University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Duchen, P;Hautphenne, S;Lehmann, L;Salamin, N

Title:
Linking micro and macroevolution in the presence of migration

Date:
2020-02-07

Citation:
Duchen, P., Hautphenne, S., Lehmann, L. & Salamin, N. (2020). Linking micro and macroevolution in the presence of migration. Journal of Theoretical Biology, 486, https://doi.org/10.1016/j.jtbi.2019.110087.

Persistent Link:
https://hdl.handle.net/11343/249543

License:
CC BY-NC-ND

Linking micro and macroevolution in the presence of migration

Pablo Duchen[1,], Sophie Hautphenne[2], Laurent Lehmann[3], Nicolas Salamin[1]

1. University of Lausanne. Department of Computational Biology. Lausanne, Switzerland.
2. University of Melbourne. School of Mathematics and Statistics. Melbourne, Australia.
3. University of Lausanne. Department of Ecology and Evolution. Lausanne, Switzerland.

Correspondence:
Pablo Duchen
email: pablo.duchenbocangel@unil.ch

Running head: OU with migration in micro and macroevolution

Abstract

Understanding macroevolutionary patterns is central to evolutionary biology. This involves the process of speciation, which starts at the microevolutionary level, for instance, when two subpopulations evolve towards different phenotypic optima. The speed at which these optima are reached is controlled by the degree of stabilising selection, which pushes the mean trait towards different optima in the different subpopulations, and ongoing migration that pulls the mean phenotype away from that optimum. Traditionally, macro phenotypic evolution is modelled by directional selection processes, but these models usually ignore the role of migration within species. Here, our goal is to reconcile the processes of micro and macroevolution by modelling migration as part of the speciation process. More precisely, we introduce an Ornstein-Uhlenbeck (OU) model where migration happens between two subpopulations within a branch of a phylogeny and this migration decreases over time as it happens during speciation. We then use this model to study the evolution of trait means along a phylogeny, as well as the way phenotypic disparity between species changes with successive epochs. We show that ignoring the effect of migration in sampled time-series data leads to a significant underestimation of the selective forces acting upon it. We also show that migration decreases the expected phenotypic disparity between species and we analyze the effect of migration in the particular case of niche filling. We further introduce a method to jointly estimate selection and migration from time-series data. Our model extends traditional quantitative genetics results of selection and migration from a microevolutionary time frame to multiple speciation events at a macroevolutionary scale. Our results further support that not accounting for gene flow has important consequences in inferences at both the micro and macroevolutionary scale.

Keywords: gene flow, Brownian motion, microevolution, niche-filling, Ornstein-Uhlenbeck, phylogenetics, selection, speciation.

1 Introduction

The study of macroevolution has proven useful in addressing key evolutionary questions about the build-up of biodiversity and the mechanisms underlying the divergence between populations (Stanley 1979; Lande 1980b; Futuyma & Agrawal 2009; Katzourakis *et al.* 2009; Campbell & Kessler 2013). These questions have been addressed by modelling, across a phylogeny, the changes in the rate of evolution of a phenotypic trait (e.g. O'Meara *et al.* 2006; Slater *et al.* 2012), the rate of diversification of species (e.g. Simpson 1944; Nee *et al.* 1992; Jablonski 2008; Silvestro *et al.* 2011; Stadler 2011; Morlon 2014), or the effect of a phenotypic trait on species diversification (e.g. Rieseberg *et al.* 2002; Cardillo *et al.* 2005; Clauset & Erwin 2008; FitzJohn 2012). Although applications of macroevolutionary models are firmly grounded in evolutionary biology (Simpson 1953), the recent theoretical developments in modelling macroevolution have helped understand the mechanisms underlying phenotypic changes across lineages (e.g. FitzJohn 2010; Landis *et al.* 2012).

One of the earliest and main focus of macroevolution has been testing hypotheses about the evolution of quantitative traits among related species (Felsenstein 1985; 2004). Along these lines, neutral trait evolution has been the standard null model for most macroevolutionary studies, and this is typically modelled by Brownian motion (BM). However, the need to incorporate biologically relevant features (e.g. Hansen 1997; Uyeda *et al.* 2011) has lead to large methodological developments (Edwards *et al.* 1964; Cavalli-Sforza & Edwards 1967; Hansen & Martins 1996; Freckleton 2012; Brawand *et al.* 2011; Duchen *et al.* 2017; Boucher *et al.* 2017). One such relevant feature is natural selection, which, in its simplest form acts to stabilise traits around an optimum value. In the presence of stochastic effects on phenotypic change, stabilising natural selection can sometimes be modelled with an Ornstein-Uhlenbeck (OU) process (e.g. Lande 1976, p. 324), which entails a linear transformation of the phenotype making the analysis generally tractable (Gardiner 2009). Modeling stabilising selection using an OU process has thus frequently been the standard approach in macroevolution (Felsenstein 1988; Hansen & Martins 1996; Cooper *et al.* 2016). However, the trait distribution obtained by different types of selective processes is varied and the OU process is not restricted to modeling stabilising selection (Hansen 1997; Cooper *et al.* 2016). And even BM can model specific cases of trait evolution under selection (Gillespie 1973; Hansen & Martins 1996).

However, variation in phenotypic data at the macro scale is often difficult to explain with just a one-dimensional OU process representing directional selection (Pennell *et al.* 2015). Additionally, if datasets contain a small number of species, an OU process tends to be incorrectly

favoured over simpler scenarios (Cooper *et al.* 2016). The application of an OU process in macroevolution therefore requires further developments and finer scrutiny. New theoretical developments should therefore start from microevolutionary dynamics, and, from this, try to derive macroevolutionary dynamics. A theoretical description of current macroevolutionary models showed that interspecific trait-covariances depend on microevolutionary forces, such as random genetic drift, stabilising selection, and mutation, at each generation (Hansen & Martins 1996).

The model of Hansen & Martins (1996) nevertheless overlooked the potential role that migration or gene flow within species plays in linking micro and macroevolutionary dynamics, and thus a more detailed connection between these processes is still needed (Salamin *et al.* 2010; Rolland *et al.* 2018). For instance, migration is determinant in setting the speed of divergence between populations, which, in turn, sets the pace at which speciation takes place (e.g. Gavrilets 2004). And more generally, microevolution is fundamentally affected by the interaction between selection and migration in populations subject to limited dispersal (e.g. Wright 1931; Hartl *et al.* 1997; Ronce & Kirkpatrick 2001; Barton *et al.* 2007). Hence, there is a need to study migration among subpopulations of a species when modelling macroevolution to understand the effects of migration on speciation and its interaction with selection.

In this paper, our goal is to connect the processes of micro and macroevolution by modelling migration between two diverging populations. Building on the stabilising selection models of Lande (1976) and Ronce & Kirkpatrick (2001), we introduce a model of phenotypic trait evolution where migration occurs between two subpopulations before speciation takes place. This model takes the form of an OU process and our approach differs from Bartoszek *et al.* (2017), who modelled migration between branches on a phylogeny, not within each branch, as proposed here. Our model lets a specific trait evolve along the branches of a phylogeny with migration decreasing the rate of gene flow through time until speciation happens. We use this model to study the evolution of trait means and phenotypic disparity between species at a macroevolutionary time scale. Further on, we show the effect of migration and phenotypic disparity in the particular case of niche filling. Finally, we analyse the effect of migration on the parameter estimates of selection by developing an estimator of the selection coefficient for cases when migration is present or absent, and we assess its accuracy with simulations. We show that not accounting for migration can drastically affect the estimation of selection in micro and macroevolutionary models, and our approach opens new avenues to better incorporate microevolutionary forces in macroevolutionary modelling.

2 Methods

2.1 Biological model

Our aim is to model the evolution of a single quantitative phenotype z along a phylogenetic tree. Such phylogenetic tree will consist of several epochs, where one epoch is defined as the time span between two successive nodes (see Supplementary Information (SI) Fig. SI-I.1). We first describe the model for one epoch, where microevolutionary forces can change the mean phenotype. We then extend the model to multiple epochs and derive expressions for the expectation and variance of the mean phenotype in each species at the end of each epoch.

2.1.1 One epoch

We assume that each epoch is of length T and that the population forming a species in any epoch is divided into two subpopulations of equal and constant sizes.

Microevolutionary time scale. For $i = 1, 2$, let $\bar{z}_i(\tau)$ denote the mean phenotype in Subpopulation i at time τ with initial phenotype $z = \bar{z}_1(0) = \bar{z}_2(0)$, where z is a normally distributed random variable with mean μ and variance σ^2. The phenotypic evolution of the two subpopulations forming one species is assumed to be characterized by the system of stochastic differential equations

$$\begin{aligned} \mathrm{d}\bar{z}_1(\tau) &= [\alpha_m(\theta_1(\tau) - \bar{z}_1(\tau)) + m_m(\tau)(\bar{z}_2(\tau) - \bar{z}_1(\tau))]\,\mathrm{d}\tau + \beta_m\,\mathrm{d}w_1(\tau) \\ \mathrm{d}\bar{z}_2(\tau) &= [\alpha_m(\theta_2(\tau) - \bar{z}_2(\tau)) + m_m(\tau)(\bar{z}_1(\tau) - \bar{z}_2(\tau))]\,\mathrm{d}\tau + \beta_m\,\mathrm{d}w_2(\tau), \end{aligned} \tag{1}$$

where $\theta_1(\tau)$ and $\theta_2(\tau)$ are deterministic functions which represent the time-dependent phenotypic optima (or the *phenotypic value* targeted by selection) in each subpopulation at time τ (where a time unit is a generation), α_m is the product of the additive genetic variance σ^2 and the strength of selection γ on the phenotype, i.e., $\alpha_m = \gamma\sigma^2$ (Lande 1979; Hansen & Martins 1996), and $m_m(\tau)$ is the rate of migration of an individual from one subpopulation to the other at time $\tau \in [0, T]$. Additionally, $w_1(\tau)$ and $w_2(\tau)$ are two independent Wiener processes (or Brownian motions) with variance $\beta_m^2 = \sigma^2/N_\mathrm{e}$, where N_e is the effective population size. As such, Eq. (1) combines elements of the quantitative genetics models of Lande (1980a, Eq. (3) & (15)) and Ronce & Kirkpatrick (2001, Eq. (2a)), and adds time dependence to the phenotypic optima and the migration rate. From a stochastic process point of view, Eq. (1) is an OU process (Gardiner 2009), and for a single isolated population this model is equivalent

to that of Hansen & Martins (1996). The optima and migration functions characterize the environment of the focal species and we assume that their dynamics are given by

$$\begin{aligned} \frac{\mathrm{d}\theta_1}{\mathrm{d}\tau} &= \frac{1}{T_\mathrm{c}} F_1(\theta_1(\tau)) \qquad \text{with } \theta_1(0) = \theta, \\ \frac{\mathrm{d}\theta_2}{\mathrm{d}\tau} &= \frac{1}{T_\mathrm{c}} F_2(\theta_2(\tau)) \qquad \text{with } \theta_2(0) = \theta, \end{aligned} \tag{2}$$

$$\frac{\mathrm{d}m_m}{\mathrm{d}\tau} = \frac{1}{T_\mathrm{c}} M_1(m_m(\tau), \theta_1(\tau), \theta_2(\tau)) \qquad \text{with } m_m(0) = 1/2, \tag{3}$$

and where T_c is a characteristic time scale over which the optima and migration rate change in each subpopulation.

Eqs. (2) describe the change of these optima in each subpopulation as a consequence of environmental change. The characteristic time T_c will take the value $T_\mathrm{c} = 1$ when environmental change occurs at the same time scale as the change in phenotype, whereas the optimum changes at a slower rate than the phenotype if $T_\mathrm{c} \gg 1$. The function $m_m(\tau)$ represents the migration rate between the two subpopulations, and Eq. (3) describes the change of migration over time between populations on their way to speciation. We assume that the time scale over which migration changes is the same as that of the optimum functions.

Macroevolutionary time scale. The process described so far concerns changes at the microevolutionary scale, that is, changes in trait values that we can observe at the level of generations. However, if we look at phenotypic change from a macroevolutionary perspective (that is, if we look at evolutionary patterns over the course of millions of years), we do not necessarily expect the behaviour of each parameter in the model shown in Eq. (2) to be the same. For instance, the selection coefficient α_m at the microevolutionary scale (which measures the strength of selection per generation) becomes, at the macroevolutionary scale, a cumulative selection coefficient over the many generations spanning the new time scale. The same will apply to the Brownian variance β_m^2, which becomes the variance over a certain period of time and not per generation. Similarly, the microevolutionary (or generational) interpretation of the evolutionary rate and the migration rate will change when looking at the behaviour of this pattern in the long term. To formally incorporate the change in time scale from generations to thousands or millions of years, we now scale the evolutionary process (Eq. (1) and Eq. (2)–(3)) to reach a longer, macroevolutionary time scale t defined as

$$t = \frac{1}{T_\mathrm{c}} \cdot \tau. \tag{4}$$

with T_c being the time scaling factor.

The rescaling of time in Eqs. (1) to (3) is done with the chain rule $\frac{dx}{d\tau} = \frac{dx}{dt} \cdot \frac{1}{T_c}$, where x represents either $\bar{z}_i$, w_i, or θ_i. The parameters of Eq. (1) and Eqs. (2)–(3) are also re-scaled such that $\alpha = T_c\alpha_m$, $\beta = T_c\beta_m$, and $m(t) = T_c m_m(\tau)$ to obtain the system of equations

$$\begin{aligned} \mathrm{d}\bar{z}_1(t) &= [\alpha(\theta_1(t) - \bar{z}_1(t)) + m(t)(\bar{z}_2(t) - \bar{z}_1(t))]\,\mathrm{d}t + \beta\,\mathrm{d}w_1(t) \\ \mathrm{d}\bar{z}_2(t) &= [\alpha(\theta_2(t) - \bar{z}_2(t)) + m(t)(\bar{z}_1(t) - \bar{z}_2(t))]\,\mathrm{d}t + \beta\,\mathrm{d}w_2(t), \end{aligned} \tag{5}$$

with corresponding phenotypic optima and migration function

$$\begin{aligned} \frac{\mathrm{d}\theta_1}{\mathrm{d}t} &= F_1(\theta_1(t)) \\ \frac{\mathrm{d}\theta_2}{\mathrm{d}t} &= F_2(\theta_2(t)) \\ \frac{\mathrm{d}m}{\mathrm{d}t} &= M_2(m(t), \theta_1(t), \theta_2(t)). \end{aligned} \tag{6}$$

Note that α in Eq. (5) accumulates the net effect of phenotypic change due to selection over multiple generations, and can thus be interpreted as a macroevolutionary selection coefficient. Finally, if we assume that selection is weak at the microevolutionary time scale, and that there is a constant but small input of mutation, then the genetic variance can be held at its mutation-drift equilibrium and $\sigma^2 = 2N_e\sigma_\mu^2$, where σ_μ^2 is the mutation variance (Lande 1980a; Hansen & Martins 1996; Walsh & Lynch 2018). Then,

$$\alpha = T_c\alpha_m = T_c N_e 2\sigma_\mu^2\gamma, \tag{7}$$

and the variance of the Wiener process is equal to

$$\beta^2 = T_c^2\beta_m^2 = T_c^2 2\sigma_\mu^2, \tag{8}$$

where, as stated above, $\beta_m^2 = \sigma^2/N_e$, and $\sigma^2 = 2N_e\sigma_\mu^2$.

Dynamics of the environment. From here on we stay only within the macroevolutionary scale, and we will refer, for simplicity, to the macroevolutionary selection coefficient α simply as the selection coefficient. We assume that there is random mixing between the two subpopulations at the beginning of each epoch. Over time, migration decreases and, thus, contributes to population divergence. More specifically, migration between the two subpopulations follows a monotonically decreasing *migration rate function* $m(t)$ such that $m(0) = 1/2$

(total random mixing) and $\lim_{t\to\infty} m(t) = 0$. It is useful to think as a *speciation event* occurring at time T if $m(T) < \epsilon$, for a chosen small value $\epsilon > 0$ (i.e. migration becomes negligible). Throughout the paper we choose $\epsilon = 10^{-4}$. Note that speciation is not necessarily required to happen as soon as $m(t)$ decreases below ϵ, but it is convenient (and biologically reasonable) to make this assumption because we are studying the effect of migration while two lineages diverge. The time of speciation can be defined more generally as the maximum between $T^{[H]}$ and T, where $T^{[H]}$ is a predefined time of speciation, and T is the smallest value such that $m(T) < \epsilon$. In Hansen (1997), $m(t) = 0$ (for all $t \geq 0$), therefore $T = 0$, and speciation happens at some predefined time $T^{[H]}$.

We assume that the optima in the two subpopulations are initially the same, $\theta_1(0) = \theta_2(0)$, but then diverge according to the *differentiation* function $d(t) := |\theta_1(t) - \theta_2(t)|$. Here, $d(t)$ represents the change of environments over time, and phenotypes differentiate as a result of this change in the environment. Since this difference is difficult to model explicitly, we use $d(t)$ as a proxy. In other words, if we used the actual difference of phenotypes, $d(t)$ would become a random variable, Eq. (5) would become quadratic, and we would no longer have an OU model. As a concrete application of our model, we consider two simple forms of the dynamics of the optima given by $\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = a_i - \theta_i$ and $\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = a_i$. For the initial condition $\theta_i(0) = \theta$ (for some initial value θ), the solution to these dynamics are given by the following parametric functions $\theta_i(t)$:

$$\theta_i(t) = a_i + (\theta - a_i)e^{-t}, \quad \theta_i(t) \to a_i \quad \text{as} \quad t \to \infty \quad (\textit{stabilising} \text{ optimum}), \text{ or} \tag{9}$$

$$\theta_i(t) = a_i t + \theta, \qquad \theta_i(t) \to \pm\infty \quad \text{as} \quad t \to \infty \quad (\textit{diverging} \text{ optimum}). \tag{10}$$

We consider migration functions of the following two types:

$$m(t) = 0.5\mathbb{1}(t \leq L) + 0.5\mathbb{1}(t \geq L)\exp(-c_1 t), \quad \text{or} \tag{11}$$

$$m(t) = 0.5\exp(-c_2 d(t) - c_3 t), \tag{12}$$

where $\mathbb{1}(\cdot)$ is the indicator function, for some constant parameters L, c_1, c_2 and c_3. Parameter L in Eq. (11) controls the length of the period during which there is total mixing between the two subpopulations before migration starts decreasing exponentially at rate c_1. In Eq. (12), the differentiation function $d(t)$ affects the decrease rate of the migration function. In other words, Eq. (11) takes into account extrinsic factors driving divergence, i.e. it can model allopatric speciation. Conversely, Eq. (12) directly links gene flow with ecological speciation, since the

larger the distance $d(t)$ between two subpopulations, the smaller $m(t)$. We present various possible trajectories of $\bar{z}_1(t)$ and $\bar{z}_2(t)$ before the first speciation event, with a stabilising optimum in Subpopulation 1 and a diverging optimum in Subpopulation 2, for different values of α, β, and $m(t)$ (Table 1, Fig. 1).

2.1.2 Multiple epochs

We now consider phenotypic dynamics over multiple epochs. To deal with this, the optimum functions can vary between different species and epochs, but we assume the migration function does not vary across the epochs and species (an alternative scenario, with variable migration functions is described in SI-C.2). The fact that the migration function is fixed implies that the speciation times are deterministic, and so is the number of branches in the phylogenetic tree at any given time: if one initially starts with a single species, then there are 2^n coexisting species during epoch n, corresponding to 2^{n+1} subpopulations ($n \geq 0$). Therefore, the following calculations are applicable to trees like the one shown in Fig. SI-I.1. Equivalent analyses for trees with different branch lengths and asynchronous speciation times are shown in section SI-C.2 and Figs. I.5 and I.6.

For $n \geq 1$, we denote by $\bar{\boldsymbol{z}}^{(n)} = (\bar{z}_i^{(n)})_{i=1,\dots,2^n}$ the random vector recording the *mean phenotype of each species* at the end of epoch $n-1$, and by $\bar{y}^{(n)} := (1/2^n)\sum_{i=1}^{2^n} \bar{z}_i^{(n)}$ the scalar random variable recording the *averaged mean phenotype* at the end of epoch $n-1$. We show in SI-C that $\bar{\boldsymbol{z}}^{(n)}$ follows a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}^{(n)}, \boldsymbol{\Sigma}^{(n)})$ whose mean vector $\boldsymbol{\mu}^{(n)}$ and covariance matrix $\boldsymbol{\Sigma}^{(n)}$ of size 2^n satisfy a first order recurrence equation,

$$\boldsymbol{\mu}^{(n)} = \exp(-\alpha T)[\boldsymbol{\mu}^{(n-1)} \otimes \mathbf{1}] + \boldsymbol{g}^{(n)}(T), \tag{13}$$

$$\boldsymbol{\Sigma}^{(n)} = \exp(-2\alpha T)[\boldsymbol{\Sigma}^{(n-1)} \otimes \mathbf{1}\cdot\mathbf{1}^\top] + I_{2^{n-1}} \otimes \mathbf{H}(T), \tag{14}$$

for $n \geq 1$, with $\boldsymbol{\mu}^{(0)} = \mu$ and $\boldsymbol{\Sigma}^{(0)} = \sigma^2$, and where $\otimes$ denotes the Kronecker product between matrices as defined in SI-A. Here, $\boldsymbol{g}^{(n)}(T)$ ($n \geq 1$) (Eq. (50)) is a sequence of vectors that depend on the optimum functions $\theta_i(t)$, and $\mathbf{H}(T)$ (Eq. (45)) is a matrix that takes into account the covariance induced by the Brownian noises acting on the mean phenotypes of the two subpopulations, and the mass exchange between these subpopulations when $m(t) > \epsilon$ (see Proposition C.1). Note that at every lineage split happening at time T the optimum functions $\theta_i(T)$ are duplicated to form the initial values of the optimum functions of the two daughter lineages (Eq. (49)).

The vector $\bar{y}^{(n)}$ follows a univariate normal distribution with mean and variance given in Eq. (58) and Eq. (59) (see SI-C.1.2). As we further show in SI-C.2, the formulas for $\boldsymbol{\mu}^{(n)}$ and $\boldsymbol{\Sigma}^{(n)}$ can be extended to the case where the migration function $m(t)$ is different for each species, leading to branches of different lengths in the phylogenetic tree.

Finally, an important descriptor of the phenotypic joint distribution is the *disparity* $D^{(n)}$ of $\bar{\boldsymbol{z}}^{(n)}$ (Harmon *et al.* 2003), which is a scalar random variable measuring the extent to which the mean phenotypes of the species present at the end of epoch $n-1$ ($n \geq 1$) differ from each other. We define disparity as

$$D^{(n)} := (1/2^n) \sum_{i=1}^{2^n} [\bar{z}_i^{(n)} - \bar{y}^{(n)}]^2 = (1/2^n) \sum_{i=1}^{2^n} (\bar{z}_i^{(n)})^2 - (\bar{y}^{(n)})^2, \tag{15}$$

and show in SI-C.1.3 that the first moment of $D^{(n)}$ is

$$\mathbb{E}[D^{(n)}] \quad = \quad (1/2^n)[\mathrm{Tr}(\boldsymbol{\Sigma}^{(n)}) + {\boldsymbol{\mu}^{(n)}}^\top \boldsymbol{\mu}^{(n)}] - (1/2^{2n})[\mathbf{1}_{2^n}^\top \boldsymbol{\Sigma}^{(n)} \mathbf{1}_{2^n} + (\mathbf{1}_{2^n}^\top \boldsymbol{\mu}^{(n)})^2], \tag{16}$$

where $\mathrm{Tr}(\boldsymbol{\Sigma}^{(n)})$ denotes the trace of the covariance matrix $\boldsymbol{\Sigma}^{(n)}$. Hence, we can evaluate the disparity in terms of Eqs. (13)–(89). Note that Eq. (15) is similar to sample variance.

2.2 Estimation of the selection coefficient and the migration parameter c

The applications of OU processes in macroevolution often aim at quantifying the amount of selection experienced by different species without considering the effects of migration. We generalize this to the case with migration and formulate estimators of α and the migration parameter $c = c_1$ in Eq. (11) when $L = 0$, that is for the case when $m(t) = 0.5 \exp(-ct)$. Our model readily lends itself to derive such estimators by assuming fixed optimum trajectories, setting $\beta = 0$ in Eq. (5), approximating these expressions as difference equations of the form $\mathrm{d}\bar{z}_1(t) \approx \bar{z}_1(t + n\,\mathrm{d}t) - \bar{z}_1(t)$, and iterating this process n times to obtain

$$\bar{z}_1(t + n\,\mathrm{d}t) = \alpha \sum_{i=0}^{n-1} [\theta_1(t + i\,\mathrm{d}t) - \bar{z}_1(t + i\,\mathrm{d}t) + m(t + i\,\mathrm{d}t)(\bar{z}_2(t + i\,\mathrm{d}t) - \bar{z}_1(t + i\,\mathrm{d}t))]\,\mathrm{d}t + \bar{z}_1(t) \ , \tag{17}$$

$$\bar{z}_2(t + n\,\mathrm{d}t) = \alpha \sum_{i=0}^{n-1} [\theta_2(t + i\,\mathrm{d}t) - \bar{z}_2(t + i\,\mathrm{d}t) + m(t + i\,\mathrm{d}t)(\bar{z}_1(t + i\,\mathrm{d}t) - \bar{z}_2(t + i\,\mathrm{d}t))]\,\mathrm{d}t + \bar{z}_2(t). \tag{18}$$

By summing Eqs. (17) and (18) and rearranging terms, the estimator of α, denoted $\hat{\alpha}$, can be written as

$$\hat{\alpha} = \frac{\bar{z}_1(t + n\,\mathrm{d}t) - \bar{z}_1(t) + \bar{z}_2(t + n\,\mathrm{d}t) - \bar{z}_2(t)}{\sum_{i=0}^{n-1} \left[\theta_1(t + i\,\mathrm{d}t) - \bar{z}_1(t + i\,\mathrm{d}t)\right] \mathrm{d}t + \sum_{i=0}^{n-1} \left[\theta_2(t + i\,\mathrm{d}t) - \bar{z}_2(t + i\,\mathrm{d}t)\right] \mathrm{d}t}. \tag{19}$$

A full step-by-step derivation of Eq. (19) is shown in section SI-G. To obtain an estimator for the migration paramter c we simply replace α in Eq. (17) or Eq. (18) with the value of Eq. (19) and solve numerically for c. If one has data only from a single isolated subpopulation, then α can be estimated from Eq. (19) using only the corresponding subpopulation, say Subpopulation 1, to estimate α:

$$\hat{\alpha} = \frac{\bar{z}_1(t + n\,\mathrm{d}t) - \bar{z}_1(t)}{\sum_{i=0}^{n-1} \left[\theta_1(t + i\,\mathrm{d}t) - \bar{z}_1(t + i\,\mathrm{d}t)\right] \mathrm{d}t}. \tag{20}$$

Here, the data consists first of: $\bar{z}_1(t+n\,\mathrm{d}t)$, which denotes the phenotype of the last sampled value, $\bar{z}_1(t)$ denoting the first sampled value, and $\bar{z}_1(t + i\,\mathrm{d}t)$ which constitute intermediate sampled points. Although two sample points would suffice to have a first estimate, the larger the sample, the more accurate the estimates (see Results). Finally, concerning the stabilising optimum $\theta_1(t)$, if this value is unknown, it is reasonable to set it to the last sampled value of $\bar{z}_1$ as an estimator of $\theta_1(t)$. Since we have only two subpopulations we cannot estimate other model parameters (that is, the parameters describing the optimum functions $\theta_1(t)$ and β) with our current setting. However, maximum likelihood estimations of these parameters are available in Butler & King (2004) (see Discussion).

2.3 Application: niche filling

An interesting application of the joint phenotypic distribution and disparity across epochs concerns niche filling. Ecologically speaking, niche filling is a phenomenon by which different populations or species "fill" the phenotypic space under two conditions: 1) the range of values a phenotype can take is bounded, and 2) two phenotypes cannot take on the same value. This happens, for instance, when there is ecological competition for resources, which prevents two populations from evolving towards the same phenotype (Price *et al.* 2014). As such, niche filling is considered a form of adaptive radiation, by which ecologically distinct species gain access to novel niche space, contrasted by non-adaptive radiation, where new species keep the ancestral niche (Reaney *et al.* 2018). In this section we aim at understanding the effect of decreasing migration in niche filling by using our OU model.

To model niche filling, we first consider the migration function given in Eq. (11) with $L = 0$, and we assume that the diverging optimum functions corresponding to each subpopulation in successive epochs, defined by the sequence $\boldsymbol{\theta}^{(n)}(t)$ (see section C.1.1) are regularly "filling" the interval $[-A, A]$ for some constant $A \geq 0$, over successive epochs of fixed length T; that is, for $0 \leq t \leq T$,

$$\boldsymbol{\theta}^{(1)}(t) = (2T)^{-1}\, t\, [A, -A]^\top \tag{21}$$

$$\boldsymbol{\theta}^{(n)}(t) = (2^{n-1}T)^{-1}\, t\, (\mathbf{1}_{2^{n-2}} \otimes [A, -A]^\top) + (\boldsymbol{\theta}^{(n-1)}(T) \otimes \mathbf{1}), \quad n \geq 2. \tag{22}$$

We refer to the left panel of Fig. 4 for a representation of the optimum functions over the first five epochs. In this particular example, if the migration function is the same for each species (Eq. (11)), then the mean disparity $\mathbb{E}[D^{(n)}]$ converges to a limiting value as $n \to \infty$, given by

$$\begin{aligned}\mathbb{E}[D^{(\infty)}] &= \frac{A^2}{3} + \frac{\beta^2}{2\alpha}\left\{\frac{1}{2} + \frac{\alpha \int_0^T \exp\{-2[\alpha(T-u) + 2\,(\bar{m}(T)T - \bar{m}(u)u)]\}\, du}{1 - \exp(-2\alpha T)}\right\} \\ &= \frac{A^2}{3} + \frac{\beta^2}{2\alpha}\left\{\frac{1}{2} + \frac{\alpha \int_0^T \exp\{-2T(\alpha + 2\bar{m}(T) + 2u(\alpha + 2\bar{m}(u))\, du}{1 - \exp(-2\alpha T)}\right\},\end{aligned} \tag{23}$$

where $\bar{m}(t) := \frac{1}{t}\int_0^t m(u)\, du$; see Proposition SI-D.1. Note that we slightly abuse notation here, because we use $\mathbb{E}[D^{(\infty)}] := \lim_{n\to\infty} \mathbb{E}[D^{(n)}]$. The term $A^2/3$ in Eq. (23) corresponds to the variance of a uniform random variable in $[-A, A]$, and the factor $\beta^2/(2\alpha)$ corresponds to the asymptotic variance of an OU process with no migration. The factor in the curly bracket accounts for migration (it reduces to 1 when there is no migration).

Next, we consider the case where the migration function depends on the differentiation function $d(t)$ (Eq. (12)). If the slopes of the optimum functions are kept the same as in the previous case, during epoch $n-1$, the differentiation function then takes the form $d^{(n)}(t) = A\, t/(2^{n-1}T)$, and the length of epoch $n-1$ is

$$T^{(n)} = \frac{-\log(2\varepsilon)}{c_2 \frac{A}{2^{n-1}T} + c_3} \longrightarrow \frac{-\log(2\varepsilon)}{c_3} \quad \text{as } n \to \infty. \tag{24}$$

Hence, the effect of differentiation disappears asymptotically. In this case, the optimum functions are not confined within the interval $[-A, A]$ (see top left of Fig. 5). The maximum

Table 1: Example of parameter combinations in the model: selection coefficient α, standard deviation of the Wiener process β, migration parameters c_1, c_2, c_3, and time span L during which there is total mixing (Eq. (11) and (12)). We indicate the biological scenarios associated to each parameter combination, and their corresponding panel in Fig. 1.

α	β	c_1	c_2	c_3	L	Biological scenario	Panel in Fig. 1
0.1	0	-	-	-	-	OU, no migration, no noise	a)
0.1	0	-	-	0.01	-	OU with migration, no noise	b)
0.01	0	-	-	0.01	-	As above with weak selection	c)
0.1	5	-	-	0.01	-	OU with migration and noise	d)
0.1	0	0.025	-	-	500	OU with migration, $L > 0$, no noise	e)
0.1	0	-	0.015	0.01	-	Migration depending on $d(t)$, no noise	f)

absolute value of the optimum functions after n epochs is given by

$$u_n = \sum_{j=1}^{n} \frac{A}{2^j T} T^{(j)},$$

whose limit, as $n \to \infty$, is finite and given by

$$u_\infty = \frac{-A}{2} \log(2\varepsilon) \sum_{j \geq 1} \frac{1}{c_2\, A + c_3\, 2^{j-1} T}.$$

Note that this value can be larger than A (see top left of Fig. 5 where it is already above 120 after five epochs while $A = 50$). It is much harder to characterize the asymptotic behaviour of the mean disparity in this setting.

3 Results

Following the same scheme as above we describe the results in three sections: results for the model, the estimators, and the application to niche filling. We give special attention to the results on a single epoch, since they can be directly translated to multiple ones.

3.1 Biological model

Without migration between the two subpopulations, the mean phenotype of each subpopulation reaches the optimum at a speed dictated by the selection coefficient α, and a strong α will result in a fast convergence of $\bar{\boldsymbol{z}}(t)$ to $\boldsymbol{\theta}(t)$ (Fig. 1a). In the presence of migration, however, the speed at which the optimum is reached is slower (Fig. 1b) and different combinations of selection and migration will counteract each other to determine the speed at which the optima will be reached (Fig. 1c). When $\beta > 0$, stochastic fluctuations alter the path of $\bar{\boldsymbol{z}}(t)$, but the

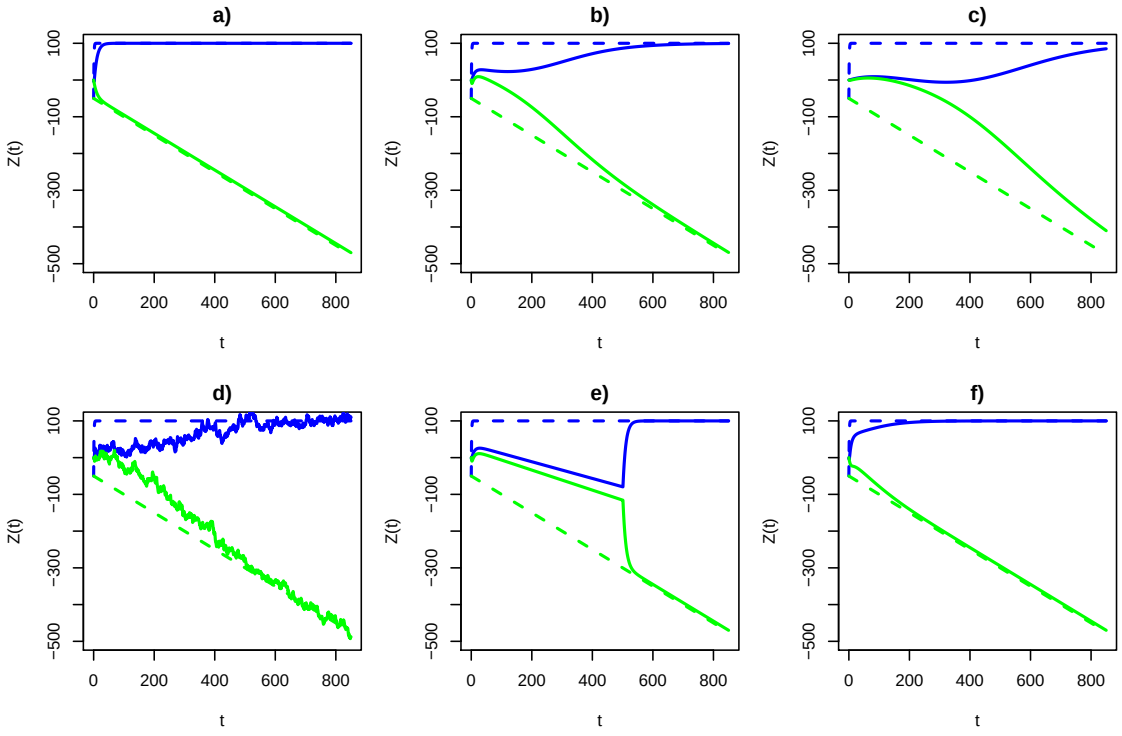


Figure 1: Behaviour of $\bar{z}_1(t)$ (blue) and $\bar{z}_2(t)$ (green) over time t under different values of the selection coefficient α and the migration parameter c_x (see Table 1 for details of each panel). Actual values of $\bar{\mathbf{z}}$ are depicted with solid lines, whereas optima are displayed with dashed lines. These results are discussed in Section 3.1.

overall trend remains (Fig. 1d). For a period of time L of total mixing, the two subpopulations behave similarly and they remain together even when the two optima differ greatly (see Eq. (11)). However, as soon as $m(t)$ starts decreasing (after time $L = 500$ in this example), the optima will be reached once again (Fig. 1e). The latter scenario reflects the introduction of a reproductive barrier at time $L = 500$ and constitutes an example of allopatric speciation. Finally, if $m(t)$ depends also on the distance $d(t)$ between θ_1 and θ_2 (Eq. (12)), then the initial approach to the optimum can be faster than in the case where $m(t)$ does not depend on $d(t)$ (Fig. 1f versus 1b).

3.2 Estimators

3.2.1 Joint estimation of α and $m(t)$

Here, we considered migration functions of the form given in Eq. (11) with $L = 0$ and we let $c = c_1$. To validate our estimator $\hat{\alpha}$ and the estimator $\hat{c}$ of the migration parameter along one epoch, we simulated 100 population trajectories following an OU process with various

combinations of the parameters α and c (with $\beta = 0.01$) for one epoch of fixed time T and various step sizes dt. We used Eq. (19) to compare the estimated α with the true value used in the simulation. Likewise, we compared the numerical solution for c with the true value used for simulations. The accuracy of the parameter estimates is directly related to the number of sampling points taken from the population trajectories, that is, inversely proportional to dt (Fig. 2). In other words, since T is fixed, a smaller dt results in more sampling points, thus increasing the accuracy of the estimation.

We also validated the estimators of α and c using the algorithm described in section SI-H, which generates individual phenotypic values rather than the mean phenotype. With no data available from a second subpopulation, we can use Eq. (20) to estimate α. In this case, however, we risk to underestimate α if there is ongoing migration from an unseen subpopulation. Consider the example in Figure SI-I.2a, where both trajectories were simulated using $\alpha = 0.05$, but the "purple" trajectory experienced migration from an unseen subpopulation. When estimating α using Eq. (20), we observe that the "black" subpopulation has a correct α estimation, whereas we underestimate α for the "purple" subpopulation (Fig. SI-I.2b).

Therefore, to disentangle the effects of selection and migration when data from a single subpopulation is available, we need to look at the actual distribution of phenotypes within the population rather than simply the mean. For instance, the distribution of phenotypes leading to the "purple" trajectory is, in fact, bimodal (Fig. SI-I.2c). On one hand, we have the bulk of the distribution that follows exactly the path of one subpopulation without migration, while we also see individuals that migrated from a second subpopulation petering out gradually as time passes by. The parameter α will be correctly estimated if the data from the bulk of the distribution is used, and this estimate will be well approximated even if only a couple of time points around the convergence value are considered (Fig. SI-I.2b, blue points). However, the robustness of $\hat{\alpha}$ decreases as the two optima become closer, since it becomes more difficult to tell apart the two subpopulations (Fig. SI-I.3a).

Finally, by taking the difference between the number of individuals corresponding to both parts of the bimodal distribution shown in Figure SI-I.2c, we can approximate the migration function $m(t)$ with high accuracy, as long as the optima of the two subpopulations are visibly different from one another (Fig. SI-I.2d, black points: actual $m(t)$; red points: estimated $m(t)$). When the two optima are more similar, the accuracy of the estimation of $m(t)$ decreases (Fig. SI-I.3b).

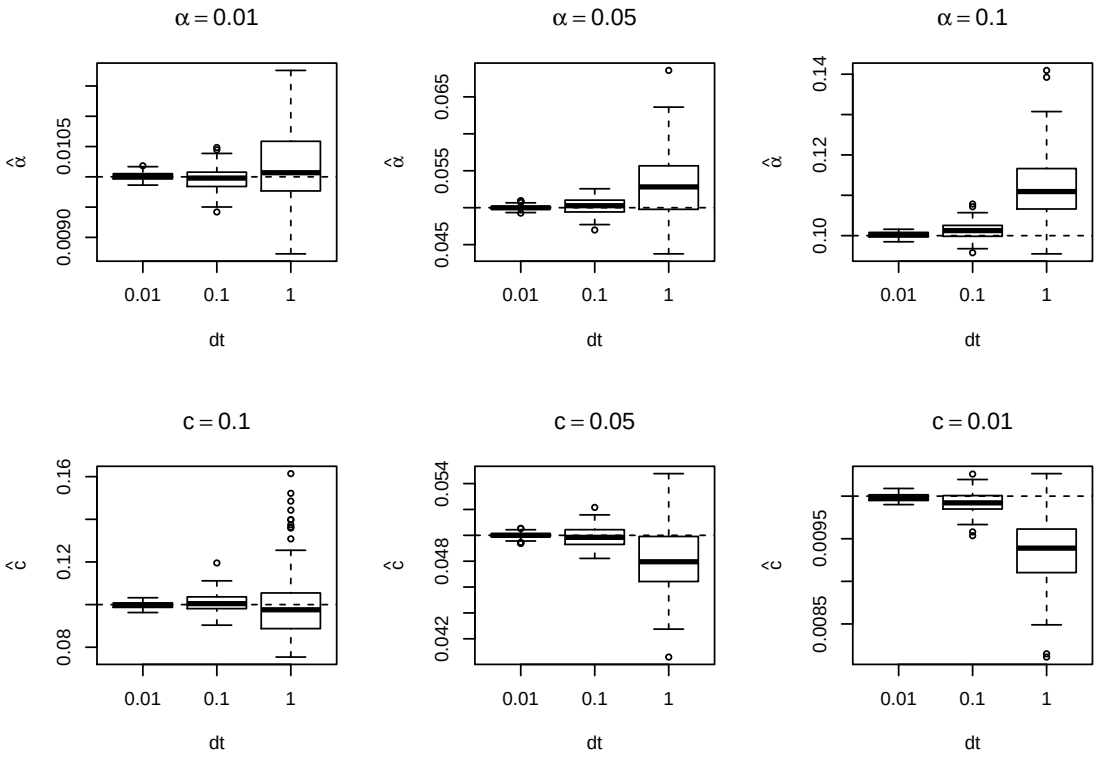


Figure 2: Empirical properties of the estimators for the selection coefficient α (upper row) and the migration parameter c (lower row). Each boxplot represents parameter estimations from 100 simulated phenotypic trajectories for two subpopulations under an OU process with $\beta = 0.01$ and three different step sizes dt. We performed three parameter combinations (corresponding to each column): $\alpha = 0.01$ with $c = 0.1$, $\alpha = 0.05$ with $c = 0.05$, and $\alpha = 0.1$ with $c = 0.01$.

3.2.2 Phenotypic disparity across epochs

We provide an expression for the mean phenotypic disparity $D^{(n)}$ while taking migration into account (Eq. (16)). Using the optima values shown in Table SI-2, we simulate the behaviour of $D^{(n)}$ for cases with and without migration for three consecutive epochs (Fig. 3). We observe that disparity is reduced when migration is present and that this difference is bigger towards the beginning of an epoch. The value of $D^{(n)}$ will however become identical when migration vanishes. This behaviour is consistent for both types of optima: a stabilising optimum $\mathcal{S}$ (Fig. 3 upper panels) and a diverging optimum $\mathcal{D}$ (Fig. 3 lower panels).

3.3 Application: niche filling

We use our model to study the filling of an ecological niche when phenotypic evolution includes migration. Niche filling occurs when there is ecological competition for limited amount of

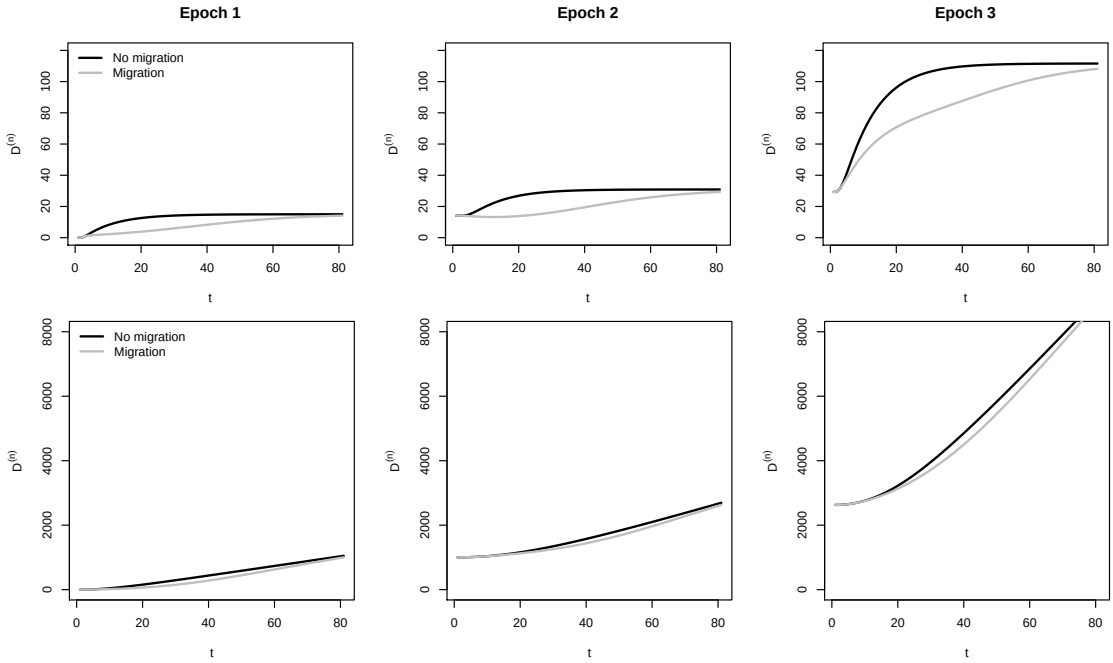


Figure 3: Phenotypic disparity $D^{(n)}$ (Eq. (15)) for three consecutive epochs. Upper panels: disparity for stabilising optima $\mathcal{S}$. Lower panels: disparity for diverging optima $\mathcal{D}$. For scenarios with migration, we use $c_1 = 0.075$ in Eq. (11) with $L = 0$. Optima per epoch were taken from Table SI-2. β was set to 0.

resources and phenotypes of two competing species cannot evolve towards the same optimum value.

We model niche filling in an interval of phenotypic values $[-A, A]$ over successive epochs of fixed length T following Eq. (21) and (22). When $A = 0$, we are in the particular case where $\boldsymbol{\theta}^{(n)}(t) = \mathbf{0}$ for all $n \geq 1$ and $t \geq 0$. When $A = 50$, we can already see the "filling" effect of a niche with optima ranging between -50 and 50 (Fig. 4 left). Over the epochs, the mean disparity increases to a limit given by Eq. (23). We investigated the role of the migration rate in this limit, assuming that $m(t)$ decreases exponentially at a rate c (Fig. 4 right). We see that there is a sharp drop in the asymptotic mean disparity as c increases from 0, then followed by a slow increase towards the constant value $A^2/3 + \beta^2/(2\alpha)$ (value of $\mathbb{E}[D^{(\infty)}]$ when $m(t) \equiv 0$ in Eq. (23)), indicating a larger mean asymptotic disparity when there is less mixing in the population. The interpretation of the drop is more difficult.

We also consider the case where the migration function depends on the differentiation function $d(t)$ (Eq. (12)). The intervals of time between speciation events are thus not constant and depend on the epochs, but converge to a limit (Eq. (24)). The slopes of the optimum functions $\boldsymbol{\theta}^{(n)}$ were kept as in Eq. (22), and their initial values at the beginning of each new

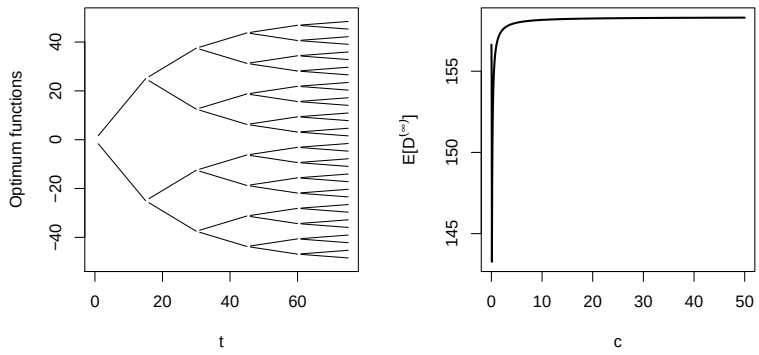


Figure 4: Left: Optimum functions $\boldsymbol{\theta}^{(n)}(t)$ as given by Equation (22) with $A = 50$ for $n = 1, \ldots, 5$ and $m(t) = 0.5\exp(-ct)$, where c is fixed such that $T = 15$. Right: asymptotic mean disparity as a function of the migration parameter c.

epoch are the continuation of their values at the end of the previous epoch. The graph of the functions $\boldsymbol{\theta}^{(n)}$ corresponding to $A = 50$ and $m(t) = 0.5\exp(-0.1d(t) - 0.05t)$ is shown in the left panel of Figure 5. The central panel of that figure highlights for three values of A the logistic increase of the mean disparity as a function of the number of epochs n. We see again that there is always a plateau and the height of the plateau increases with A, but the limit is harder to characterize in this case. Finally, the mean disparity is not necessarily monotonically increasing with A when the migration function depends on the differentiation function $d(t)$ (Fig. 5 right panel). Indeed, for each epoch n there is some threshold value of A such that the mean disparity increases for A less than this threshold due to the fact that the optima are more spread out, and it decreases for A larger than the threshold due to the fact that the differentiation becomes larger, and therefore the speciation times are smaller and there is less mixing before speciation.

Comparison with other macroevolutonary models Using Equations (21), (22), and (1) we generated phenotypic trajectories following the optima and time intervals represented in Fig. 4. We then used the tip data from such trajectories and associated tree as input to the function *dtt* from the R package *geiger* (Pennell *et al.* 2014) to generate disparity-through-time (DTT) plots. These plots show the average disparity for every subclade present at any given time, thus, as phenotypic distances between sister lineages decrease so will disparity. As shown by the DTT plots we found that phenotypic data generated under a model with niche filling with migration will generate lower disparity when compared to other macroevo-

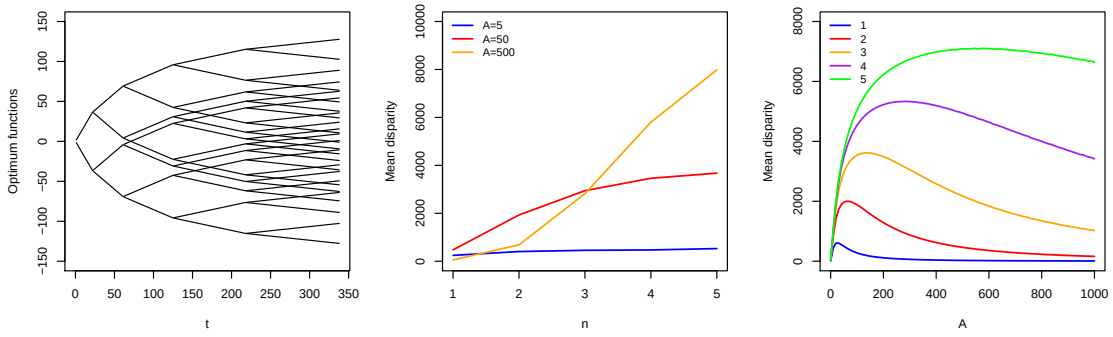


Figure 5: Left: Optimum functions as given by Eq. (22) with $A = 50$ for $n = 1, \ldots, 5$, and $m(t) = 0.5 \exp(-0.1d(t) - 0.05t)$, where $d(t)$ is the differentiation function. Center: mean disparity over five epochs for three values of A. Right: mean disparity at the end of five epochs (labeled 1 to 5), as a function of A.

lutionary models, such as niche filling without migration (OU), bounded BM and late burst (Fig. I.7, a-d). Recall that a similar pattern is observed when comparing disparity under OU with and without migration (Fig. 3, where the presence of migration slows down the accumulation of disparity). Conversely, early burst models (that is, models where the evolutionary rate increases with time) can mimic the disparity generated by niche filling with migration (Fig. I.7 e,f).

4 Discussion

Model We introduced a model with migration between subpopulations (of a single species or a single branch on a phylogeny) that can be applied to macroevolution and can be represented as an OU process. This model combines features of the quantitative genetic models of Lande (1980a) (see also Lande (1976)) and Ronce & Kirkpatrick (2001), but allows for decreasing gene flow between subpopulations. A similar model but where gene flow stays constant has been proposed by Bulmer (1971), although he focused more on the fate of polymorphic alleles under those conditions. In our case, we introduced decreasing gene flow to capture speciation caused by subpopulations tending towards different phenotypic optima as a result of selection. We showed that, as expected, migration reduces the speed at which speciation takes place (e.g. Gavrilets 2004), and we highlighted the counteracting effect of migration on selection (see Fig. 1 for an illustration). Additionally, we showed that such effect (also described in e.g. Wright 1931; Ehrlich & Raven 1969; Stanley 1979; Charlesworth *et al.* 1982; Hartl *et al.* 1997; Ronce & Kirkpatrick 2001; Gavrilets 2004) is valid not only for one epoch but also for

multiple ones and thus crucially extends to macroevolutionary levels. For instance, we show that the (macroevolutionary) selection coefficient α can be vastly underestimated if migration is overlooked (a similar conclusion has been reached in Bartoszek *et al.* (2017)), and migration slows down the speed at which phenotypic disparity among species is reached (Fig. 3).

Migration has also recently been incorporated in macroevolution by Bartoszek *et al.* (2017). Their approach and ours have in common the study of phenotypic evolution under an OU process with migration. However, the biological applications, model assumptions, and mathematical implementations are different. First, they introduced migration *between* branches of a phylogeny, not *within* branches like in our study. A second important difference is that they assume a constant optimum and migration rates. In our case, optima and migration rates are functions of time, where migration decreases over time, as it happens with diverging populations. This is why we modeled migration within branches, to capture this process. Constant migration rates between branches (even for trees representing populations within a species) cannot be linked to speciation. This brings us to a third important difference: in our model we allow for speciation times to be related to migration rates. Therefore, the shape of the phylogenetic tree may depend on the phenotypic evolution observed. In Bartoszek *et al.* (2017), on the other hand, the authors superimpose their model on a phylogeny assumed to be known. Additionally, our specific setting allows us to obtain explicit recursive solutions for the mean and covariance matrices of $\bar{\boldsymbol{z}}^{(n)}$, without making specific assumptions on the eigenstructure of the migration matrix. Lastly, we aimed here at describing a model that links micro and macroevolutionary time scales, so that the interpretation of model parameters is clear. In summary, the biological applications of these two models complement each other, but the conclusions reached concerning the effect of migration on selection estimates are very similar.

Finally, our model assumes (for simplicity) that the two speciating subpopulations have the same size. Since we are mainly studying the behaviour of the mean phenotype then the effect of the population size is small, as long as the mean phenotype approximates the true population mean.

Micro and macroevolution. Bridging the micro and macroevolutionary scales has been a concern for evolutionary biologists since Darwin, and different ways of connecting these two levels have been proposed (Arnold *et al.* 2001; Reznick & Ricklefs 2009; Pennell & Harmon 2013). For instance, natural selection, which is a microevolutionary force, has itself an im-

pact on macroevolutionary patterns. This can be seen in small isolated populations, which can experience fast phenotypic change at the microevolutionary scale if selection on the traits is strong (Lande 1980a), and this isolation can eventually turn into cladogenesis. The microevolutionary parameter that captures the strength of selection is α_m (Eq. (1)). While α_m measures the selection coefficient generation after generation, its macroevolutionary counterpart α (Eq. (5)) is a cumulative selection coefficient over thousands of generations. Thus, the biological interpretation of the selection coefficient after the time transformation is different. The evolutionary rate σ_m^2 at the microevolutionary scale is also subject to the same time transformation, resulting in the macroevolutionary σ^2. In this case, the statistical properties of this parameter remain the same after the transformation, but the biological interpretation is still unclear. What we do know, however, is that σ^2 (in the long term) will depend on the influx of new mutations to the populations, as it is shown in Eq. (8) (see also Harmon 2018). Other microevolutionary forces, such as genetic drift and mutation, can also impact macroevolutionary patterns (Hansen & Martins 1996). For instance, under drift-mutation balance, the covariance between species phenotypes decreases with time and equates $2G_m t_z$, with G_m being the mutation variance and t_z phylogenetic time (Hansen & Martins 1996), which is proportional to our Eq. (7). However, migration has not been considered yet. Gene flow can span over the two time scales and constitutes an important link between micro and macroevolution. Here, we propose that the two time scales are linked by the characteristic time T_c, which defines the time over which the optima change in each subpopulation. A small T_c leads to changes occurring at a microevolutionary scale, while a large T_c indicates changes occurring at a macroevolutionary scale.

One epoch. Our model assumes that a population (or species) consists of two subpopulations with initial random mixing. Each subpopulation then evolves towards distinct optima and the migration rate decreases until it becomes negligible. Speciation occurs when migration stops, leading to ecologically dependent reproductive isolation (EDRI, Hendry 2004). Selection against migrants contributes to EDRI, but also a number of other factors including: 1) reduced mating between individuals from both subpopulations (e.g. Higgie *et al.* 2000; Kirkpatrick 2001; Nosil *et al.* 2003); 2) habitat preferences (e.g. Rice 1984; Bush 1994; Via 1999); and 3) return to a specific breeding location (Hendry *et al.* 2004). Our results show that migration will slow down the speed at which the mean phenotype will reach an optimum value (Fig. 1b,c). When individuals migrate into a subpopulation, the mean phenotype of the

latter is pushed towards the phenotype of the new migrants (Fig. 1e). The effect on the population optimum is initially strong, but will decrease with the reduced number of migrants over time. A similar conclusion was reached when the effect of migration rate on two connected populations of equal sizes was studied (Ronce & Kirkpatrick 2001).

Multiple epochs. A similar pattern was observed across multiple epochs: the speed at which the different optima are reached depends on the antagonising effects of the selection coefficient α and the migration function $m(t)$. Here, we showed that the underestimation of α also happens across multiple epochs if migration is not considered. We observed this while analysing the mean disparity $D^{(n)}$ along successive epochs while accounting for migration (Fig. 3). Recall that disparity is an important and widely-used statistic in ecology and phylogenetics (e.g. Harmon *et al.* 2003; O'Meara *et al.* 2006; Slater *et al.* 2010; Harmon *et al.* 2010; Lumbsch *et al.* 2010). The lag in $D^{(n)}$ due to migration happens because gene flow reduces the phenotypic variance, since the addition of migrants in the subpopulations will reduce the differences between the subpopulation mean phenotypes. Here, let us keep in mind that we assume the same migration function $m(t)$ for all branches and, thus, the tree topology and branch lengths are deterministic. We have also studied the case of having different $m(t)$ for different lineages, where branches can have different lengths within each epoch (section SI-C.2). Although phylogenies with different branch lengths are more common in nature, cases of synchronous and parallel speciation often arise in adaptive radiations (Nagalingum *et al.* 2011), host-parasite co-speciation (Hoberg *et al.* 1997; Brändle *et al.* 2005), or when strong selective pressure affects independent lineages and trigger simultaneous reproductive isolation from their ancestral populations (Hudson *et al.* 2010; Ostevik *et al.* 2012). Our model is also flexible regarding the branch lengths and asynchrony of speciation times, and it can accommodate different migration rates and differentiation functions, yielding more realistic phylogenies (Figs. SI-I.5 and SI-I.6).

Niche filling. We used the multiple-epoch version of the model to study the effect of exponentially decreasing migration (with rate c) on niche filling. Biologically speaking, niche filling constitutes a case of adaptive radiation where new species will occupy novel niches (e.g. Price *et al.* 2014). Conversely, the non-adaptive version happens when newly-formed species retain the ancestral niche. Both scenarios can be modelled through OU processes (e.g. Reaney *et al.* 2018), and the final disparity can be estimated. In our case, we found a sharp drop and subsequent increase in the asymptotic mean disparity across increasing

values of c. While this increase indicates that disparity becomes larger with less gene flow, the interpretation of the drop is more difficult (Fig. 4). Another interesting pattern arises when introducing the differentiation function $d(t)$ in the model. Recall that in such case, the migration function will depend on the current distance between the mean phenotypic values of the two diverging subpopulations (Eq. (12)). Thus, this case relates to ecological speciation, which establishes that divergent selection due to different ecological environments (which, in turn increase $d(t)$) create reproductive barriers. In our case, $d(t)$ also affects the rate at which the mean asymptotic disparity is reached and it increases the final disparity value when compared to a no-differentiation case with the same number of species (Fig. 5 vs. Fig. 4). Additionally, we compared phenotypic data generated under a niche-filling scenario plus migration, with simulated data under BM and models of adaptive radiation. We found that niche filling with migration generates lower disparity values when compared to BM and late burst models. The only scenario that has a similar signal than that of niche filling would be cases of early burst models, where the evolutionary rate increases with time, thus representing adaptive radiations.

Estimation of α and migration. Inferring selection and migration has been a primary interest for evolutionary biologists (e.g. Lynch 1993; Kingsolver *et al.* 2001; Aitken *et al.* 2008). Our approach to inferring the selection coefficient α at the macroevolutionary scale takes the phenotype evolutionary trajectory into account, and the corresponding estimator $\hat{\alpha}$ will be accurate around a time window of a couple tens of generations before and after the optimum value has been reached (and this approach can also be applied at the microevolutionary scale, to infer α_m, Eq. (1)). After that, as we let the phenotypes evolve continuously around the same optimum for more generations, $\hat{\alpha}$ will slowly decrease (Fig. SI-I.2). This means that having phenotypic data sampled before or around the optimum is ideal when estimating selection. In any case, the estimator is robust even in cases when only a few data points have been sampled along the trajectory, and as few as six time points leads to an accurate $\hat{\alpha}$ (Fig. SI-I.2b). If there is only data from before or after the optimum value has been reached the estimation is still accurate, but it will rather reflect an increased (or decreased) selective strength corresponding to the timing of the sampling, instead of the overall selection force associated to the entire process. Finally, a more robust estimation of α in the presence of migration happens when the optima of the two subpopulations are different from each other. If they become too close then the estimation accuracy decreases (Fig. SI-I.3a). A similar case has been reported in

Bartoszek *et al.* (2017) where they state that the estimation of α is improved when there are multiple optima.

The migration function is more difficult to estimate unless one has time-series data. Such estimation is possible either if data from two subpopulations is available (Eq. (19)), or from one subpopulation only (Eq. (20)). Estimating migration with data from one subpopulation is possible because migrating individuals often display different phenotypic values if the two subpopulations have been already diverging. However, the difference in phenotypic optima between the two subpopulations has to be large enough to detect any effects. If the optima are close or similar to one another, an empirical estimation of the migration function becomes more difficult (Fig. SI-I.3). Contrary to Bartoszek *et al.* (2017) we are able to disentangle the effects of migration and selection, but only as long as time-series data is available. Current methods for estimating selection and migration jointly apply exclusively to genetic data (Hey & Nielsen 2004; Hey 2006; Mathieson & McVean 2013), whereas methods for estimating the selection coefficient α in phenotypic data do not take the effects of migration into account (Hansen & Martins 1996; Hansen 1997). Here, we show the importance of accounting for this evolutionary force since failing to do so results in a significant underestimation of the actual selective strength (Fig. SI-I.2).

Finally, in order to estimate θ and β we would need to observe more than two populations. An alternative approach was proposed by Butler & King (2004), where they use Hansen's model to compute maximum likelihood estimators of α, θ, and β. However, these estimators are based on piecewise constant optimum functions belonging to a finite set. Also, they assume that phenotypes are observed only once, while here we assume the phenotypes are observed at multiple regular intervals and, thus, our estimators have explicit forms.

Advantage of OU processes to model phenotypic evolution. A direct (mathematical) advantage of our microevolutionary model with selection is that it is an OU process. Namely, it involves a linear phenotypic transformation, which makes the re-scaling of time straightforward, keeping the entire structure of the model maintained when we go from the micro to the macroevolutionary time scale (compare Eq. (1) with Eq. (5)). Therefore, the selection coefficient at the microevolutionary scale α_m becomes a cumulative selection coefficient α at the macroevolutionary scale, which amalgamates the effects of selection over multiple generations. This generalization should be further studied to investigate if the properties of the OU model at the microevolutionary scale can be extrapolated across speciation events and

towards macroevolutionary time scales.

Conclusions. We developed a model of phenotypic evolution with migration within species, which constitutes an extension of the OU model of Hansen & Martins (1996), and we propose a way to link the time scales of micro and macroevolution. We show that, as expected at the microevolutionary scale, migration counteracts selection when populations diverge towards different optima for the quantitative trait, but our model allows us to extend these results across multiple speciation events. The effect of migration is, therefore, important even for modelling trait evolution at the macroevolutionary scale and not accounting for this process can have important consequences for the estimation of key parameters such as selection levels typically considered in macroevolution.

Acknowledgements

The authors would like to thank the University of Lausanne and the Swiss National Foundation (SNF grant 31A00_163428) for funding this project. SH would like to thank the Australian Research Council (ARC) for support through her Discovery Early Career Researcher Award DE150101044.

Author's contributions

NS and LL conceived the ideas. PD and SH developed the project. PD and SH led the writing of the manuscript with substantial input from NS and LL. All authors contributed critically to the drafts and gave final approval for publication.

References

Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T. & Curtis-McLane, S. (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

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

Barton, N., Briggs, D., Eisen, J., Goldstein, D. & Patel, N. (2007) *Evolution*. Cold Spring Harbor Laboratory Press.

Bartoszek, K., Glémin, S., Kaj, I. & Lascoux, M. (2017) Using the Ornstein–Uhlenbeck process to model the evolution of interacting populations. *Journal of Theoretical Biology*, **429**, 35–45.

Boucher, F.C., Démery, V., Conti, E., Harmon, L.J. & Uyeda, J. (2017) A general model for estimating macroevolutionary landscapes. *Systematic biology*, **67**, 304–319.

Brändle, M., Knoll, S., Eber, S., Stadler, J. & Brandl, R. (2005) Flies on thistles: support for synchronous speciation? *Biological Journal of the Linnean Society*, **84**, 775–783.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M. *et al.* (2011) The evolution of gene expression levels in mammalian organs. *Nature*, **478**, 343.

Bulmer, M. (1971) Stable equilibria under the two-island model. *Heredity*, **27**, 321.

Bush, G.L. (1994) Sympatric speciation in animals: new wine in old bottles. *Trends in Ecology & Evolution*, **9**, 285–288.

Butler, M.A. & King, A.A. (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist*, **164**, 683–695.

Campbell, S.A. & Kessler, A. (2013) Plant mating system transitions drive the macroevolution of defense strategies. *Proceedings of the National Academy of Sciences*, p. 201213867.

Cardillo, M., Mace, G.M., Jones, K.E., Bielby, J., Bininda-Emonds, O.R., Sechrest, W., Orme, C.D.L. & Purvis, A. (2005) Multiple causes of high extinction risk in large mammal species. *Science*, **309**, 1239–1241.

Cavalli-Sforza, L.L. & Edwards, A.W. (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*, **21**, 550–570.

Charlesworth, B., Lande, R. & Slatkin, M. (1982) A neo-Darwinian commentary on macroevolution. *Evolution*, **36**, 474–498.

Clauset, A. & Erwin, D.H. (2008) The evolution and distribution of species body size. *Science*, **321**, 399–401.

Cooper, N., Thomas, G.H., Venditti, C., Meade, A. & Freckleton, R.P. (2016) A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society*, **118**, 64–77.

Duchen, P., Leuenberger, C., Szilágyi, S.M., Harmon, L., Eastman, J., Schweizer, M. & Wegmann, D. (2017) Inference of evolutionary jumps in large phylogenies using Lévy processes. *Systematic biology*, **66**, 950–963.

Edwards, A., Cavalli-Sforza, L., Heywood, V. & McNeill, J. (1964) Phenetic and phylogenetic classification. *Systematic Association Publication*, pp. 67–76.

Ehrlich, P.R. & Raven, P.H. (1969) Differentiation of populations. *Science*, pp. 1228–1232.

Felsenstein, J. (1985) Phylogenies and the comparative method. *The American Naturalist*, **125**, 1–15.

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

Felsenstein, J. (2004) *Inferring phylogenies*, volume 2. Sinauer associates Sunderland, MA.

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

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

Freckleton, R.P. (2012) Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution*, **3**, 940–947.

Futuyma, D.J. & Agrawal, A.A. (2009) Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences*, pp. pnas–0904106106.

Gardiner, C. (2009) *Stochastic Methods*, volume 4. Springer Berlin.

Gavrilets, S. (2004) *Fitness landscapes and the origin of species (MPB-41)*, volume 41. Princeton University Press.

Gillespie, J.H. (1973) Natural selection with varying selection coefficients–a haploid model. *Genetics Research*, **21**, 115–120.

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

Hansen, T.F. & Martins, E.P. (1996) Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, **50**, 1404–1417.

Harmon, L. (2018) *Phylogenetic comparative methods: learning from trees*. EcoEvoRxiv.

Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W., Kozak, K.H., McPeek, M.A., Moreno-Roark, F., Near, T.J. *et al.* (2010) Early bursts of body size and shape evolution are rare in comparative data. *Evolution: International Journal of Organic Evolution*, **64**, 2385–2396.

Harmon, L.J., Schulte, J.A., Larson, A. & Losos, J.B. (2003) Tempo and mode of evolutionary radiation in iguanian lizards. *Science*, **301**, 961–964.

Hartl, D.L., Clark, A.G. & Clark, A.G. (1997) *Principles of population genetics*, volume 116. Sinauer associates Sunderland.

Hendry, A.P. (2004) Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evolutionary Ecology Research*, **6**, 1219–1236.

Hendry, A.P., Castric, V., Kinnison, M.T., Quinn, T.P., Hendry, A. & Stearns, S. (2004) The evolution of philopatry and dispersal. *Evolution illuminated: salmon and their relatives*, pp. 52–91.

Hey, J. (2006) Recent advances in assessing gene flow between diverging populations and species. *Current opinion in genetics & development*, **16**, 592–596.

Hey, J. & Nielsen, R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of drosophila pseudoobscura and d. persimilis. *Genetics*, **167**, 747–760.

Higgie, M., Chenoweth, S. & Blows, M.W. (2000) Natural selection and the reinforcement of mate recognition. *Science*, **290**, 519–521.

Hoberg, E.P., Brooks, D.R. & Siegel-Causey, D. (1997) Host-parasite co-speciation: history, principles, and prospects. *Host–Parasite Evolution: General Principles and Avian Models*, pp. 212–235.

643 Hudson, A.G., Vonlanthen, P. & Seehausen, O. (2010) Rapid parallel adaptive radiations from
644 a single hybridogenic ancestral population. *Proceedings of the Royal Society B: Biological*
645 *Sciences*, **278**, 58–66.

646 Jablonski, D. (2008) Biotic interactions and macroevolution: extensions and mismatches across
647 scales and levels. *Evolution: International Journal of Organic Evolution*, **62**, 715–739.

648 Katzourakis, A., Gifford, R.J., Tristem, M., Gilbert, M.T.P. & Pybus, O.G. (2009) Macroevo-
649 lution of complex retroviruses. *Science*, **325**, 1512–1512.

650 Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C., Hoang,
651 A., Gibert, P. & Beerli, P. (2001) The strength of phenotypic selection in natural popula-
652 tions. *The American Naturalist*, **157**, 245–261.

653 Kirkpatrick, M. (2001) Reinforcement during ecological speciation. *Proceedings of the Royal*
654 *Society of London B: Biological Sciences*, **268**, 1259–1263.

655 Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolu-*
656 *tion*, **30**, 314–334.

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

659 Lande, R. (1980a) Genetic variation and phenotypic evolution during allopatric speciation.
660 *The American Naturalist*, **116**, 463–479.

661 Lande, R. (1980b) Microevolution in relation to macroevolution. *Paleobiology*, **6**, 233–238.

662 Landis, M.J., Schraiber, J.G. & Liang, M. (2012) Phylogenetic analysis using lévy processes:
663 finding jumps in the evolution of continuous traits. *Systematic biology*, **62**, 193–204.

664 Lumbsch, H.T., Parnmen, S., Rangsiruji, A. & Elix, J.A. (2010) Phenotypic disparity and
665 adaptive radiation in the genus cladia (lecanorales, ascomycota). *Australian Systematic*
666 *Botany*, **23**, 239–247.

667 Lynch, M. (1993) Evolution and extinction in response to environ mental change. *Biotic*
668 *interactions and global change*, pp. 234–250.

669 Mathieson, I. & McVean, G. (2013) Estimating selection coefficients in spatially structured
670 populations from time series data of allele frequencies. *Genetics*, pp. genetics–112.

671 Morlon, H. (2014) Phylogenetic approaches for studying diversification. *Ecology letters*, **17**,
672 508–525.

673 Nagalingum, N., Marshall, C., Quental, T.B., Rai, H., Little, D. & Mathews, S. (2011) Recent
674 synchronous radiation of a living fossil. *Science*, **334**, 796–799.

675 Nee, S., Mooers, A.O. & Harvey, P.H. (1992) Tempo and mode of evolution revealed from
676 molecular phylogenies. *Proceedings of the National Academy of Sciences*, **89**, 8322–8326.

677 Nosil, P., Crespi, B. & Sandoval, C. (2003) Reproductive isolation driven by the combined
678 effects of ecological adaptation and reinforcement. *Proceedings of the Royal Society of*
679 *London B: Biological Sciences*, **270**, 1911–1918.

680 O'Meara, B.C., Ané, C., Sanderson, M.J. & Wainwright, P.C. (2006) Testing for different
681 rates of continuous trait evolution using likelihood. *Evolution*, **60**, 922–933.

682 Ostevik, K.L., Moyers, B.T., Owens, G.L. & Rieseberg, L.H. (2012) Parallel ecological speci-
683 ation in plants? *International Journal of Ecology*, **2012**.

684 Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., FitzJohn, R.G.,
685 Alfaro, M.E. & Harmon, L.J. (2014) geiger v2. 0: an expanded suite of methods for fitting
686 macroevolutionary models to phylogenetic trees. *Bioinformatics*, **30**, 2216–2218.

687 Pennell, M.W., FitzJohn, R.G., Cornwell, W.K. & Harmon, L.J. (2015) Model adequacy and
688 the macroevolution of angiosperm functional traits. *The American Naturalist*, **186**, E33–
689 E50.

690 Pennell, M.W. & Harmon, L.J. (2013) An integrative view of phylogenetic comparative meth-
691 ods: connections to population genetics, community ecology, and paleobiology. *Annals of*
692 *the New York Academy of Sciences*, **1289**, 90–105.

693 Price, T.D., Hooper, D.M., Buchanan, C.D., Johansson, U.S., Tietze, D.T., Alström, P.,
694 Olsson, U., Ghosh-Harihar, M., Ishtiaq, F., Gupta, S.K. *et al.* (2014) Niche filling slows the
695 diversification of himalayan songbirds. *Nature*, **509**, 222.

696 Reaney, A.M., Saldarriaga-Córdoba, M. & Pincheira-Donoso, D. (2018) Macroevolutionary
697 diversification with limited niche disparity in a species-rich lineage of cold-climate lizards.
698 *BMC evolutionary biology*, **18**, 16.

Reznick, D.N. & Ricklefs, R.E. (2009) Darwin's bridge between microevolution and macroevolution. *Nature*, **457**, 837.

Rice, W.R. (1984) Disruptive selection on habitat preference and the evolution of reproductive isolation: a simulation study. *Evolution*, **38**, 1251–1260.

Rieseberg, L.H., Widmer, A., Arntz, A.M. & Burke, J.M. (2002) Directional selection is the primary cause of phenotypic diversification. *Proceedings of the National Academy of Sciences*, **99**, 12242–12245.

Rolland, J., Silvestro, D., Litsios, G., Faye, L. & Salamin, N. (2018) Clownfishes evolution below and above the species level. *Proceedings Biological sciences*, **285**.

Ronce, O. & Kirkpatrick, M. (2001) When sources become sinks: migrational meltdown in heterogeneous habitats. *Evolution*, **55**, 1520–1531.

Salamin, N., Wüest, R.O., Lavergne, S., Thuiller, W. & Pearman, P.B. (2010) Assessing rapid evolution in a changing environment. *Trends in ecology & evolution*, **25**, 692–698.

Silvestro, D., Schnitzler, J. & Zizka, G. (2011) A bayesian framework to estimate diversification rates and their variation through time and space. *BMC evolutionary biology*, **11**, 311.

Simpson, G.G. (1944) *Tempo and mode in evolution*. 15. Columbia University Press.

Simpson, G.G. (1953) The baldwin effect. *Evolution*, **7**, 110–117.

Slater, G.J., Harmon, L.J., Wegmann, D., Joyce, P., Revell, L.J. & Alfaro, M.E. (2012) Fitting models of continuous trait evolution to incompletely sampled comparative data using approximate bayesian computation. *Evolution: International Journal of Organic Evolution*, **66**, 752–762.

Slater, G.J., Price, S.A., Santini, F. & Alfaro, M.E. (2010) Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 3097–3104.

Stadler, T. (2011) Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences*, **108**, 6187–6192.

Stanley, S.M. (1979) *Macroevolution, pattern and process*. Johns Hopkins University Press.

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*, p. 201014503.

Via, S. (1999) Reproductive isolation between sympatric races of pea aphids. i. gene flow restriction and habitat choice. *Evolution*, **53**, 1446–1457.

Walsh, B. & Lynch, M. (2018) *Evolution and selection of quantitative traits*. Oxford University Press.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Supplementary Information

A Notation

In this paper we make use of the Kronecker product between matrices, which is defined as follows: if $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is a $p \times q$ matrix, then the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $mp \times nq$ block matrix defined by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\, B & \cdots & A_{1n}\, B \\ \vdots & \ddots & \vdots \\ A_{m1}\, B & \cdots & A_{mn}\, B \end{bmatrix}. \tag{25}$$

The symbol $^\top$ denotes the matrix transposition. We let $\mathbf{1} := [1,1]^\top$, $\mathbf{J} := \mathbf{1} \cdot \mathbf{1}^\top$, $\boldsymbol{e} = [1,-1]^\top$, and $\mathbf{E} = \boldsymbol{e} \cdot \boldsymbol{e}^\top$. We denote by $\mathbf{I}$ the identity matrix of size two, and we use the notation $\mathbf{1}_x$ for the column vector of 1's of size x.

B Trait evolution along one epoch

We first focus on the joint evolution of the phenotype of two subpopulations forming one species between the birth of the species at time $t = 0$ until the next speciation event.

B.1 Dynamics along a single lineage

The system of stochastic differential equations characterizing the phenotypic evolution of the two subpopulations forming one species with common initial phenotype, $\bar{z}_1(0) = \bar{z}_2(0) = z$, is given by

$$d\bar{z}_1(t) = [\alpha(\theta_1(t) - \bar{z}_1(t)) + m(t)(\bar{z}_2(t) - \bar{z}_1(t))]dt + \beta dw_1(t), \tag{26}$$

$$d\bar{z}_2(t) = [\alpha(\theta_2(t) - \bar{z}_2(t)) + m(t)(\bar{z}_1(t) - \bar{z}_2(t))]dt + \beta dw_2(t), \tag{27}$$

where $w_1(t), w_2(t)$ are two independent Wiener processes (Brownian motion), α denotes the strength of selection, and β describes the rate of stochastic evolution away from the optimum. Letting

$$\bar{\boldsymbol{z}}(t) = [\bar{z}_1(t), \bar{z}_2(t)]^\top, \quad \boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t)]^\top, \quad \text{and} \quad d\boldsymbol{w}(t) = [dw_1(t), dw_2(t)]^\top, \tag{28}$$

Eq. (26,27) may be rewritten in matrix form as

$$\bar{\boldsymbol{z}}(0) = z\,\mathbf{1}, \tag{29}$$

$$d\bar{\boldsymbol{z}}(t) = [\alpha\boldsymbol{\theta}(t) - \mathbf{A}(t)\bar{\boldsymbol{z}}(t)]dt + \beta d\boldsymbol{w}(t), \tag{30}$$

where

$$\mathbf{A}(t) = \alpha\mathbf{I} + m(t)\mathbf{E}. \tag{31}$$

We let $\boldsymbol{\theta}(0) = \theta\mathbf{1}$ for some constant parameter θ, because the optimum is initially the same for the two subpopulations forming a new species. Eq. (30) describes a multivariate inhomogeneous time-dependent Ornstein-Uhlenbeck (OU) process (Gardiner 2009, Section 4.5).

We now solve (29,30) for $t \in [0, T]$ following (Gardiner 2009, Sections 4.5.8 and 4.5.9). The homogeneous equation corresponding to (30) is the deterministic equation

$$d\bar{\boldsymbol{z}}(t) = -\mathbf{A}(t)\bar{\boldsymbol{z}}(t)dt, \tag{32}$$

which is soluble since $\mathbf{A}(t)\mathbf{A}(u) = \mathbf{A}(u)\mathbf{A}(t)$ for any $t, u \geq 0$, and has the solution

$$\bar{\boldsymbol{z}}(t) = \exp[-\bar{\mathbf{A}}(t)t]\bar{\boldsymbol{z}}(0), \tag{33}$$

where $\bar{\mathbf{A}}(t)$ is given by

$$\bar{\mathbf{A}}(t) := \frac{1}{t}\int_0^t \mathbf{A}(u)du = \alpha\mathbf{I} + \bar{m}(t)\mathbf{E}, \qquad t \geq 0, \tag{34}$$

with

$$\bar{m}(t) := \frac{1}{t}\int_0^t m(u)du. \tag{35}$$

The general solution of (29,30) is given by

$$\bar{\boldsymbol{z}}(t) = \exp[-\bar{\mathbf{A}}(t)t]\, z\,\mathbf{1} + \boldsymbol{g}(t) + \beta\exp[-\bar{\mathbf{A}}(t)t]\int_0^t \exp[\bar{\mathbf{A}}(u)u]d\boldsymbol{w}(u), \qquad t \geq 0, \tag{36}$$

where $\boldsymbol{g}(t)$ is the deterministic 2×1 vector

$$\boldsymbol{g}(t) = \alpha\exp[-\bar{\mathbf{A}}(t)t]\int_0^t \exp[\bar{\mathbf{A}}(u)u]\boldsymbol{\theta}(u)du, \qquad t \geq 0. \tag{37}$$

If z is deterministic or normally distributed, then $\bar{\boldsymbol{z}}(t)$ is normally distributed for any $t \geq 0$.

Thanks to the particular form (34) of the matrix $\bar{\mathbf{A}}(t)$, its exponential can be simplified,

as we show in the next lemma.

Lemma B.1. *For any $t \geq 0$,*

$$\exp[\pm\bar{\mathbf{A}}(t)t] = \exp(\pm\alpha t)\mathbf{I} + \{\exp[\pm\alpha t \pm 2\bar{m}(t)t] - \exp(\pm\alpha t)\}(\mathbf{E}/2). \tag{38}$$

In particular, $\exp[-\bar{\mathbf{A}}(t)t]\mathbf{1} = \exp(-\alpha t)\mathbf{1}$.

Proof. First observe that for $k \geq 1$, $\mathbf{E}^k = 2^{k-1}\mathbf{E}$. Then, using the binomial theorem for commuting matrices,

$$\begin{aligned}
\bar{\mathbf{A}}^n(t) &= (\alpha\mathbf{I} + \bar{m}(t)\mathbf{E})^n \\
&= \sum_{k=0}^{n} \binom{n}{k} (\alpha\mathbf{I})^{n-k} (\bar{m}(t)\mathbf{E})^k \\
&= \alpha^n\mathbf{I} + 2^{-1}\sum_{k=1}^{n} \binom{n}{k} \alpha^{n-k} (2\bar{m}(t))^k\mathbf{E} \\
&= \alpha^n\mathbf{I} + [(\alpha + 2\bar{m}(t))^n - \alpha^n](\mathbf{E}/2).
\end{aligned}$$

It follows that

$$\begin{aligned}
\exp[\pm\bar{\mathbf{A}}(t)t] &= \sum_{n\geq 0} \frac{(\pm)^n \bar{\mathbf{A}}^n(t)t^n}{n!} \\
&= \sum_{n\geq 0} \frac{(\pm)^n\{(\alpha t)^n\mathbf{I} + [(\alpha t + 2\bar{m}(t)t)^n - (\alpha t)^n](\mathbf{E}/2)\}}{n!} \\
&= \exp(\pm\alpha t)\mathbf{I} + \{\exp[\pm\alpha t \pm 2\bar{m}(t)t] - \exp(\pm\alpha t)\}(\mathbf{E}/2),
\end{aligned}$$

and since $\mathbf{E}\mathbf{1} = \mathbf{0}$, we obtain the result. □

As a consequence of Lemma B.1, the vector $\boldsymbol{g}(t)$ can be rewritten as

$$\begin{aligned}
\boldsymbol{g}(t) = \frac{\alpha}{2}\Bigg\{ &\int_0^t \exp[-\alpha(t-u)](\theta_1(u) + \theta_2(u))\, \mathrm{d}u\, \mathbf{1} \\
&+ \int_0^t \exp[-\alpha(t-u)]\, \exp[-2(\bar{m}(t)t - \bar{m}(u)u](\theta_1(u) - \theta_2(u))\, \mathrm{d}u\, \boldsymbol{e}.\Bigg\}
\end{aligned} \tag{39}$$

This expression can be further simplified if $\boldsymbol{\theta}(u)$ takes some special form. For instance,

- if $\boldsymbol{\theta}(u) = \theta(u)\mathbf{1}$, that is, if the optimum functions are the same for the two subpopulations forming a species, then by (39) we obtain

$$\boldsymbol{g}(t) = \alpha \int_0^t \exp[-\alpha(t-u)]\theta(u)du\mathbf{1}; \tag{40}$$

- if $\theta_1(u) = a + bu$ and $\theta_2(u) = a - bu$ (opposite linear functions with origin a and slope b), then $\boldsymbol{g}(t)$ simplifies to

$$\boldsymbol{g}(t) = F(t)\, a\, \mathbf{1} + J(t)\, b\, \boldsymbol{e}, \tag{41}$$

where

$$F(t) := 1 - \exp(-\alpha t), \qquad J(t) := \alpha \int_0^t \exp[-\alpha(t-u)]\, \exp[-2(\bar{m}(t)t - \bar{m}(u)u]\, u\, du. \tag{42}$$

B.2 Dynamics of the mean and variance

We assume that the common phenotype z at time 0 is normally distributed with mean μ and variance σ^2. We are now in a position to fully characterise the solution $\bar{\boldsymbol{z}}(t)$ of (30).

Proposition B.2. *For any time $0 \leq t \leq T$, the random vector of mean phenotypes $\bar{\boldsymbol{z}}(t)$ follows a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}^{(1)}(t), \boldsymbol{\Sigma}^{(1)}(t))$, with 2×1 mean vector $\boldsymbol{\mu}^{(1)}(t)$ and 2×2 covariance matrix $\boldsymbol{\Sigma}^{(1)}(t)$ given by*

$$\boldsymbol{\mu}^{(1)}(t) = \mu \exp(-\alpha t)\mathbf{1} + \boldsymbol{g}(t), \tag{43}$$

$$\boldsymbol{\Sigma}^{(1)}(t) = \sigma^2 \exp(-2\alpha t)\, \mathbf{J} + \mathbf{H}(t), \tag{44}$$

where $\boldsymbol{g}(t)$ is given by (39)*, and*

$$\mathbf{H}(t) = \frac{\beta^2}{2} \int_0^t \exp[-2\alpha(t-u)]\{\mathbf{J} + \exp[-4\,(\bar{m}(t)t - \bar{m}(u)u)]\, \mathbf{E}\}\, du. \tag{45}$$

Proof. To obtain the expression for the mean, we take the expectation of the right-hand-side of (36), noting that $\mathbb{E}[d\boldsymbol{w}(t)] = 0$. To obtain the covariance matrix, we take the expectation of $\bar{\boldsymbol{z}}(t)\bar{\boldsymbol{z}}(t)^\top$ using (36) again, noting that $\bar{\mathbf{A}}(t) = \bar{\mathbf{A}}^\top(t)$ and that $\mathbb{E}[d\boldsymbol{w}(t)\, d\boldsymbol{w}(s)] = dt\, \mathbb{1}(s = t)$, which leads to

$$\boldsymbol{\Sigma}^{(1)}(t) = \sigma^2 \exp[-\bar{\mathbf{A}}(t)t]\, \mathbf{J}\, \exp[-\bar{\mathbf{A}}(t)t] + \mathbf{H}(t), \tag{46}$$

where

$$\mathbf{H}(t) = \beta^2 \exp[-2\bar{\mathbf{A}}(t)t] \int_0^t \exp[2\bar{\mathbf{A}}(u)u] du. \tag{47}$$

The final expressions (43) and (44) are then derived after some algebraic manipulations using Lemma B.1. □

The first term in $\boldsymbol{\Sigma}^{(1)}(t)$, $\sigma^2 \exp(-2\alpha t)\, \mathbf{J}$, takes into account the covariance induced by the

common initial value z of $\bar{z}_1(t)$ and $\bar{z}_2(t)$, while the second term, $\mathbf{H}(t)$, takes into account the covariance induced by the Brownian noises acting on the two variables, and the mass exchange between the branches when $m(t) > \epsilon$.

Remark B.3. *In this setting we assumed that $w_1(t)$ and $w_2(t)$ are independent. The result can be generalized to the case where the two Wiener processes are not independent. In that case, we define $\rho(t) := \mathrm{Cov}(w_1(t), w_2(t))$, and we can show that $\mathrm{Cov}(w_1(t), w_2(s)) = \rho(\min(s,t))$ for all $s, t \geq 0$. The matrix $\mathbf{H}(t)$ then becomes*

$$\mathbf{H}(t) = \beta^2 \int_0^t \exp[-(\bar{\mathbf{A}}(t)t - \bar{\mathbf{A}}(u)u)] \begin{bmatrix} du & d\rho(u) \\ d\rho(u) & du \end{bmatrix} \exp[-(\bar{\mathbf{A}}(t)t - \bar{\mathbf{A}}(u)u)].$$

In particular, if $w_1(t) = w_2(t)$, then $\rho(t) = t$, and using Lemma B.1, we obtain

$$\mathbf{H}(t) = \frac{\beta^2}{2\alpha}(1 - e^{-2\alpha t})\,\mathbf{J}. \tag{48}$$

C Trait evolution along the entire phylogenetic tree

We now consider the full process starting at time $t = 0$ with one species with mean phenotype $z \sim N(\mu, \sigma^2)$ that splits into two subpopulations, and where migration occurs at rate $m(t)$. In our model, each branch segment of the phylogenetic tree corresponds to two subpopulations evolving according to a two-dimensional OU process. There is a first speciation event at time $T = \inf\{t : m(t) \leq \epsilon\}$, for a chosen value of ϵ. After that time, there is negligible mixing between the two subpopulations which give rise to two new species evolving independently of each other, *conditionally* on their initial mean phenotype $\bar{z}_1(T)$ and $\bar{z}_2(T)$. Again, each new species is made up of two subpopulations, and the process continues.

C.1 Constant migration function $m(t)$

C.1.1 Dynamics of the mean and variance

If the migration function $m(t)$ is deterministic and identical for all species, the next speciation events will happen at times $2T$, $3T$, etc. At time nT (corresponding to the end of epoch $n-1$), there will be exactly 2^n species in the process (for $n \geq 1$). The tree topology and branch lengths are therefore deterministic. However, the joint phenotypic distribution of the species at the end of each epoch is random and follows a multivariate normal distribution that

we specify below.

For $n \geq 1$, let $\bar{\boldsymbol{z}}^{(n)}(t)$, $t \in [0,T]$, be the 2^n-dimensional OU process describing the phenotypic evolution during epoch $n-1$, that is, between time $(n-1)T$ and time nT. We assume that we are given a sequence of (deterministic) functional vectors $\{\boldsymbol{\theta}^{(n)}(t)\}_{n\geq 1}$, of respective sizes $2^n \times 1$, defined for $t \in [0,T]$, and containing the optimum functions corresponding to each epoch in the tree. That is, for $n \geq 1$, $\boldsymbol{\theta}^{(n)}(t)$ is the vector corresponding to $\bar{\boldsymbol{z}}^{(n)}(t)$. In order to ensure the continuity of the optimum function along each lineage, the vectors $\boldsymbol{\theta}^{(n)}(t)$ must satisfy

$$\begin{aligned} \boldsymbol{\theta}^{(1)}(0) &:= \theta \mathbf{1} \\ \boldsymbol{\theta}^{(n)}(0) &= (\boldsymbol{\theta}^{(n-1)}(T) \otimes \mathbf{1}) \quad \text{for } n \geq 2. \end{aligned} \tag{49}$$

In addition to the sequence of vectors $\{\boldsymbol{\theta}^{(n)}(t)\}_{n\geq 1}$, we define the related sequence of vectors $\{\boldsymbol{g}^{(n)}(t)\}_{n\geq 1}$ of size $2^n \times 1$ as follows:

$$\begin{aligned} \boldsymbol{g}^{(n)}(t) &= \alpha \int_0^t \left(\mathbf{I}_{2^{n-1}} \otimes \exp[\bar{\mathbf{A}}(u)u - \bar{\mathbf{A}}(t)t]\right) \boldsymbol{\theta}^{(n)}(u) du \\ &= \mathbf{I}_{2^{n-1}} \otimes \frac{\alpha}{2} \left\{ \int_0^t \exp[-\alpha(t-u)](\theta_1^{(n)}(u) + \theta_2^{(n)}(u))\, du\, \mathbf{1} \right. \\ &\quad \left. + \int_0^t \exp[-\alpha(t-u)]\, \exp[-2(\bar{m}(t)t - \bar{m}(u)u](\theta_1^{(n)}(u) - \theta_2^{(n)}(u))\, du\, \boldsymbol{e}, \right\} \end{aligned} \tag{50}$$

where $\mathbf{I}_{2^{n-1}}$ denotes the identity matrix of size 2^{n-1}.

Let $\bar{\boldsymbol{z}}^{(n)} := \bar{\boldsymbol{z}}^{(n)}(T)$ denote the random vector of phenotypes at the end of epoch $n-1$.

Proposition C.1. *For $n \geq 1$, $\bar{\boldsymbol{z}}^{(n)}$ follows a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}^{(n)}, \boldsymbol{\Sigma}^{(n)})$ of which the $2^n \times 1$ mean vector $\boldsymbol{\mu}^{(n)}$ and the $2^n \times 2^n$ covariance matrix $\boldsymbol{\Sigma}^{(n)}$ can be expressed recursively as*

$$\boldsymbol{\mu}^{(n)} = \exp(-\alpha T)[\boldsymbol{\mu}^{(n-1)} \otimes \mathbf{1}] + \boldsymbol{g}^{(n)}(T) \tag{51}$$

$$\boldsymbol{\Sigma}^{(n)} = \exp(-2\alpha T)[\boldsymbol{\Sigma}^{(n-1)} \otimes \mathbf{J}] + \mathbf{I}_{2^{n-1}} \otimes \mathbf{H}(T), \tag{52}$$

with $\boldsymbol{\mu}^{(0)} = \mu$ and $\boldsymbol{\Sigma}^{(0)} = \sigma^2$.

Proof. The recursion works by updating the initial (random) mean phenotype value of each species at the start of each epoch, which corresponds to the mean phenotype value of each subpopulation at the end of the previous epoch. The Kronecker products reflect the independent evolution of the mean phenotypes along each branch segment of the tree, conditional on

their initial value. □

Corollary C.2. *For $n \geq 1$, $\boldsymbol{\mu}^{(n)}$ and $\boldsymbol{\Sigma}^{(n)}$ take the following explicit forms*

$$\boldsymbol{\mu}^{(n)} = \exp(-\alpha n T)\, \mu\, \mathbf{1}_{2^n} + \sum_{i=1}^{n} \exp[-\alpha(n-i)T]\, [\boldsymbol{g}^{(i)}(T) \otimes \mathbf{1}_{2^{n-i}}] \tag{53}$$

$$\boldsymbol{\Sigma}^{(n)} = \exp(-2\alpha n T)\, \sigma^2\, \mathbf{1}_{2^n} \cdot \mathbf{1}_{2^n}^\top + \sum_{i=1}^{n} \exp[-2\alpha(n-i)T]\, [\mathbf{I}_{2^{i-1}} \otimes \mathbf{H}(T) \otimes \mathbf{1}_{2^{n-i}} \cdot \mathbf{1}_{2^{n-i}}^\top]. \tag{54}$$

C.1.2 Evolution along a random lineage

Recall that each branch segment of the phylogenetic tree corresponds to two subpopulations evolving according to a two-dimensional OU process. One lineage of length n in the tree is thus one particular sequence of n branch segments controlled by a bivariate OU process, where at each branching point an optimum function $\theta_i(\cdot)$ (i.e. a direction) is chosen. Picking one lineage at random in a tree with n epochs is equivalent to selecting one of the 2^n leaves uniformly at random. The phenotype $\bar{y}^{(n)}$ of this selected individual at time nT is given by

$$\bar{y}^{(n)} = (1/2^n) \sum_{i=1}^{2^n} \bar{z}_i^{(n)} = (1/2^n) \mathbf{1}_{2^n}^\top \bar{\boldsymbol{z}}^{(n)}. \tag{55}$$

The random variable $\bar{y}^{(n)}$ is thus simply the average mean phenotype at time nT, and is normally distributed with mean and variance

$$\mu_{\bar{y}^{(n)}} = (1/2^n) \mathbf{1}_{2^n}^\top \boldsymbol{\mu}^{(n)} \tag{56}$$

$$\sigma^2_{\bar{y}^{(n)}} = (1/2^{2n}) \mathbf{1}_{2^n}^\top \boldsymbol{\Sigma}^{(n)} \mathbf{1}_{2^n}, \tag{57}$$

which satisfy a simple recursion: $\mu_{\bar{y}^{(0)}} = \mu$ and $\sigma^2_{\bar{y}^{(0)}} = \sigma^2$, and for $n \geq 1$,

$$\mu_{\bar{y}^{(n)}} = \exp(-\alpha T) \left(\mu_{\bar{y}^{(n-1)}} + \alpha \int_0^T \exp(\alpha u)\, \bar{\theta}^{(n)}(u) du \right) \tag{58}$$

$$\sigma^2_{\bar{y}^{(n)}} = \exp(-2\alpha T)\, \sigma^2_{\bar{y}^{(n-1)}} + \frac{\beta^2}{\alpha\, 2^{n+1}} [1 - \exp(-2\alpha T)], \tag{59}$$

where $\bar{\theta}^{(n)}(u) := (1/2^n) \mathbf{1}_{2^n}^\top \boldsymbol{\theta}^{(n)}(u)$ denotes the average optimum function during the nth epoch.

Asymptotically, as $n \to \infty$, the variance vanishes, $\sigma^2_{\bar{y}^{(\infty)}} = 0$, and $\bar{y}^{(n)}$ converges towards a constant

$$\mu_{\bar{y}^{(\infty)}} = \frac{\alpha \int_0^T \exp[-\alpha(T-u)]\, \bar{\theta}^{(\infty)}(u) du}{1 - \exp(-\alpha T)} \tag{60}$$

where $\bar{\theta}^{(\infty)}(u) = \lim_{n\to\infty} \bar{\theta}^{(n)}(u)$.

C.1.3 Disparity of the phenotypic distribution

The disparity of the multivariate vector $\bar{\boldsymbol{z}}^{(n)}$, denoted by $D^{(n)}$, is a scalar random variable which measures how much the mean phenotypes of the 2^n species present at the end of the nth epoch differ from each other. We define it as

$$D^{(n)} = (1/2^n)\sum_{i=1}^{2^n}[\bar{z}_i^{(n)} - \bar{y}^{(n)}]^2 \quad (61)$$

$$= (1/2^n)\sum_{i=1}^{2^n}(\bar{z}_i^{(n)})^2 - (\bar{y}^{(n)})^2, \quad (62)$$

where $\bar{y}^{(n)}$ is given by (55). The disparity $D^{(n)}$ is not to be confused with the variance of $\bar{y}^{(n)}$, $\sigma^2_{\bar{y}^{(n)}}$, which measures the variability of the (random) average phenotype. The disparity corresponds to the sample variance of the mean phenotypes.

The first moment of $D^{(n)}$ is given by

$$\mathbb{E}[D^{(n)}] = (1/2^n)\sum_{i=1}^{2^n}(\boldsymbol{\Sigma}_{ii}^{(n)} + \mathbb{E}[\bar{z}_i^{(n)}]^2) - (\sigma^2_{\bar{y}^{(n)}} + \mu^2_{\bar{y}^{(n)}}) \quad (63)$$

$$= (1/2^n)[\mathrm{Tr}(\boldsymbol{\Sigma}^{(n)}) + {\boldsymbol{\mu}^{(n)}}^\top \boldsymbol{\mu}^{(n)}] - (1/2^{2n})[\mathbf{1}_{2^n}^\top \boldsymbol{\Sigma}^{(n)} \mathbf{1}_{2^n} + (\mathbf{1}_{2^n}^\top \boldsymbol{\mu}^{(n)})^2], \quad (64)$$

where $\mathrm{Tr}(\boldsymbol{\Sigma}^{(n)})$ denotes the trace of the covariance matrix $\boldsymbol{\Sigma}^{(n)}$. In specific cases, such as the niche filling example, it is possible to characterise the asymptotic mean disparity $\mathbb{E}[D^{(\infty)}] := \lim_{n\to\infty} \mathbb{E}[D^{(n)}]$, as we show in section D.

C.2 Variable migration function $m(t)$

Here we assume that $m(t)$ is deterministic but potentially different along each branch segment of the tree. That is, we define a sequence of functions $\{\boldsymbol{m}^{(n)}(t)\}_{n\geq 1}$, where $\boldsymbol{m}^{(n)}(t)$ is a vector of size $2^{n-1}\times 1$ that contains the migration functions corresponding to the 2^{n-1} systems of OU equations describing the phenotypic evolution during epoch $n-1$ in the tree.

The definition of the sequence of migration functions $\{\boldsymbol{m}^{(n)}(t)\}_{n\geq 1}$ induces a sequence of corresponding speciation times $\{\boldsymbol{T}^{(n)}\}_{n\geq 1}$ which are such that $m_k^{(n)}(T_k^{(n)}) = \varepsilon$ for $k = 1,\ldots,2^{n-1}$; the branch segments of the tree can then have different lengths.

To each function $m_k^{(n)}(t)$ now corresponds a matrix

$$\bar{\mathbf{A}}_k^{(n)}(t) := \alpha\mathbf{I} + \bar{m}_k^{(n)}(t)\mathbf{E}, \tag{65}$$

with

$$\bar{m}_k^{(n)}(t) := \frac{1}{t}\int_0^t m_k^{(n)}(u)du. \tag{66}$$

We define the related sequences of vectors $\{\boldsymbol{g}^{(n)}\}_{n\geq 1}$ of size $2^n \times 1$ and matrices $\{\mathbf{H}^{(n)}\}_{n\geq 1}$ of size $2^n \times 2^n$ as follows:

- $\boldsymbol{g}^{(n)}$ contains 2^{n-1} block vectors of size 2×1, where the kth block vector, $\boldsymbol{g}_{[k]}^{(n)}$, is defined as

$$\boldsymbol{g}_{[k]}^{(n)} = \alpha \exp[-\bar{\mathbf{A}}_k^{(n)}(T_k^{(n)})T_k^{(n)}]\int_0^{T_k^{(n)}} \exp[\bar{\mathbf{A}}_k^{(n)}(u)u]\boldsymbol{\theta}_{[k]}^{(n)}(u)du, \tag{67}$$

  where $\boldsymbol{\theta}_{[k]}^{(n)}(u)$ is the kth block-vector of size 2×1 in the $2^n \times 1$ vector $\boldsymbol{\theta}^{(n)}(u)$, $1 \leq k \leq 2^{n-1}$;

- $\mathbf{H}^{(n)}$ is $2^{(n-1)} \times 2^{(n-1)}$ block-diagonal, where the kth block matrix of size 2×2 on the diagonal, $\mathbf{H}_{[k]}^{(n)}$, is defined as

$$\mathbf{H}_{[k]}^{(n)} \quad = \quad \beta^2 \exp[-2\bar{\mathbf{A}}_k^{(n)}(T_k^{(n)})T_k^{(n)}]\int_0^{T_k^{(n)}} \exp[2\bar{\mathbf{A}}_k^{(n)}(u)u]du. \tag{68}$$

Like in (39) and (45), these expressions can be simplified using Lemma B.1. Let $\mathrm{Diag}[\exp(-\alpha\boldsymbol{T}^{(n)})]$ be the $2^{n-1}\times 2^{n-1}$ diagonal matrix whose (i,i)th entry is $\exp(-\alpha T_i^{(n)})$. As before, $\bar{\boldsymbol{z}}^{(n)}$ denotes the vector of phenotypes at the end of epoch $n-1$. Note however that species in that epoch may now be born at different times. The random vector $\bar{\boldsymbol{z}}^{(n)}$ follows a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}^{(n)}, \boldsymbol{\Sigma}^{(n)})$ of which the mean vector and covariance matrix can be expressed recursively:

$$\boldsymbol{\mu}^{(n)} \quad = \quad \{\mathrm{Diag}[\exp(-\alpha\boldsymbol{T}^{(n)})]\boldsymbol{\mu}^{(n-1)} \otimes \mathbf{1}\} + \boldsymbol{g}^{(n)} \tag{69}$$

$$\boldsymbol{\Sigma}^{(n)} \quad = \quad \{\mathrm{Diag}[\exp(-\alpha\boldsymbol{T}^{(n)})]\boldsymbol{\Sigma}^{(n-1)}\mathrm{Diag}[\exp(-\alpha\boldsymbol{T}^{(n)})] \otimes \mathbf{1}\cdot\mathbf{1}^\top\} + \mathbf{H}^{(n)} \tag{70}$$

for $n \geq 1$, with $\boldsymbol{\mu}^{(0)} = \mu$ and $\boldsymbol{\Sigma}^{(0)} = \sigma^2$.

We refer to Figure I.6 for a simulation of the phenotypic trajectories over the first three epochs for an example with variable migration functions (Eq. (11), with $L = 0$) with values taken from Table 2.

Note that in the setting of variable migration functions, it is not possible to obtain a

recursive expression for the mean and variance of the average mean phenotype $\bar{y}^{(n)}$.

D Niche filling

In the niche filling example in the interval $[-A, A]$ with fixed migration function $m(t)$, the speciation time T is fixed. Let $\boldsymbol{a}^{(n)}$ and $\boldsymbol{b}^{(n)}$ denote the $2^{n-1} \times 1$ vectors containing the origins and slopes of the optimum functions $\boldsymbol{\theta}^{(n)}$, $n \geq 1$. These vectors satisfy

$$\boldsymbol{a}^{(1)} = 0, \quad \boldsymbol{a}^{(n)} = \boldsymbol{a}^{(n-1)} \otimes \mathbf{1} + \frac{A}{2^{n-1}}(\mathbf{1}_{2^{n-2}} \otimes \boldsymbol{e}), \quad \text{and} \quad \boldsymbol{b}^{(n)} = \frac{A}{2^n T}\mathbf{1}_{2^{n-1}}.$$

By (41) and (50), we then have $\boldsymbol{g}^{(n)}(T) = F(T)(\boldsymbol{a}^{(n)} \otimes \mathbf{1}) + J(T)(\boldsymbol{b}^{(n)} \otimes \boldsymbol{e})$, where $F(t)$ and $J(t)$ are defined in (42). Therefore, the mean vector and covariance matrix (53) and (54) of $\bar{\boldsymbol{z}}^{(n)}$ become

$$\boldsymbol{\mu}^{(n)} = \exp(-\alpha T)(\boldsymbol{\mu}^{(n-1)} \otimes \mathbf{1}) + F(T)(\boldsymbol{a}^{(n)} \otimes \mathbf{1}) + J(T)(\boldsymbol{b}^{(n)} \otimes \boldsymbol{e}), \tag{71}$$

$$\boldsymbol{\Sigma}^{(n)} = \exp(-2\alpha T)[\boldsymbol{\Sigma}^{(n-1)} \otimes \mathbf{1} \cdot \mathbf{1}^\top] + \mathbf{I}_{2^{n-1}} \otimes \mathbf{H}(T). \tag{72}$$

The asymptotic mean disparity takes a simple form, as we now show.

Proposition D.1.

$$\mathbb{E}[D^{(\infty)}] = \frac{A^2}{3} + \frac{\beta^2}{2\alpha}\left\{\frac{1}{2} + \frac{\alpha \int_0^T \exp\{-2[\alpha(T-u) + 2\,(\bar{m}(T)T - \bar{m}(u)u)]\}\,du}{1 - \exp(-2\alpha T)}\right\}. \tag{73}$$

Proof. As $n \to \infty$, due to symmetry, we have $\mu_{\bar{y}^{(n)}} \to 0$, and we also have $\sigma^2_{\bar{y}^{(n)}} \to 0$. Therefore,

$$\begin{aligned}
\mathbb{E}(D^{(\infty)}) &:= \lim_{n\to\infty} \mathbb{E}(D^{(n)}) \\
&= \lim_{n\to\infty} (1/2^n)\mathrm{Tr}(\boldsymbol{\Sigma}^{(n)}) && (74) \\
&\quad + \lim_{n\to\infty} (1/2^n)\boldsymbol{\mu}^{(n)\top}\boldsymbol{\mu}^{(n)}. && (75)
\end{aligned}$$

We first treat (74) and let $X := \lim_{n\to\infty}(1/2^n)\mathrm{Tr}(\boldsymbol{\Sigma}^{(n)})$. By (72), we have

$$\mathrm{Tr}(\boldsymbol{\Sigma}^{(n)}) = 2e^{-2\alpha T}\mathrm{Tr}(\boldsymbol{\Sigma}^{(n-1)}) + 2^{n-1}\mathrm{Tr}(\mathbf{H}(T)).$$

Dividing both sides by 2^n, and taking $n \to \infty$, we get $X = e^{-2\alpha T} X + \mathrm{Tr}(\mathbf{H}(T))/2$, leading to

$$X = \lim_{n\to\infty} (1/2^n)\mathrm{Tr}(\mathbf{\Sigma}^{(n)}) = \frac{\mathrm{Tr}(\mathbf{H}(T))}{2(1-e^{-2\alpha T})}.$$

Next, we evaluate (75) and let $W := \lim_{n\to\infty}(1/2^n)\boldsymbol{\mu}^{(n)\top}\boldsymbol{\mu}^{(n)}$. Using (71), and the fact that $\mathbf{1}^\top\mathbf{1} = \boldsymbol{e}^\top\boldsymbol{e} = 2$, $\boldsymbol{e}^\top\mathbf{1} = \mathbf{1}^\top\boldsymbol{e} = 0$, we get

$$\begin{aligned}
\boldsymbol{\mu}^{(n)\top}\boldsymbol{\mu}^{(n)} &= \{e^{-\alpha T}(\boldsymbol{\mu}^{(n-1)\top} \otimes \mathbf{1}^\top) + F(T)(\boldsymbol{a}^{(n)\top} \otimes \mathbf{1}^\top) + J(T)(\boldsymbol{b}^{(n)\top} \otimes \boldsymbol{e}^\top)\} && (76)\\
&\quad \cdot\{e^{-\alpha T}(\boldsymbol{\mu}^{(n-1)} \otimes \mathbf{1}) + F(T)(\boldsymbol{a}^{(n)} \otimes \mathbf{1}) + J(T)(\boldsymbol{b}^{(n)} \otimes \boldsymbol{e})\} && (77)\\
&= 2e^{-2\alpha T}\boldsymbol{\mu}^{(n-1)\top}\boldsymbol{\mu}^{(n-1)} + 4e^{-\alpha T}F(T)\boldsymbol{\mu}^{(n-1)\top}\boldsymbol{a}^{(n)} && (78)\\
&\quad + 2F(T)^2\boldsymbol{a}^{(n)\top}\boldsymbol{a}^{(n)} + 2J(T)^2\boldsymbol{b}^{(n)\top}\boldsymbol{b}^{(n)}. && (79)
\end{aligned}$$

Dividing both sides by 2^n, and taking $n \to \infty$, we get

$$\begin{aligned}
W &= e^{-2\alpha T}W\\
&\quad + 2e^{-\alpha T}F(T)\lim_{n\to\infty}(1/2^{n-1})\boldsymbol{\mu}^{(n-1)\top}\boldsymbol{a}^{(n)} && (80)\\
&\quad + F(T)^2\lim_{n\to\infty}(1/2^{n-1})\boldsymbol{a}^{(n)\top}\boldsymbol{a}^{(n)} && (81)\\
&\quad + J(T)^2\lim_{n\to\infty}(1/2^{n-1})\boldsymbol{b}^{(n)\top}\boldsymbol{b}^{(n)}. && (82)
\end{aligned}$$

It remains to treat (80)–(82), which we do separately.

- Eq. (82): $\boldsymbol{b}^{(n)\top}\boldsymbol{b}^{(n)} = \dfrac{A}{2}$, therefore (82) $= 0$.
- Eq. (81): Using the recursion for $\boldsymbol{a}^{(n)}$,

$$\begin{aligned}
\boldsymbol{a}^{(n)\top}\boldsymbol{a}^{(n)} &= \{\boldsymbol{a}^{(n-1)\top} \otimes \mathbf{1}^\top + \frac{A}{2^{n-1}}(\mathbf{1}_{2^{n-2}}^\top \otimes \boldsymbol{e}^\top)\} && (83)\\
&\quad \cdot\{\boldsymbol{a}^{(n-1)} \otimes \mathbf{1} + \frac{A}{2^{n-1}}(\mathbf{1}_{2^{n-2}} \otimes \boldsymbol{e})\} && (84)\\
&= 2\boldsymbol{a}^{(n-1)\top}\boldsymbol{a}^{(n-1)} + 2\frac{A^2}{2^{2n-2}}2^{n-2} && (85)\\
&= \boldsymbol{a}^{(n-1)\top}\boldsymbol{a}^{(n-1)} + \frac{A^2}{2^{n-1}}. && (86)
\end{aligned}$$

This is a first order recurrence equation with $\boldsymbol{a}^{(1)\top}\boldsymbol{a}^{(1)} = 0$ whose solution is

$$\boldsymbol{a}^{(n)\top}\boldsymbol{a}^{(n)} = \frac{A^2}{3}\frac{(2^{2n-2}-1)}{2^{n-1}},$$

so we have

$$\lim_{n\to\infty}(1/2^{n-1}){\boldsymbol{a}^{(n)}}^\top \boldsymbol{a}^{(n)} = \lim_{n\to\infty}\frac{A^2}{3}\frac{(2^{2n-2}-1)}{2^{2n-2}} = \frac{A^2}{3},$$

905 therefore $(81) = F(T)^2A^2/3$.

906 • Eq. (80):

$$\begin{aligned}
{\boldsymbol{\mu}^{(n-1)}}^\top \boldsymbol{a}^{(n)} &= \{e^{-\alpha T}({\boldsymbol{\mu}^{(n-2)}}^\top \otimes \mathbf{1}^\top) + F(T)({\boldsymbol{a}^{(n-1)}}^\top \otimes \mathbf{1}^\top) + J(T)({\boldsymbol{b}^{(n)}}^\top \otimes \boldsymbol{e}^\top)\} \\
&\quad \cdot\{\boldsymbol{a}^{(n-1)} \otimes \mathbf{1} + \frac{A}{2^{n-1}}(\mathbf{1}_{2^{n-2}} \otimes \boldsymbol{e})\} \\
&= 2e^{-\alpha T}{\boldsymbol{\mu}^{(n-2)}}^\top \boldsymbol{a}^{(n-1)} + 2F(T){\boldsymbol{a}^{(n-1)}}^\top \boldsymbol{a}^{(n-1)} + 2\frac{A}{2^{n-1}}J(T){\boldsymbol{b}^{(n)}}^\top \mathbf{1}_{2^{n-2}} \\
&= 2e^{-\alpha T}{\boldsymbol{\mu}^{(n-2)}}^\top \boldsymbol{a}^{(n-1)} + 2F(T){\boldsymbol{a}^{(n-1)}}^\top \boldsymbol{a}^{(n-1)} + J(T)\frac{A^2}{2^{n-1}T}.
\end{aligned}$$

Let $Z := \lim_{n\to\infty}(1/2^{n-1}){\boldsymbol{\mu}^{(n-1)}}^\top \boldsymbol{a}^{(n)}$. Dividing both sides by 2^{n-1}, and taking $n \to \infty$, we get $Z = e^{-\alpha T}Z + F(T)A^2/3$, therefore

$$\lim_{n\to\infty}(1/2^{n-1}){\boldsymbol{\mu}^{(n-1)}}^\top \boldsymbol{a}^{(n)} = \frac{F(T)A^2}{3(1-e^{-\alpha T})},$$

and

$$(80) = \frac{2e^{-\alpha T}F(T)^2A^2}{3(1-e^{-\alpha T})}.$$

Coming back to the equation for W, we therefore have

$$W = e^{-2\alpha T}W + \frac{2e^{-\alpha T}F(T)^2A^2}{3(1-e^{-\alpha T})} + F(T)^2\frac{A^2}{3},$$

which, using the fact that $F(T) = 1 - e^{-\alpha T}$, simplifies to $W = e^{-2\alpha T}W + (A^2/3)(1 - e^{-2\alpha T})$, giving

$$W = \lim_{n\to\infty}(1/2^n){\boldsymbol{\mu}^{(n)}}^\top \boldsymbol{\mu}^{(n)} = \frac{A^2}{3}.$$

907 Summarizing, we have

$$\mathbb{E}(D^{(\infty)}) = \frac{\mathrm{Tr}(\mathbf{H}(T))}{2(1-e^{-2\alpha T})} + \frac{A^2}{3}, \tag{87}$$

908 where,

$$\begin{aligned}
\mathrm{Tr}(\mathbf{H}(T)) &= \beta^2\left\{\int_0^T e^{-2\alpha(T-u)}du + \int_0^T e^{-2[\alpha(T-u)+2\,(\bar{m}(T)T-\bar{m}(u)u)]}\,du\right\} \\
&= \frac{\beta^2}{2\alpha}(1-e^{-2\alpha T}) + \beta^2\int_0^T e^{-2[\alpha(T-u)+2\,(\bar{m}(T)T-\bar{m}(u)u)]}\,du,
\end{aligned}$$

909 leading to (73). □

	$n=1$	$n=2$	$n=3$
$\boldsymbol{m}^{(n)}(t)$	$c=0.75$	$c=0.4$	$c=0.2$
		$c=0.55$	$c=0.3$
			$c=0.4$
			$c=0.5$
$\boldsymbol{\theta}^{(n)}(t)$	$\mathcal{S}$, $a_1=10$	$\mathcal{S}$, $a_1=50$	$\mathcal{D}$, $a_1=5$
	$\mathcal{D}$, $a_2=-20$	$\mathcal{D}$, $a_2=-10$	$\mathcal{S}$, $a_2=100$
		$\mathcal{D}$, $a_3=25$	$\mathcal{S}$, $a_3=-300$
		$\mathcal{S}$, $a_4=-30$	$\mathcal{D}$, $a_4=-5$
			$\mathcal{D}$, $a_5=10$
			$\mathcal{D}$, $a_6=0$
			$\mathcal{S}$, $a_7=-100$
			$\mathcal{S}$, $a_8=0$

Table 2: **Variable migration functions**: Parameters in the migration vectors $\boldsymbol{m}^{(n)}(t)$ and optimum vectors $\boldsymbol{\theta}^{(n)}(t)$ corresponding to the first three epochs of a phylogenetic tree. In $\boldsymbol{\theta}^{(n)}(t)$, $\mathcal{S}$ means *stabilising* (form (9)), and $\mathcal{D}$ means *diverging* (form (10)).

	$n=1$	$n=2$	$n=3$
$\boldsymbol{\theta}^{(n)}(t)$	$\mathcal{D}$, $a_1=10$	$\mathcal{S}$, $a_1=-10$	$\mathcal{D}$, $a_1=5$
	$\mathcal{D}$, $a_2=10(1+p)$	$\mathcal{S}$, $a_2=-10(1+p)$	$\mathcal{D}$, $a_2=5(1+p)$
		$\mathcal{D}$, $a_3=30$	$\mathcal{S}$, $a_3=-300$
		$\mathcal{D}$, $a_4=30(1+p)$	$\mathcal{S}$, $a_4=-300(1+p)$
			$\mathcal{D}$, $a_5=-5$
			$\mathcal{D}$, $a_6=-5(1+p)$
			$\mathcal{S}$, $a_7=100$
			$\mathcal{S}$, $a_8=100(1+p)$

Table 3: Parameters in the optimum vectors $\boldsymbol{\theta}^{(n)}(t)$ corresponding to the first three epochs of a phylogenetic tree. In $\boldsymbol{\theta}^{(n)}(t)$, $\mathcal{S}$ means *stabilising* (form (9)), and $\mathcal{D}$ means *diverging* (form (10)).

E Additional experiment

In this experiment, we fix $m(t) = 0.5\exp(-c_1 t)$ with $c_1 = 0.5678$ so that $T = 15$ (with $\epsilon = 10^{-4}$), and we consider the optimum function $\boldsymbol{\theta}^{(n)}(t)$ depending on one parameter $0 \leq p \leq 5$ as given in Table 3. These functions are illustrated in Figure I.4 for the extreme values $p = 0$ (Fig. I.4a) and $p = 5$ (Fig. I.4b). When $p = 0$ we are in the particular case where $\theta_1^{(n)}(t) = \theta_2^{(n)}(t)$ for all $n \geq 1$ and $t \geq 0$. The parameter p controls the percentage of increase in the parameter a_2 of $\theta_2^{(n)}(t)$ with respect to a_1 in $\theta_1^{(n)}(t)$. The mean disparity $\mathbb{E}[D^{(n)}]$ is plotted in Figure I.4c as a function of p. As expected, the mean disparity increases with p, but we also see that there is a threshold value p^* such that $\mathbb{E}[D^{(3)}] > \mathbb{E}[D^{(2)}]$ for $p < p^*$, and $\mathbb{E}[D^{(3)}] < \mathbb{E}[D^{(2)}]$ for $p > p^*$.

F Comparison with Hansen's model

The novelty of our model lies in the introduction of the migration function $m(t)$, which itself determines the speciation times. Other models in the literature, such as for example Hansen's model in (Hansen & Martins 1996; Hansen 1997), do not consider any migration function. In Hansen's model the phylogenetic tree and all times between speciation events (lengths of branch segments) are assumed to be known, and the lineage of each species j can then be described by an univariate OU process $X_j(t)$ with, for instance, a constant optimum function $\theta_j(t)$ (Butler & King 2004). Based on the observed phenotypic multivariate distribution at present time (multivariate normal), the parameters α, β and the constant optima along each branch segment can then be inferred. In contrast, in our model, each branch segment of the tree corresponds to a bivariate OU process.

In what follows we make a conection between our model and Hansen's model. The parameters of Hansen's model are:

- $\alpha, \beta, \mu, \sigma$, like in our model;
- a sequence of speciation times $\{\boldsymbol{T}^{(n)}\}_{n\geq 1}$, where $\boldsymbol{T}^{(n)}$ if a vector of size 2^{n-1}, like in our model;
- a sequence of optima $\{\boldsymbol{\theta}^{(n)[H]}\}_{n\geq 1}$, where $\boldsymbol{\theta}^{(n)[H]}$ is a vector of size 2^{n-1}, while in our model, the vector $\boldsymbol{\theta}^{(n)}(t)$ is of size 2^n. We use the superscript $^{[H]}$ to differentiate Hansen's model from ours.

In Hansen's model, the vector of phenotypes at the end of epoch $n-1$, denoted by $\bar{\boldsymbol{z}}^{(n)[H]}$, is of size 2^{n-1} (versus 2^n in our model), and is distributed according to a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}^{(n)[H]}, \Sigma^{(n)[H]})$ of which the mean vector and covariance matrix can be expressed recursively:

$$\boldsymbol{\mu}^{(n)[H]} = \text{Diag}[\exp(-\alpha \boldsymbol{T}^{(n)})](\boldsymbol{\mu}^{(n-1)[H]} \otimes \mathbf{1}) + \text{Diag}[1 - \exp(-\alpha \boldsymbol{T}^{(n)})]\boldsymbol{\theta}^{(n)[H]} \quad (88)$$

$$\Sigma^{(n)[H]} = \text{Diag}[\exp(-\alpha \boldsymbol{T}^{(n)})](\Sigma^{(n-1)[H]} \otimes \mathbf{1} \cdot \mathbf{1}^\top)\text{Diag}[\exp(-\alpha \boldsymbol{T}^{(n)})] \quad (89)$$

$$+(\beta^2/2\alpha)\text{Diag}[[1 - \exp(-2\alpha \boldsymbol{T}^{(n)})] \quad (90)$$

for $n \geq 1$, with $\boldsymbol{\mu}^{(0)[H]} = \mu$ and $\Sigma^{(0)[H]} = \sigma^2$.

To make the comparison between Hansen's model and ours, we specify the parameters of our model so that they match Hansen's model:

- $m(t) = 0$ for all t, which implies that $\bar{\mathbf{A}}(t) = \alpha \mathbf{I}$, and the speciation time is obtained

as the maximum between $T = 0$ and a predefined speciation time $T^{[H]}$ (see section *Dynamics of the environment* in the main text),

- $\boldsymbol{\theta}^{(n)}(t) = (\boldsymbol{\theta}^{(n)[H]} \otimes \mathbf{1})$ (where $\mathbf{1}$ is a vector of size 2, and $t \geq 0$),
- $W_1(t) = W_2(t)$ for each species.

Consequently, $\boldsymbol{g}^{(n)}$ and $\mathbf{H}^{(n)}$ simplify to

$$\boldsymbol{g}^{(n)} = \{\text{Diag}[1 - e^{-\alpha \boldsymbol{T}^{(n)}}]\boldsymbol{\theta}^{(n)[\mathbf{H}]} \otimes \mathbf{1}\} \tag{91}$$

$$\mathbf{H}^{(n)} = \begin{bmatrix} \mathbf{H}(T_1^{(n)}) & & \\ & \ddots & \\ & & \mathbf{H}(T_{2^{n-1}}^{(n)}) \end{bmatrix} \tag{92}$$

where $\mathbf{H}(\cdot)$ is as in (48). We can then show that

$$\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(n)[H]} \otimes \mathbf{1} \tag{93}$$

$$\Sigma^{(n)} = \Sigma^{(n)[H]} \otimes \mathbf{1} \cdot \mathbf{1}^\top, \tag{94}$$

so that $\boldsymbol{\mu}^{(n)[H]}$ and $\Sigma^{(n)[H]}$ can be re-obtained from our model as

$$\boldsymbol{\mu}^{(n)[H]} = \frac{1}{2}(I \otimes \mathbf{1}^\top)\boldsymbol{\mu}^{(n)} \tag{95}$$

$$\Sigma^{(n)[H]} = \frac{1}{4}(I \otimes \mathbf{1}^\top)\Sigma^{(n)}(I \otimes \mathbf{1}). \tag{96}$$

G Estimation of the selection coefficient and the migration rate

As stated before, one of the main goals of this study is to analyze the bias on the estimation of α when failing to account for intraspecific migration. The motivation behind is that standard OU applications in macroevolution often aim at quantifying the amount of selection experienced by different species but they do not consider the effect that intraspecific gene flow has in these estimations. Therefore, given that in real phenotypic samples the selection coefficient α is unknown, we aim here at formulating two estimators: an estimator of α and an estimator of the migration rate parameter $c = c_1$ in Eq. (11) when $L = 0$. Our model readily lends itself to derive such estimators by setting $\beta = 0$ in Eq. (5) and approximating these expressions as

965 difference equations as follows:

$$\bar{z}_1(t+\mathrm{d}t) = [\alpha(\theta_1(t)-\bar{z}_1(t))+m(t)(\bar{z}_2(t)-\bar{z}_1(t))]\,\mathrm{d}t+\bar{z}_1(t), \quad (97)$$

$$\bar{z}_2(t+\mathrm{d}t) = [\alpha(\theta_2(t)-\bar{z}_2(t))+m(t)(\bar{z}_1(t)-\bar{z}_2(t))]\,\mathrm{d}t+\bar{z}_2(t). \quad (98)$$

966 Following the same logic, in a second time step we have

$$\bar{z}_1(t+\mathrm{d}t+\mathrm{d}t) = [\alpha(\theta_1(t+\mathrm{d}t)-\bar{z}_1(t+\mathrm{d}t))+m(t+\mathrm{d}t)(\bar{z}_2(t+\mathrm{d}t)-\bar{z}_1(t+\mathrm{d}t))]\,\mathrm{d}t+\bar{z}_1(t+\mathrm{d}t), \quad (99)$$

967 where the last term $\bar{z}_1(t+\mathrm{d}t)$ is given in Eq. (97). So Eq. (99) becomes

$$\begin{aligned}\bar{z}_1(t+2\,\mathrm{d}t) &= [\alpha(\theta_1(t+\mathrm{d}t)-\bar{z}_1(t+\mathrm{d}t))+m(t+\mathrm{d}t)(\bar{z}_2(t+\mathrm{d}t)-\bar{z}_1(t+\mathrm{d}t))]\,\mathrm{d}t+\bar{z}_1(t+\mathrm{d}t)\\ &\quad +[\alpha(\theta_1(t)-\bar{z}_1(t))+m(t)(\bar{z}_2(t)-\bar{z}_1(t))]\,\mathrm{d}t+\bar{z}_1(t).\end{aligned}$$

After n steps, we obtain

$$\bar{z}_1(t+n\,\mathrm{d}t) = \alpha\sum_{i=0}^{n-1}[\theta_1(t+i\,\mathrm{d}t)-\bar{z}_1(t+i\,\mathrm{d}t)+m(t+i\,\mathrm{d}t)(\bar{z}_2(t+i\,\mathrm{d}t)-\bar{z}_1(t+i\,\mathrm{d}t))]\,\mathrm{d}t+\bar{z}_1(t) \quad (100)$$

$$\bar{z}_2(t+n\,\mathrm{d}t) = \alpha\sum_{i=0}^{n-1}[\theta_2(t+i\,\mathrm{d}t)-\bar{z}_2(t+i\,\mathrm{d}t)+m(t+i\,\mathrm{d}t)(\bar{z}_1(t+i\,\mathrm{d}t)-\bar{z}_2(t+i\,\mathrm{d}t))]\,\mathrm{d}t+\bar{z}_2(t). \quad (101)$$

968 Adding Eq. (17) and (18) leads to

$$\begin{aligned}\bar{z}_1(t+n\,\mathrm{d}t)+\bar{z}_2(t+n\,\mathrm{d}t) &= \alpha\sum_{i=0}^{n-1}[\theta_1(t+i\,\mathrm{d}t)-\bar{z}_1(t+i\,\mathrm{d}t)]\,\mathrm{d}t+\bar{z}_1(t)\\ &\quad +\alpha\sum_{i=0}^{n-1}[\theta_2(t+i\,\mathrm{d}t)-\bar{z}_2(t+i\,\mathrm{d}t)]\,\mathrm{d}t+\bar{z}_2(t).\end{aligned}$$

969 By rearranging terms, an estimator of α can be written in terms of the mean phenotype in
970 the two subpopulations at times $t, t+\mathrm{d}t, \ldots, t+n\,\mathrm{d}t$ as

$$\hat{\alpha} = \frac{\bar{z}_1(t+n\,\mathrm{d}t)-\bar{z}_1(t)+\bar{z}_2(t+n\,\mathrm{d}t)-\bar{z}_2(t)}{\sum_{i=0}^{n-1}[\theta_1(t+i\,\mathrm{d}t)-\bar{z}_1(t+i\,\mathrm{d}t)]\,\mathrm{d}t+\sum_{i=0}^{n-1}[\theta_2(t+i\,\mathrm{d}t)-\bar{z}_2(t+i\,\mathrm{d}t)]\,\mathrm{d}t}. \quad (102)$$

971 To obtain an estimator of the migration rate c we simply replace α in Eq. (17) or Eq. (18)
972 with the value of Eq. (19) and solve numerically for c.

973 If there is no data on a possible second subpopulation from which migration could be

taking place then α can be estimated from Eq. (19) using only the terms corresponding to Subpopulation 1:

$$\hat{\alpha} = \frac{\bar{z}_1(t + n\,\mathrm{d}t) - \bar{z}_1(t)}{\sum_{i=0}^{n-1} \left[\theta_1(t + i\,\mathrm{d}t) - \bar{z}_1(t + i\,\mathrm{d}t)\right] \mathrm{d}t}. \tag{103}$$

H Simulations

This subsection describes a simulator that uses a standard OU process to generate individual phenotype trajectories with known α values. This simulator is used to 1) check the accuracy of $\hat{\alpha}$, and 2) check the effect of gene flow in the estimation of α (accomplished by adding migrants from a second population).

Direct OU simulations To check the accuracy of our estimator $\hat{\alpha}$ we simulated full population phenotype trajectories by using a standard OU process with the following steps:

1. With arbitrary values of $\bar{z}(0)$, α, and θ (where θ is here a constant) we generated a trajectory of trait means $\bar{z}(t)$ for $t = 200$ generations using equation $d\bar{z}(t) = \alpha(\theta - \bar{z}(t))dt + \beta dw(t)$. For the purpose of these simulations we set $\beta = 0$ and worked with the first term of the expression as a difference equation, as it is usually done when programming differential equations. The chosen parameter values were: $\bar{z}(0) = 2$, $\alpha = 0.05$, and $\theta = 10$.

2. The vector of phenotype means $\bar{z}(t)$ generated in step 1 for discrete values of t was then used to draw 100 values from normal distributions $\mathcal{N}(\bar{z}(t), s)$ with means $\bar{z}(t)$ and some arbitrary variance ($s = 3$) which represent the phenotypic values of a subpopulation of size $N = 100$ at any given t. This results in a population with a known α, a step that is necessary to validate the estimator described in Section G. These trajectories are stored in a matrix with 100 rows and 200 columns (note that since all 100 values were placed randomly in each column, rows of this matrix do not necessarily follow individual trajectories, but the number of columns still represents t). A plot of the mean value of each column (mean phenotype) can be seen in Fig. I.2a (black curve).

3. To check for the effect on $\hat{\alpha}$ of migrants from outside we simulated a second subpopulation with the same parameters, varying only the optimum $\theta = -5$. This second simulated subpopulation is also stored in a 100×200 matrix. To recreate migration, we exchanged individuals (phenotypic values) between the two matrices column-wise. That is, for each column (generation), the two subpopulations exchanged $Nm(t)$ migrants (rounded to the

nearest integer) chosen uniformly at random among the entries of the column. Here, $m(t)$ is given by Eq. (11) with $L = 0$, and t here represents the columns of the matrices. Recall that since $m(t)$ decreases exponentially over time, the number of migrants exchanged in successive columns is also decreasing. The above procedure does not modify the size of the matrices of phenotypes. Notice here why it was necessary to simulate full population phenotypes and not only mean phenotypes, so that the two subpopulations can actually exchange migrants. A plot of the mean value of each column (mean phenotype) can be seen in Fig. I.2a (purple curve), and a plot of the full distribution per column in Fig. I.2c.

I Supplementary figures

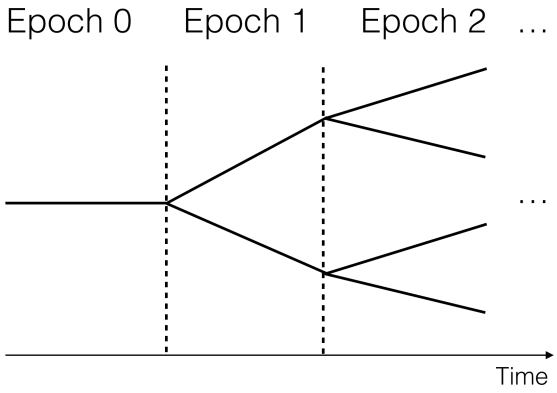


Figure I.1: Schematic diagram of a phylogenetic (species) tree with three epochs. Here, there are 2^n coexisting species during epoch n, corresponding to 2^{n+1} subpopulations ($n \geq 0$). Thus, in epoch $n = 0$ there are $2^0 = 1$ species, in epoch $n = 1$ there are $2^1 = 2$ species, and so on. In our model, within each branch of this tree there are two subpopulations that exchange migrants at rate $m(t)$ (Eqs. (11) and (12)).

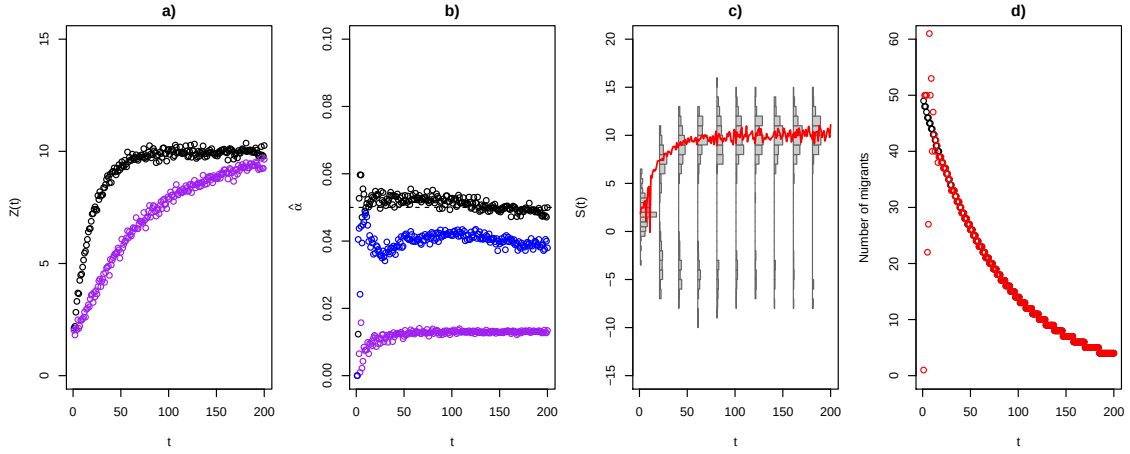


Figure I.2: Estimation of α and empirical estimation of $m(t)$. a) Trajectory of a mean phenotype Z following an OU process (simulated following the algorithm of section H) with $\alpha = 0.05$ without incoming migration (black) and with migration from a second population (purple). b) α values estimated with Eq. (20) for the trajectory with migration (purple), without migration (black), and without migration but when data is sampled at only 6 time points around the convergence value (blue). c) Actual distributions of the "purple" phenotypes shown in panel a); the red line follows the mean phenotype along the distribution part with the highest density. d) True migration function of the trajectory shown with the purple circles in panel a) (black), versus the estimated migration function obtained from the distributions in panel c) (red).

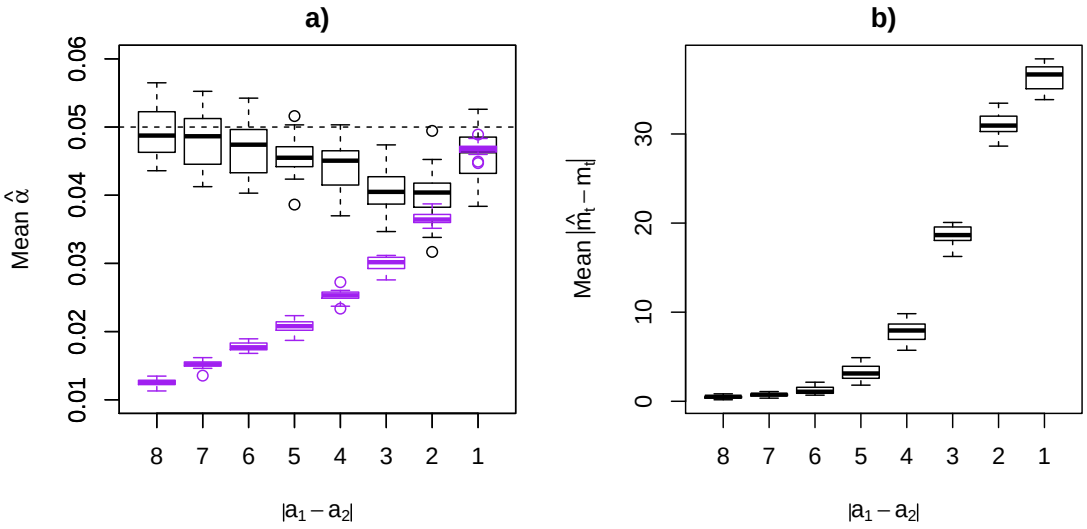


Figure I.3: a) Robustness of $\hat{\alpha}$ relative to the distance between the optima a_1 and a_2 of Subpopulations 1 and 2, respectively. Black boxplots represent the mean of $\hat{\alpha}$ after 20 repetitions when accounting for migration. Purple boxplots represent the mean of $\hat{\alpha}$ after 20 repetitions when not accounting for migration. The black horizontal dashed line shows the true α value. b) Robustness of $\hat{m}(t)$ relative to the distance between the optima of Subpopulations 1 and 2, measured as the mean absolute difference between the real an estimated migration function $m(t)$ after 20 repetitions.

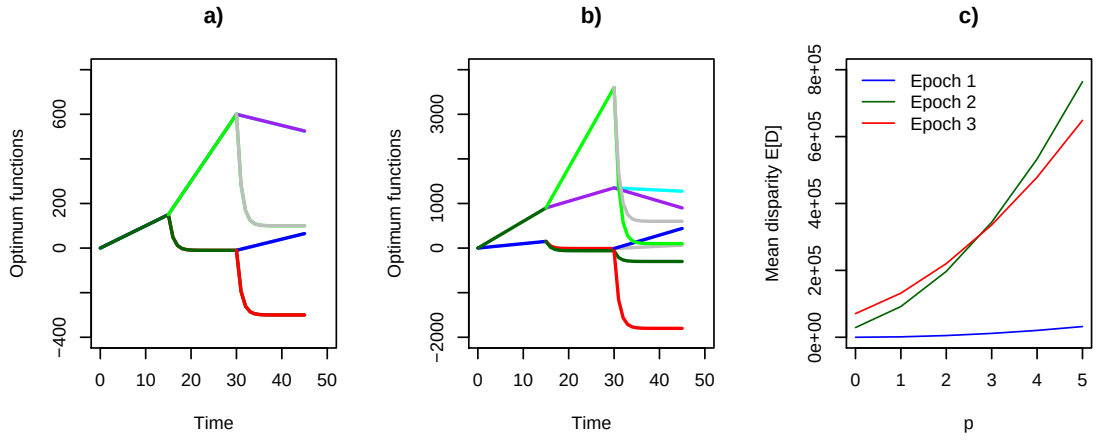


Figure I.4: Optimum functions as given in Table 3 for $p = 0$ (a), and $p = 5$ (b). c) Mean disparity $E[D^{(n)}]$ as a function of p for $n = 3$ epochs.

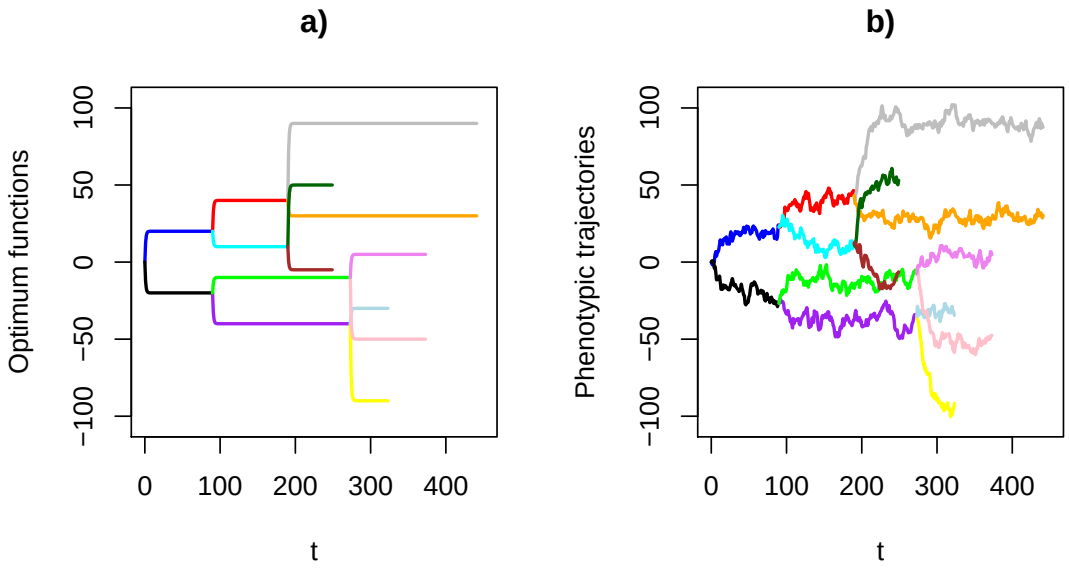


Figure I.5: Optimum functions (a) and phenotypic trajectories (b) (following the optimum functions) for $n = 3$ epochs, but when $m(t)$ depends on the differentiation function $d(t)$ (Eq. (12)). Given that the optima shown in panel a) are different from one another, and that migration rates also differ, this results in different speciation times. Thus, branch lenghts are different and epochs are no longer synchronous. Migration rates and optima used are the following: Epoch 1) $c = 0.05$, $a_1 = 20$, $a_2 = -20$; Epoch 2) $c = [0.055, 0.03]$, $a_1 = [40, -40]$, $a_2 = [10, -10]$; Epoch 3) $c = [0.01, 0.051, 0.05, 0.03]$, $a_1 = [90, 50, -90, -50]$, $a_2 = [30, -5, -30, 5]$. All optima were set to mode "stabilising" (Eq. (9)).

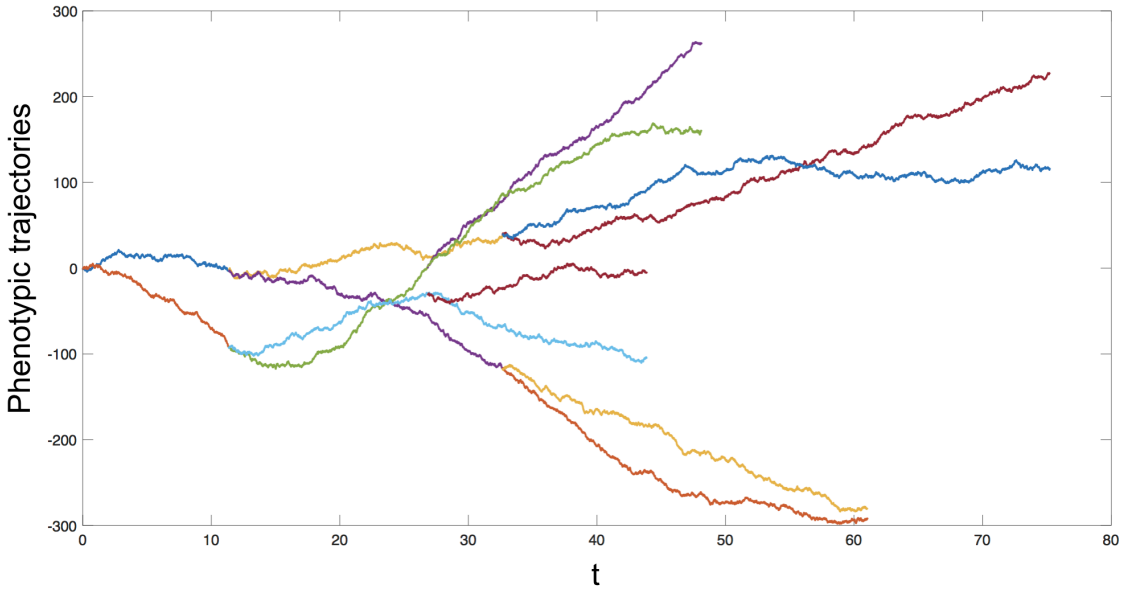


Figure I.6: Phenotypic trajectories simulated with variable migration functions, thus resulting in trees where epochs are no longer synchronous. Parameter values for the simulation were taken from Table 2.

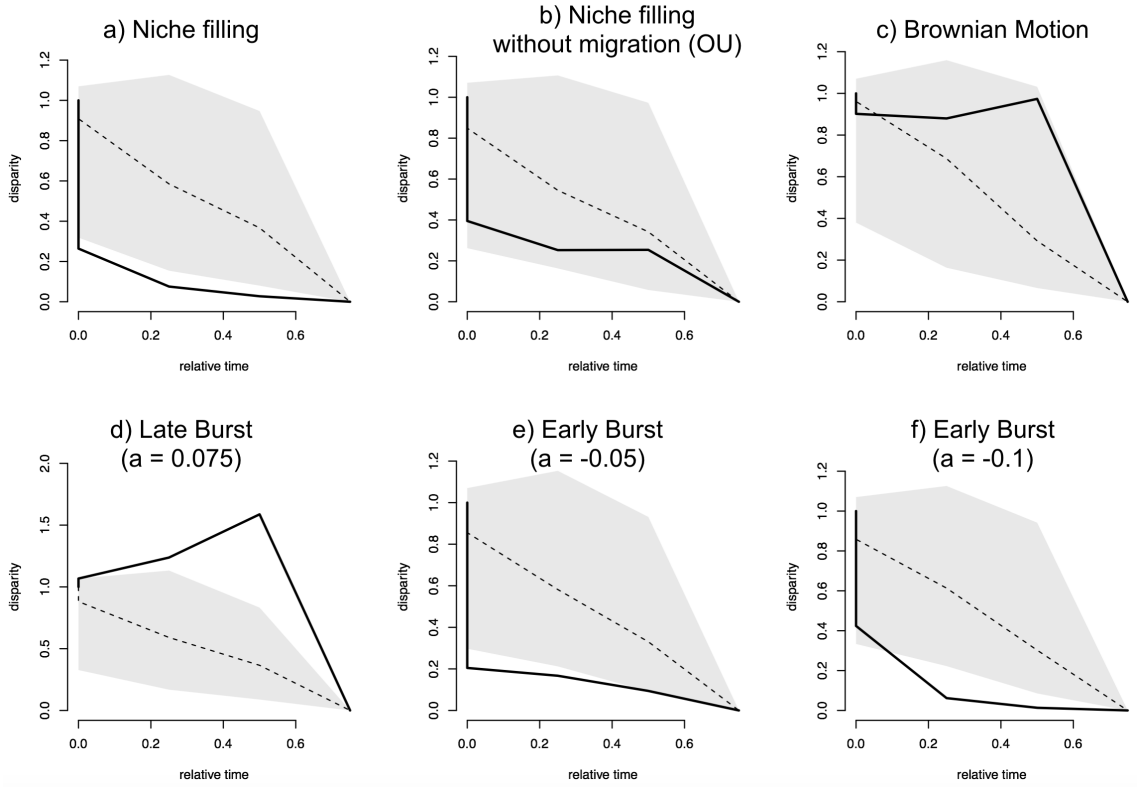


Figure I.7: Disparity-through-time (DTT) plots as calculated by the *dtt* function of the R package *geiger*. The dotted line is the expected disparity under BM and the grey area represents the 95% confidence interval on the disparity (where the *dtt* argument nsim was set to 100). In Panel a) we depict the DTT for phenotypic data generated under the niche-filling scenario presented in Fig. 4, with migration rate $c = 5.678 \times 10^{-6}$. Panel b) represents also niche filling, but without migration, thus resulting in a standard OU model. Subsequent panels use the same tree, but with phenotypic data simulated under other macroevolutionary models including BM, late burst, and early burst. The data simulated under all models is bounded within the range $[-A, A]$ (where $A = 50$). Late or early burst models, exponentially decrease or increase the rate of evolution with rate a.