Meara 2012; Pennell & Harmon 2013) and are thus ‘Fishierian’ in concept because they are based on expectations of how a population with a single mean will behave in macroevolutionary timescales under drift or selection. This is illustrated by looking at how a simple Brownian motion model is fit to a stratigraphic sequence of samples (Hunt 2006). Brownian motion is a pure random walk in which the direction of trait change at any step (generation) is independent of previous changes and has equal probability of being negative or positive (Felsenstein 1985, 1988; Bookstein 1987). When a Brownian motion process is repeated over many lineages, it produces a predictable pattern of outcomes in which the trait values at the ends of the lineages are, on average, equal to the ancestral value and their variance equals the variance of the changes at each step (\(\sigma^2\), which is the rate parameter) times the number of steps \(\sqrt{\sigma^2}\) (Fig. 2B). The fit of the changes observed in real palaeontological data with respect to those expected from a Brownian motion process is evaluated by fitting a
normal distribution with a mean ($\mu$) of zero (which is appropriate for Brownian motion because positive and negative changes are equally likely) and an unknown variance (i.e. rate $\sigma^2$) that is estimated from the trait data. Maximum likelihood algorithms like that of Hunt (2006) yield a likelihood value for all estimates of $\sigma^2$, including the one that maximizes the likelihood and thus gives the best estimate for the evolutionary rate given the data and the Brownian motion model.

Other evolutionary models, such as directional or stabilizing selection (the latter including stasis), can be fit using a similar likelihood algorithm and an appropriate set of parameters. Stabilizing selection, for example, is an OU adaptive peak model with parameters for the location of the optimum ($\theta$) and width of the peak ($\omega$) (Fig. 2A, B). Using the Akaike Information Criterion (AIC) to adjust for the differences in parameters, the relative strength of fit can be used to select the model that best describes the observed

FIG. 2. Statistical evolutionary models. A, characteristics of Lande’s (1976) adaptive peak model for selection and drift; see text for details. B, simulations of evolution in a stratigraphic lineage, each for 100 steps under Brownian motion, directional, and stabilizing models of evolution; each model was simulated 500 times (grey lines) with one run highlighted (black line) to illustrate an actual outcome against the statistical expectation of the model. C, an example of a multiple-peak model of evolution that shifts from stabilizing to directional selection as lineages move from one peak to another; see Polly (2018b) for an animated version of the multiple-peak model.
data (Hunt, 2006). Such procedures are frequently used to infer the mode of evolution in a lineage or clade (e.g. whether the evolutionary process had a random, directional, stabilizing, or static pattern), or to reconstruct ancestral trait values (Martins & Hansen 1997; Polly 2004; Revell et al. 2008; Pennell & Harmon 2013; Slater & Harmon 2013; Pennell et al. 2014; Hunt & Rabosky 2014). While statistical models of evolution have many variants, such as multiple rate and multiple peak models (Fig. 2C; Butler & King 2004; O’Meara et al. 2006; Hunt 2008; Hunt et al. 2015) they are all ‘Fisherian’ in that the statistical outcomes are derived from expectations of how evolutionary processes affect the trait mean of a single population.

In this paper, I will present examples of how spatially explicit ‘Wrightian’ processes can cause long-term patterns of trait evolution to depart from those expected under ‘Fisherian’ models. For example, pure drift processes combined with range expansions can produce species-level patterns that resemble punctuated equilibrium models, and stabilizing selection processes in which the optimum varies geographically can produce patterns that resemble directional selection. The evolutionary effects of spatial dynamics will be illustrated using computational modelling of trait change in geographically explicit settings over geological time scales. Computational modelling allows control of the parameters for population size, selection, dispersal probability, gene flow and probability of extinction in the local populations of a species (metapopulation) over thousands or millions of generations, either as a single unbranching lineage or as a branching phylogenetic tree (Polly 2004; Polly et al. 2016). We will first look at how drift processes that are normally associated with Brownian motion can produce punctuated bursts of evolutionary change driven solely by changes in the geographical range of the evolving species. We will then look at how spatially structured selection along environmental gradients can produce patterns of trait evolution that mimic directional selection, stasis and punctuated equilibrium. I will apply Hunt’s (2006, 2007, 2008) likelihood method for evolutionary model selection to estimate what modes of evolution we might have inferred if we were ignorant of the role spatial processes played in generating the simulated data. Finally, I will review two examples of real data from the fossil record where spatial processes might logically have played a role and discuss how ‘Fisherian’ and ‘Wrightian’ explanations of the patterns might differ.

BROWNIAN MOTION IN A CYCLICAL ENVIRONMENT

Brownian motion is arguably the simplest and most predictable statistical model for quantitative trait evolution (Felsenstein 1985, 1988; Bookstein 1987; Roopnarine 2001; Hunt 2006). As described above, Brownian motion is a pure random walk process in which each step (ΔZ) is drawn from a random distribution that is symmetrical around a mean (µ) zero with a variance or step rate (σ²), which produces outcomes that are normally distributed with a mean equal to the ancestral value and a variance equal to the tσ² (Fig. 2B; Felsenstein 1988; Hunt 2006; Revell et al. 2008). Pure genetic drift is a random sampling process where σ² is equal to G/N (Fig. 2A; Lande 1976). Note that pure drift can be distinguished in palaeontological data from selectional drift, which is a Brownian motion process involving stochastically fluctuating selection (Kimura 1954; Felsenstein 1988; Revell & Harmon 2008), by measuring the phenotypic variance P and combining it with credible estimates of h² and N to determine whether the estimate of σ² is substantially greater than expected (Polly 2004; Wójcik et al. 2006; Polly et al. 2013). Note also that the long-term expectations of drift and selection assume that heritable variance is replenished. If rate of change exceeds the heritable variance available in the population, then the rate becomes limited by the accumulation of new mutations that contribute to genetic variance (Kimura 1965; Turelli & Barton 1990). In such cases, long-term divergence will be slower than expected based on P, h² and N alone, but divergence under a pure genetic drift process should not exceed the expectation of Lande’s equation.

Genetic drift in a ‘Wrightian’ spatial context behaves more like a variable-rate model of evolution than like typical Brownian motion when it is accompanied by geographical range changes. It is probably safe to say that geographical range changes have been the norm for most species, because of post-speciation expansion from the founder populations and because of changes in habitat availability driven by Earth systems processes like sea-level changes, glacial-interglacial cycles, or long-term climatic trends (Hewitt 2004; Stigall 2013; Maguire et al. 2015; Patzkowsky 2017). Range-expansion processes create ‘non-random’ patterns of traits among the local populations and range contractions create a disconnection between drift in local populations (which have sizes N) and the behaviour of the trait mean of the total species metapopulation (which has size N = ∑Ni).

Computational model

The impact of geographical range changes on drift is illustrated here using a computational model in which an island metapopulation expands and contracts in response to sea-level cycles (Fig. 3). This model was set on an island platform that was entirely exposed during lowstands with three peaks 2 m above platform height that remained exposed as isolated islands during highstands
(Fig. 3A). The entire island platform was gridded into 5000 cells (50 × 100), each of which could potentially be occupied by a local population.

Eustasy was modelled as a sine wave through 2.5 cycles using the equation $-5 \sin(0.005\pi x - \pi) - 4$, where $x$ is time measured in model steps (Fig. 3B). This equation causes sea-level to start 3 m below platform height (thus exposing the entire platform at the beginning of each model run), cresting at 2 m above platform height at highstands (thus inundating everything except the very peaks of the three islands) and dropping to 8 m below platform height during lowstands.

An evolving species was modelled through space and time using a metapopulation concept. At the beginning of each run, a single founder population was randomly placed on the platform with a local population size ($N$) of 10, a trait value of 0, a phenotypic variance ($P$) of 0.002, and a heritability ($h^2$) of 0.5. Recalling that the rate of genetic drift is $h^2/P/N$, drift in local populations (demes) was modelled as a random change drawn from a normal distribution with a mean of zero and a variance of 0.001. The heritability value was chosen realistic for morphological traits (Cheverud & Buikstra 1982) but the other two values were arbitrarily chosen (see discussion below about consequences of higher or lower rates of drift). At each model step each local population: (1) reproduced a new generation in its own cell, undergoing a drift event at the same time; (2) had a chance of expanding its range by giving rise to a new local population in any or all of the adjacent cells ($P = 0.8$ for each cell); and becoming locally extinct ($P = 0.2$). Each new population, like its parent, had $N = 10$, $P = 0.2$, $h^2 = 0.5$, and a trait mean equal to the parent’s after a new drift event.

If a population expanded into an already occupied cell, the two populations were merged through reproduction by averaging their phenotypes and resetting $N = 10$. The averaging of traits in cohabiting populations creates gene flow since a phenotype in one area can spread via dispersal and reproduction to another. If the cell was covered by water the probability of extirpation rose to 1.0.

The geographical range of the species expands when populations move into empty cells and contracts when populations are extirpated. Sea-level thus changes the geographical range and the number of extant local populations. At lowstands the species tends to grow to approximately 5000 local populations that cover the entire platform (Fig. 2C). But in fact, the patterns are erratic behaviour, the tempo and direction of which changes in phases along with the rise and fall of sea-level (Fig. 3D). The trait values shown in Figure 3D are the average over all the local populations and thus represent the average trait value of the species as a whole. Because it is an emergent property of the underlying metapopulation processes of local drift, gene flow and changes in the geographical range, the species mean behaves differently than we would expect under ordinary Brownian motion. It fluctuates wildly during highstands (i), tends to go through rapid directional shifts during regressive expansion phases (ii), and has long periods of what looks like stasis during lowstands (iii). While the variance between model runs generally increases through time, as one would expect of a normal Brownian motion process (Fig. 2B), it does not do so with the regular predictability of ordinary Brownian motion.

Visually, the patterns produced by the spatial model appear to be like those produced by a multiple adaptive peak model involving shifts between stabilizing and directional selection when sea level rises above or falls below the platform level (Fig. 2C). But in fact, the patterns are produced by a multiple-rate random walk. Application of Hunt’s complex model of trait change (Hunt & Rabosky 2014; Hunt et al. 2015) to the model outputs demonstrates that this is so. The simple version of Hunt’s models uses maximum-likelihood and the parameter-corrected Akaike Information Criterion (AICc) to estimate the rate and mode of evolution from a temporal series of samples, such as those modelled here (Hunt 2006, 2007). The modes used here were unbiased random walks (i.e. ordinary Brownian motion), directional random walks, and stasis (i.e. stabilizing selection or OU process). Hunt’s algorithm estimates which evolutionary model best fits empirical data using the AICc parameter-adjusted maximum likelihood. In the complex application of the models, the data sequence is subdivided into segments, each of which potentially can evolve under a different rate and mode or have a different adaptive peak optimum (Hunt et al. 2015). The number of model parameters increases with the number of segments, so the AICc criterion is again used to choose between simpler models in which the sequence is generated by a process with a single rate model runs can be found in Polly (2018b, S2 (data), S8 (raw output)). Models were run on Indiana University’s Karst high-throughput computer cluster.

Results and discussion

The outcomes of this model, which is a pure drift process albeit with a spatial component, differ considerably from the standard expectations of Brownian motion. The trait mean, when measured as a species-wide average, has an erratic behaviour, the tempo and direction of which
and mode and more complex models where the process varies between segments of the lineage.

I used Hunt’s strategy to test for differences in rate and mode associated with three phases of the sea level cycle: the high-stand phase (i) when populations are isolated on three small islands; the regressive phase (ii) when sea-level is retreating across the platform and populations are expanding; and the low-stand phase (iii) when sea-level is below the platform edge and the geographic area is stable (Fig. 2). Fitting nine alternative evolutionary models of increasing complexity (Table 1) found that the one that best fits these data is a series of six unbiased random
walks (one per sea-level phase in the computational model), each with a different rate parameter (including separate rates for the four different expansion phases). This multiple-rate random walk model not only fit the run highlighted in Figure 3D (Table 1) but all nine runs (Polly 2018b, S4). In a few cases an individual segment of a model run fit a directional or stasis model better than an unbiased random walk. Of the 81 individual segments found collectively in the computational model runs (nine segments (i–iii) in nine runs), only three supported a stasis model and 15 supported a directional model (Polly 2018b, S4). The overall support for a multi-rate random walk process in these models makes sense, because computation model is driven by pure drift at the local population level that produces emergent but related patterns when measured as change in the species as a whole. Two spatial processes in particular influence the species-level pattern.

The first process that makes these spatial models depart from ordinary random walk is that the overall population size of the species changes during the course of the runs, as does the total amount of trait variance. As described above, genetic drift is a sampling phenomenon in which the mean trait value of an offspring generation varies from its parent generation because of randomness in which individuals reproduce (Wright 1969; Lande 1976). For a single population, the rate of drift (\(h^2 \cdot P/N\)) is the evolutionary rate \(\sigma^2\) and the Brownian motion process involves repeated resamplings from a normal distribution with \(\mu = 0\) and \(\sigma^2 = h^2 \cdot P/N\), in other words a single rate random walk model. However, the total population size of the species as a whole \(N_t = \sum N_i\), where \(N_i\) is size of the ith local population (\(N_i = 10\) for all local populations in these models). At hightimestamp \(N_t\) drops to about 120, but at lowstands it may approach 50 000 when the species covers the entire platform. Similarly, the trait means of the local populations vary such that that variance of the metapopulation (i.e. the species as a whole) is different from the variance within a given local population and is a function of the rates of drift in the local populations (which causes them to differentiate, thus increasing the overall variance) vs gene flow between them (which homogenizes them, thus decreasing the total variance) both relative to the rate at which the geographical range expands and contracts. The rate of drift at the population level \(\sigma_{ip}^2\) was set at \(1.0 \times 10^{-5}\), which is about four orders of magnitude higher than the rate for the species as a whole at hightstands (\(\sigma_t^2 = 3.20 \times 10^{-7}\) in the model highlighted in Fig. 3D), five orders of magnitude higher than during expansion phases (\(\sigma_{ii}^2 = 1.44 \times 10^{-8}\), \(\sigma_{ii}^2 = 3.17 \times 10^{-9}\), \(\sigma_{ii}^2 = 1.68 \times 10^{-8}\), \(\sigma_{id}^2 = 7.04 \times 10^{-8}\)) and six orders of magnitude higher than at lowstands (\(\sigma_{ii}^2 = 4.14 \times 10^{-9}\)). While the rate of drift at the species level is a complex function of the rates of drift, gene flow and geographical range area, \(\sigma_{ip}^2\) will not be constant if \(N_t\) and the overall trait variance are not.

The second spatial process that makes trait evolution depart from ordinary Brownian motion is non-random resampling of peripheral populations during range expansion. Because range expansion happens by multiplication of local populations at the edge of the range, the idiosyncratic variants in those populations can be multiplied across the newly occupied area. This process can produce the patchworks of spatial differentiation that are seen in Figure 3C (and even more clearly in the animations found in Polly (2018b)). The effect is even more pronounced if the range contracts into geographically isolated refugia, like the three islands in the computational model, where they will differentiate by drift during the period of separation. Formerly minor variants can quickly become common during such expansion events, rapidly

### Table 1. Results of evolutionary model fitting for the computational model run highlighted in Figure 3D.

<table>
<thead>
<tr>
<th>Model no.</th>
<th>Model type</th>
<th>(\mu_{step})</th>
<th>(\sigma_{step}^2)</th>
<th>(\theta)</th>
<th>(\omega)</th>
<th>(K)</th>
<th>(l)</th>
<th>AICc</th>
<th>Akaike weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Random walk</td>
<td>–</td>
<td>Same</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td>6523.8, –13045.5</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>Directional</td>
<td>Same</td>
<td>Same</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>6523.8, –13043.6</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>Stasis</td>
<td>–</td>
<td>Same</td>
<td>Same</td>
<td>Same</td>
<td>–</td>
<td>2</td>
<td>4452.9, –8901.8</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>Random walk</td>
<td>–</td>
<td>Diff (×3)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3</td>
<td>7393.5, –13041.5</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>Directional</td>
<td>Diff (×3)</td>
<td>Diff (×3)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>6</td>
<td>7389.5, –14766.9</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>Stasis</td>
<td>–</td>
<td>Diff (×3)</td>
<td>Diff (×3)</td>
<td>–</td>
<td>6</td>
<td>2299.7, –4587.4</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Random walk</td>
<td>Diff (×6)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>6</td>
<td>12203.0, –24393.9</td>
<td>1.000</td>
</tr>
<tr>
<td>8</td>
<td>Directional</td>
<td>Diff (×6)</td>
<td>Diff (×6)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>12</td>
<td>7472.9, –14921.4</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>Stasis</td>
<td>–</td>
<td>Diff (×6)</td>
<td>Diff (×6)</td>
<td>–</td>
<td>12</td>
<td>5119.6, –10214.9</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Models 1–3 constrain the entire lineage to have the same rate (\(\mu_{step}\), \(\sigma_{step}^2\)) or adaptive peak (\(\theta\), \(\omega\)) parameters, models 4–6 allowed independent rate or adaptive peak parameters for each of the three sea-level phases of the computational model (i–iii) and models 7–9 allowed independent parameters for each phase, including separate parameters for the four individual expansion phases (ii). \(K\) is the number of parameters in the evolutionary model, \(l\) is the log likelihood of the model, AICc is the Akaike adjusted likelihood, and Akaike Weights is the relative support for each model. Bold indicates the preferred model. See Polly (2018b) for results from all nine runs of the computational model.
changing the trait mean of the species as a whole as seen by the large, seemingly directional trait changes during phase ii intervals (Fig. 3D). The effects of this process have been extensively documented in spatial patterning of molecular markers, notably in post-glacial recolonization events of extant terrestrial plant and animal species (Hewitt 1996, 1999; Avise et al. 1998; Magri 2008; Dapporto 2010), but also in marine species (Colson & Hughes 2007; Maggs et al. 2008). The genetic mechanism by which a rare allele becomes common as a result of range expansion, a specialized kind of founder effect (Ibrahim et al. 1996; Hewitt 2000; Excoffier et al. 2009; Slatkin & Excoffier 2012), has been referred to as genetic ‘surfing’ because a random gene rides the expanding wave of the range (Excoffier & Ray 2008).

Both processes are likely to affect the evolution of phenotypic traits as measured in the fossil record. While sampling is often coarse compared to the extant world, geographical range changes not only logically occurred in the fossil record but they have been demonstrated in many cases (Hellberg et al. 2001; Maguire & Stigall 2008; Blois & Hadly 2009; Stigall 2010; Feurdean et al. 2013). Refugial expansion and cyclic changes in the geographical ranges of species in response to Quaternary climate cycles are especially well documented (Roy et al. 1995; Lyons 2003; Stuart et al. 2004; Stewart 2009). Analyses of geographical variation in bones and teeth in extant and fossil species indicates that the ‘surfing’ phenomenon also in morphological traits in much the same way as it does in genetic markers (Polly 2001, 2007; Szuma 2007; Polly et al. 2013). Spatially modified pure drift processes could therefore drive seemingly non-random trait changes in palaeontological sequences, at least in principle, a possibility that should be considered among other alternative models of evolution.

Note that the computational algorithm and the parameters used in this paper were specifically chosen to illustrate the potential for spatial processes to produce geographic differentiation and evolutionary rate shifts. For a wide-ranging evaluation of how the balance between drift, migration, local extinction and geographical ranges affects differentiation within a species see Slatkin (1977), Wade & McCauley (1988), Lande (1992), Ibrahim et al. (1996) and Whitlock (1999).

**SELECTION IN SPATIALLY HETEROGENEOUS ENVIRONMENTS**

Selection and speciation in a spatially heterogeneous environment can also cause notable departures from standard evolutionary models, even in unchanging environments. Characteristics that affect the performance of an organism with respect to its environment are known as functional traits. The morphology of teeth (King et al. 2005), the strength of crocodilian skulls (Rayfield 2011) or the shape of angiosperm leaves (Nicotra et al. 2011) are examples of traits that affect the performance and therefore the relative fitness of an organism in different environments (Arnold 1983). Directional and stabilizing selection (OU) models of evolution are often used to test for this kind of trait-environment relationship, but they operate in a ‘ Fisherian’ manner that assumes that there is a single optimal value for the entire species (or more precisely, that \( \Delta Z \) for the species is responding as though it were a single population with a single \( G \) value experiencing the same selection gradient \( \beta \), and thus responding to the same adaptive peak \( W \)).

Even though directional selection causes change in a trait and stabilizing selection prevents it from changing, the two modes of evolution are both produced by standard adaptive peak (OU) models like Lande’s (Fig. 2) and thus deserve comment. An OU model produces directional change when traits are far away from the optimum but results in stabilizing selection or stasis when traits are near the optimum (Butler & King 2004). Palaeontologists in particular are concerned with the difference between directional selection and stasis, in part because of interest in testing Eldredge & Gould’s (1972) punctuated equilibrium model of evolution, and therefore tend to use two separate statistical models for the two modes in order to distinguish between them. Hunt (2006, 2007, 2008), for example, represented directional selection with a generalized random walk model (GRW) in which the selectional events are drawn from a normal distribution with a non-zero mean (\( \mu_{\text{rep}} \)) that produces a directional bias because the selection events are, on average, in the same positive or negative direction. In contrast, Hunt’s unbiased random walk model (URW) describes Brownian motion because its events are described by a normal distribution with a mean of zero, which represents an equal chance of moving in a positive or negative direction. Note that Hunt’s directional model describes a directional trend that has no reference to an optimum and could, in principle, continue infinitely, but it adequately describes the expectation for a directional trend through a stratigraphical sequence of finite duration. To represent stasis or stabilizing selection, Hunt (2006, 2007, 2008) used a simplified OU model that followed Sheets & Mitchell’s (2001) ‘white noise’ model with only two parameters for a trait optimum \( \theta \) and adaptive peak width \( \phi \) (cf. Fig. 2) to which the \( \Delta Z \) of a real data set is fit, thus capturing only the stabilizing part of the process without any directional component. Hunt’s models are used below because they have the power to distinguish directional trends from static ones, but readers should note their differences from other OU implementations (Hansen 1997; Butler & King 2004) that have directional...
components when the trait is far from the optimum, stabilizing components when it is near the optimum, and Brownian motion components when it is at the optimum.

A species evolving in a spatial mosaic of different selective environments does not respond to selection in the ‘Fisherian’ manner just described. If the range of the species extends across a gradient of environments, its constituent local populations will experience functional selection toward different local optima that will tend to produce geographical variation in traits (Fig. 1). Selective differentiation among local populations is partially counterbalanced by gene flow from other parts of the species range. If speciation occurs by vicariance in a geographically differentiated species, the offspring species will be different from one another. This effect is most extreme in speciation by peripheral isolation, which is considered to be the most common mechanism (Mayr 1942; Coyne & Orr 2004) because the peripheral founder population is likely to have a trait value that is a substantial outlier from the mean of the parent populations, thus producing a punctuated jump between parent and offspring (Eldredge & Gould 1972). Expansion and contraction of the species’ range adds to the ‘Wrightian’ effects because it alters the number of local populations exposed to a particular local selective optimum and thus changes the evolutionary trajectory of the trait’s mean at the species level. The evolutionary outcomes of a spatially distributed species thus depend on the balance between local selection, gene flow, range changes and speciation (Endler 1977; Slatkin 1977; Wade & McCauley 1988; Lande 1992; Frey 1993).

**Computational model**

The impact of spatial selection gradients and speciation are illustrated using runs from the computational model described by Polly et al. (2016). The model was set in a mid-latitude island continent with a north–south trending mountain ranges, latitudinal and altitudinal temperature gradients, a rain shadow effect east of the mountains, and four different vegetation biomes (Fig. 4A). Like with the genetic drift model above, the model space was gridded into approximately 25 x 40 cells (with irregular continental margins, the exact number of cells was 822).

A univariate trait (tooth crown height) was simulated in a clade (mammalian herbivores) whose fitness depends on the match between tooth type and local conditions: high crown teeth are favoured in arid, gritty, grassy environments (trait value = 3) and low crown teeth in moist forested environments (trait value = 1). The optimal value of the trait in each cell is a function the local combination of climatic and environmental conditions (Fig. 4B).

As with the genetic drift computational model above, each species consisted of a set of local populations that experienced drift ($N = 100$, $P = 0.05$, and $r^2 = 0.5$, except where otherwise stated), had a chance of dispersing into adjacent cells ($P = 0.5$), experienced gene flow when two or more conspecific local populations occupied the same grid cell, and had a chance of becoming locally extinct (function of the distance between the local optimum and the trait value in the local population). Unlike the genetic drift model, the local populations were also under selection using Lande’s adaptive peak model (described above) where the optimum was a function of the local conditions (Fig. 4B) and the adaptive peak width (strength of selection) was set at 0.5 (except as otherwise noted). This model also included four speciation events in which each extant species gave rise to a new species arbitrarily every 100 models steps to yield a total of 30 species (two species in steps 1–100, four in steps 101–200, eight in steps 201–300, and 16 in steps 301–400) except where lineages may have become prematurely extinct. Speciation occurred by peripheral isolation in which the population that was most distant from the geographical centre of the species’ range became the founder of the new species.

Runs lasted 400 steps, with 100 steps between speciation events (including one at the first step). Note that with 822 grid cells and up to 16 contemporaneous species, there can be more than 13 000 evolutionary events per step at the local population level late in each run. The computational model runs discussed here (Fig. 4C–F) are a subset of experiments in which key parameters (strength of selection, phenotypic variance, extirpation probability and ancestral starting location) were systematically varied (see Polly et al. 2016 for a full description). The runs were selected because they illustrate the range of effects that spatial processes can have on trait evolution.

This computational model is fully described in Polly et al. (2016) and its Mathematica code and outputs are archived with that paper. The runs were performed on the Karst high-throughput computing cluster at Indiana University.

**Results and discussion**

The outcomes of this computational model, which used simple adaptive peak processes at the level of local populations, produces a wide variety of patterns in trait evolution at the species level. The outcomes of four runs are shown in Figure 4C–F, chosen because they illustrate the range of patterns that can arise from the spatial interactions of selection, gene flow, extirpation and speciation. Each run was part of a larger computational experiment in which most parameters
were held constant at a mid-level value and one was systematically varied from low to high values (Polly et al. 2016).

**Weak selection.** The run in Figure 4C was from an experiment in which the intensity of selection (i.e. the width of the adaptive peak) was varied. This run had a weak selection (i.e. an adaptive peak that was 40 times wider than the phenotypic variance) which lessened the effects of selection relative to gene flow and drift and lessened the probability of extirpation. The lineages appeared to undergo a form of parallelism in which they all were steadily selected toward higher trait values. Speciation consistently produced small punctuation toward lower trait values in this run. This pattern emerged because each species was able to expand its range across the entire continent (i.e. into all available environments) due to the weakening of the extirpation effect (see Polly et al. 2016 for spatially explicit maps for all the runs discussed here). Furthermore, gene flow dominated local selection making the trait values fairly homogeneous over all local populations. Each species was selected toward a trait value of 1.08, which is the net mean of all the local selective optima on the continent (i.e. the average of the values shown in Fig. 4B). Punctuation events occurred because the populations on the periphery of each species range received less input from gene flow than those at the centre, therefore the peripheral founders of each new species were different from the mean of their parent species. Speciation always produced a consistently lower trait value because each species was spread across the entire continent so that the peripheral populations was consistently located in local environments that selected for low trait values (Fig. 3B).
To show how this pattern might be interpreted using standard evolutionary model fitting, Hunt's (2006) method was applied to a path through the tree from tip to ancestor (highlighted in Fig. 4C). This path is analogous to the kind of real palaeontological data that might be assembled for a stratigraphic sequence of fossils when the record is sparse enough for speciation events to go unnoticed. The data shown here best support a Brownian motion model (URW) with reasonable alternative support for a directional model (GRW) (Table 2). Even though the highlighted path has a strongly directional trend, the punctuation event at step 200 is inconsistent enough with the GRW model to tip the support toward Brownian motion. The punctuation events can be ignored by separately analysing its 30 individual branches (16 tip branches plus 14 internal ones). These paths represent the ideal palaeontological data in which only a single species lineage is included. All 30 branches support a directional model (GRW; Table 2). The data generated by this run thus appear to support Brownian motion or directional models of evolution, even though they are produced by what are fundamentally OU processes. In real life such processes could play out over any number of temporal and spatial scales, some of which would be resolvable in the fossil record and some of which would not.

Strong extirpation. The run in Figure 4D shows what happens when the probability of local extirpation is high. All of the branches were selected toward a low trait value because the species were unable to expand out of their ancestral environments (which happened to favour low trait values in this run). This outcome was produced by the same factors as in the weak selection run, except that the species was exposed to a narrower range of local conditions meaning that less change was needed to reach the net optimum. Selection intensity was also stronger in this run which helped each species to converge on the net optimum faster. The highlighted lineage (Fig. 4D) supported a Brownian motion model (URW) because the punctuation events were strongly inconsistent with directional or stasis models and branch-by-branch analysis favoured directional evolutionary models (GRW) for 29 out of 30 branches (Table 2).

High phenotypic variance. The run in Figure 4E had high phenotypic variance $P$ in the local populations, which produces a high rate of genetic drift and a strong response to selection equations (as explained above). With all other parameters set to mid-level values, the result was that all species were able to expand their ranges to the entire continent (as happened in the weak selection run) and had strongly differentiated local populations (in contrast to the weak selection run) because of the combined effects of high rates of drift and strong selection toward the local optimum. The geographical differentiation made the punctuation events larger than in the previous two runs because the founder populations were more differentiated. The run converged on the same net optimum trait value as in the weak selection run (1.08), but here each lineage was selected back to the net optimum more quickly in about 100 model steps. The combination of large punctuations at speciation and quick return to the net optimum produced a much different pattern of evolution than in the weak selection run. The highlighted lineage in this run also supported a Brownian motion model of evolution (URW), largely because of the punctuation events were inconsistent with the two alternative models, and branch-by-branch analysis supported directional selection (GRW) in 27 out of 28 species (Table 2).

Note that parameters used in this run have connections, albeit opposite ones, to Sheldon's (1996) plus ça change hypothesis of evolution. Sheldon argued that wildly fluctuating environments produce generalists whose traits do not track the changes but instead hover about a mean optimum. This implies that these generalists respond slowly to selective changes and that they are able to flourish in a wide variety of environments. In terms of the parameters used in this computational model, one way of getting a slow response to selection is by having a small phenotypic variance (or low heritability) and one way of surviving in a wide range of environments is to have a low probability of extirpation. See Polly et al. (2016) for a companion simulation that was parameterized with low phenotypic variance and is thus similar to the expectation of the plus ça change hypothesis.

Extreme ancestral environment. The most unique outcome of the four runs is one in which all parameters were set to mid-level values and the ancestral population was placed in an 'extreme' environment where the optimal trait value was much higher than the starting point of other runs (Fig. 4F). 'Extreme' here means 'uncommon' because the cells favouring high trait values are far less numerous than those favouring low values (Fig. 4B). As the run unfolded, the species repeatedly colonized regions that favoured lower trait values so the overall pattern of trait evolution appeared to be directional. However, the pattern of punctuation events and within-branch anagenetic changes were much more idiosyncratic than in the other runs. In this run there was a nearly equal balance between the strength of local selection, the rate of drift, the rate of gene flow, and the probability of extirpation which made chance events more important for determining the final range size of each species, and thus the number of local environments they occupied. Some species became specialists for a narrow range of environments and some because...
generalists over a wide range, resulting in a wider range of traits at the tips of the tree by step 400 than in the other runs. The trends in this run were strong enough that the highlighted lineage (Fig. 4F) favoured a directional evolutionary model (GRW) despite the punctuation event (Table 2). The branch-by-branch analysis supported directional evolution for 28 out of 29 species but note that the direction in this run was variable in contrast to the uniformly positive directions of the other runs. The individual branches in this run also had a stronger random component due to greater influence of random processes (drift and extirpation) compared to the smoothly deterministic branches seen in the other runs.

**General principles.** The departures of these computational model runs from the patterns of trait evolution expected from standard evolutionary models can be seen by comparing Figures 2B and 4C–F. Technically speaking, all of the runs in Figure 4C–F are governed by OU processes with a small component of Brownian motion drift. The departure from the standard evolutionary models is due to the spatial selection gradient and the differences between the runs is due to the balance between the strength of selection, the rate of gene flow, and the probability of extirpation. Generally speaking, a spatial gradient of selection will produce geographical variation within a species. The species-level trait mean is therefore an emergent property of the trait means in the local populations, unlike a ‘Fisherian’ species in which the mean is a property of a single pool of variation. In a ‘Wrightian’ context, trait evolution at the species-level therefore depends on whether a species has a range that extends into more than one selective environment (weaker selection and less probable extinction make this more likely), whether it is geographically differentiated (stronger local selection and slower gene flow make this more likely) and how rapidly it colonizes new selective environments (higher probability of extirpation and steeper environmental gradients make this a slower process). Species whose ranges encompass many environments and have high rates of gene flow tend to evolve toward a mean trait value that converges on the average of all the local optima in its range (but note that the fact that it converges specifically on the arithmetic mean is dependent on how gene flow was implemented in this computational model). If such species reach the net optimum slowly, as would happen if selection is weak or gene flow is slow, then trait evolution will take on a directional trend from its ancestral state towards the net value (Fig. 4C). However, if selection is strong and gene flow and dispersal happen quickly, then the species will quickly reach the net optimum, closer to the expectation of a standard OU model (Fig. 4E). Only when the drift component is more substantial than selection or when selection, gene flow and extirpation are balanced enough that chance events determine the outcome will a ‘Wrightian’ system on a selection gradient be truly consistent with Brownian motion (but see

| Table 2. Results of evolutionary model fitting for the computational model runs shown in Figure 4C–F. |
|-------------------------------|---------------------------------|---------------------------------|-------------------------------|-------------------------- |
| Best model | Weak selection | Strong extinction | High variance | Ancestral environment |
| N | URW | URW | URW | GRW |
| 404 | 404 | 404 | 404 |
| $\mu_{\text{step}}$ | 0.001 | 0.000 | 0.002 | $-0.003$ |
| $\sigma^2_{\text{step}}$ | 0.001 | 0.0001 | 0.002 | 0.001 |
| $\theta$ | 0.605 | 0.293 | 0.9 | 1.41 |
| $\omega$ | 0.019 | 0.015 | 0.082 | 0.262 |
| AICc URW | $-1848$ | $-2609.4$ | $-1280.1$ | $-1596.1$ |
| AICc GRW | $-1846.6$ | $-2607.5$ | $-1278.5$ | $-1597.8$ |
| AICc stasis | $-447.78$ | $-549.28$ | $137.36$ | 607.81 |
| wt URW | 0.67 | 0.73 | 0.69 | 0.29 |
| wt GRW | 0.33 | 0.28 | 0.31 | 0.71 |
| wt stasis | 0.00 | 0.00 | 0.00 | 0.00 |
| N branches URW | 30 | 29 | 27 | 28 |
| N branches GRW | 0 | 0 | 0 | 0 |
| N branches stasis | 0 | 0 | 0 | 0 |

For each of the four tip-to-ancestor lineages highlighted in Figure 4, the best estimates of the rate ($\sigma^2_{\text{step}}$), direction ($\mu_{\text{step}}$), and selection parameters ($\theta$, $\omega$) are reported along with the AICc values and associated Akaike weights that determine which evolutionary model is best supported by the data. The last three lines of each table report the number of branches (out of max 30 except where lineages became extinct) that support each evolution model of evolution when the test is repeated branch by branch. Abbreviations: URW, unbiased random walk; GRW, generalized random walk (directional selection); stasis, stabilizing selection (OU process).
the genetic drift section above for the behaviour of a ‘Wrightian’ system under Brownian motion).

The magnitude of punctuated events in this system is controlled by how much geographical variation is present in the parent species. In homogeneous species (ones with high gene flow or weak local selection) peripheral populations are like the parent, which produces little trait change at speciation (Fig. 4C, D). But in geographically differentiated species (strong local selection, rapid response to selection, or weak gene flow) peripheral populations are likely to be uniquely different and larger punctuation events are likely to occur (Fig. 4E). If punctuated speciation events are inadvertently included in an evolutionary model fitting analysis, which might happen if cladogenetic events are not recognized in a stratigraphic time series, the excursions they produce may result in poor fits with directional or stabilizing models and thus cause the data to support a Brownian motion process (Table 2).

**INTERPRETING REAL DATA**

The potential implications of spatial processes for evolutionary model fitting in palaeontology can be illustrated by examining alternate interpretations of two real-world data sequences. Figure 5 shows the change in area of the lower first molar as the natural log of length × width for two lineages of mammalian carnivores, *Viverravus laytoni*–*V. acutus* and *Didymictis proteus–D. protenus* (Polly 2018b, S5).

These data are from the late Paleocene and early Eocene sequences of the Bighorn and Clarks Fork basins in Wyoming and pass through the Paleocene–Eocene thermal maximum (PETM), which was a rise in global temperature of 5–8°C that lasted less than 20 kyr (Zachos et al. 2003, 2006). Profound biogeographical changes were associated with the PETM, including immigration of many new taxa into Europe and North America from some unknown, probably Asian source, as well as speciation events and evolutionary changes in body size and other traits (McInerney & Wing 2011; Hooker & Collinson 2012). Locally in the Bighorn and Clarks Fork basins there was a pronounced change in vegetation from broad-leaved tropical forests before and after the PETM to a more open vegetation characteristic of dry, tropical savannahs or shrublands (McInerney & Wing 2011). Many mammalian taxa in the Bighorn Basin sequence had much smaller body size during the Wa0 subage of the Wasatchian North American Land Mammal Age that corresponds with the PETM (Gingerich 2006). The response to the PETM involved a complex mixture of geographical range changes, evolutionary change, and community reorganization (Clyde & Gingerich 1998; Rankin et al. 2015).

Both of these carnivoran lineages have unknown geographical range sizes and putatively contain a cryptic speciation event. Paleocene and Eocene *Viverravus* and *Didymictis* are primarily known from the Bighorn and nearby basins in western North America, but isolated occurrences are found elsewhere, including the late Paleocene of Mississippi (Beard & Dawson 2009) and post-PETM western Europe (Hooker 2010; Solé et al. 2016). Whether these extralimital occurrences belong to different species or are indicative of widespread but largely undocumented ranges of the Bighorn Basin species is poorly understood. Because good sampling of these taxa is geographically limited, we are forced to treat them in a ‘Fisherian’ way by presuming that the traits measured in Wyoming are characteristic of the average for the entire species. Polly (1997) interpreted each of these lineages as
representing two species (ancestor and descendant) separated by a cladogenetic event early in the sequence that gave rise to a second descendant species that is not shown here. However, previous and subsequent authors have interpreted these taxa as being divided into either more or fewer species, with and without cladogenic events (cf. Gingerich & Winkler 1985; Secord et al. 2006; Secord 2008). The uncertainty whether data represent a single evolving lineage or a more complex set of closely related taxa is typical of most stratigraphic sequences of palaeontological data (Benton & Pearson 2001). These data therefore make interesting examples for exploring ‘Fisherian’ and ‘Wrightian’ interpretations because either kind of process would be credible.

The molars of both lineages show a general trend toward becoming larger with seemingly erratic excursions around the time of the PETM, which is also about the time of the putative speciation events (Fig. 5). The data best support a Brownian motion model of evolution for Viverravus and an OU model for Didymictis (Table 3) but support for an alternative directional model is high in Viverravus (0.34) and support for an alternative Brownian motion model is low but credible for Didymictis (0.10). A purely ‘Fisherian’ face-value interpretation of these data might therefore be that, with Brownian motion in one taxon and stasis in the other, the PETM had little effect on the evolution of either lineage. But several features of these data sequences are similar to ones illustrated above. Both lineages seem to have a pronounced excursion in trait values at what might have been speciation events, which are followed by more gradual directional trends toward larger molar size. Given both the environmental perturbations and the range changes that have been documented at the PETM, these data are, in principle, consistent with punctuation events associated with speciation in two geographically widespread and differentiated species, followed by a directional trend driven by range expansion, either under a pure Brownian motion process like the ‘surfing’ effects seen in the computational model of drift (Fig. 3) or under an adaptive peak selection process accompanied by gene flow as seen in the computational model of environmental gradients (Fig. 4).

**CONCLUSIONS**

Spatial dynamics can produce unexpected departures from standard evolutionary statistical models, even when the evolutionary processes at the level of local populations are fully consistent with the same formulations of drift and adaptive peak selection that form the basis of the...
statistical models. Discrepancies arise because the expectations of virtually all evolutionary models make the ‘Fisherian’ assumption that drift and selection are acting uniformly on a species-wide trait distribution. But ‘Wrightian’ metapopulation processes involving local populations, drift, selection gradients, gene flow and range changes can produce trait patterns at the species level that are unexpected from the perspective of most evolutionary statistical models. The examples discussed in this paper are summarized in Table 4. Ignoring spatial processes may lead to misleading or incorrect interpretations when palaeontological data are assessed using standard evolutionary model fitting.

A ‘Wrightian’ system of trait evolution on a spatial selectional gradient is likely to have affected many clades that are preserved in the palaeontological record. Most of the objects preserved in the fossil record (teeth, shells, bones, leaves) play some sort of functional role in the life of the organism and are thus likely to be subject to selection. Whether the selection is weak or strong, directional or stabilizing, balanced by gene flow, or distributed uniformly across the range of a species is context dependent, but that it occurred can probably assumed for the majority of palaeontological traits. It is also likely that most species in the fossil record had geographical ranges that encompassed at least some variety of environments, if for no other reasons than commonness in the fossil record is associated with geographical range and abundance (Jernvall & Fortelius 2004), because geographical range is positively correlated with temporal duration (Jablonski 1987), and because geographical variation in traits is common in extant species with large ranges (Fig. 1). The ‘Wrightian’ processes discussed in this paper are therefore likely to influence species-level trait evolution as preserved in the fossil record and may cause departures from standard models that require some independent means of detection.

Pennell et al. (2015) demonstrated that standard statistical evolutionary models are often inadequate to describe real data. They assessed the adequacy of Brownian motion, single-peak OU, and early burst models to explain the evolution of more than 300 functional traits in angiosperms. Their analysis was based on phylogenies and traits in tip taxa rather than stratigraphic sequences, but their first two models are essentially the same as the ones considered here (the early burst model is not meaningful for a stratigraphic sequence). Only 133 of 337 traits were fully consistent with their best supported model, meaning that nearly two-thirds of the data sets were not adequately explained by any of these standard models. One reason for the inadequacy of standard models may well be that spatial dynamics produce evolutionary patterns that do not fit ‘Fisherian’ expectations, even when the processes of drift and selection at the local population-level are effectively the same as standard Brownian motion and OU models. Using an approach that compared data to models in both absolute and relative terms, Voje et al. (2018) found that more than one-third of the fossil data sets they examined were not adequately explained by the statistical models they evaluated.

Unfortunately, few tools are available to help us distinguish between the ‘Fisherian’ and ‘Wrightian’ interpretations of comparative data. Already, a new generation of evolutionary statistical models based on more complex scenarios of trait evolution is emerging, although none of them explicitly incorporate the spatial effects discussed in this paper. Most of these are implemented only for phylogenetic trees rather than palaeontological sequences. They include variable rate models (Mooers & Schluter 1999; O’Meara et al. 2006; Eastman et al. 2011; Uyeda et al. 2011; Smaers et al. 2016; Landis & Schraiber 2017) and multiple peak OU models (Hansen et al. 2008). Hunt and colleagues have done the most to elaborate model fitting techniques for stratigraphic sequences by allowing for hypotheses involving changes in tempo and mode (Hunt & Rabosky 2014; Hunt et al. 2015; see Table 1 for example).

Even though spatially explicit statistical models for trait evolution are not readily available, many palaeontologists have studied spatial processes in the fossil record. The mechanisms of punctuated equilibrium have frequently been analysed in a spatially explicit context, including speciation by peripheral isolation (Eldredge & Gould 1972), mechanisms for stasis (Lieberman & Dudgeon 1996) and related mechanisms for turnover in clades and communities (Vrba 1993; Lieberman et al. 2007). Spatial gradients of selection (Hoffmeister & Kowalewski 2001), phylogeographical turnover within species (Polly 2001), extinction (Foote et al. 2007; Jablonski 2008) and metacommunity spatial dynamics (Lyons 2003; Liow & Stenseth 2007; Blois & Hadly 2009; Stigall 2013) have been studied in the fossil record, as have contributions of spatial heterogeneity to speciation, trait distributions and biodiversity (Badgley & Finarelli 2013; Lyons & Smith 2013). The role of spatial processes in trait evolution is an active area that invites new methodological advances in modelling of trait evolution.

Acknowledgements. Thank you to Gene Hunt for verifying my Mathematica code for evolutionary model fitting against his original R code and for offering insightful comments about model fitting. Thanks also to Paul Smith and Mark Sutton for organizing the 2017 Palaeontological Association Symposium where this work was originally presented and to Sally Thomas and Andrew Smith for their patience and editorial help with this manuscript. I have benefited from discussions about spatial processes with many insightful people over the years, including Catherine Badgley, Tony Barnosky, Jessica Blois, Andrea Cardini, Pavel Borodin, Jussi Eronen, Mikael Fortelius, Marianne Fred, Islam Gündüz, Jeremy Hooker, David Jablonski, Christine Janis,
Jacques Hauser, Michelle Lawing, Adrian Lister, Richard Nichols, Svetlana Pavlova, Andrei Polyakov, Danielle Schrewe, Jeremy Searle, Nils Stenseth, John Stewart, Chris Stringer, Elwira Szuma, Marvalee Wake, and Jan Wójcik, as well as Graham Slater and two anonymous reviewers (none of whom are to be blamed for any shortcomings in this paper). This work was supported by US National Foundation Grant EAR-1338298. The use of the Karst high-throughput computing cluster was by the Indiana University Pervasive Technology Institute and the Indiana MACt Initiative, both of which received partial support from the Lilly Endowment, Inc.

DATA ARCHIVING STATEMENT

Data for this study are available in the Dryad Digital Repository: https://doi.org/10.5061/dryad.99pj691

Editor. Andrew Smith

REFERENCES


and WÓCICK, J. M. In press. Geometric morphometric tests for phenotypic divergence between chromosome races. In


