Analysis of diversity-dependent species evolution using concepts in population genetics

Ingemar Kaj, Sylvain Glémin, Daniah Tahir, Martin Lascoux

HAL Id: hal-03157283
https://hal.science/hal-03157283

Mathematical Biology

Analysis of diversity-dependent species evolution using concepts in population genetics

Ingemar Kaj[1] · **Sylvain Glémin[2]** · **Daniah Tahir[1]** · **Martin Lascoux[3]**

Abstract

In this work, we consider a two-type species model with trait-dependent speciation, extinction and transition rates under an evolutionary time scale. The scaling approach and the diffusion approximation techniques which are widely used in mathematical population genetics provide modeling tools and conceptual background to assist in the study of species dynamics, and help exploring the analogy between trait-dependent species diversification and the evolution of allele frequencies in the population genetics setting. The analytical framework specified is then applied to models incorporating diversity-dependence, in order to infer effective results from processes in which the net diversification of species depends on the total number of species. In particular, the long term fate of a rare trait may be analyzed under a partly symmetric scenario, using a time-change transform technique.

1 Introduction

Models of species richness based on simple birth-death mechanisms with constant speciation and extinction rates suffer from the classical dichotomy of supercritical branching processes; species must either go extinct or grow in number without bound. Such models applied to characterizing divergence of living organisms in terms of number of species in taxas and families, consequently fail to produce what is typically

✉ Ingemar Kaj
 ikaj@math.uu.se

Extended author information available on the last page of the article

observed. The situation is similar for diversification models of multi-trait species based on multi-type, linear branching process theory. To qualify as a species model, at least one trait should have a strictly positive net growth rate, which again leads to the same supercritical dichotomy as for a single trait. See e.g., Haccou et al. (2005) for a general account of branching processes with emphasis on variation, growth and extinction of populations.

In reality, species numbers derived from fossil data or estimated from analysis of phylogenetic trees often tend to vary over long periods of time in a mode of stationarity or quasi-stationarity within a finite range of realistic values. It is, thus, rather natural to seek to implement in the modeling set-up, some form of population size control or diversity-dependent regulation. While logistic-type population-size dependence is successfully implemented in branching process theory and easily interpretable in term of population dynamics, the nature and causes of diversity-dependent diversification are still debated, cf. Lambert (2006), Rabosky (2013). The diversity-independent models studied in e.g., Maddison et al. (2007) and Tahir et al. (2019), assume that the per-lineage rates of speciation, extinction, and transition of traits are trait-dependent but constant in the sense that they do not depend on the current number of species of a particular trait, or the current total number of species. On the other hand, systems involving diversity-dependence retard supercritical growth by regulating the increase in species numbers through birth and death rates in various ways (Quental and Marshall 2010). In the papers of Parsons and Quince (2007a, b), that consider population-level dynamics, the birth rates are assumed to be decreasing functions of the total number of individuals in the population, thus, imposing a maximal carrying capacity of the system in case the birth rates vanish at some level. In logistic branching models, additional deaths are imposed in proportion to the square of the number of individuals to slow down population growth, Lambert (2005). Similarly, Parsons et al. (2008, 2010) apply population size dependent mortality rates. Fournier and Méléard (2004) consider an ecological system where individuals are characterized by their location, and the mortality rate depends on the local population density. In an environment of say two competing traits, both speciation and extinction rates could be diversity-dependent, and the rates could be allowed to have different sensitivity to competition (Mallet 2012). Further variations have been proposed, such as using time varying, instead of fixed, carrying capacities (Marshall and Quental 2016), and including adaptive radiations by decoupling of the diversity-dependent dynamics of a sub-clade (Etienne and Haegeman 2012). Also proposed are models in which positive net diversification rates are followed by negative net diversification rates to produce 'waxing and waning' diversity dynamics (Morlon 2014), and models where the speciation rates are decreasing functions of time, such that the magnitude of the decline in speciation increases as time approaches the environment's carrying capacity (Rabosky and Lovette 2008). Typical examples are discussed in e.g., Rabosky (2013) and Etienne et al. (2012), which summarize the parametrizations and properties of models involving density-dependence. In Rabosky (2013), Darwinian diversity-dependence and asymptotic diversity-dependence are contrasted. The former entails that a slowdown in speciation rates with diversity is in agreement with inter-species competition and Darwin's principle of divergence. The latter represents patterns of species saturation and long term stability of diversity trajectories, but does not specify clear mechanisms of cause and effect. In general, the time scales rele-

vant for diversity-dependence or diversity-independence are unspecified, thus leaving open whether regulatory effects act on microevolutionary rates over shorter ecological time scales, or on macroevolutionary rates averaged over much longer geological time spans (cf. Rabosky (2013). See also, Benton and Emerson (2007)).

We present here, a framework to address diversity-independent versus diversity-dependent dynamics for a family of species with a binary trait, for which the rates of creation and extinction of species as well as the rates of transition of the trait between species, are trait-dependent. Formal equivalence can be made between species level and population level dynamics, where number of species is equivalent to number of individuals, diversification rate is equivalent to growth rate, and transition in species traits is equivalent to mutation (e.g. Chevin (2016), Vellend (2010)). So, useful insights can be gained from using concepts of population genetics to species diversification analysis. We first rigorously investigate the mechanisms in species tree models which are analogous to those of population genetics models, such as mutation, selection, and genetic drift in allelic frequency models, and later, we apply the population genetics concepts to study diversity-dependent processes. The fraction of species in the family which carries one of the traits evolves in a manner directly comparable with the evolution of allele frequencies in the population genetics framework. The transition of a species of one trait to a species of the other trait resembles mutational change from one allelic type to another. Similarly, the trait-dependence present in creation rates, extinction rates, and the associated turnover rates, cause selection effects as well as frequency dependent genetic drift coefficients. Furthermore, various forms of population size dependence may be cast as density-dependent selection mechanisms. Despite, this formal equivalence, there are relevant biological differences between species level and population level models. For instance, whereas mutation is often much lower than selection in population genetics, there is no a priori reason to assume that rates of trait transition are of different magnitude than speciation and extinction rates. This makes species traits less "heritable" than individual traits. Besides, the number of species relevant to the species diversification dynamics is likely far less than the effective size of a species (100, 1000 for the former versus usually much more than 10^4 for the latter). This makes diversification dynamics more sensitive to stochastic effects (see below).

The central technique in our approach is a diffusion approximation of a species tree Markov chain, which runs on a suitable time scale of evolutionary time units. In particular, our scaling method incorporates the point of view that "macroevolutionary speciation rates can involve processes associated with both the splitting and extinction of populations over ecological and demographic timescales", as discussed in Rabosky (2013). From the complementary viewpoint of diffusion processes where the dynamics is provided by solutions of a PDE, our contribution relates closely to the pioneering work by Gillespie (1974). Using these links we study the effective population size in the model, and shed some light on an apparent paradoxical conclusion reported in Gillespie (1974). Several simplified, partly symmetric, parameter settings are identified in the stochastic model leading up to a reference case, or neutral case, which corresponds to fully neutral evolution in population genetics modeling.

Research has previously been carried out on similar topics, in which diffusion approximation methods were used to study density-dependent processes. Lambert

(2006) provides a branching process approach to studying probability of fixation for two-allele population models, including density-dependence mechanisms. Parsons and Quince (2007a) extended a logistic growth model to a supercritical model of two competing types, and using diffusion approximation methods, estimated the probability of fixation for both types. They proposed a so called 'non neutral' model, in which the ratios of birth to death rates were assumed to be different for the two types, and showed that the type with higher birth to death rate ratio eventually takes over the entire population. In a subsequent analysis by Parsons and Quince (2007b), the fixation probability was again approximated using diffusion processes, but for 'quasi neutral' populations in which both types had equal birth to death ratios. In this case, the type with the higher birth rate shows an increase in numbers during the growth phase at densities below the carrying capacity, whereas near the carrying capacity, the type with lower birth rate is favored, due to smaller fluctuations in population density – a phenomenon termed as 'r vs K selection' (see e.g., Pianka (1970)). Abu Awad and Coron (2018) use a scaling approach similar to ours to study effective population mass, absorption and extinction times, etc., for a population controlled by a carrying capacity. Chevin (2016) utilized diffusion approximation methods to analyze the evolution of binary discrete and continuous traits, and interpreted diversification models in terms of population genetics concepts of species selection and random drift. In our current work, we develop a further direction by providing a thorough mathematical basis and an analytical framework for the analogy between binary trait species diversification and population genetics models. For completeness and clarity, we re-derived some of these previous results using our proposed analytical framework.

2 The species branching model

We consider the dynamics of the size of a species family which carries a binary trait, marked 0 or 1, and undergoes trait-dependent splitting, extinction, and transition. We apply a two-type, continuous time Markov process $X = (X_u)_{u \geq 0}$ with components $X = (K, L)$, such that $X_u = (K_u, L_u)$ and K_u and L_u represent the number of species with trait 0 and trait 1, respectively, at time u. The two-type branching events are

$$
(k, \ell) \mapsto
\begin{cases}
\begin{array}{ll}
\text{target state} & \text{branching rate} \\
(k + 1, \ell) & \lambda_0 k \\
(k, \ell + 1) & \lambda_1 \ell \\
(k - 1, \ell) & \mu_0 k \\
(k, \ell - 1) & \mu_1 \ell \\
(k - 1, \ell + 1) & \delta_{01} k \\
(k + 1, \ell - 1) & \delta_{10} \ell
\end{array}
\end{cases}
\tag{1}
$$

in terms of splitting rates λ_0, λ_1, extinction rates μ_0, μ_1, and transition rates δ_{01} and δ_{10}. For the analysis in this work, however, it will be convenient to use the equivalent set of parameters consisting of net diversification rates d_i and turnover rates τ_i, $i = 0, 1$, given by

$$d_0 = \lambda_0 - \mu_0, \quad d_1 = \lambda_1 - \mu_1, \quad \tau_0 = \lambda_0 + \mu_0, \quad \tau_1 = \lambda_1 + \mu_1,$$

in addition to the transition rates δ_{01} and δ_{10}. Important aspects of the stochastic behavior of X_u are determined by the eigenvalues, γ_- and γ_+, $\gamma_- \leq \gamma_+$, of the 2×2 mean matrix

$$A = \begin{bmatrix} d_0 - \delta_{01} & \delta_{10} \\ \delta_{01} & d_1 - \delta_{10} \end{bmatrix}.$$

For the case $\gamma_+ \leq 0$ of subcritical ($<$) and critical ($=$) branching, the species family will go extinct at a random time η_0, the extinction time, which is finite with probability one and such that $X_u = (0, 0)$ for all $u \geq \eta_0$. Under the supercritical assumption $\gamma_+ > 1$, with a positive probability strictly less than one the process goes extinct, otherwise the species family survives and grows in size without bound. The species model is a two-type branching process $X = (X_u)_{u \geq 0}$ with linear jump rates, such that the dynamics of $X = (K, L)'$ (here viewed as column vector) satisfy

$$X_u = X_0 + \int_0^u A X_s \, ds + M_u, \quad u \geq 0, \tag{2}$$

where X_0 is the initial state of the model (e.g., $X_0 = (1, 0)'$) and $(M_u)_{u \geq 0}$ is a stochastic (martingale) term with mean $\mathbb{E}(M_u) = 0$. A general class of two-type branching processes, including population size dependent models, is characterized in Kaj and Tahir (2019) as unique solutions of such stochastic equations.

We are particularly interested in the representation (P, R) of X, given by

$$P_u = \frac{K_u}{K_u + L_u}, \quad R_u = K_u + L_u, \quad u \geq 0$$

so that P is the fraction of trait 0 among the species, and R is the total number of species, regardless of trait. Here, $R_u = 0$, $u \geq \eta_0$, and we note that P_u is well-defined for all u, with either $P_u = 0$ or $P_u = 1$ for $u \geq \eta_0$. Conversely, given (P, R), we obtain (K, L) from

$$K_u = P_u R_u, \quad L_u = (1 - P_u) R_u, \quad u \geq 0.$$

By using Eq. (2) and Itô's formula for pure jump processes, one obtains a stochastic equation for (P, R), see Section 4 in Kaj and Tahir (2019).

As a simple illustration, Fig. 1 shows a simulation of the two-type process $X = (K, L)$ and the corresponding representation (P, R). The choice of parameters is supercritical, such that the net growth rate is strictly positive for trait 0, $\gamma_+ = 6$, and strictly negative for trait 1, $\gamma_- = -5$. Thus, trait 1 on its own would go extinct. Here, however, the number of trait-1 species counted by L is sustained and growing by transitions from trait 0.
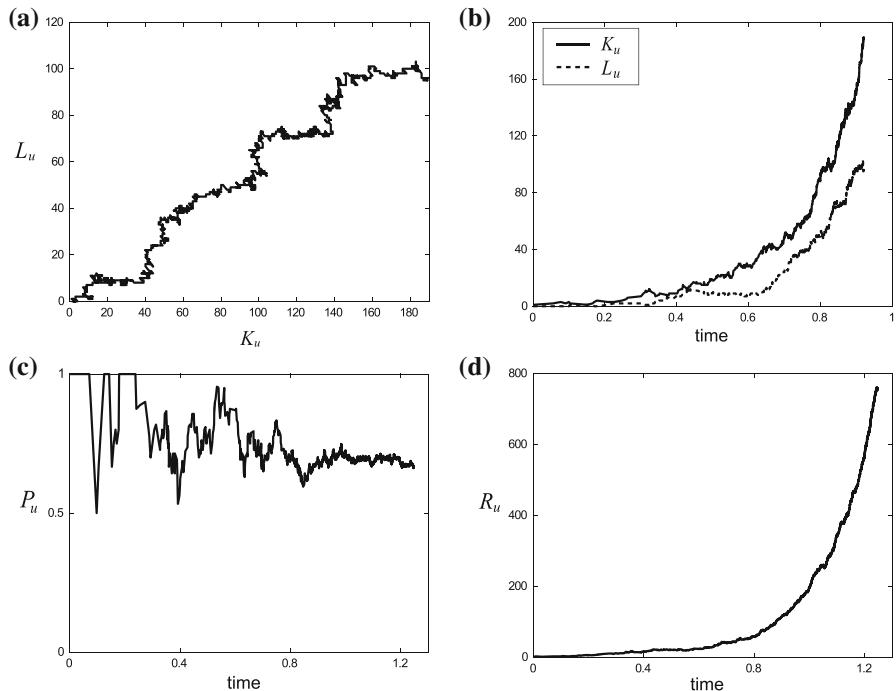
Fig. 1 Upper panels: simulation of the two-type branching process X; L versus K in a) and (K_u, L_u) versus time u in b). Lower panels: simulation of P versus time in c) and R versus time in d). Parameter values used for the simulations are $d_0 = 7$, $d_1 = 4$, $\tau_0 = 17$, $\tau_1 = 16$, $\delta_{01} = 4$ and $\delta_{10} = 6$

2.1 Law of large numbers

To understand the average behavior of (K, L) and of (P, R), it is useful to consider a system which starts at time $u = 0$ with m species, which is equivalent to summing m i.i.d. copies of the original model starting with one initial species each. Write $X^{(m)} = (K^{(m)}, L^{(m)})'$ for such a system, with e.g. $X_0^{(m)} = (m, 0)'$, and let $\widehat{X}^m = \frac{1}{m} X^{(m)}$ be the resulting average. The (strong) law of large numbers applies, so for each $u \geq 0$,

$$\widehat{X}_u^m \to \widehat{x}_u = (k_u, \ell_u)' \ \ a.s., \quad m \to \infty,$$

where the limit is the expected value $\widehat{x}_u = \mathbb{E}(X_u)$, which solves the linear ODE

$$\frac{d}{du} \widehat{x}_u = A \widehat{x}_u, \quad \widehat{x}_0 = (1, 0)'.$$

It follows from the law of large numbers for $\widehat{X}^m = (\widehat{K}^m, \widehat{L}^m)$, that the process (P^m, R^m), defined by

$$P_u^m = \frac{\widehat{K}_u^m}{\widehat{K}_u^m + \widehat{L}_u^m}, \quad R_u^m = \widehat{K}_u^m + \widehat{L}_u^m, \quad u \geq 0,$$

converges, as $m \to \infty$, to the solution (p, r), where $p_u = k_u/(k_u + \ell_u), r_u = k_u + \ell_u$, of the deterministic limiting ODE

$$\frac{d}{du}p_u = (d_0 - d_1)p_u(1 - p_u) - \delta_{01}p_u + \delta_{10}(1 - p_u), \quad p_0 = 1$$

$$\frac{d}{du}r_u = r_u\big(d_0 p_u + d_1(1 - p_u)\big), \quad r_0 = 1.$$

The limit equation for the fraction p, is a first indication of a connection to population genetics modeling, as it resembles the deterministic part of allele frequency dynamics (with mutation intensities δ_{01}, δ_{10} and selection intensity $d_0 - d_1$). In this model the net transition rate of trait 0, given by the term $-\delta_{01}p_u + \delta_{10}(1 - p_u)$, arises from so called anagenetic transitions between the two types. Anagenetic transitions occur along the branches of the species tree separate from speciation events. In contrast, a cladogenetic transition is a change in trait combined with the survival of the original type, hence associated with two-type speciation. To account for cladogenetic transitions occurring with probabilities a_0 and a_1 respectively for the two types, we modify the speciation and transition rates in (1) as

$$(k, \ell) \mapsto \begin{cases} (k + 1, \ell) & \lambda_0 k + a_1\delta_{10}\ell \\ (k, \ell + 1) & \lambda_1\ell + a_0\delta_{01}k \\ (k - 1, \ell + 1) & (1 - a_0)\delta_{01}\, k \\ (k + 1, \ell - 1) & (1 - a_1)\delta_{10}\, \ell, \end{cases}$$

while the extinction rates remain the same (Tahir et al. 2019). Then the stated ODE for (p, r) holds with d_0 and d_1 replaced by $\tilde{d}_0 = d_0 + a_0\delta_{01}$ and $\tilde{d}_1 = d_1 + a_1\delta_{10}$. Hence the relative impact of cladogenetic transitions as measured by a_0 and a_1 affects the "selection intensity" $\tilde{d}_0 - \tilde{d}_1$. This type of transition is analogous to a 'reduction in species level heritability of the trait' (Chevin 2016).

3 Evolutionary time scaling

In addition to the short time scale of "species generations", we introduce an evolutionary time scale and apply diffusion approximation methods to help understand the dynamics of the species model over the long time scale. Indeed, some of the most powerful tools of mathematical population genetics rely on approximation with diffusion processes, which allow for efficient computation of quantities such as fixation probabilities, expected times to fixation, and expected frequency spectra. The starting point for this approach is to relate the Markov chain X, with the population genetics *pre-limit* process, such as the Wright–Fisher model or Moran model, using the time scale of generations.

3.1 Scaled Wright–Fisher and Moran models

To see the relevance of time scales, we recall the haploid version of the standard biallelic Wright–Fisher model in discrete time $k \geq 0$, with fixed population size N, selection coefficient s representing reproductive weights $1 + s$ for trait 0 and 1 for

trait 1, and probabilities p_{01} and p_{10} for mutations from allele 0 to 1 and 1 to 0, respectively. Letting Z_0 be the initial number of trait-0 alleles and Z_k, $k \geq 1$, the number of 0-alleles after selective sampling of k new generations allowing for mutational change of traits in each step, this is the Markov chain $Z = (Z_k)_{k \geq 0}$. The standard diffusion approximation of the Wright–Fisher model involves a re-scaling of both time and population size with N, as well as a scaling of the model parameters. For this aim, let γ, ρ_{01}, and ρ_{10} be the scaling parameters which control the rate at which the strength of selection and mutation tends to 0 with increasing N, as

$$s_N = \gamma/N, \quad p_{01}^{(N)} = \rho_{01}/N, \quad p_{10}^{(N)} = \rho_{10}/N.$$

More generally, these relations can be understood as limit relations, for example $s_N \sim \gamma/N$, i.e., $\lim_{N \to \infty} N s_N = \gamma$. We write $Z^{(N)}$, for the associated scaled Wright–Fisher model, and consider the frequency process $\xi^N = (\xi_t^N)_{t \geq 0}$ on the evolutionary time scale of Nt generations, defined as

$$\xi_t^N = \frac{1}{N} Z_{[Nt]}^{(N)}, \quad t \geq 0.$$

The Wright–Fisher diffusion, $\xi = (\xi_t)_{t \geq 0}$, is the limit process of ξ^N as $N \to \infty$, and is known to be the unique, strong solution of the stochastic differential equation

$$d\xi_t = \gamma \xi_t (1 - \xi_t)\, dt - \rho_{01} \xi_t\, dt + \rho_{10}(1 - \xi_t)\, dt + \sqrt{\xi_t(1 - \xi_t)}\, dB_t, \quad \xi_0 = x,$$

$$(3)$$

with B a Brownian motion. Written in the form $d\xi_t = b(\xi_t)\, dt + \sigma(\xi_t)\, dB_t$, with

$$b(y) = \gamma y(1 - y) - \rho_{01} y + \rho_{10}(1 - y), \quad \sigma^2(y) = y(1 - y),$$

we may refer to b as the infinitesimal mean or the diffusion drift function and to σ^2 as the infinitesimal variance, the diffusion variance, or, the genetic drift function. The Moran model applies a birth-death mechanism for reproduction of alleles, rather than the characteristic binomial sampling of the Wright–Fisher model. Since birth events, where alleles copy and spread, are always compensated for by deaths of randomly sampled individuals, the population size is still maintained at a constant level N. The corresponding rescaling and diffusion approximation as $N \to \infty$, yield the same limit process except for a factor 2 in the variance function, $\sigma^2(y) = 2y(1 - y)$. The further parameterized case, $\sigma^2 = y(1 - y)/N_{\text{eff}}$, can be associated with the Wright–Fisher approximation of a system modulated by an effective population size, N_{eff}. For the general theory we refer to Etheridge (2011).

3.2 Gillespie approach

The infinitesimal generator of the Wright–Fisher diffusion process is the differential operator G, defined for a class of sufficiently regular functions f on the unit interval

by

$$Gf(y) = \frac{1}{2}y(1-y)f''(y) + \gamma y(1-y)f'(y), \quad 0 \le y \le 1,$$

for which the function $u(t, x) = \mathbb{E}_x[f(\xi_t)]$ solves Kolmogorov's backward equation

$$\frac{\partial}{\partial t}u_t(x) = Gu_t(x), \quad t > 0. \quad u_0(x) = f(x), \quad 0 < x < 1.$$

The population genetics model of Gillespie (1974) allows the population size to vary following an offspring distribution with mean $1 + \mu_i$ and variance σ_i^2, for two alleles A_i, $i = 1, 2$. Writing p for the frequency of allele A_1 and r for the total population size, this approach suggests the two-dimensional Markov generator

$$
\begin{aligned}
Gf(p, r) &= \left(\mu_1 - \mu_2 + \frac{1}{r}(\sigma_2^2 - \sigma_1^2)\right)\frac{\partial f}{\partial p} + \frac{p(1-p)}{r}(p\sigma_2^2 + (1-p)\sigma_1^2)\frac{1}{2}\frac{\partial^2 f}{\partial p^2} \\
&\quad + r(p\mu_1 + (1-p)\mu_2)\frac{\partial f}{\partial r} + r(p\sigma_1^2 + (1-p)\sigma_2^2)\frac{1}{2}\frac{\partial^2 f}{\partial r^2} \\
&\quad + p(1-p)(\sigma_1^2 - \sigma_2^2)\frac{\partial^2 f}{\partial p \partial r}.
\end{aligned}
\tag{4}
$$

Here, we are using r for population size as opposed to the more classical population genetics notation n as in Gillespie (1974), to avoid mix-up with our general scaling parameter n used below. The associated Kolmogorov's backward equation is dual to the corresponding forward equation given as Equation 3 in Gillespie (1974). In particular, we read out that $r/(p\sigma_2^2 + (1-p)\sigma_1^2)$ acts as a varying effective population size. It is perhaps worth stressing that this measure of effective population size varies not only with the size r but also, in general, with the frequency p, due to possibly different variance parameters σ_1^2, σ_2^2. Earlier references on random effective population size and stochastic population size models can be found in Kaj and Krone (2003), Sjödin et al. (2005), together with developments on coalescent models for such situations. An additional result in Gillespie (1974), in particular Equations 6,7, refers to the special case $\mu_1 = \mu_2$ saying that the probability of fixation of allele A_1, given an initial frequency p, equals

$$u(p) = \frac{\sigma_2^2 p}{\sigma_1^2(1-p) + \sigma_2^2 p} \tag{5}$$

The diffusion approximation model of species richness we develop next will be closely related to the model described by Eq. (4). We will recover and discuss further the role of both the random effective population size $r/(p\sigma_2^2 + (1-p)\sigma_1^2)$ and the fixation probability $u(p)$.

3.3 Scaled species model

Our investigation takes the view that, given non-extinction of the continuous time branching process X, the component $K = (K_u)_{u \geq 0}$ in the species model relates naturally to the discrete time Markov chain $(Z_k)_{k \geq 0}$, of the population genetics model. Of course these quantities are not immediately comparable since the total number of species R varies randomly, while the population size N is fixed from one generation to the next. For this reason, we introduce a species family process $X^{(n)} = (K^{(n)}, L^{(n)})'$ marked by a separate scaling variable n. The two-type branching process is driven by an initial number of species of magnitude n, and has scaled parameters $d_0^{(n)}, d_1^{(n)}, \tau_0^{(n)}$, $\tau_1^{(n)}, \delta_{01}^{(n)}$ and $\delta_{10}^{(n)}$, which control the rates of diversification, turnover, and transition on the population timescale. For fixed t, the fraction $n^{-1} X_{[nt]}^{(n)}$, is the typical configuration of trait frequencies in the species family seen over the span of t time units that elapse upon completing $[nt]$ generations. The precise behavior of the rates as functions of n is not crucial as long as they scale for large n with the following macroevolutionary rates which act on the new time scale,

$$d_i^{(n)} \sim \frac{\beta_i}{n}, \quad \tau_i^{(n)} \sim \tau_i, \quad i = 0, 1 \qquad \delta_{01}^{(n)} \sim \frac{\rho_{01}}{n}, \quad \delta_{10}^{(n)} \sim \frac{\rho_{10}}{n}. \qquad (6)$$

Here, τ_i measures variance in the production of daughter species and is the reference turnover parameter for extinction and/or speciation events of trait i, β_i is the macroevolutionary net diversification rate of trait i, seen as the long-time average net effect of speciation and extinction, and ρ_{01} and ρ_{10} are the resulting macroevolutionary rates of exchange of traits. For later reference we also adopt from Gillespie (1977) the *new evolutionary principle* that trait success (fitness in population genetics terminology) not only increases with offspring mean but also may decrease with offspring variance inversely proportional to population size. Using macroevolution terminology, τ refers to the species turn-over rate, a quantity which has received increasing attention in the empirical literature aiming at interpreting patterns of species diversity (Gamisch and Comes 2019; Han et al. 2020; Nakov et al. 2019). In the current situation of developmental stochastic dynamics, as opposed to environmental stochastic influences, β_i (or rather $1 + \beta_i$) represent offspring mean and τ_i offspring variance. Then, following Gillespie (1974, 1977) the fitness is properly measured by $f_i(r)$, where

$$f_i(r) = \beta_i - \frac{\tau_i}{r}, \quad i = 0, 1. \qquad (7)$$

Conditionally on the set of paths of X which do not go extinct, let

$$P_t^n = \frac{K_{nt}^{(n)}}{K_{nt}^{(n)} + L_{nt}^{(n)}}, \qquad R_t^n = \frac{1}{n}(K_{nt}^{(n)} + L_{nt}^{(n)}), \quad t \geq 0.$$

The ratio P^n will play the role of the frequency process ξ^N, even though the scaled Wright Fisher process ξ^N is derived from a population of fixed size N whereas P^n is a frequency with respect to the stochastically varying species richness process R^n.

The final step towards a generic species trait richness model is finding the limit processes in the scaling regime as $n \to \infty$, expecting the remaining macroevolutionary rates to be τ_0, τ_1 as rates of change for the speciation/extinction process, β_0, β_1 diversification rates for net growth (if positive) or net decline (if negative), and ρ_{01}, ρ_{10} controlling trait interchange. We expect in the limit to obtain a bi-variate, coupled diffusion process, $\mathcal{X} = (\mathcal{X}_t)_{t \geq 0}$, with representations $(\mathcal{K}, \mathcal{L})'$ and $(\mathcal{P}, \mathcal{R})$ say. To assist in identifying the limit model we will exploit the existing theory of measure-branching processes, briefly laid out in Sect. 5 below. In fact, it is straightforward to re-interpret $X^{(n)}$ as a measure-valued branching process on trait space $E = \{0, 1\}$ with scaled parameters. In this view our macroevolutionary scaling is consistent with a super-process approximation, for which the spatial motion is the on-off process on E with jump rates ρ_{01}, ρ_{10}. As a consequence, $(n^{-1} X^{(n)}_{[nt]})_{t \geq 0}$, has a weak limit which is a super-on/off process with binary branching.

Proposition 1 *Suppose that*

$$n^{-1} X_0^{(n)} \to \mathcal{X}_0, \quad n \to \infty,$$

where $\mathcal{X}_0$ is a nonnegative deterministic column vector. As $n \to \infty$ the scaled branching process $(n^{-1} X_{nt}^{(n)})_{t \geq 0}$, converges weakly to the continuous state branching process $\mathcal{X} = (\mathcal{K}, \mathcal{L})'$, which is the unique solution of the SDE

$$\mathcal{X}_t = \mathcal{X}_0 + \int_0^t \begin{bmatrix} \beta_0 - \rho_{01} & \rho_{10} \\ \rho_{01} & \beta_1 - \rho_{10} \end{bmatrix} \mathcal{X}_s \, ds + \binom{1}{0} \int_0^t \sqrt{\tau_0 \mathcal{K}_s} \, dB_s^0$$

$$+ \binom{0}{1} \int_0^t \sqrt{\tau_1 \mathcal{L}_s} \, dB_s^1, \quad t \geq 0,$$

where B^0, B^1 are independent, standard Brownian motions. Furthermore, (P^n, R^n) converges weakly to $(\mathcal{P}, \mathcal{R})$, the unique solution of

$$d\mathcal{P}_t = \mathcal{P}_t(1 - \mathcal{P}_t)(f_0(\mathcal{R}_t) - f_1(\mathcal{R}_t)) \, dt - \rho_{01} \mathcal{P}_t \, dt + \rho_{10}(1 - \mathcal{P}_t) \, dt$$

$$+ \sqrt{\mathcal{P}_t(1 - \mathcal{P}_t)(\tau_0(1 - \mathcal{P}_t) + \tau_1 \mathcal{P}_t) \frac{1}{\mathcal{R}_t}} \, dB_t^-, \qquad (8)$$

$$d\mathcal{R}_t = \mathcal{R}_t(\beta_0 \mathcal{P}_t + \beta_1(1 - \mathcal{P}_t)) \, dt + \sqrt{\mathcal{R}_t(\tau_0 \mathcal{P}_t + \tau_1(1 - \mathcal{P}_t))} \, dB_t^+, \qquad (9)$$

where $f_0(r)$ and $f_1(r)$ are the fitness measures in Eq. (7) and B_t^-, B_t^+ are standard Brownian motions. The Brownian motions are independent when $\tau_0 = \tau_1$. In general, the quadratic covariation processes of (B^-, B^+) and $(\mathcal{P}, \mathcal{R})$ are given by

$$\langle\!\langle B^-, B^+ \rangle\!\rangle_t = \left(\frac{\tau_0}{\tau_1} - 1\right) \int_0^t \frac{\sqrt{\mathcal{P}_s(1 - \mathcal{P}_s)}}{\sqrt{(\tau_0(1 - \mathcal{P}_s)/\tau_1 + \mathcal{P}_s)(\tau_0 \mathcal{P}_s/\tau_1 + 1 - \mathcal{P}_s)}} \, ds,$$

and

$$\langle\!\langle \mathcal{P}, \mathcal{R} \rangle\!\rangle_t = (\tau_0 - \tau_1) \int_0^t \mathcal{P}_s(1 - \mathcal{P}_s)\, ds. \tag{10}$$

3.4 Effective population size: resolving Gillespie's paradox

The infinitesimal mean and infinitesimal variance terms of the diffusions $\mathcal{P}$ and $\mathcal{R}$, except for the mutation terms involving ρ_{01} and ρ_{10}, as well as the covariation in Eq. (10), have matching terms in the generator G in Eq. (4). In Eq. (8) we recognize a Wright–Fisher diffusion with selection and mutation, and with trait-dependent genetic drift (or 'species drift') $\mathcal{P}_t(1 - \mathcal{P}_t)/\mathcal{N}_t$ in the diffusion variance of B^-, such that

$$\mathcal{N}_t = \frac{\mathcal{R}_t}{\tau_0(1 - \mathcal{P}_t) + \tau_1 \mathcal{P}_t} \tag{11}$$

acts as a stochastically varying effective population size for the differential dynamics of $d\mathcal{P}_t$. This observation is parallel to that in the original work Gillespie (1974), where the corresponding quantity appears in the coefficient of $\partial^2 f / \partial p^2$ in Eq. (4). The species drift increases as the number of species decreases, analogous to genetic drift of the allele frequency in population genetics. Similarly, the species drift increases with the turn over rates τ_0 and τ_1, which represent the species offspring variance. By applying Ito's formula to $\mathcal{N}_t = g(\mathcal{P}_t, \mathcal{R}_t), t \geq 0$, with $g(p, r) = r/(\tau_0(1-p)+\tau_1 p)$, we obtain directly the dynamics of $(\mathcal{P}, \mathcal{N})$ as an alternative to that of $(\mathcal{P}, \mathcal{R})$. For simplicity, we consider the case where the diversification rates are equal, $\beta_0 = \beta_1 = \beta$, and the rates of trait exchange vanish, $\rho_{01} = \rho_{10} = 0$. Then

$$d\mathcal{P}_t = -\frac{(\tau_0 - \tau_1)\mathcal{P}_t(1 - \mathcal{P}_t)}{\tau_0(1 - \mathcal{P}_t) + \tau_1 \mathcal{P}_t} \frac{1}{\mathcal{N}_t}\, dt + \sqrt{\mathcal{P}_t\left(1 - \mathcal{P}_t\right)\frac{1}{\mathcal{N}_t}}\, dB_t^-,$$

$$d\mathcal{N}_t = \beta \mathcal{N}_t\, dt + \frac{(\tau_0 - \tau_1)^2 \mathcal{P}_t(1 - \mathcal{P}_t)}{(\tau_0(1 - \mathcal{P}_t) + \tau_1 \mathcal{P}_t)^2}\, dt$$
$$+ \sqrt{\mathcal{N}_t \frac{\tau_0 \mathcal{P}_t + \tau_1(1 - \mathcal{P}_t)}{\tau_0(1 - \mathcal{P}_t) + \tau_1 \mathcal{P}_t}}\, dB_t^+ + \sqrt{\mathcal{N}_t \frac{(\tau_0 - \tau_1)\mathcal{P}_t(1 - \mathcal{P}_t)}{\tau_0(1 - \mathcal{P}_t) + \tau_1 \mathcal{P}_t}}\, dB_t^-$$

The benefit of this additional representation of $\mathcal{P}$ in comparison with the previous stochastic equation (8), is that the effective population size $\mathcal{N}$ in the species drift of $\mathcal{P}$ now appears in classical form. The disadvantage is the increased complexity of the diffusion drift for $\mathcal{P}$ as well as diffusion drift and variance of $\mathcal{N}$, as compared to those of $\mathcal{R}$ in (9).

The discussion in Gillespie (1974) also points to a potential drawback of the appearance of $\mathcal{N}$ in the present form in (11). If trait 1 is rare or near elimination from the species family, so that $\mathcal{P} \approx 1$, then formally $\mathcal{N} \approx \mathcal{R}/\tau_1$ which suggests that the now-absent trait 1 controls the effective size through the parameter τ_1. Quoting Gillespie (1974): *This rather uncomfortable conclusion suggests that the concept of effective*

population size loses a good deal of its value in the context of the present model. Using the wider frame of linked diffusion processes, however, this apparent paradoxical conclusion may be avoided. The above relation for $d\mathcal{N}_t$ suggests that if trait 1 would be eliminated and hence $\mathcal{P}$ reach fixation at 1, then the resulting effective size $(\mathcal{N}_t^0)_{t \geq 0}$, would satisfy

$$d\mathcal{N}_t^0 = \beta \mathcal{N}_t^0 \, dt + \sqrt{\mathcal{N}_t^0 (\tau_0/\tau_1)} \, dB_t^+,$$

and hence depend on the relative turnover ratio τ_0/τ_1 and β, rather than just τ_1. Of course, for the same parameter settings, if we let $\mathcal{P}$ tend to 1 in (9), the resulting richness $\mathcal{R}_t^0$ of trait 0 satisfies

$$d\mathcal{R}_t^0 = \beta \mathcal{R}_t^0 \, dt + \sqrt{\mathcal{R}_t^0 \tau_0} \, dB_t^+, \tag{12}$$

consistent with the relation $\mathcal{N}^0 = \mathcal{R}^0/\tau_1$, which appears to have been the origin of the "uncomfortable conclusion" discussed in Gillespie (1974). The diffusions $\mathcal{N}^0$ and $\mathcal{R}^0$, are Feller processes with linear drift belonging to the larger class of continuous state branching processes. It is a classical result in branching process theory that, if $\mathcal{R}_0^0 = r$, and accordingly $\mathcal{N}_0^0 = r/\tau_1$, then the extinction probabilities are given by

$$\mathbb{P}_r(\lim_{t \to \infty} \mathcal{R}_t^0 = 0) = e^{-2r\beta/\tau_0}$$

and

$$\mathbb{P}_{r/\tau_1}(\lim_{t \to \infty} \mathcal{N}_t^0 = 0) = e^{-2(r/\tau_1) \cdot \beta/(\tau_0/\tau_1)} = e^{-2r\beta/\tau_0}.$$

Naturally, they no more depend on the parameter τ_1. In response to a question raised by an anonymous referee, we also mention an alternative approach to this topic. By changing variables from (p, r) to (p, n) with $n = r/(\sigma_1^2(1 - p) + \sigma_2^2 p)$ one obtains a new generator $Gf(p, n)$ parallel to $Gf(p, r)$ in (4). The coefficients in $Gf(p, n)$ will now have matching terms to those in the SDE for $(\mathcal{P}_t, \mathcal{N}_t)$.

3.5 Further interpretation of Eqs. (8–10)

The species 'selection coefficient' in Eq. (8) is the fitness difference

$$\gamma_t = f_0(\mathcal{R}_t) - f_1(\mathcal{R}_t) = \beta_0 - \beta_1 - \frac{\tau_0 - \tau_1}{\mathcal{R}_t}, \quad t \geq 0,$$

dependent on net diversification rates, turnover rates and species richness. If $\beta_0 > \beta_1$, so that trait 0 has larger net diversification, then trait 0 has a "fitness" advantage over trait 1, $\gamma_t > 0$ at time t, if either $\tau_1 > \tau_0$ or $\tau_0 > \tau_1$ and $\mathcal{R}_t > (\tau_0 - \tau_1)/(\beta_0 - \beta_1)$. In other words, species-level selection is reduced in small populations and enhanced in large. If $\tau_1 > \tau_0$, the opposite holds. Hence, traits with higher diversification rates

and slower turnover rates are favored by species selection. The time average of γ up to time t is

$$\frac{1}{t}\int_0^t \gamma_s\,ds = (\beta_0 - \beta_1)(1 - F_t),$$

where

$$F_t = \frac{\tau_0 - \tau_1}{\beta_0 - \beta_1}\frac{1}{t}\int_0^t \frac{1}{\mathcal{R}_s}\,ds$$

can be seen as the relative contribution of $\tau_0 - \tau_1$ to species selection (Chevin 2016).

The total species richness is measured by $\mathcal{R}$ in Eq. (9), which is a continuous state branching process with linear drift and trait-dependence in both the diffusion drift and variance functions. Conditional on non-extinction, $\mathcal{R}$ will grow exponentially as $t \to \infty$. Hence, the diffusion variance term for $d\mathcal{P}_t$ in (8), driven by B^-, vanishes. The diffusion drift function for $d\mathcal{P}_t$ also simplifies and we expect $\mathcal{P}_t \to p_\infty$ as $t \to \infty$, where p_∞ is the solution of the second order equation

$$p_\infty(1 - p_\infty)(\beta_0 - \beta_1) - \rho_{01}p_\infty + \rho_{10}(1 - p_\infty) = 0. \tag{13}$$

The covariation of $\mathcal{P}$ and $\mathcal{R}$ in Eq. (10) measures the degree to which these quantities vary simultaneously, in the sense that the expected change over a small time interval $[t, t + h)$, conditional on the present state $(\mathcal{P}_t, \mathcal{R}_t)$, is given by

$$\mathbb{E}[(\mathcal{P}_{t+h} - \mathcal{P}_t)(\mathcal{R}_{t+h} - \mathcal{R}_t)|(\mathcal{P}_t, \mathcal{R}_t)] = (\tau_0 - \tau_1)\mathcal{P}_t(1 - \mathcal{P}_t)h + o(h), \quad h \to 0.$$

We observe that, while the covariation of $\mathcal{P}$ and $\mathcal{R}$ is proportional to the difference in turnover rates, $\tau_0 - \tau_1$, the analogous covariation of the two underlying Brownian motions, B^- driving $\mathcal{P}$ and B^+ driving $\mathcal{R}$, is a function only of the turnover-ratio τ_0/τ_1. In both cases, the dependence is positive in case $\tau_0 > \tau_1$ and negative otherwise. We conclude that the species richness co-varies positively with the trait with the highest turnover. The turnover ratio τ_0/τ_1 determines the strength of the covariation which is intrinsic at the level of Brownian noise, whereas both of τ_0 and τ_1 are required in order to find the covariation of the state variables $\mathcal{P}$ and $\mathcal{R}$.

3.6 Special cases

(a) Symmetric net diversification rates, $\beta_0 = \beta_1$

Let $\beta_0 = \beta_1 = \beta$. Then, from Eq. (8) and Eq. (9), we have

$$d\mathcal{P}_t = \mathcal{P}_t(1 - \mathcal{P}_t)\frac{-(\tau_0 - \tau_1)}{\mathcal{R}_t}\,dt - \rho_{01}\mathcal{P}_t\,dt + \rho_{10}(1 - \mathcal{P}_t)\,dt$$

$$+ \sqrt{\mathcal{P}_t(1 - \mathcal{P}_t)(\tau_0(1 - \mathcal{P}_t) + \tau_1\mathcal{P}_t)\frac{1}{\mathcal{R}_t}}\,dB_t^-,$$

$$dR_t = \beta R_t \, dt + \sqrt{R_t \big(\tau_0 P_t + \tau_1 (1 - P_t) \big)} \, dB_t^+.$$

By Eq. (10) the covariance structure remains the same as in the general model. Here, R is a supercritical Feller-type diffusion process with diffusion variance function modulated by the trait proportions P and $1 - P$. The SDE for P can be viewed as a Wright–Fisher diffusion with population size dependent infinitesimal drift and variance functions. The selection coefficient $-(\tau_0 - \tau_1)/R$ shows that the species associated with the trait of the smallest of the two turnover rates, i.e. trait 0 if $\tau_0 < \tau_1$ and trait 1 if $\tau_0 > \tau_1$, are selected for instead of species with the higher turnover. Biologically, this means that, all other things being equal, more long-lived species are selected for, as increasing the species generation time will automatically reduce τ (Lin et al. 2012).

(b) Symmetric turnover rates, $\tau_0 = \tau_1$

Let $\tau_0 = \tau_1 = \tau$. Then Eq. (8) and Eq. (9) simplify as

$$dP_t = P_t(1 - P_t)(\beta_0 - \beta_1) \, dt - \rho_{01} P_t \, dt + \rho_{10}(1 - P_t) \, dt$$
$$+ \sqrt{\frac{\tau P_t (1 - P_t)}{R_t}} \, dB_t^-,$$
$$dR_t = R_t \big(\beta_0 P_t + \beta_1 (1 - P_t) \big) \, dt + \sqrt{\tau R_t} \, dB_t^+,$$

with independent Brownian motions B^- and B^+. The equation for P is the SDE of the Wright–Fisher diffusion process with selection coefficient $\beta_0 - \beta_1$ and mutation rates ρ_{01} and ρ_{10} as in Eq. (3), except that now the genetic drift term is inversely proportional to R, which acts as a randomly varying effective population size. At the same time, the population richness process R is a Feller-type diffusion process with diffusion drift function modulated by the trait proportions P and $1 - P$.

(c) Neutral evolution: symmetric net diversification rates, $\beta_0 = \beta_1$ and turn over rates, $\tau_0 = \tau_1$

Let $\tau_0 = \tau_1 = \tau$ and $\beta_0 = \beta_1 = \beta$ be nonnegative parameters. By combining the cases a) and b), we obtain

$$dP_t = -\rho_{01} P_t \, dt + \rho_{10}(1 - P_t) \, dt + \sqrt{\frac{\tau P_t (1 - P_t)}{R_t}} \, dB_t^-,$$
$$dR_t = \beta R_t \, dt + \sqrt{\tau R_t} \, dB_t^+.$$

Here, the first equation of P is comparable to the Wright–Fisher diffusion process Eq. (3) with only mutation and no selection, hence the term 'neutral evolution'. The total size process is a supercritical Feller diffusion process, which we encountered previously in (12) as a limiting case under trait fixation. In the present context, we

use once more that for a given initial size $\mathcal{R}_0 = r$, the extinction probability that $\mathcal{R}_t \to 0$ as $t \to \infty$, equals $e^{-2\beta r/\tau}$. From this we retrieve the known result in stochastic demography that higher turn-over rate increases the extinction risk. On the complementary set of nonextinction, the total size process tends to infinity almost surely, $\mathcal{R}_t \to \infty$ a.s.

(d) Quasi-neutral rates

Quasi-neutrality, as proposed and discussed in Parsons and Quince (2007b), Parsons et al. (2008), is a rate symmetry condition, which in our notation would mean

$$\frac{\tau_0}{\beta_0} = \frac{\tau_1}{\beta_1} = \kappa.$$

According to Parsons et al. (2008), the ratio $2\beta_i/\tau_i$ is an alternative measure of relative success or relative fitness of trait i. In this framework, therefore, quasi-neutral traits are equally fit. Eq. (8) and Eq. (9) become

$$d\mathcal{P}_t = \mathcal{P}_t(1 - \mathcal{P}_t)\Big((\beta_0 - \beta_1)\big(1 - \frac{\kappa}{\mathcal{R}_t}\big)\Big) dt - \rho_{01}\mathcal{P}_t \, dt + \rho_{10}(1 - \mathcal{P}_t) \, dt$$
$$+ \sqrt{\mathcal{P}_t(1 - \mathcal{P}_t)(\beta_0(1 - \mathcal{P}_t) + \beta_1\mathcal{P}_t)\frac{\kappa}{\mathcal{R}_t}} \, dB_t^-,$$
$$d\mathcal{R}_t = \mathcal{R}_t(\beta_0\mathcal{P}_t + \beta_1(1 - \mathcal{P}_t)) \, dt + \sqrt{\kappa\mathcal{R}_t(\beta_0\mathcal{P}_t + \beta_1(1 - \mathcal{P}_t))} \, dB_t^+.$$

4 Application of the population genetics approach to diversity-dependent models

We discuss the effect of diversity dependence on the species richness model. The interaction is imposed by letting the instantaneous rates of speciation and extinction in (6) depend on the scaled total size R^n. A variety of interaction schemes can be incorporated into the measure-branching model formalism, which we use to derive Proposition 1. For an introduction to some of the principles of interaction in measure-branching and superprocesses, see Etheridge (2000). The relevant approach for us is the use of an analogue of the Girsanov theorem that allows for the introduction of "non-linear branching", and in the end logistic versions of the superprocess limits obtained by an absolutely continuous change-of-drift. Our analysis is therefore similar in many ways to the development of the spatial model in Fournier and Méléard (2004), considering our trait space as the mark of location. As in Fournier and Méléard (2004) we apply superprocess techniques but at this stage our interaction schemes are less general. It is also known that logistic Feller diffusion processes arise under much more general branching mechanisms than ours (Lambert 2005), indicating more general results.

Suppose the current state of the species model at time t is (P_t^n, R_t^n). A constant c will have the role of a 'carrying capacity'. We consider in Sect. 4.1 the diversity-dependent species models which have jump rates at t, given by

$$\tau_i^{(n)} = \tau_i, \quad d_i^{(n)} = \frac{\beta_i}{n}\left(1 - \frac{R_t^n}{c}\right).$$

- In greater generality, the carrying capacity is determined by a pair of parameters, c_0, c_1, and the strength in reduction of the net growth is trait-dependent. In Sect. 4.2 we consider the corresponding model with jump rates

$$\tau_i^{(n)} = \tau_i, \quad d_i^{(n)} = \frac{\beta_i}{n}\left(1 - \frac{R_t^n}{c_i}\right), \quad i = 0, 1.$$

Further generalization can also be made by assuming that the two kinds of species both affect and are affected differently by species richness:

$$d_i^{(n)} = \frac{\beta_i}{n}\left(1 - R_t^n\left(\frac{P_t^n}{c_{0i}} + \frac{1 - P_t^n}{c_{1i}}\right)\right), \quad i = 0, 1.$$

This parallels Lotka-Volterra competition models but this latter case will not be developed further here and is left for future work. We have applied classical logistic type diversity-dependence acting on the macroevolutionary net growth rate $nd_i^{(n)}$. Starting from $R_0^n < c$, the net effect of speciations and extinctions is reduced with increasing species richness and turns into a subcritical regime above the level c, in case $R_t^n > c$.

To simplify notation in the rest of this section we introduce the function

$$\psi(p) = \tau_0(1 - p) + \tau_1 p, \quad 0 \le p \le 1. \tag{14}$$

4.1 Diversity-dependent macroevolutionary diversification rates

The limit process as $n \to \infty$ is obtained as in Proposition 1, with β_0 and β_1 at time t replaced by the species richness dependent functions $\beta_0(1 - \mathcal{R}_t/c)$ and $\beta_1(1 - \mathcal{R}_t/c)$. The system in Eqs. (8)–(9) becomes

$$d\mathcal{P}_t = \mathcal{P}_t(1 - \mathcal{P}_t)\big(f_0^c(\mathcal{R}_t) - f_1^c(\mathcal{R}_t)\big)\,dt$$
$$- \rho_{01}\mathcal{P}_t\,dt + \rho_{10}(1 - \mathcal{P}_t)\,dt + \sqrt{\mathcal{P}_t(1 - \mathcal{P}_t)\psi(\mathcal{P}_t)\frac{1}{\mathcal{R}_t}}\,dB_t^-, \tag{15}$$

$$d\mathcal{R}_t = \mathcal{R}_t\left(1 - \frac{\mathcal{R}_t}{c}\right)\big(\beta_0\mathcal{P}_t + \beta_1(1 - \mathcal{P}_t)\big)\,dt + \sqrt{\mathcal{R}_t\psi(1 - \mathcal{P}_t)}\,dB_t^+,$$

with trait fitness now of the form

$$f_i^c(r) = \beta_i\left(1 - \frac{r}{c}\right) - \frac{\tau_i}{r}.$$

In particular, if we impose the additional assumptions of neutral rates discussed as special case c) in Sect. 3.6, the process $\mathcal{R}$ is the solution of

$$d\mathcal{R}_t = \beta\mathcal{R}_t(1 - \mathcal{R}_t/c)\,dt + \sqrt{\tau\mathcal{R}_t}\,dB_t^+, \quad t \ge 0.$$

This is the logistic Feller diffusion process, which has non-negative paths and goes extinct with probability one. The logistic Feller diffusion is known to have a quasi-stationary distribution in the sense of a Yaglom limit, namely that the law of $\mathcal{R}_t$ conditioned on $\{\mathcal{R}_t > 0\}$ converges as $t \to \infty$ to a probability measure (Cattiaux et al. 2009). Such a Yaglom limit, however, is not known in explicit terms and need not even be unique. For our model, we note that Cattiaux et al. (2009) Thm 8.2, yields uniqueness of the Yaglom limit for the case $c < 1$, whereas under biologically relevant conditions we expect $c > 1$.

For the general, non-neutral, case, letting $\widetilde{\mathcal{R}} = \mathcal{R}/c$ be the total species richness in relation to the carrying capacity c, the SDE system for the state variables is

$$dP_t = P_t(1 - P_t)\Big((\beta_0 - \beta_1)(1 - \widetilde{\mathcal{R}}_t) - \frac{\tau_0 - \tau_1}{c\widetilde{\mathcal{R}}_t}\Big) dt$$

$$- \rho_{01}P_t \, dt + \rho_{10}(1 - P_t) \, dt + \sqrt{P_t(1 - P_t)\psi(P_t)\frac{1}{c\widetilde{\mathcal{R}}_t}} \, dB_t^-,$$

$$d\widetilde{\mathcal{R}}_t = \widetilde{\mathcal{R}}_t(1 - \widetilde{\mathcal{R}}_t)\big(\beta_0 P_t + \beta_1(1 - P_t)\big) dt + \sqrt{\frac{1}{c}\widetilde{\mathcal{R}}_t\psi(1 - P_t)} \, dB_t^+.$$

The two SDEs with respect to B^- and B^+ have second order moments proportional to $1/c$. This can be used to show formally that the diffusion terms vanish in the subsequent limit of large carrying capacity, $c \to \infty$, revealing the deterministic limit equation

$$dP_t = P_t(1 - P_t)(\beta_0 - \beta_1)(1 - \widetilde{\mathcal{R}}_t) \, dt - \rho_{01}P_t \, dt + \rho_{10}(1 - P_t) \, dt,$$
$$d\widetilde{\mathcal{R}}_t = \widetilde{\mathcal{R}}_t (1 - \widetilde{\mathcal{R}}_t)\big(\beta_0 P_t + \beta_1(1 - P_t)\big) dt,$$

that is, the ODE system

$$p_t' = p_t(1 - p_t)(1 - \tilde{r}_t)(\beta_0 - \beta_1) - \rho_{01} p_t + \rho_{10}(1 - p_t),$$
$$\tilde{r}_t' = \tilde{r}_t(1 - \tilde{r}_t)\big(\beta_0 p_t + \beta_1(1 - p_t)\big).$$

With initial value $(p_0, \tilde{r}_0) \in (0, 1)^2$, the equilibrium solution as t tends to infinity, is $p_\infty = \rho_{10}/(\rho_{10} + \rho_{01})$, $\tilde{r}_\infty = 1$. We see that species selection, acting on the net diversification rates β_0 and β_1, is effective only during the growth phase but vanishes as the number of species approaches the carrying capacity, i.e., when $\tilde{r}_t$ approaches 1. In ecology, this is said to be a model with only r-selected traits but no K-selected traits (where r is the growth rate and K is the carrying capacity), that is, the diversification of the two traits is regulated in the same way as $\tilde{r}_t$ approaches 1 (Pianka 1970). This scenario would correspond to a trait allowing adaptive radiation with a rapid diversification of species possessing this trait when many niches are still available, until an almost neutral dynamics when carrying capacity is reached.

4.2 Trait- and diversity-dependent macroevolutionary diversification rates

The density dependent mechanism is again to reduce the growth of the species family with increasing diversity, but now the efficiency of the nonlinear influence is tuned for each trait, with the resulting trait fitness

$$f_i^{c_i}(r) = \beta_i(1 - r/c_i) - \tau_i/r, \quad i = 0, 1.$$

By adapting the derivation leading to Eq. (8)–(9) to this case and putting $\widetilde{\mathcal{R}} = \mathcal{R}/c$ as before, then in the limit $n \to \infty$,

$$d\mathcal{P}_t = \mathcal{P}_t(1 - \mathcal{P}_t)\Big(f_0^{c_0}(c\widetilde{\mathcal{R}}_t) - f_1^{c_1}(c\widetilde{\mathcal{R}}_t)\Big) dt - \rho_{01}\mathcal{P}_t\, dt + \rho_{10}(1 - \mathcal{P}_t)\, dt$$

$$+ \sqrt{\mathcal{P}_t(1 - \mathcal{P}_t)\psi(\mathcal{P}_t)\frac{1}{c\widetilde{\mathcal{R}}_t}}\, dB_t^-,$$

$$d\widetilde{\mathcal{R}}_t = \widetilde{\mathcal{R}}_t\Big\{\beta_0\mathcal{P}_t\Big(1 - \frac{c}{c_0}\widetilde{\mathcal{R}}_t\Big) + \beta_1(1 - \mathcal{P}_t)\Big(1 - \frac{c}{c_1}\widetilde{\mathcal{R}}_t\Big)\Big\} dt$$

$$+ \sqrt{\frac{\widetilde{\mathcal{R}}_t}{c}\psi}(1 - \mathcal{P}_t)\, dB_t^+.$$

A special case of interest is obtained by taking $\beta_0/c_0 = \beta_1/c_1 = \alpha/c$, $\tau_0 = \tau_1 = \tau$, using a new parameter α. For simplicity, let us also take $\rho_{01} = \rho_{10} = 0$. Then

$$d\mathcal{P}_t = (\beta_0 - \beta_1)\mathcal{P}_t(1 - \mathcal{P}_t)\, dt + \sqrt{\mathcal{P}_t(1 - \mathcal{P}_t)\frac{\tau}{c\widetilde{\mathcal{R}}_t}}\, dB_t^-,$$

$$d\widetilde{\mathcal{R}}_t = \widetilde{\mathcal{R}}_t\Big\{\beta_0\mathcal{P}_t + \beta_1(1 - \mathcal{P}_t) - \alpha\mathcal{R}_t\Big\} dt + \sqrt{\tau\widetilde{\mathcal{R}}_t/c}\, dB_t^+.$$

In this example we notice that the growth of species richness is diversity- and density-dependent while the selection term is now density-independent.

Let us now come back to the general model studied in this section and suppose that as $c \to \infty$ then both of c_0 and c_1 also tend to infinity, such that $c/c_i \to \alpha_i$, $i = 0, 1$, for some positive constants α_0 and α_1. Then formally

$$p_t' = p_t(1 - p_t)\big(\beta_0(1 - \alpha_0\tilde{r}_t) - \beta_1(1 - \alpha_1\tilde{r}_t)\big) - \rho_{01}p_t + \rho_{10}(1 - p_t),$$
$$\tilde{r}_t' = \tilde{r}_t\big(\beta_0 p_t(1 - \alpha_0\tilde{r}_t) + \beta_1(1 - p_t)(1 - \alpha_1\tilde{r}_t)\big). \tag{16}$$

The case $\beta_0 = \beta_1 = \beta > 0$ gives the system of ODEs

$$p_t' = \beta(\alpha_1 - \alpha_0)\, p_t(1 - p_t)\tilde{r}_t - \rho_{01}p_t + \rho_{10}(1 - p_t),$$
$$\tilde{r}_t' = \beta\tilde{r}_t\big(1 - \tilde{r}_t(\alpha_0 p_t + \alpha_1(1 - p_t))\big).$$

The equation for p above shows that selection, positive or negative, strengthens with increasing species richness. In ecology, this would be equivalent to K-selection,

whereby the trait less sensitive to competition is favored. For example, if $c_0 > c_1$ so that trait 0 is less constrained than trait 1 then the selective rate of growth of trait 0 is positive, $\beta(\alpha_1 - \alpha_0) > 0$. For both "r-like" and "K-like" cases, selection is diversity-dependent. Returning to (16) and the special case $\alpha_0 \beta_0 = \alpha_1 \beta_1 = \alpha$, although total species richness is regulated, species selection becomes diversity-independent and the selection coefficient reduces to $\beta_0 - \beta_1$. This would correspond to a scenario where the total diversity is regulated from mechanisms unrelated to the focal traits.

In macroevolution, several traits have been supposed to be evolutionary dead-ends, such as asexuality (Maynard-Smith 1978) or self-fertilization (Igic and Busch 2013). The classical view is that negative diversification is associated with such traits but they are continuously reintroduced through asymmetrical (or even unidirectional) shifts from traits associated with positive diversification (Goldberg et al. 2010). The above formulation suggests a more elaborate scenario whereby the dead-end trait has not necessarily a basal negative diversification rate but is more sensitive to diversity dependence. Such traits could lead to initial diversification in initially species-poor environments (i.e. much lower than the species carrying capacity) but diversification would then decrease and eventually be negative through time, as the number of species would increase. This is in agreement with recent empirical observations in the plant genus *Capsella* that selfing species are more sensitive to competition than outcrossing ones (Petrone Mendoza et al. 2018; Yang et al. 2018), and a plausible explanation for the hypothesis that selfing and asexual lineages "senesce" in diversification rates (Ho and Agrawal 2017).

4.3 Fixation and extinction of traits

To study trait fixation probabilities using a similar approach as in Parsons et al. (2008) we let $\rho_{01} = \rho_{10} = 0$ in Eq. (15), which means that species are unable to change traits. The initial composition of traits in the scaled species family at time 0 is a fraction x of trait 0 and the remaining fraction $1 - x$ consisting of trait 1 species, $0 < x < 1$. As in Parsons et al. (2008), we strive to analyze the fate of the trait 0 element over time as a function of x. For this aim we must restrict further our considerations to the density dependent case with symmetric net growth rates, $\beta_0 = \beta_1 = \beta$, and a single carrying capacity c, the same for both traits. Then

$$d\mathcal{P}_t = -\mathcal{P}_t(1 - \mathcal{P}_t)(\tau_0 - \tau_1)\frac{1}{\mathcal{R}_t}\,dt + \sqrt{\mathcal{P}_t(1 - \mathcal{P}_t)\psi(\mathcal{P}_t)\frac{1}{\mathcal{R}_t}}\,dB_t^-,$$

$$d\mathcal{R}_t = \beta\mathcal{R}_t\Big(1 - \frac{\mathcal{R}_t}{c}\Big)\,dt + \sqrt{\mathcal{R}_t\psi(1 - \mathcal{P}_t)}\,dB_t^+, \qquad (17)$$

which is a logistic version of the model listed as special case (a) in Sect. 3.6. The crucial aspect of the resulting SDE for $\mathcal{P}$, which is required for our method of proof to work, is that all dependence on $\mathcal{R}$ is mediated through a single function $g(\mathcal{R})$, in this case $g(r) = 1/r$, which appears as a multiple in both the diffusion drift term and the diffusion variance term. This is the reason why the closely related special case

$\beta_0 = \beta_1, c_0 \neq c_1$ is not covered by our method, nor is $\beta_0 \neq \beta_1, c_0 = c_1$ or $\tau_0 = \tau_1$, $\beta_0 \neq \beta_1$.

In the present situation, however, general properties of logistic branching processes show that the species richness $\mathcal{R}$ goes extinct with probability one as $t \to \infty$. We may then consider the sequence of stopped processes $(\mathcal{P}_{t \wedge \delta_n}, \mathcal{R}_{t \wedge \delta_n})_{t \geq 0}$, where $\delta_n = \inf\{t > 0 : \mathcal{R}_t = 1/n\}$, and have a well-defined solution of (17) with positive richness component on $\{0 \leq t \leq \delta_n\}$ for each fixed n. Yet, the mixture of traits as captured by $(\mathcal{P}_{t \wedge \delta_n})_{t \geq 0}$ along the path to extinction of the species family, as $n \to \infty$, would perhaps be of limited interest, indicating that the fixation probability concept might have limited relevance. On the other hand, the quasi-stationary behavior of $\mathcal{R}$ mentioned previously means that the time to extinction can be very long. This allows us to circumvent species extinction by modeling the species family rather using a process $\mathcal{R}^+$, say, which is $\mathcal{R}$ conditioned on the event $\{\mathcal{R}_t > 0, t > 0\}$ of ultimate nonextinction. In the following we study trait absorption with respect to such non-extinct paths. Let $(\mathcal{P}^+, \mathcal{R}^+)$ denote the solution of (17) conditioned on the event of ultimate nonextinction. The boundaries 0 and 1 are both absorbing for $\mathcal{P}^+$. Let us denote by η_0 the fixation time of trait 1, that is, the random time at which $\mathcal{P}^+$ first hits the lower boundary 0, if ever. Similarly, η_1 is the fixation time of trait 0, the time at which the upper boundary point 1 is first hit. If η_0 is finite, then all species are trait 1 from that time and onward, meaning that the upper limit is never reached, $\eta_1 = \infty$. Similarly, if $\eta_1 < \infty$ all species end up as trait 0, and $\eta_0 = \infty$. The absorption time $\eta = \min(\eta_0, \eta_1)$ is the time of extinction or fixation, whichever occurs first. Next we recover the fixation probability (5) within our setting, and derive a result on the "species frequency spectrum".

Theorem 1 *The fixation probability* $q(x) = \mathbb{P}_x(\eta_1 < \infty)$ *of* $\mathcal{P}^+$ *rendering all species trait 0 eventually, given initial frequency* x *of trait 0, equals*

$$q(x) = \frac{\tau_1 x}{\tau_0(1 - x) + \tau_1 x}.$$

Moreover, for bounded functions f *defined on the unit interval* $[0, 1]$,

$$\mathbb{E}_x\left[\int_0^{\eta^+} \frac{f(\mathcal{P}_s^+)}{\mathcal{R}_s^+} \, ds\right] = \frac{2}{\tau_0(1-x) + \tau_1 x}\left\{x \int_x^1 \frac{f(y)}{y} \, dy + (1-x) \int_0^x \frac{f(y)}{1-y} \, dy\right\}.$$

Proof Let $X = (X_t)_{t \geq 0}$ be a diffusion process, which solves

$$dX_t = -X_t(1 - X_t)(\tau_0 - \tau_1) \, dt + \sqrt{X_t(1 - X_t)\psi(X_t)} \, dB_t^-. \tag{18}$$

The SDE for X is the same as that of $\mathcal{P}^+$ under the enforced condition $\mathcal{R}^+ \equiv 1$. It turns out that the fixation probabilities of $\mathcal{P}^+$ and X respectively, coincide. The reason for this is that the distribution of X can be extracted from that of $\mathcal{P}^+$, via a random time change. Since $\mathcal{R}_t^+ > 0$ for all t, the transform of $\mathcal{R}^+$, given by the function

$$B_t = \int_0^t \frac{1}{\mathcal{R}_s^+} \, ds, \quad t > 0,$$

is a strictly increasing random time change with time change rate $1/\mathcal{R}_t^+$, such that its left-inverse

$$A_t = \inf\{s > 0 : B_s > t\}, \quad t > 0,$$

is continuous, and $A_{B_t} = t$ for all $t \geq 0$. Thus,

$$q(x) = \mathbb{P}_x(\eta_1^+ < \infty) = \mathbb{P}_x(B_{\eta^+} < \infty), \quad B_{\eta^+} = \inf\{t > 0 : \mathcal{P}_{A_t}^+ = 1\}.$$

The desired function $q(x)$ is now obtained as the fixation probability of the time-changed process $(\mathcal{P}_{A_t}^+)_{t \geq 0}$, which turns out to be closely connected to X. Indeed, it is a consequence of the time change result Theorem 8.5.1 in Øksendal (2007), that, for each $t \geq 0$, $\mathcal{P}_{A_t}^+$ and X_t have the same distribution. Let η_1^X denote the fixation time of X. Since

$$\mathbb{P}_x(\eta_1^X > t) = \mathbb{P}_x(X_t = 1) = \mathbb{P}(\mathcal{P}_{A_t}^+ = 1) = \mathbb{P}_x(B_{\eta^+} > t),$$

it follows that η_1^X and $B_{\eta_1^+}$ have the same distribution, and so $q(x) = \mathbb{P}(\eta_1^X < \infty)$. Similarly, η^X and B_{η^+} have the same distribution.

The SDE for X in Eq. (18) is

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t^-, \quad b(x) = (\tau_1 - \tau_0)x(1-x), \quad \sigma^2(x) = x(1-x)\psi(x),$$

with $\psi(x) = \tau_0(1-x) + \tau_1 x$ introduced in Eq. (14), and where $X_0 = x, 0 < x < 1$, and the boundary points $\{0, 1\}$ are absorbing. The scale function $S(x)$ and speed function $m(x)$ associated with $(X_t)_{t \geq 0}$, $X_0 = x$, are defined by

$$S(x) = \int_0^x s(y)\,dy, \quad m(x) = \frac{1}{\sigma^2(x)s(x)}, \quad 0 < x < 1,$$

where

$$s(x) = \exp\left\{-\int_0^x \frac{2b(y)}{\sigma^2(y)}\,dy\right\},$$

Here,

$$S(x) = \frac{\tau_0 x}{\tau_0(1-x) + \tau_1 x}, \quad m(x) = \frac{\tau_0(1-x) + \tau_1 x}{\tau_0^2 x(1-x)}.$$

By using Feller's boundary classification and the theory of one-dimensional diffusion processes (Karlin and Taylor 1981; Etheridge 2011), it follows that the boundary points 0 and 1 are both accessible from the interior state space as exit boundaries. The

extinction probability is determined by the scale function (using the normalization $S(0) = 0$), as

$$q(x) = \frac{S(x)}{S(1)} = \frac{\tau_1 x}{\tau_0(1-x) + \tau_1 x}.$$

Moreover, for bounded functions f,

$$\mathbb{E}_x \left[\int_0^{\eta^X} f(X_s) \, ds \right] = 2q(x) \int_x^1 (S(1) - S(y)) m(y) f(y) \, dy$$

$$+ 2(1 - q(x)) \int_0^x S(y) m(y) f(y) \, dy.$$

It is straightforward to check that the right hand side equals

$$\frac{2}{\tau_0(1-x) + \tau_1 x} \left\{ x \int_x^1 \frac{f(y)}{y} \, dy + (1-x) \int_0^x \frac{f(y)}{1-y} \, dy \right\}.$$

To complete the proof, we first use once again that $\mathcal{P}_{A_t}^+$ and X_t have the same distribution for each fixed t, and then make the change-of-variables $s = B_r$, to obtain

$$\mathbb{E}_x \left[\int_0^{\eta^X} f(X_s) \, ds \right] = \mathbb{E}_x \left[\int_0^\infty 1_{\{0 < X_s < 1\}} f(X_s) \, ds \right]$$

$$= \mathbb{E}_x \left[\int_0^\infty 1_{\{0 < \mathcal{P}_{A_s}^+ < 1\}} f(\mathcal{P}_{A_s}^+) \, ds \right]$$

$$= \mathbb{E}_x \left[\int_0^{B_{\eta^+}} f(\mathcal{P}_{A_s}^+) \, ds \right] = \mathbb{E}_x \left[\int_0^{\eta^+} \frac{f(\mathcal{P}_r^+)}{\mathcal{R}_r^+} \, dr \right].$$

$\square$

Application of Theorem 1: fixation and extinction of a rare trait

Suppose we have a family in which all species carry a single trait, and we introduce a species with a second trait of the same net growth rate into the population. What are the chances of the new trait getting fixed or lost? What are the implications if the new trait causes a shift in turnover rate? For example, all else being equal, this would correspond to a shift in life span, lifespan being inversely proportional with respect to turnover rate. As another example, recently, polyploidy (the doubling of the genome) has been shown to have no effect on diversification ($\beta_0 = \beta_1$) but to increase turnover rate ($\tau_0 > \tau_1$) (Zenil-Ferguson et al. 2019).

To address these questions we observe that if a small fraction x of trait-0 species are inserted into a population otherwise consisting entirely of trait-1 species, then as $x \to 0$,

$$q(x) = \frac{\tau_1 x}{\tau_0(1-x) + \tau_0 x} \sim \frac{\tau_1}{\tau_0} x.$$

In addition, by an application of the second statement in Theorem 1 with $f(y) = 1$, $y \in [0, 1]$, the expected absorption time weighted by the inverse species richness process, satisfies

$$\mathbb{E}_x\left[\int_0^{\eta^+} \frac{1}{\mathcal{R}_s^+} ds\right] = \frac{-2x\ln(x) - 2(1-x)\ln(1-x)}{\tau_0(1-x) + \tau_1 x} \sim \frac{-2x\ln(x) + 2x}{\tau_0},$$

as $x \to 0$. Keeping the turnover rate τ_1 of the original trait fixed, we see that the probability of fixation of the new trait is inversely proportional to the turnover rate τ_0 of the new trait. Thus, traits which cause a burst in both speciation and extinction are suppressed whereas traits with low turnover rates are favored, with respect to possible fixation. Similarly, the expected weighted absorption time decreases with increasing turnover rate τ_0, consistent with shorter time to extinction.

An alternative interpretation of the result in Theorem 1 suggests a notion of trait frequency spectrum in analogy to the allele frequency spectrum in population genetics. For this we assume, with reference to the space and time scaling parameter n, that in each of the n time steps forming one evolutionary time unit, trait injections occur at rate $\theta > 0$. When one of these events happens, a fraction $1/n$ of the family gets a new trait of turnover rate τ_0. The remaining fraction has turnover rate τ_1. We think of the successive injection events representing each time a new trait unrelated to previous ones. Each new trait traces out its own path $\mathcal{P}^+$ but all relate to the same $\mathcal{R}^+$. Most of the new traits quickly go extinct. A few might survive for a while and even get fixed, eventually. The scaled trait 0 fixation probability per time unit is the large n limit

$$nq(1/n) \to \frac{\tau_1}{\tau_0}.$$

A possible interpretation is that with larger turnover rate τ_0 of an "invading" trait, the smaller is the fixation rate τ_1/τ_0, and hence, the more efficient is the existing species family in purging such an intruder. For non-negative bounded functions f on the interval $[0, 1]$ with $f(0) = 0$, which satisfies the integrability condition $\int_0^1 f(y) y^{-1} dy < \infty$, we find

$$\lim_{n\to\infty} \theta n \mathbb{E}_{1/n}\left[\int_0^{\eta^+} \frac{f(\mathcal{P}_r^+)}{\mathcal{R}_r^+} dr\right] = \frac{\theta \tau_1}{\tau_0} \int_0^1 f(y) \frac{2}{\tau_1 y} dy = \int_0^1 f(y) \frac{2\theta}{\tau_0 y} dy.$$

In particular,

$$\lim_{n\to\infty} \theta n \mathbb{E}_{1/n}\left[\int_0^{\eta^+} \frac{\mathcal{P}_r^+}{\mathcal{R}_r^+} dr\right] = \frac{2\theta}{\tau_0}.$$

The weight function $2\theta \tau_0^{-1}/y$ which arises in the above scheme of limiting expected values, plays a similar role as the stationary allele frequency spectrum in population

genetics. While such a frequency weight function is not integrable over $y \in (0, 1)$, and hence does not allow a probability density interpretation, it does define an intensity measure for a well-defined Poisson random measure on $(0, 1)$. For the Poisson random field approach in population genetics, see Sawyer and Hartl (1992) and e.g. Kaj and Mugal (2016), Section 2.3. Intuitively, for each $y \in (0, 1)$, $(2\theta/\tau_0)y^{-1}$ is in this sense the stationary intensity that $(\mathcal{P}_t)_{t \geq 0}$ occupies the frequency band near y, as measured by the size of $(\mathcal{P}_t^+/\mathcal{R}_t^+)_{t \geq 0}$ within the time frame during which the trait remains in the family.

The case $\tau_0 = \tau_1$ is the fully neutral case with fixation probability $q(x) = x$ and trait spectrum intensity $2\theta/y$. The case $\tau_0 \neq \tau_1$ represents a form of trait selection, which only affects the relative magnitude of trait frequencies present in the species family, but not the shape of the trait frequency spectrum.

5 The super-process representation of the limit process

In this final section, we return to the two-type branching process X with rates (1) running on the scale of species generations. The aim is to provide a sketch of the proof of the limit result Proposition 1, shifting to the evolutionary time scale.

As an alternative and from a wider perspective, we may consider X as a measure-branching process with spatial motion on the trait space $E = \{0, 1\}$ and spatially dependent binary branching. Each lineage in the branching tree changes its current species trait according to independent copies of the space motion $J = (J_u)$, which is the two-state Markov jump process with jump rates δ_{01} and δ_{10} for transitions from 0 to 1 and vice versa. In addition to the on/off motion caused by J, the number of species of each trait develop independently as linear binary branching processes with generating functions

$$F(z; j) = \frac{\lambda_j z^2 + \mu_j}{\lambda_j + \mu_j} = \frac{d_j + \tau_j}{2\tau_j} z + \frac{\tau_j - d_j}{2\tau_j}, \quad j = 0, 1.$$

The branching offspring distribution for daughter species of trait j has mean d_j/τ_j, and variance $(\tau_j - d_j)/\tau_j$. Now let us put $\langle X_u, f \rangle = f(0)K_u + f(1)L_u$, $u \geq 0$, where $f = (f(0), f(1))$ is a function (row vector) on E. We write $\mathbb{E}_j$ for the expected value conditional on $J_0 = j$, that is $X_0 = (1, 0)$ when $j = 0$ and $X_0 = (0, 1)$ when $j = 1$. Then X is the measure-branching process for which the function v_u defined by

$$v_u(j) = \mathbb{E}_j[e^{\langle X_u, f \rangle} - 1],$$

is the unique solution of the integral equation

$$v_u(j) = \mathbb{E}_j[e^{f(J_u)} - 1] + \mathbb{E}_j \int_0^u \left(F(1 + v_{u-r}(J_r); J_r) - 1 - v_{u-r}(J_r) \right) \tau_{J_r} \, dr,$$

(19)

for $j \in E$, cf. Li (2011), Ch. 4.1. It is well-known that suitably scaled measure-valued branching processes converge in the limit of many particles, long times, and small masses to an associated superprocess, in this case a super-on/off process. More exactly, the superprocess is the weak limit as n independent copies $X^{(k,n)}$, $k = 1, \ldots, n$, of X with spatial motion $J^{(n)}$ and branching mechanism $F^{(n)}(z, ;)$ scaled according to (6), are averaged over the time span $u = nt$. This, however, is the species model under evolutionary time scaling, and hence the superprocess will be the limit $\mathcal{X}$ of $\mathcal{X}_t^{(n)} = n^{-1} X_{nt}^{(n)}$, $t \geq 0$, in the sense of weak convergence in path space as $n \to \infty$. To study the limit process we put

$$V_t^{(n)}[f](j) = n\mathbb{E}_j[e^{\langle X_{nt}^{(n)}, f/n \rangle} - 1],$$

so that $V_t^{(1)}[f](j) = v_t(j)$. Then, for large n,

$$\ln \mathbb{E}_j[e^{\langle \mathcal{X}_t^{(n)}, f \rangle}] = n \ln \mathbb{E}_j\left[\exp\{\langle X_{nt}^{(n)}, f/n \rangle\} \right]$$

$$= n \ln \left(1 + \frac{n\mathbb{E}_j[e^{\langle X_{nt}^{(n)}, f/n \rangle} - 1]}{n} \right) \sim V_t^{(n)}[f](j).$$

By using once more the scaling assumption (6), we have

$$P(J_{n(t+h)}^{(n)} = 1 | J_{nt}^{(n)} = 0) = \delta_{01}^{(n)} nh + \frac{o(nh)}{n} = \rho_{01} h + o(h),$$

and the analogous relation for converse transitions from 1 to 0. This means that the limit process of $J_{nt}^{(n)}$, $t \geq 0$, is again an on/off process on E but now with jump rates ρ_{01}, ρ_{10}. For simplicity we retain the notation J in the limit. Also, a calculation shows,

$$n^2\big(F^{(n)}(1 + v/n; j) - 1 - v/n\big)\tau_j^{(n)} = \frac{n^2}{2}\Big(\tau_j^{(n)}(v/n)^2 + d_j^{(n)}(2(v/n) + (v/n)^2)\Big)$$

$$\sim \beta_j v + \tau_j \frac{v^2}{2}.$$

As we combine these observations with Eq. (19), it follows that the function $V^{(n)}[f](j)$ satisfies

$$V_t^{(n)}[f](j) = \mathbb{E}_j[n(e^{f(J_t)/n} - 1)]$$

$$+ \mathbb{E}_j \int_0^t \big(\beta_{J_r} V_{t-r}^{(n)}[f](J_r) + \tau_{J_r} V_{t-r}^{(n)}[f](J_r)^2/2\big) dr + H_t^{(n)},$$

where the remainder term $H^{(n)}$ can be controlled and shown to vanish in the limit $n \to \infty$, see Ch. 4.2 in Li (2011). With reference to the theory of superprocesses we can now conclude that $\mathcal{X}^{(n)}$ possesses a weak limit process $\mathcal{X}$, such that $V_t[f](j) = \ln \mathbb{E}_j[e^{\langle \mathcal{X}_t, f \rangle}]$ is the unique solution of the integral equation

$$V_t[f](j) = \mathbb{E}_j[f(J_t)] + \int_0^t \mathbb{E}_j\big[\beta_{J_r} V_{t-r}[f](J_r) + \tau_{J_r} V_{t-r}[f](J_r)^2/2\big] dr, \quad j \in E,$$

cf. Prop. 4.5 in Li (2011). The associated stochastic equation for the superprocess is

$$\langle \mathcal{X}_t, f \rangle = \langle \mathcal{X}_0, f \rangle + \int_0^t \langle \mathcal{X}_s, \mathcal{G}f \rangle \, ds + \int_0^t \langle \mathcal{X}_s, \beta f \rangle \, ds + M_t(f). \qquad (20)$$

Here, $\mathcal{G}f$ is the infinitesimal generator of J, defined by

$$\mathcal{G}f(0) = (f(1) - f(0))\rho_{01}, \quad \mathcal{G}f(1) = (f(0) - f(1))\rho_{10},$$

and $M(f)$ is a continuous martingale with representation

$$M_t(f) = f_0 \int_0^t \sqrt{\tau_0 \mathcal{K}_s} \, dB_s^0 + f_1 \int_0^t \sqrt{\tau_1 \mathcal{L}_s} \, dB_s^1,$$

using a 2-dimensional standard Brownian motion (B^0, B^1) with independent components. Hence the quadratic variation of $M(f)$ is

$$\langle\!\langle M(f), M(f) \rangle\!\rangle_t = \int_0^t \langle \mathcal{X}_s, \tau f^2 \rangle \, ds.$$

By observing that $\langle \mathcal{X}, f \rangle_t$ is the scalar product $f \cdot \mathcal{X}_t$ of the vectors f and $\mathcal{X}_t$ and that

$$\langle \mathcal{X}_s, \mathcal{G}f \rangle + \langle \mathcal{X}_s, \beta f \rangle = f \cdot A \mathcal{X}_s,$$

it follows that the equation for $\mathcal{X}$ in Proposition 1 is an alternative representation of Eq. (20), cf. Li (2011), Ch. 7.5.

Conditional on ultimate survival, the results for $(\mathcal{P}, \mathcal{R})$ now follow from Ito's formula with

$$B_t^+ = \int_0^t \sqrt{\frac{\tau_0 P_s}{\tau_0 P_s + \tau_1(1 - P_s)}} \, dB_s^0 + \int_0^t \sqrt{\frac{\tau_1(1 - P_s)}{\tau_0 P_s + \tau_1(1 - P_s)}} \, dB_s^1,$$

and

$$B_t^- = \int_0^t \sqrt{\frac{\tau_0(1 - P_s)}{\tau_0(1 - P_s) + \tau_1 P_s}} \, dB_s^0 - \int_0^t \sqrt{\frac{\tau_1 P_s}{\tau_0(1 - P_s) + \tau_1 P_s}} \, dB_s^1.$$

6 Conclusions

In this work, we presented a stochastic modeling framework for a binary trait-dependent species family. The model provides insights on the joint evolution of trait frequencies and species diversity. Guided by the long-term successful use of stochastic

techniques in population genetics, we applied various probabilistic methods to study the effects of diversity-dependence and related properties in the species family, hence developing a phylogenetic methodology. At the core of the study is the interpretation that phylogenetic trait frequency and population genetics allele frequency are closely related from the view point of stochastic models. We showed that evolutionary time rescaling is a powerful tool to reveal relevant analogies and distinctions, and to provide a bridge between two seemingly distant application areas.

We applied methods, which rely on the analysis of trajectories of solutions, to stochastic differential equations of Wright–Fisher type with diversity dependent selection and genetic drift coefficients. Our main structural result, Proposition 1, derived from an embedding argument into an abstract two-type superprocess, provides the precise dynamics of trait frequency jointly with species richness. In parallel, we also discussed the corresponding Markov generator dynamics, referred to as the Gillespie approach. In doing so, we were able to resolve a paradoxical conclusion in Gillespie (1974), dealing with the effective population size. To put these abstract results into context, we discussed a number of special cases with partly symmetric parameter settings leading up to a fully symmetric, or neutral, case. To help clarify the role of parameters, we identified trait success (or trait fitness) functions, which are further utilized for the application of the model to diversity-dependent interaction in terms of carrying capacity. A particular logistic version of the diversification-dependent model was the subject of an in-depth study of fixation and extinction of traits. The main result, Theorem 1, provides the fixation probability as function of the initial frequency of a newly injected trait, as well as a type of trait frequency spectrum, which is analogous to allele frequency spectrum in population genetics.

References

Abu Awad D, Coron C (2018) Effects of demographic stochasticity and life-history strategies on times and probabilities to fixation. Heredity 121:374–386

Benton MJ, Emerson BC (2007) How did life become so diverse? The dynamics of diversification according to the fossil record and molecular phylogenetics. Paleontology 50:23–40

Cattiaux P, Collet P, Lambert A, Martinez S, Méléard S, San Martin J (2009) Quasi-stationary distributions and diffusion models in population genetics. Ann Prob 37:1926–1969

Chevin ML (2016) Species selection and random drift in macroevolution. Evolution 70:513–525

Etheridge A (2000) An introduction to superprocesses, vol 20. University lecture series. American Mathematical Society, Providence

Etheridge A (2011) Some mathematical models from population genetics, vol 2012. Lecture notes in mathematics. Springer, Berlin

Etienne RS, Haegeman B (2012) A conceptual and statistical framework for adaptive radiations with a key role for diversity-dependence. Am Nat 180:E75–E89

Etienne RS, Haegeman B, Stadler T, Aze T, Pearson PN, Purvis A, Phillimore AB (2012) Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc R Soc Lond B Biol Sci 297:1300–1309

Fournier N, Méléard S (2004) A microscopic probabilistic description of a locally regulated population and macroscopic approximation. Ann Appl Probab 14:1880–1919

Gamisch A, Comes HP (2019) Clade-age-dependent diversification under high species turnover shapes species richness disparities among tropical rainforest lineages of Bulbophyllum (orchidaceae). BMC Evol Biol 19:93

Gillespie JH (1974) Natural selection for within-generation variance in offspring number. Genetics 76:601–606

Gillespie JH (1977) Natural selection for variances in offspring numbers: a new evolutionary principle. Am Nat 111:1010–1014

Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igic B (2010) Species selection maintains self-incompatibility. Science 330:493–495

Haccou P, Jagers P, Vatutin VA (2005) Branching processes: variation, growth, and extinction of populations. Cambridge University Press, Cambridge

Han TS, Zheng QJ, Onstein RE, Rojas-Andres BM, Hauenschild F, Muellner-Riehl AN, Xing YW (2020) Polyploidy promotes species diversification of Allium through ecological shifts. New Phytol 225:571–583

Ho EKH, Agrawal AF (2017) Aging asexual lineages and the evolutionary maintenance of sex. Evolution 71:1865–1875

Igic B, Busch JW (2013) Is self-fertilization an evolutionary dead end? New Phytol 198:386–397

Kaj I, Krone SM (2003) The coalescent process in a population with stochastically varying size. J Appl Prob 40:33–48

Kaj I, Mugal CF (2016) The non-equilibrium allele frequency spectrum in a Poisson random field framework. Theor Pop Biol 111:51–64

Kaj I, Tahir D (2019) Stochastic equations and limit results for some two-type branching models. Stat Probab Lett 150:35–46

Karlin S, Taylor HM (1981) A second course in stochastic processes. Academic Press, New York

Lambert A (2005) The branching process with logistic growth. Ann Appl Probab 15:1506–1535

Lambert A (2006) Probability of fixation under weak selection: a branching process unifying approach. Theor Popul Biol 69:419–441

Li Z (2011) Measure-valued branching Markov processes. Springer, Berlin

Lin YT, Kim H, Doering CR (2012) Features of fast living: on the weak selection for longevity in degenerate birth-death processes. J Stat Phys 148:647–663

Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. Syst Biol 56:701–710

Mallet J (2012) The struggle for existence: how the notion of carrying capacity, K, obscures the links between demography, Darwinian evolution, and speciation. Evol Ecol Res 14:627–665

Marshall CR, Quental TB (2016) The uncertain role of diversity dependence in species diversification and the need to incorporate time-varying carrying capacities. Phil Trans R Soc B 371:20150217

Maynard-Smith J (1978) The evolution of sex. Cambridge University Press, Cambridge

Morlon H (2014) Phylogenetic approaches for studying diversification. Ecol Lett 17:508–525

Nakov T, Beaulieu JM, Alverson AJ (2019) Diatoms diversify and turn over faster in freshwater than marine environments. Evolution 73:2497–2511

Øksendal B (2007) Stochastic differential equations, an introduction with applications. Springer, Berlin

Parsons TL, Quince C (2007a) Fixation in haploid populations exhibiting density dependence I: the non-neutral case. Theor Popul Biol 72:121–135

Parsons TL, Quince C (2007b) Fixation in haploid populations exhibiting density dependence II: the quasi-neutral case. Theor Popul Biol 72:468–479

Parsons TL, Quince C, Plotkin JB (2008) Absorption and fixation times for neutral and quasi-neutral populations with density dependence. Theor Popul Biol 74:302–310

Parsons TL, Quince C, Plotkin JB (2010) Some consequences of demographic stochasticity in population genetics. Genetics 185:1345–1354

Petrone Mendoza S, Lascoux M, Glémin S (2018) Competitive ability of Capsella species with different mating systems and ploidy levels. Ann Bot 121:1257–1264

Pianka ER (1970) On r- and K- selection. Am Nat 104:592–597

Quental TB, Marshall CR (2010) Diversity dynamics: molecular phylogenies need the fossil record. Trends Ecol Evol 25:434–441

Rabosky DL (2013) Diversity-dependence, ecological speciation, and the role of competition in macroevolution. Annu Rev Ecol Evol Syst 44:481–502

Rabosky DL, Lovette IJ (2008) Density-dependent diversification in North American wood warblers. Proc R Soc B 275:2363–2371

Sawyer SA, Hartl DL (1992) Population-genetics of polymorphism and divergence. Genetics 132:1161–1176

Sjödin P, Kaj I, Krone S, Lascoux M, Nordborg M (2005) On the meaning and existence of an effective population size. Genetics 169:1061–1070

Tahir D, Glémin S, Lascoux M, Kaj I (2019) Modeling a trait-dependent diversification process coupled with molecular evolution on a random species tree. J Theor Biol 461:189–203

Vellend M (2010) Conceptual synthesis in community ecology. Q Rev Biol 85:183–206

Yang X, Lascoux M, Glémin S (2018) Variation in competitive ability with mating system, ploidy and range expansion in four Capsella species. bioRxiv 214866, ver 5 peer-reviewed and recommended by PCI Evolutionary Biology

Zenil-Ferguson R, Burleigh JG, Freyman W, Igic B, Mayrose I, Goldberg EE (2019) Interaction among ploidy, breeding system and lineage diversification. New Phytol 224:1252–1265

Affiliations

Ingemar Kaj[1] · **Sylvain Glémin[2]** · **Daniah Tahir[1]** · **Martin Lascoux[3]**

Sylvain Glémin
sylvain.glemin@univ-rennes1.fr

Daniah Tahir
daniahtahir@gmail.com

Martin Lascoux
martin.lascoux@ebc.uu.se

[1]   Department of Mathematics, Uppsala University, Box 480, Uppsala 751 06, Sweden

[2]   CNRS, Université Rennes 1, ECOBIO (écosystèmes, biodiversité, évolution) – UMR 6553, Avenue du général Leclerc, 35042 Rennes, France

[3]   Program in Plant Ecology and Evolution, Department of Ecology and Genetics, EBC, Uppsala University, Norbyvägen 18D, Uppsala 752 36, Sweden