DR STEPHEN DE LISLE (Orcid ID : 0000-0001-9587-8665)

Extinction and the temporal distribution of macroevolutionary bursts

Stephen P. De Lisle[1]*, David Punzalan[2]*, Njal Rollinson[3,4], & Locke Rowe[3,5]

*Equal contribution

Corresponding author: Stephen P. De Lisle
Email: stephen.de_lisle@biol.lu.se
Phone: +46 70 892 8731

1.   Evolutionary Ecology Unit
        Department of Biology
        Lund University
        Sölvegatan 37, Lund Sweden 223 62

2.   Department of Biology
        University of Victoria
        Cunningham 202, 3800 Finnerty Rd
        Victoria, BC, Canada V8P 5C2

3.   Department of Ecology and Evolutionary Biology
        University of Toronto

25 Willcocks St. Toronto, Canada M5S 3B2

4.   School of the Environment

        University of Toronto

        33 Willcocks St. Toronto, Canada, M5S 3B8

5.   Swedish Collegium for Advanced Study,

        Thunbergsvägen 2, SE-752 38

        Uppsala, Sweden

RUNNING TITLE: Extinction resolves the stasis paradox

WORD COUNT ABSTRACT/TEXT: 186/5610

NUMBER OF REFERECES: 45

FIGURES/TABLES: 4/0

ABSTRACT

Phenotypic evolution through deep time is slower than expected from microevolutionary rates. This is the paradox of stasis. Previous models suggest stasis occurs because populations track adaptive peaks that remain relatively stable on million-year intervals, raising the equally perplexing question of why these large changes are so rare. Here, we consider the possibility that peaks can move more rapidly than populations can adapt, resulting in extinction. We model peak movement with explicit population dynamics, parameterized with published microevolutionary estimates. Allowing extinction greatly increases the parameter space of peak movements that yield the appearance of stasis observed in real data through deep time. Extreme peak displacements, regardless of their frequency, will rarely result in an equivalent degree of trait evolution because of extinction. Thus, larger peak displacements will rarely be inferred using trait data from extant species or observed in fossil records. Our work highlights population ecology as an important contributor to macroevolutionary dynamics, presenting an alternative perspective on the paradox of stasis, where apparent constraint on phenotypic evolution in deep time reflects our restricted view of the subset of earth's lineages that were fortunate enough to reside on relatively stable peaks.

INTRODUCTION

Phenotypic change can be rapid at microevolutionary timescales (Hendry & Kinnison 1999; Kinnison & Hendry 2001; Gotanda *et al.* 2015), consistent with strong selection (Endler 1987; Conner 2001) and abundant genetic variance (Mousseau & Roff 1987). Yet these microevolutionary processes do not appear to accrue to influence inferred evolutionary rates in deep time (Eldredge & Gould 1972). This apparent disparity in evolutionary rate across timescales is often referred to as the paradox of macroevolutionary stasis, where 'stasis' is relative; observed rates of evolution in deep time are low *relative* to expectations from microevolutionary theory and data. Recent analyses of an impressive dataset combining microevolutionary, fossil, and comparative data on vertebrate body size have recapitulated the problem of stasis (Gingerich 1993, 2001; Estes & Arnold 2007; Uyeda *et al.* 2011). These new analyses fit models of stabilizing selection towards shifting trait optima, and indicate that although evolution can be rapid for a handful of generations, this period of rapid change is followed by a protracted period of relatively slow evolution lasting for over one million years, prior to striking bursts of divergence in deeper time (Uyeda *et al.* 2011; Arnold 2014). These new results highlight a fundamental and largely unaddressed question in contemporary evolutionary biology: what processes may account for discrepancies in our inference of evolutionary dynamics between analyses of recent and deep time?

The pattern of limited phenotypic change accruing over long timescales followed by macroevolutionary bursts suggests that some process must constrain the accrual of phenotypic change within populations, as well as among closely related populations. Concomitantly, these results suggest that selection in the wild is net stabilizing, with large changes in the optimum phenotype occurring with exceptional rarity. Despite some indications of considerable temporal variation in the direction of selection (Siepielski et al. 2009, 2011), little is known about the behavior of adaptive peaks, and previous treatments have typically assumed relatively slow movement (Estes and Arnold 2007). Identifying why phenotypic change appears constrained over long timescales, and hence resolving the 'paradox of stasis', remains one of the most important problems to arise from the modern synthesis (Hansen & Houle 2004; Futuyma 2010; Arnold 2014;

Pujol *et al.* 2018; Voje *et al.* 2018), embodying the fundamental challenge of identifying the microevolutionary processes that can explain patterns of divergence at greater timescales.

Estes and Arnold (2007) analyzed a dataset (Gingerich 1993, 2001) of contemporary and historical rates of vertebrate microevolution, and found that the only models of phenotypic evolution that seem to explain observed body size evolution across timescales depended primarily on movement of an adaptive peak (or optimum), whereas standing genetic variance (i.e. trait heritability) had little explanatory power (Arnold 2014). Subsequently, Uyeda et al. (2011) performed a similar analysis after incorporating phylogenetic and additional data. Focusing on several models of peak movement, they concluded that the best-supported model is one of multiple peak displacements, where displacements of varying magnitude occur relatively infrequently (Uyeda *et al.* 2011). A recent study by Landis and Schraiber (2017) also supports models of episodic peak movement as the best model to explain patterns of vertebrate body size evolution. A problem that emerges when considering a model with peak movement is that, in real populations, displacements from the optimum phenotype will often be accompanied by reductions in mean absolute fitness and, thus, population size (Haldane 1937). In the extreme case, frequent or extreme displacements will result in an extinction event. Even under weak selection and modest but persistent peak movement, population size can fall below the replacement rate and the population will eventually go extinct (Lynch & Lande 1993; Bürger & Lynch 1995; Gomulkiewicz & Holt 1995). Although Uyeda et al. (2011) acknowledged that extinction could play a role generating the patterns observed in their data, they did not explicitly consider extinction in their likelihood-based approach to estimating model fits and parameters, or consider how extinction may influence estimates of peak displacements. A previous treatment (Estes & Arnold 2007) made similar restrictive assumptions regarding the magnitude of between-generation peak movement, on the grounds that displacements of excessively large magnitudes would result in the extinction of lineages. Yet the importance of extinction for patterns of phenotypic diversity are broadly recognized (Jablonski 2008), and moreover, the observation that the vast majority of biodiversity is now extinct (van Valen 1973) suggests extinction has played a critical role in shaping the patterns of diversity we see both today and in the past. This also highlights an important shortcoming of many studies attempting to draw conclusions regarding the behaviour of adaptive peaks, simply from phenotypic (both neontological and paleontological) divergence data; these provide an incomplete view, restricted to those instances in which a lineage was able to persist despite changes in selective regimes.

In the present study we attempt to reconcile patterns of morphological stasis with observed variation in selection in the wild, by incorporating microevolutionary processes and population dynamics into macroevolutionary models. Our goal was not to reanalyze the previous studies, critique any specific model of evolution, or fit models explicitly to empirical data. Rather our modeling approach attempts to highlight the effects of a general feature of biological populations—extinction—which has rarely been considered in models of phenotypic macroevolution (but see Bokma 2008 ; FitzJohn 2010). We focus on a subset of models of peak movement previously identified as the prime candidates for the observed data (Arnold 2014). We simulated anagenetic evolution using empirically-derived estimates of peak displacement, while simultaneously considering the effect of these displacements on the likelihood of extinction by tracking population dynamics explicitly. Although non-random extinction is known to influence our ability to interpret macroevolutionary models (e.g., when trait values correlate with extinction; Maddison 2006), and demography is known to impose constraint on microevolutionary change (Gomulkiewicz & Houle 2009), the potential effects of extinction on our ability to estimate the dynamics of the adaptive landscape are currently unclear. We show that 1) available estimates of phenotypic selection coefficients from temporally-replicated studies in the wild indicate that displacements of the optimum are far more commonplace and severe than inferred from previous analyses of phenotypic change, and 2) incorporating extinction explicitly into macroevolutionary models of peak movement indicate that lineage loss could contribute substantially to apparent patterns of morphological stasis. Put simply, extinction can lead to low probability that any large peak displacement can result fully in phenotypic change, even if such shifts are commonplace. Our work suggests that inference on the movement of adaptive peaks using observed phenotypic data alone fail to capture the fact that lineage loss may erase the history of rapid or severe peak displacements. Moreover, our work demonstrates that explicit integration of population ecology may shed light on patterns of phenotypic evolution in deep time.

METHODS

Basic approach to simulating phenotypic evolution and population size

We simulated the evolution of a quantitative trait in replicate populations, where populations were subject to several different scenarios determined by a range of evolutionary genetic parameters. We focus on anagenetic evolutionary change (our model contains no speciation process). Our simulations were demographically explicit, although for all parameter values we contrast two

categories of model: one where population size is permitted to decrease to extinction under maladaptation, and another where population size was 'rescued' artificially and kept above a threshold minimum size. This later treatment, although biologically unrealistic, allows us to contrast the effects of extinction across otherwise identically-parameterized simulations. We note this later treatment is analogous to typical approaches to modelling phenotypic macroevolution, which are typically agnostic to the demographic effects of maladaptation. We assumed phenotypic selection acting on a single, continuously distributed trait, z, with a population mean phenotype, $\bar{z}$, experiencing selection approximated by a Gaussian fitness function (roughly corresponding to the 'adaptive landscape'; Lande 1979; Arnold *et al.* 2001) with a width, ω, with the position of a single optimum located at θ, and described by:

$$W(z) = \exp\left[-(z-\theta)^2 / 2\omega^2\right] \tag{1}$$

and we can approximate evolutionary change on this fitness surface via

$$\Delta\bar{z} \approx h^2\sigma^2(\bar{z}-\theta)/(\sigma^2+\omega^2) \tag{2}$$

(see (Lande 1979; Estes & Arnold 2007)) where σ^2 equals the phenotypic variance. We simulated evolution of $\bar{z}$ for up to 10,000,000 generations while allowing θ to vary in position on a generation-by-generation basis (i.e. assuming discrete time with non-overlapping generations). The behavior of θ was governed by processes simulating either Brownian motion of the optimum, or peak displacement with (potentially) multiple bursts (see Uyeda *et al.* 2011 and below). In addition to tracking phenotypic evolution, at every generation, t, we allowed population size (N) to change according to the average fitness of the population, $\overline{W}_t$, which depends on the phenotypic distribution relative to the optimum following

$$\overline{W}_t = W_{max} \exp\left[-d^2 / 2(\sigma^2+\omega^2)\right] \tag{3}$$

where W_{max} corresponds to maximum absolute fitness, σ^2 equals the phenotypic variance, and $d^2 = (\bar{z}-\theta)^2$ (Gomulkiewicz & Holt 1995, Eqn. 6; Estes & Arnold 2007, Eqn. 2) is the distance a population is from its optimum trait value. This yields density-independent population growth,

$$N_{t+1} = N_t\overline{W}_t \tag{4}$$

In our simulations of density-independent population growth, we capped population size at 10 million to prevent geometric expansion to numbers unrealistic for most metazoans.

To model the more realistic scenario of density-dependent population growth associated with phenotypic evolution, we assumed (logistic) population growth, r, is a function of the maximum growth rate $\ln(W_{max})$ and the distance to the new optimum, described by:

$$r = \ln(W_{max}) * (1 - N/K) - (d^2 + \sigma^2)/2(\omega^2 + \sigma^2) \qquad (5)$$,

which can be used to describe demographic change in our model by

$$N_{t+1} = N_t \exp(r) \qquad (6)$$

after Lynch and Lande (1993, Eqn 2), and incorporating load on population growth introduced by phenotypic variance (σ^2) (see Kirkpatrick and Barton (1997, Eqn. 7)).

We assumed that, under a given selection scenario (i.e. for all replicates for a set of parameters, for the duration of the simulated interval) the shape of the fitness function remained constant. In the models considered, the response to selection depended on the available genetic variance (expressed as heritability), h^2.

We focused on two classes of models of peak movement: one that invokes movement of the optimum at relatively constant rate: the Brownian motion of the optimum (Felsenstein 1988) and another that envisions more sporadic movement: the peak displacement or 'burst' model (sensu Estes & Arnold 2007; Uyeda et al. 2011). To estimate the impact of extinction, we contrasted each model of peak movement when extinction was allowed and when extinction was prevented. In our simulations, extinction was prevented by simply re-seeding the population with some minimum N if the population size fell to zero, or below some critical threshold. Because the multiple burst model has been a focus of past work and because similar conclusions of extinction effects are reached under a Brownian motion model, we present the Brownian motion results in the supplemental material.

Although past workers have made use of explicit likelihood functions for alternative models of peak movement that allow comparisons with observed data (Estes & Arnold 2007; Uyeda et al. 2011), we aim to explore the qualitative effects of peak movement on population dynamics to examine different scenarios that succeed or fail to generate the characteristic pattern of relative stasis, followed by bursts of change (i.e. qualitatively resembling the 'blunderbuss' pattern observed by Uyeda et al. 2011). In the models of Brownian motion of the optimum, at each time step ($t + 1$), the position of θ was equal to its position at t, plus a deviation with expected mean = 0 and variance = σ_θ^2. Brownian motion models with (BME) and without extinction (BM)

differed only in that the latter allowed the populations to be rescued whenever the population fell below $N = 50$.

To model a scenario where the optimum experiences displacement less frequently than at every generation, the value of θ at a given time step was determined by its previous position but potentially also by a displacement that occurred with some probability. As in Uyeda et al. (2011), the probability of a displacement event occurring was modeled as a Poisson process, determined by the parameter, λ, or the average expected number of displacement events per generation. In instances where a displacement event occurred, its magnitude was drawn from a normal distribution with mean = 0 and variance = σ_θ^2. As above, displaced optimum models with (DOE) and without extinction (DO) were distinguished by allowing the latter populations to be rescued whenever $N < 50$. Note that both the probability of a peak displacement λ and variance of peak displacements σ_θ^2 together determine the total rate of peak movement in this model. The distribution of peak movements after a given amount of time is $N(0, m\sigma_\theta^2)$, where m is the expected number of peak displacements occurring during a time period (Uyeda *et al.* 2011). As such, peak displacements that are large but extremely rare result in a slow average rate of peak movement through time, while frequent displacements of moderate magnitude would result in a higher average rate of peak movement.

Model parameters

Whenever possible and appropriate, we used parameter values that are similar or the same as in past work (Estes & Arnold 2007; Uyeda *et al.* 2011). One major difference, however, is that we used a dataset of phenotypic selection gradients (Lande and Arnold, 1983) in temporally replicated studies compiled by Siepielski et al. (2011) to calculate the empirical distribution of displacement distances. For all studies in the database that recorded both standardized linear, β, and negative nonlinear (corresponding to approximately stabilizing) selection gradients, γ, we calculated the distance to the optimum within each selective episode as:

$$d = |\bar{z} - \theta| \approx |\beta / (-\gamma)| \tag{7}$$

(Phillips & Arnold 1989, Eqn 11; Estes & Arnold 2007, Eqn 7). By convention, gradients were normalized to a sample-specific mean of 0 and standard deviation of 1 (Lande & Arnold 1983). The standardized selection gradients allow estimation of the relative distance between $\bar{z}$ and θ

within a given i^{th} temporal replicate or 'episode', but provide no direct information on the actual values/coordinates of θ. Nonetheless, under the assumption that at the ith episode, $\bar{z}$ is closer to its optimum than it was at episode i - 1, the difference in absolute distances, Δd, between a given pair of successive replicates provides a proxy for the total displacement of the optimum between episodes. The variance of the distribution of Δd, σ_θ^2, was used to parameterize peak movement in the simulations, assuming a normal distribution of possible displacements as in the burst models of Uyeda et al. (2011). We relaxed this assumption of normality by also running the same models but with peak displacements drawn from the empirical estimates of Δd. Implicit in our use of the empirical data to approximate displacements of θ is that variation in selection reflects movement of the optimum, and not temporal variability in the phenotypic distribution. We recognize that the Siepielski et al. (Siepielski *et al.* 2009) dataset may reflect both sources of variation and results in unbiased, albeit imperfect, estimates of peak displacements. As a caveat, our approximation of Δd assumes a quadratic fitness surface, and does not explicitly incorporate measurement error (Morrissey & Hadfield 2012), which is unreported for many studies. Following these caveats and Estes and Arnold's (Estes & Arnold 2007) treatment of similar data, we excluded several observations corresponding to questionably large values of θ (i.e. > |35| standard deviations) that most likely reflect gradients associated with large estimation errors.

Each simulation began (i.e. at t = 0) with a population mean phenotype, $\bar{z}$ = 0, and phenotypic variance, σ_p^2 = 1. We arbitrarily set W_{max} = 1.5, which corresponds to rapid exponential population growth (Malthusian r = 0.41) for populations residing at the optimum. Variation in W_{max} had little effect on conclusions under density dependent population growth, as this growth rate is rarely experienced (e.g., a small population residing at its optimum phenotype), and also had little effect under density-independent growth (Figure S1). For each simulated population, carrying capacity (K) was randomly drawn from a range of $10561 < K < 12259$, based on estimates for vertebrates (see Table 2 in Reed *et al.* 2003), and each simulation began with a population at carrying capacity (i.e. $N = K$). Preliminary runs when carrying capacity was allowed to vary between values an order of magnitude lower or higher ($1000 < K < 120000$) generated plots virtually indistinguishable from those obtained using the empirically informed range, suggesting that variance in lineage-specific carrying capacity has little bearing on patterns of divergence.

For each population, heritability (h^2) was drawn from another empirical dataset on vertebrate body size (Hansen *et al.* 2011), and we assumed a constant h^2 throughout a given

simulation replicate. For all simulations, we assumed an adaptive landscape with a width of $\omega^2 =$ 3, comparable to median values from previous synthetic studies (Estes & Arnold 2007). Although we did not exhaustively explore the entire parameter space, the qualitative results of the simulations (i.e. the shape of the plots; see supplemental material) were generally robust to changes in parameters, though notable exceptions and their interpretation are addressed in the Results and Discussion. We note our scale of divergence (in units of phenotypic standard deviation) differs from the one used by Uyeda et al. (2011), who reported divergence in units of log-differences in body size. We chose a variance-standardized scale in order to relate our models to contemporary estimates of selection, which are reported in units of phenotypic standard deviations.

Obtaining an empirical benchmark for the frequency and magnitude of shifts in the optimum is difficult. Indeed, the very premise of this paper is that estimates of λ and σ_θ^2 from some empirical datasets suggest peak displacements are extremely rare, seemingly at odds with estimates of selection in the wild. Because the expected distribution of phenotypic divergence under a multiple burst model is N(0, $m\sigma_\theta^2$), where the number of bursts m in a given timescale t is given by $m = \sum_t \lambda$, λ and σ_θ jointly govern evolutionary rate in the model. We bracketed values of the rate parameters in our simulations based on two empirically-informed upper and lower bounds. For the lower bound, we assumed $\lambda = 10^{-7}$ and $\sigma_\theta = 3$, approximating the values estimated in other studies (Uyeda *et al.* 2011; Arnold 2014), which corresponds to an average waiting time of 10 million years for a modest shift in the optimum trait value. Note that because we are working in time units of generations, as opposed to years as in these past studies, our lower bound is conservative; most vertebrates have generation times far in excess of 1 year. Although no empirical benchmark for an upper bound for λ exists in the literature, an approximation can be obtained from our empirical sample of Δd, where values of $\Delta d > 3$ sds across a generation are observed in 48 out of 227 cases, indicating λ on the order of 10^{-1}. Because this upper bound is likely biased upwards due to sampling error and regardless is approaching a Brownian Motion model (see supplementary material) we capped our upper bound of λ at 10^{-4}. Finally, we note that although rate parameters determine the magnitude of the effect of extinction on patterns of divergence across lineages, for any specific lineage the magnitude of a peak shift (when it occurs) alone determines the severity of the bias arising from extinction (see results).

We compared the distributions (variances) of phenotypic divergence at the end of each simulation run using Levene's test for equality of variances, and we also plot the accrual of

phenotypic variance through time. For each model, we simulated 500 replicate lineages (100 for simulations run to 10^7 generations). Simulations were implemented using the R base package (R Core Team 2017), and examples of code for our 4 main classes of models (BM, BME, DO, DOE) are available in the Supplementary Material. The subsets of data analyzed to generate estimates of Δd and from which h^2 was sampled are available in the supplemental material.

RESULTS

Our analysis of the empirical dataset of temporally-replicated selection gradients resulted in 369 estimates of θ from 138 studies. From these we obtained 231 estimates (from 87 studies) of temporal displacements of the optimum, Δd. The distribution of displacements (around zero) of the optimum was roughly symmetric (Figure 1) around zero, with a Laplace distribution with a standard deviation of 12.98. Although the median magnitude ($|\Delta d|$) of 1.21σ (mean = 3.92σ, standard error = 0.82) is comparable to the estimates for θ in Estes and Arnold (2007), our empirical estimate of the standard deviation of Δd is far higher than estimates of the variation in Δd inferred in past empirical studies. Specifically, ML estimates of the magnitude of peak displacement from ref 8 correspond to a standard deviation of Δd of approximately 3 phenotypic standard deviations (Arnold 2014), an order of magnitude less than the empirically-observed value.

Next, we modeled discrete peak displacements that occur with varying frequency, from our lower empirically-motivated benchmark of infrequent and modest shifts (i.e. $\lambda = 10^{-7}$, $\sigma_\theta = 3$; see Uyeda et al. (2011), to three orders of magnitude higher ($\lambda = 10^{-4}$) and with magnitude proportional to the empirically-approximated distribution of Δd obtained above . The model yields a narrow band of initial divergence that becomes considerably more distinct when extinction is permitted (Figure 2, left sub-panels). When peak displacements are assumed to be rare and at our lower empirical benchmark (Fig. 2A), extinction has substantial consequences on the amount of divergence observed in deep time (Levene's test: $F_{1,198} = 11.44$, $P = 0.0008$, Fig. 2A right panel). Increasing the value of lambda by one order of magnitude while decreasing the timescale of simulation by two orders of magnitude further illustrates that rare (in this case, relative to the timescale of observation) shifts in the optimum can have striking effects on accumulation of phenotypic change (Levene's test: $F_{1,998} = 7.85$, $P = 0.005$, Fig. 2C). Further increasing the frequency of peak displacements has severe demographic consequences, in turn restricting early divergence (Fig. 2E-H), and illustrating the potential for extinction to generate a

pattern of apparent evolutionary constraint (all Levene's test: $P < 0.0001$). When we assume that peak displacements still occur sporadically but with even higher frequency ($\lambda = 10^{-4}$), some key insights are revealed. Extinction becomes critical to shaping the temporal pattern of phenotypic evolution, as divergence becomes otherwise unconstrained over longer timescales (Levene's test: $F_{1,998} = 737$, $P < 2.2*10^{-16}$, Fig. 2C, right panel). Shorter waiting times between displacements lead predictably to a shortening of the period of relative stasis, and a shorter average longevity of lineages (Figure 2E-G). More frequent displacement events also result in a coarse transition between the period of stasis to the eventual 'bursts' of evolution, implying that longer periods of favorable conditions, and therefore more stable population sizes, are more conducive to the formation of patterns to those that have been empirically observed. Noteworthy is that the assumed model of population growth, density dependent vs independent, has little effect on these observed patterns (cf Fig. 2A-C, Fig. 2D-F). Moreover, although the effects of extinction are most pronounced when peak displacements are frequent and large in magnitude, we observed statistically significant and striking effects of extinction even at our lower benchmark values of λ and σ_θ (Fig 2A; Fig. 3A, Fig4 C), and these effects persist unless population growth is unbounded towards exceptionally large population sizes (Figure 2B). The substantial effect of including extinction can be readily observed in the accrual of phenotypic variance through time, which is drastically reduced when extinction is permitted (Fig. 3A-D), even when peak displacements are exceptionally rare ($\lambda = 10^{-7}$) and the distribution of Δd is reduced to 3 phenotypic standard deviations (Figure 3A). These general effect of extinction on reducing the apparent evolutionary rate holds under a Brownian motion model of peak movement (See supplemental material).

These effects of extinction on curtailing the perceived rate of peak movement can be illustrated by considering short-term survival probability of lineages that experience a single displacement in their optimum phenotype. Displacements beyond approximately 2 within-population standard deviations are associated with reduced lineage survival over the span of 100 generations (Figure 4A). Lineages that fail to survive have also failed to reach their new optimum (Figure 4B), and so these severe displacements do not result in an equivalent degree of phenotypic change. The probability that a displacement will result in complete phenotypic change approaches zero for displacements beyond approximately 7 standard deviations. Note that these effects are expected regardless of the model of peak movement; a rarity of large peak displacements does not change the likelihood that large displacements lead to extinction when they do occur. Incorporating these survival probabilities into a displaced optimum model of peak movement

illustrates extinction's effect on expected wait times to observe a peak shift in trait data (Figure 4C).

Our conclusions from the multiple-burst model were essentially unchanged when re-running the model with displacements randomly drawn from the (exponential) empirical distribution of Δd instead of a normal distribution (Supplementary Material, Figure S3). Although the empirical estimates used to parameterize σ^2_θ may be biased upwards by sampling error, equivalent conclusions are obtained assuming a value σ^2_θ that is an order of magnitude lower than the observed value (Levene's test, $F_{1,998} = 78$, $P < 2.2 * 10^{-16}$), and the accrual of variance through time is significantly reduced by extinction at long timescales even when σ_θ is reduced to 3 standard deviations (Fig. S2). Only when decreasing σ^2_θ by over two orders of magnitude from the empirically observed value, to the order of a single phenotypic standard deviation, does the effect of extinction on macroevolutionary divergence disappear (Levene's test, $F_{1,998} = .98$, $P = .32$). Yet such a small value of σ^2_θ is inconsistent with the empirical observation that the divergence is confined to an interval that is several within-population phenotypic standard deviations wide (Uyeda *et al.* 2011; Arnold 2014). Thus, our conclusions regarding the importance of extinction in shaping macroevolutionary patterns appear robust to any potential biases in our empirical estimate of σ^2_θ .

DISCUSSION

Our analyses illustrate an important and previously underappreciated role for extinction and population dynamics in generating patterns of apparent constraint in macroevolution. We show that in the presence of rapidly moving optima, populations may face extinction while attempting to track their adaptive peak, preventing these rapid peak displacements from being recorded as phenotypic change. Large displacements in trait optima, regardless of their frequency, have exceedingly low probability of resulting in complete phenotypic evolution, thus limiting our ability to infer movements of peaks from trait data alone. As a specific example of extinction's effect on our ability to use trait data to infer movement of corresponding trait optima, we can imagine a clade where the average wait time for a shift in a lineage's optimum of 11 standard deviations is 100 generations. Although this would be expected to be borne out as a very high evolutionary rate, the reduced survival probability associated with such an event (p $\approx$ 0.0001; Figure 4A) leads to an approximate expected waiting time of one million generations before such an event is captured in trait data. This general effect is expected regardless of whether trait data

come from fossil lineages or extant time series. Although the effects of demographic constraint on microevolutionary change are well known (Lynch & Lande 1993; Gomulkiewicz & Houle 2009), our work illustrates how these constraints can manifest survivorship bias that leads to apparent patterns of stasis at the macroevolutionary scale. The observation of morphological stasis over macroevolutionary time may not reflect a lack of movement in phenotypic optima, but instead reflect selection bias in phenotypic timeseries generated by non-random extinction, leaving a subset of observable lineages fortunate to have resided on relatively stable optima.

We have focused on two simple models of peak movement, displaced optimum and Brownian motion. These are popular phenomenological models, co-opted from evolutionary genetics, that avoid the true complexities that must underlie the dynamics of actual adaptive surfaces. A salient conclusion from both models is that the greater the frequency and magnitude of peak shifts, the greater a role extinction will play in shaping patterns of phenotypic evolution across lineages; a conclusion expected to also apply to more complex characterizations of peak movement. Thus, observed patterns of phenotypic divergence in a variety of taxa (Estes & Arnold 2007; Uyeda *et al.* 2011; Arnold 2014) could be partly accounted for by an accrual of changes in many long-lived lineages fortunate to have experienced only small peak movements, combined with a subset of lineages that, due to abrupt displacements, went extinct before much phenotypic divergence had accumulated.

Using a dataset of temporally-replicated estimates of phenotypic selection from the wild, we show that the frequency and distribution of peak displacements in extant natural populations is inconsistent with the rarity and modesty of such events as inferred from analysis of observed macroevolutionary phenotypic change that ignore extinction, providing quantitative support to the idea that these past results are indeed paradoxical (as suggested by Uyeda *et al.* 2011). Concomitantly, such frequent and extreme shifts lead to predicted (in the absence of extinction) macroevolutionary rates that are far more extreme than those observed in empirical data. However, incorporating extinction into these macroevolutionary simulations results in temporal patterns of phenotypic divergence similar to empirical observations, and in all cases resulted in a decrease in the variance in phenotypic change observed in deep time (Fig. 2, right panels, Fig. 3). Thus, the catastrophic consequences of large peak displacements that were considered "biologically untenable" by Estes and Arnold (2007, p. 237) may, in fact, be the key to reconciling microevolutionary processes with macroevolutionary pattern. Our results suggest that non-random extinction plays a crucial role in resolving the stasis paradox; although we frequently

observe substantial peak displacements in extant populations (Feya *et al.* 2015), many of these populations would be expected to face extinction over longer timescales in the face of such maladaptation (e.g. Fig. 4), leaving a pattern of apparent morphological stasis over macroevolutionary time despite strong selection observed in extant populations. The relative insensitivity of our results to assumptions of heritability further suggest that extinction can play an important role in patterns of long-term phenotypic evolution, even when selected traits are evolvable.   Moreover, the apparent insensitivity to the form of population growth suggests that our conclusions hold for finite populations.

Extinction will contribute to the appearance of stasis whenever it scales with the frequency or magnitude of peak displacements. Compared to the demographic rescue model in which extinction does not occur, models allowing extinction consistently result in a delay and reduction in phenotypic divergence. This pattern appears to hold under a wide range of values of stabilizing selection ($1.5 > \omega^2 > 20$), though very strong curvature severely restricts evolution to a narrow range for the entire 'lifetime' of the lineage (Figure S5). Similarly, varying the range of heritability did not qualitatively change the importance of extinction. For example, the effect of varying heritability from $h^2 = 0.9$ to $h^2 = 0.1$ was, as expected, primarily to expand or reduce (respectively) the extent of divergence at all timescales (Figure S6). We also explored a "white noise" parameter in a subset of our simulations, in the same manner described by Estes and Arnold (Estes & Arnold 2007). This too had no impact on our conclusions, and omitting this parameter is conservative as it exacerbates the effects of extinction (Figure S7). Thus, a role for extinction in reducing apparent macroevolutionary rate do not appear to be qualitatively dependent on values of heritability, the curvature of the fitness surface, white noise, or even the macroevolutionary/phenomenological model of peak movement. The relatively small role of genetic parameters for explaining broad patterns is consistent with the findings of Estes and Arnold (Estes & Arnold 2007).

Our results are consistent with the observation that extinction is an important contributor to the distribution of Earth's diversity, as most of life that has existed has gone extinct, and suggests that considering extinction events may be critical to understanding shifts in phenotypic optima through the history of life. The extent to which ignoring extinction will impact estimates of peak movements is illustrated in the effects of extinction on the observed variance in phenotypic change, and the accrual of phenotypic change (Fig. 2, 4, left panels, Fig. 5). Because both Brownian Motion and Displaced Optimum models rely on the observed distribution of phenotypic

divergence in time to obtain estimates of peak movement, where the distribution is treated as N(0, $t\sigma_\theta^2$) (BM) or N(0, $m\sigma_\theta^2$) (DO), any effect of extinction on the realized distribution of phenotypic change, if left unaccounted for, will be borne out in parameter estimates for peak movement that are lower than reality.

Nonetheless, our results suggest the importance of extinction on the inference of peak displacements in data comprising large amalgamations of lineages will depend upon the value of σ_θ^2 and the frequency with which peak displacements occur, because these parameters determine the frequency of large peak displacements that will fail to be fully represented as phenotypic evolution. Yet even at our lower benchmark value of $\lambda = 10^{-7}$ and $\sigma_\theta = 3$, informed by past work (Uyeda *et al.* 2011; Arnold 2014) we find that extinction can play a striking role in limiting the ability of even this slow rate of peak movement to be fully reflected in phenotypic evolution. Thus, our results suggest that estimating rates of peak movement from trait data alone may be problematic at best, even if the true rate of peak movement is very slow relative to our expectations from contemporary microevolutionary studies.

A major challenge in reconciling micro- and macroevolution lies in the disparity between processes occurring in a single population versus the dynamics of the suite of interconnected populations constituting a species. Our simulations take the simplistic approach of characterizing evolutionary change assuming a single population per lineage, which has the benefit of allowing existing estimates of microevolutionary parameters (which are measured at the population level) to be incorporated (see also Arnold *et al* 2001, Arnold 2014, Estes & Arnold 2007, Uyeda *et al.* 2011). Yet, metapopulation structure across a species range likely has consequences for macroevolutionary dynamics. For example, gene flow among populations within a species can potentially act as a constraint on species-wide macroevolutionary change (Futuyma 1987, Eldridge *et al.* 2005). Further, how selection within populations manifests total selection across a species is unclear (Lieberman & Dudgeon 1996). Regardless, extinction due to the demographic effects of maladaptation would be expected to result in an apparent reduction of evolutionary rates.

Much of the phenotypic diversity that has ever emerged was probably quickly lost, and extinction has long been recognized as a potentially important force in phenotypic macroevolution; for example, explaining apparent disparities in speciation rates estimated for different timescales in (the 'Ephemeral Speciation Model'; Futuyma 1987; Rosenblum *et al.* 2012; Auilée *et al.* 2018). Our study complements this view, but with an important distinction: our models do not invoke speciation and only permit anagenetic change. While the Ephemeral

Speciation model (and other models of speciation-dependent trait evolution; Duchen *et al.* 2020) emphasizes divergence that was aborted, our models emphasize the importance of extinction for the divergence that cannot accrue in the first place, representing simply one example of the general challenge that extinction imposes for comparative biology. As our ability to infer microevolutionary fitness surface is limited to the distribution of phenotypes we observe, our understanding of the shape and dynamics of macroevolutionary adaptive landscapes is limited by a reliance upon those lineages that have not gone extinct. Thus, extinction poses a challenge for all comparative methods that exploit phenotypic variation to infer nuances of the adaptive landscape, or even estimate rates of phenotypic evolution. Although this issue of extinction has been made especially clear for evolutionary rates of discrete traits (Maddison 2006), it seems especially germane when the very goal of the evolutionary model is to make a statement on (mal)adaptation (Hansen 1997; Bartoszek *et al.* 2012) or the dynamics of macroevolutionary adaptive peaks (Uyeda & Harmon 2014). The only potential resolution to the challenge extinction poses to comparative biology is to incorporate, whenever possible (e.g., Maddison *et al.* 2007), the dynamics of lineage accumulation explicitly into macroevolutionary models of trait evolution. Our simulation models represent a first step in this effort to estimate peak movement. Future work further integrating selection, population and macroevolutionary dynamics may present an important way forwards in resolving the paradox of macroevolutionary stasis.

REFERENCES

Arnold, S.J. (2014). Phenotypic evolution: the ongoing synthesis. *The American Naturalist*, 183, 729-746.

Arnold, S.J., Pfrender, M.E. & Jones, A.G. (2001). The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica*, 112/113, 9-32.

Auilée, R., Gascuel, F., Lambert, A. & Ferriere, R. (2018). Clade diversification dynamics and the biotic and abiotic controls of speciation and extinction rates. *Nature Communications*, 9, 3013.

Bartoszek, K., Pienaar, J., Mostad, P., Andersson, S. & Hansen, T.F. (2012). A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology*, 314, 204-215.

Bokma, F. (2008). Detection of 'punctuated equilibrium' by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. *Evolution*, 62, 2718-2726.

Bürger, R. & Lynch, M. (1995). Evolution and extinction in a changing environment: a quantitative-genetic analysis. *Evolution*, 49, 151-163.

Conner, J.K. (2001). How strong is natural selection? *Trends in Ecology and Evolution*, 16, 215-217.

Duchen, P., Alfaro, M.L., Rolland, J., Salamin, N., Silvestro, D. 2020. On the effect of asymemetrical trait inheritance on models of trait evolution. *Systematic Biology* (in press)

Eldredge, N. & Gould, S.J. (1972). Punctuated equilibria: an alternative to phyletic gradualism. In: *Models in Paleobiology* (ed. Schopf, TJM). Freeman Cooper Sanfrancisco.

Eldredge, N., Thompson, J.N., Brakefield, P.M., Gavrilets, S., Jablonski, D., Jackson, J.B.C., Lenski, R.E., Lieberman, B.S., McPeek, M.A., & Miller, W. III. (2005). The dynamics of evolutionary stasis. Paleobiology 31:133-145

Endler, J.A. (1987). Predation, light intensity and courtship behavior in *Poecilia reticulata* (Pisces: Poecilidae). *Animal Behaviour*, 35, 1376-1385.

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

Felsenstein, J. (1988). Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, 19, 445-471.

Feya, S.B., Siepielski, A.M., Nussléd, S., Cervantes-Yoshidad, K., Hwand, J.L., Huberd, E.R. *et al.* (2015). Recent shifts in the occurrence, cause, and magnitude of animal mass mortality events. *Proceedings of the National Academy of Sciences*, 112, 1083-1088.

FitzJohn, R.G. (2010). Quantitative traits and diversification. *Systematic Biology*, 59, 619-633.

Futuyma, D.J. (1987). On the role of species in anagenesis. *American Naturalist*, 130, 465-473.

Futuyma, D.J. (2010). Evolutionary constraint and ecological consequences. *Evolution*, 64, 1865-1884.

Gingerich, P.D. (1993). Quantification and comparison of evolutionary rates. *American Journal of Science*, 293, 453-478.

Gingerich, P.D. (2001). Rates of evolution on the time scale of the evolutionary process. *Genetica*, 112, 127-144.

Gomulkiewicz, R. & Holt, R.D. (1995). When does evolution by natural selection prevent extinction? *Evolution*, 49, 201-207.

Gomulkiewicz, R. & Houle, D. (2009). Demographic and genetic constraints on evolution. *The American Naturalist*, 174, E218-229.

Gotanda, K.M., Correa, C., Turcotte, M.M., Rolshausen, G. & Hendry, A.P. (2015). Linking macro-trends and micro-rates: re-evaluating microevolutionary support for Cope's rule. 69, 1345-1354.

Haldane, J.B.S. (1937). The effect of variation on fitness. *The American Naturalist*, 71, 337-349.

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

Hansen, T.F. & Houle, D. (2004). Evolvability, stabilizing selection, and the problem of stasis. In: *The Evolutionary Biology of Complex Phenotypes* (eds. Pigliucci, M & Preston, K). Oxford University Press.

Hansen, T.F., Pélabon, C. & Houle, D. (2011). Heritability is not Evolvability. *Evolutionary Biology*, 38, 258-277.

Hendry, A.P. & Kinnison, M. (1999). Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution*, 53, 1637-1653.

Jablonski, D. (2008). Extinction and the spatial dynamics of biodiversity. *Proceedings of the National Academy of Sciences*, 105, 11528-11535.

Kinnison, M. & Hendry, A.P. (2001). The pace of modern life II: from rates of contemporary microevolution to pattern and process. *Genetica*, 112-113, 145-164.

Kirkpatrick, M. & Barton, N. (1997). Evolution of species' range. *The American Naturalist*, 150, 1-23.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, 33, 402-416.

Lande, R. & Arnold, S.J. (1983). The measurement of selection on correlated characters. *Evolution*, 37, 1210-1226.

Landis, M.J. & Schraiber, J.G. (2017). Pulsed evolution shaped modern vertebrate body sizes. *Proceedings of the National Academy of Sciences*, 12, 13224-13229.

Liebermann, B.S., Dudgeon, S. (1996). Evaluating stabalizing selection as a mechanism for stasis. Palaeogeography, Palaeoclimatology, Palaeoecology 127:229-238.

Lynch, M. & Lande, R. (1993). Evolution and extinction in response to environmental change. In: *Biotic Interactions and Global Change* (eds. Kareiva, P, Kingsolver, JG & Huey, R). Sinauer Assosciates Sunderland, MA.

Maddison, W.P. (2006). Confounding asymmetries in evolutionary diversification and character change. *Evolution*, 60, 1743-1746.

Maddison, W.P., Midford, P.E. & Otto, S.P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, 56, 701-710.

Morrissey, M.B. & Hadfield, J.D. (2012). Directional seleciton in temporally replicated studies is remarkably consistent. *Evolution*, 66, 435-442.

Mousseau, T.A. & Roff, D.A. (1987). Natural selection and the heretibility of fitness components. *Heredity*, 59, 181-197.

Phillips, P.C. & Arnold, S.J. (1989). Visualizing multivariate selection. *Evolution*, 43, 1209-1222.

Pujol, B., Blanchet, S., Charmantier, A., Danchin, E., Facon, B., Marrot, P. *et al.* (2018). The missing response to selection in the wild. *Trends in Ecology and Evolution*, 33, 337-346.

R Core Team (2017). R: A language and environment for statistical computing.

Reed, D.H., O'Grady, J.J., Brook, B.W., Ballous, J.B. & Frankham, R. (2003). Estimates of minimum viable population sizes for vertebrates and factors influencing those estimates. *Biological Conservation*, 113, 23-34.

Rosenblum, E.B., Sarver, B.A.J., Brown, J.W., Des Roches, S., Hardwick, K.M., Hether, T.D. *et al.* (2012). Goldilocks Meets Santa Rosalia: An Ephemeral Speciation Model Explains Patterns of Diversification Across Time Scales. *Evolutionary Biology*, 39, 255-261.

Siepielski, A.M., DiBattista, J.D. & Carlson, S.M. (2009). It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, 12, 1261-1276.

Siepielski, A.M., DiBattista, J.D., Evans, J.A. & Carlson, S.M. (2011). Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. *Proceedings of the Royal Society of London Series B.*, 278, 1572-1580.

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

Uyeda, J.C. & Harmon, L.J. (2014). A novel bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic Biology*, 63, 902-918.

van Valen, L. (1973). A new evolutionary law. *Evol. Theory*, 1, 1-30.

Voje, K.L., Starrfelt, J. & Liow, L.H. (2018). Model adequacy and microevolutionary explanations for stasis in the fossil record. *The American Naturalist*, 191, 509-523.
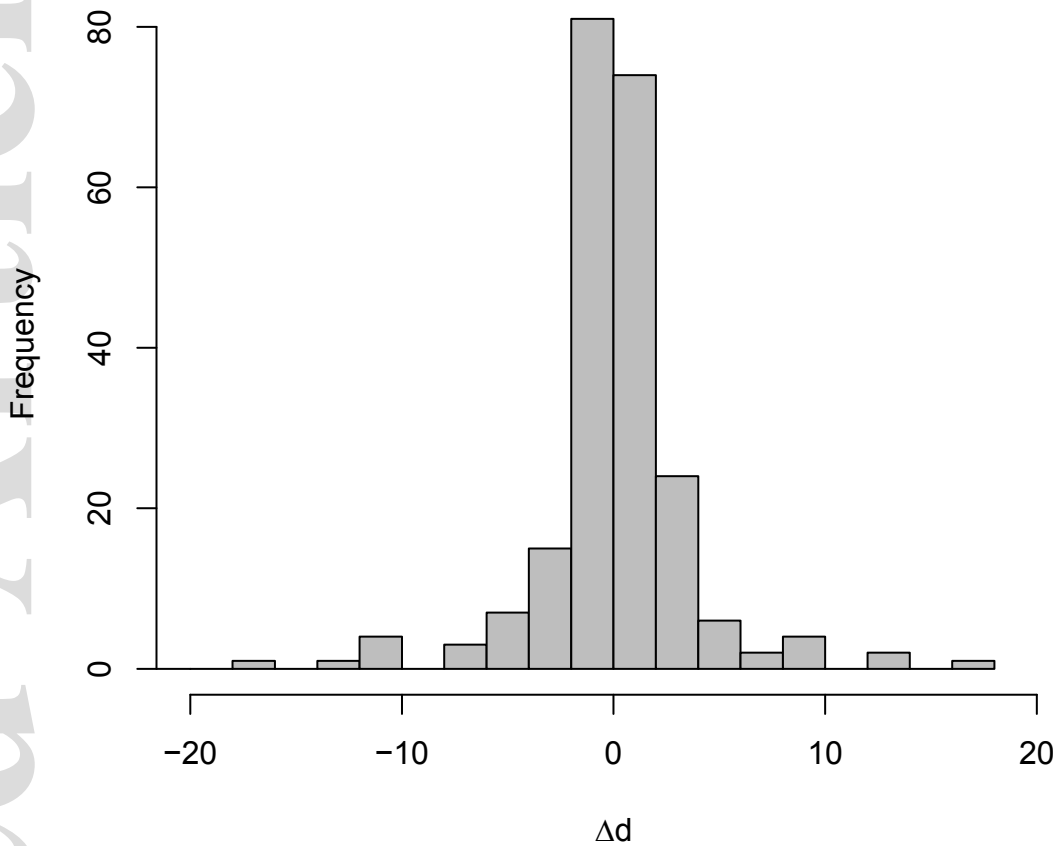
Figure 1. Histogram of estimated temporal (per-year) changes, Δd, in the displacement of the population mean phenotype from the phenotypic optimum $d = |\bar{z} - \theta| = |\beta / (-\gamma)|$. Thus Δd reflects, in part, among year changes in the optimum phenotype observed in extant wild populations. Based on the dataset of selection gradients from temporally replicated studies in Siepielski et al. (2009), restricting data to observed instances of negative nonlinear (stabilizing) selection. To aid in visualization, estimates of $|\Delta d| > 20$ are excluded.
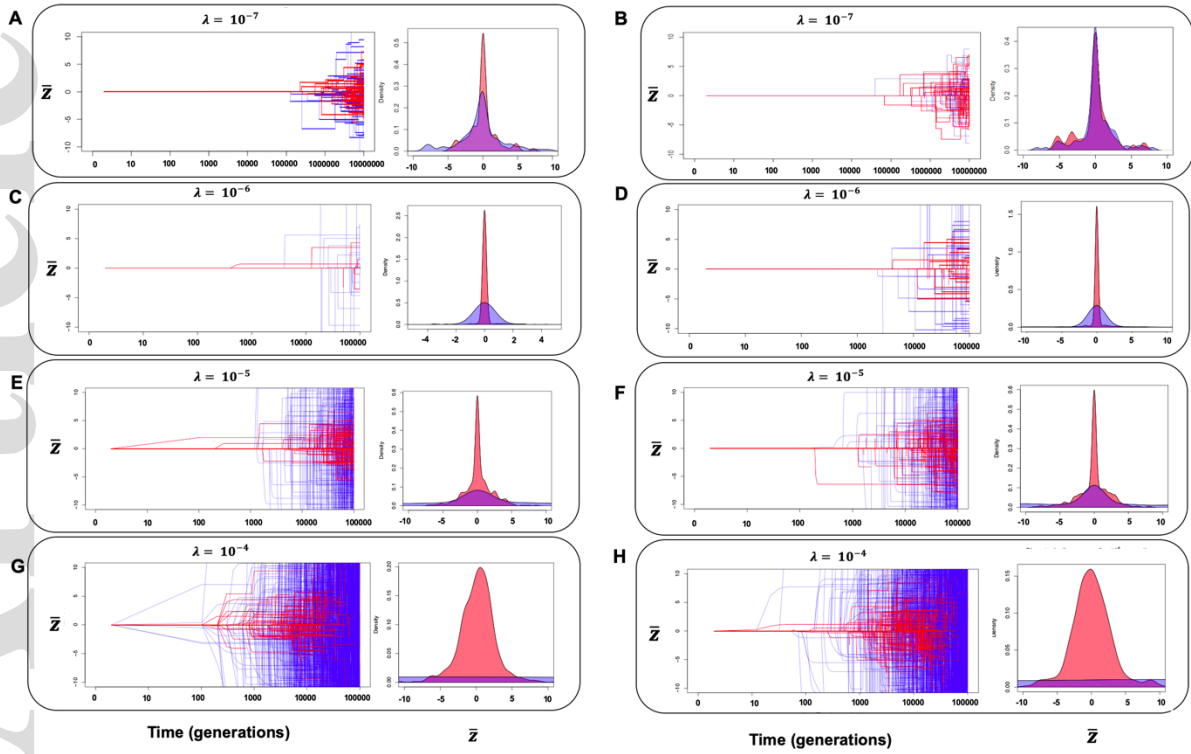
Figure 2. Displaced optimum models (DO, blue; and DOE, red), assuming $\sigma_\theta = 3$ (Panel A, B) or $\sigma_\theta = 13$ (the empirically observed value) at different assumed probabilities of a peak displacement, λ. Lambda represents the probability that the optimum shifts in a given generation, with $1/\lambda$ giving the expected wait time for an optimum shift in units of generations .In panels A, C, E and G population growth was assumed to be density-dependent, in panels B, D, F, and H population growth was assumed to be density-independent. DOE models allow the possibility of extinction for maladapted populations, while populations in DO simulations are 'rescued' from potential extinction (see text). Right sub-panels indicate the phenotypic distributions at the end of the simulations; either at extinction or 10^5 generations (10^7 for panels A and B). Note scale differences in x axes of right sub-panels.
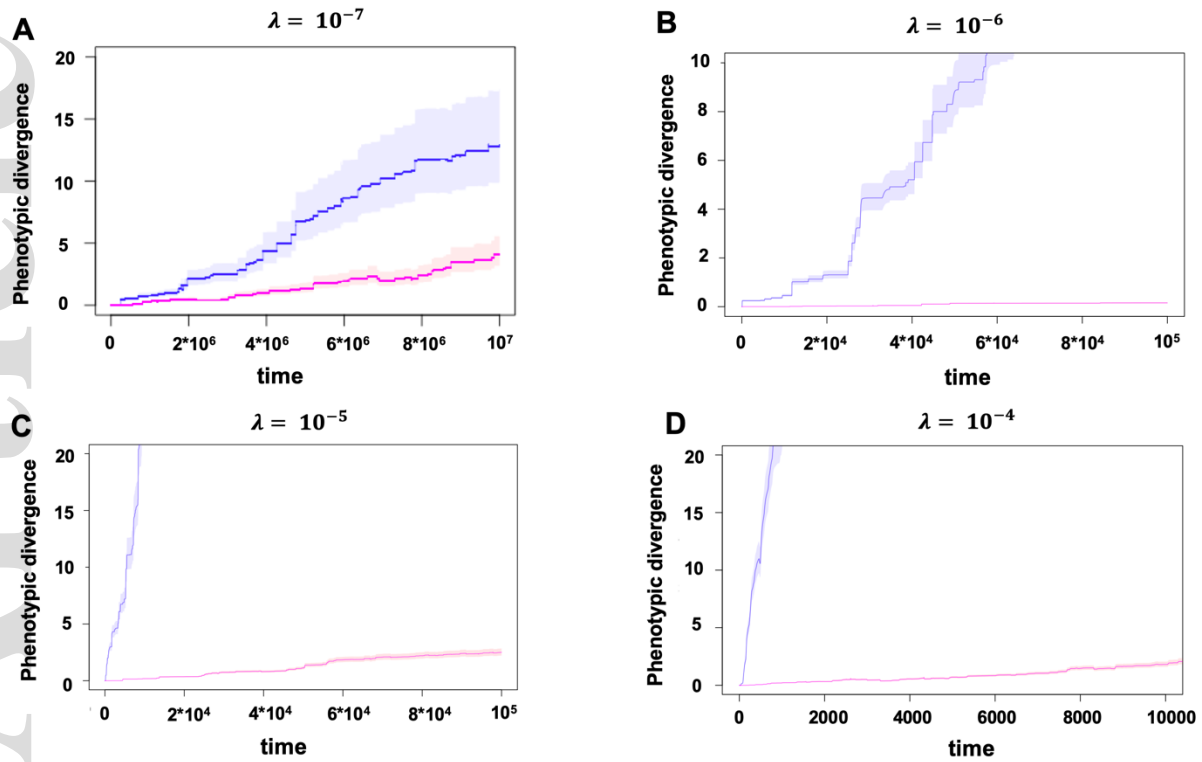
F**igure 3.** Variance accrual through time in displaced optimum models (DO, blue; and DOE, red), assuming our lower empirically-informed benchmark of $\sigma_\theta = 3$ and $\lambda = 10^{-7}$ (panel A), $\sigma_\theta = 13$ (panel B -C) at different assumed probabilities of a peak displacement, λ. DOE models allow the possibility of extinction for maladapted populations, while populations in DO simulations are 'rescued' from potential extinction (see text).
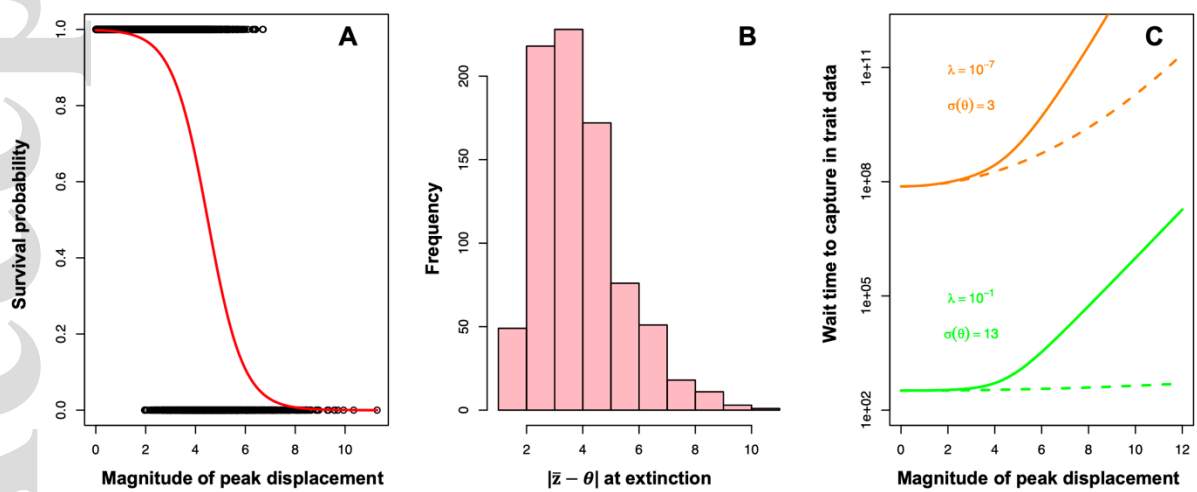
Figure 4. Lineage loss is biased towards large displacements of the optimum, and this can lead to an increase in the wait time to observe peak shifts in trait data. Panel A shows the relationship between lineage survival (to 100 generations after the displacement) and the magnitude of the displacement. Panel B shows the distribution of distances d each extinct lineage was from its optimum upon extinction, illustrating that large peak displacements are not fully reflected in phenotypic change. Based on simulations where lineages were started at their carrying capacity with a displacement from their optimum sampled from a normal distribution with $\sigma^2_\theta = 9$ (changing this distribution has little effect on the observed pattern). Simulations were otherwise parameterized as described in text and assuming density-dependent growth; the same pattern was found under geometric population growth. Panel C shows the probabilities in Panel A converted to expected wait times under a displaced optimum model with parameter values set at our upper and lower benchmarks (see text), where wait times are 1 / (probability a peak shift * probability of a given magnitude * probability of survival). Dashed lines indicate expectations without extinction. Note log scale of y axis.