Protracted speciation revitalizes the neutral theory of biodiversity

February 20, 2010

James Rosindell, Stephen J. Cornell, Stephen P. Hubbell, Rampal S. Etienne

1. James Rosindell, (Corresponding author) Institute of Integrative and Comparative Biology, University of Leeds, Leeds, United Kingdom LS2 9JT, James@Rosindell.org , +44 7968 278 436

2. Stephen J. Cornell, Institute of Integrative and Comparative Biology, University of Leeds, Leeds, United Kingdom LS2 9JT. S.J.Cornell@leeds.ac.uk, +44 113 343 2899

3. Stephen P. Hubbell, Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 900954, USA. shubbell@eeb.ucla.edu , +1 310 206 8165

4. Rampal S. Etienne, Community and Conservation Ecology Group, Centre for Ecological and Evolutionary Studies, University of Groningen, BOX 14, 9750 AA Haren, The Netherlands. R.S.Etienne@rug.nl , +31 50 363 2230

Running head: Protracted speciation and neutral theory

Keywords: neutral theory, neutral model, speciation, species longevity, species abundance, log-series, log-normal, incipient species

Type of article: Letter

Abstract

Understanding the maintenance and origin of biodiversity is a formidable task, yet many ubiquitous ecological patterns are predicted by a surprisingly simple and widely studied neutral model that ignores functional differences between species. However, this model assumes that new species arise instantaneously as singletons and consequently makes unrealistic predictions about species lifetimes, speciation rates and number of rare species. Here we resolve these anomalies — without compromising any of the original model's existing achievements and retaining computational and analytical tractability — by modeling speciation as a gradual, protracted, process rather than an instantaneous event. Our model also makes new predictions about the diversity of 'incipient' species and rare species in the metacommunity. We show that it is both necessary and straightforward to incorporate protracted speciation in future studies of neutral models, and argue that non-neutral models should also model speciation as a gradual process rather than an instantaneous one.

Introduction

Hubbell's (2001) neutral model has been shown to give a surprisingly good fit to ecological data such as species abundance distributions (SADs) (Hubbell, 2001; Etienne, 2005) and species area curves (Rosindell & Cornell, 2007, 2009) for many systems ranging from tropical forests to river dwelling organisms and microbial communities (Muneepeerakul *et al.*, 2008; Volkov *et al.*, 2007). However, several other predictions of the model are highly unrealistic (Ricklefs, 2003; Nee, 2005; Ricklefs, 2006): (i) new species arise too frequently; (ii) many species have a global abundance of just one or two; and (iii) the average species lifetime is just a few generations. Because a theory's success ultimately depends on its ability to make useful and realistic predictions (Friedman, 1966), these problems are more immediate than the validity of the neutrality assumption itself (Alonso *et al.*, 2006), which may be defended either as a first-order approximation to a more complicated reality (Volkov *et al.*, 2005; Alonso *et al.*, 2006), or as an emergent consequence of community evolution (Holt, 2006; Hubbell, 2006; Scheffer & van Nes, 2006).

Neutral theory's success stems partly from its prediction that the SAD at large 'metacommunity' scales follows Fisher's log-series (Fisher *et al.*, 1943). The log-series can produce good fits to large-scale empirical samples of species and genera (Hubbell, 2001; Hubbell *et al.*, 2008), but it also predicts an extremely large number of very rare species in the complete, un-sampled, metacommunity—a prediction which has found little empirical support (McGill, 2003b,a). These rare and hence extinction-prone species imply rapid turnover, and lead to the aforementioned problems with species lifetimes and speciation rates. Hence, the classic neutral model's prediction of a log-series for the metacommunity SAD enables it to fit some sampled data well but is associated with poor predictions of other patterns. There are many alternative models for SADs (McGill *et al.*, 2007; Sizling *et al.*, 2009; Kurka *et al.*, 2010) that produce fewer rare species than a log-series. Probably the best-known example is the log-normal distribution which predicts fewer rare species and is supported by some empirical studies, for example in corals (Connolly *et al.*, 2005) and in birds Gregory (1994, 2000), but it has no clear mechanistic basis (Williamson & Gaston, 2005). The distributions observed are often left-skewed log-normal distributions and there is some debate regarding whether these are sampling artifacts or phenomena that demand

a biological explanation (Nee *et al.*, 1991; McGill, 2003a).

The classic neutral *local* community model reconciles the log-series and log-normal models to some extent by incorporating dispersal limitation that restricts the number of rare species in the sampled local community (Hubbell, 2001). However, this model only explains log-normal-like SADs in the local community, but not in the metacommunity species pool, and still has the aforementioned problems regarding speciation rates and species longevities.

In neutral theory, a log-series meta-community always results from the assumption that new species arise instantaneously as singletons from a point mutation ('point mutation speciation') (Hubbell, 2001, 2003). Alternative speciation mechanisms have been proposed to resolve the aforementioned difficulties with species lifetimes, speciation rates and rare species (Hubbell, 2001; Hubbell & Lake, 2002; Hubbell, 2003; Allen & Savage, 2007), but point mutation speciation remains the only mode that provides compelling fits to locally sampled species abundance data (Hubbell, 2001; Etienne *et al.*, 2007) (also see Etienne and Haegeman under review).

It is generally thought that the SAD on its own does not contain sufficient information to distinguish between a large number of competing models and this calls for biodiversity models that predict more than just abundance distributions (McGill *et al.*, 2007). Fortunately, neutral theory makes many testable predictions besides species abundances (Etienne, 2007; Rosindell & Cornell, 2009; Jabot & Chave, 2009; O'Dwyer & Green., 2009), but to be a credible theory of community ecology, a speciation mode is called for that resolves the problems with speciation rates, species longevities and rare species without compromising the fit to SAD data. In this paper we propose such a speciation mode which we call 'protracted speciation' and show that it does indeed solve these problems and also makes new predictions. We make empirical comparisons with metacommunity and local community abundance data from tropical forest trees, corals and reef fish and in all cases find support for protracted speciation rather than point mutation speciation. As we will show, protracted speciation leads to a new expression for metacommunity abundance distributions which retains the desirable properties of point mutation speciation and the log-series, but does not suffer from any of the aforementioned problems associated with the classic neutral model.

Protracted speciation

The precise mechanisms behind speciation are complex and hotly debated, but there is certainly general agreement that speciation is not simply a phenomenon lasting for a single generation (Gavrilets, 2004; Coyne & Orr, 2004) as has been assumed in all previous neutral models (Hubbell, 2003; Allen & Savage, 2007; Etienne *et al.*, 2007). Our model incorporates this seemingly trivial but essential fact by no longer interpreting point mutation as an instantaneous event but as the initiation of a drawn-out (protracted) speciation process that only produces a recognisable new species after a transition period of τ generations has elapsed (Schluter & Weir, 2007). We study a Moran model which means that generations overlap. We use the convention that one 'generation' is defined as the average lifetime of an individual as opposed to the alternative arising from a comparison with a Wright-Fisher model (with non overlapping generations), in which case the generation length would be divided by a factor of 2. The transition period τ in protracted speciation implicitly captures the outcome of what in reality are complex, ecological and genetic processes, which given enough time lead to the birth of a new species (Schluter, 2009).

The classic neutral model is a good model for lineage branching (Hubbell, 2003), but some alternative modes of speciation violate this by selecting conspecific individuals at random to form a novel species (Hubbell, 2001, 2003; Allen & Savage, 2007). In contrast, protracted speciation maintains the original model's relationship between lineage and species identity, because the individuals that form a new species are necessarily complete groups of closely related individuals (figure 1). The model keeps track of relatives and the topology of the resulting genealogy is unaffected by speciation. The mode of speciation determines the species identities of individuals on this genealogy and hence influences the phylogeny (figure 1).

Protracted speciation can be analysed in practice by first considering the point mutation mode of speciation, then ignoring speciation events that occurred within the most recent τ generations. This is because any speciation process which started during the last τ generations will not be complete at the present day, and will produce 'incipient' species which for the purposes of analysis are considered as conspecific with their would-be sister species. Species identities

are therefore the same as if the recent speciation-initiation event never took place. We do not ignore speciation processes that finish during this period, because these must have started more than τ generations before the present day and are therefore captured by point mutation. By considering the genealogy, it can be seen that this approach is exact, does not require any further biological or mathematical assumptions and enables the use of powerful analytical techniques.

Protracted speciation is a seemingly subtle change to the existing neutral models (Hubbell, 2001, 2003; Allen & Savage, 2007), but represents a major conceptual advance because it singles out the crucial property of speciation: that it takes time. We will show that many of the major problems with the classic neutral model are solved with protracted speciation. We will argue that ignoring protracted speciation can easily lead to ecological misinterpretations, for example regarding extinction rates and rare species in the meta-community.

Speciation rate

Under protracted speciation, the speciation rate per individual, ν, is smaller than the speciation-initiation rate μ, because a lineage undergoing speciation may drift to extinction before speciation is complete. The probability of successful speciation is simply the probability of a single neutrally drifting lineage surviving for at least τ generations, which is $\frac{1}{1+\tau}$ (Box 1 ; Leigh, 2007) and hence

$$\nu = \frac{\mu}{1+\tau}. \tag{1}$$

Thus, an important prediction of the protracted speciation model is that the speciation rate is smaller than the speciation-initiation rate by a factor of $1+\tau$, which is likely to be large in biologically realistic scenarios. Protracted speciation reduces to point mutation speciation ($\nu = \mu$) in the special case $\tau = 0$, where speciation is instantaneous.

The expected initial abundance of a new species under protracted speciation is $1+\tau$ (see box 1), and the distribution of initial abundances (equation. (8) in box 1) is displayed together with alternative modes in figure 2. One of the greatest strengths of the original neutral model is that

it is based on mechanistic processes: births, deaths etc. Likewise, in the new model the initial abundances of new species are derived as a consequence of births and deaths under protracted speciation. By contrast, other alternative speciation modes (Hubbell, 2003; Allen & Savage, 2007) simply assume a convenient ad-hoc form for the distribution of initial abundances.

Metacommunity dynamics

Under the protracted speciation model, the expected number of species with abundance j in a large metacommunity (where any incipient species is lumped together with its parent species), is (see box 1)

$$\mathbb{E}[S(j)] = \frac{\theta}{j}\left(\left(1 - \frac{\mu}{1+\tau\mu}\right)^j - \left(1 - \frac{1}{1+\tau}\right)^j\right) \tag{2}$$

where θ is Hubbell's fundamental biodiversity number $\theta = \frac{\mu J_M}{1-\mu} \approx \mu J_M$—which now depends on the speciation-initiation rate μ rather than the speciation rate ν. We use the term 'difference-log-series' (DLS) to describe $\mathbb{E}[S(j)]$, because it is the difference between two log-series terms. The DLS is reminiscent of Preston's (1981) 'Diffonential' distribution for species abundances; the main distinction being that Preston used the difference between two exponential functions without the prefactor of $\frac{1}{j}$ that is present in the log-series, and offered no biological interpretation of its parameters.

When $\tau = 0$, the DLS reduces to the standard log-series $\mathbb{E}[S(j)] = \frac{\theta}{j}(1-\mu)^j$ of the original point mutation speciation model. When τ is not zero, the log-series sill arises for sufficiently large j, because the second term in equation 2 decays much more rapidly than the first. The DLS therefore follows a log-series distribution at large abundances, but predicts few rare species like the log-normal. At intermediate abundances a negligible increase in the number of species can be seen, but this is too small to be visible in standard visual representations of the data (figure 3). For large values of τ the distribution can appear qualitatively very similar to a log-normal (figure 3). The DLS can thus be used as a simple and mechanistically justifiable alternative to the log-series and log-normal metacommunity models. Although other neutral models have

achieved something similar (Hubbell, 2001; Etienne & Alonso, 2005; Allen & Savage, 2007), the DLS is the most parsimonious and captures the essence of speciation as a non-instantaneous process.

The terms in the DLS are determined by the speciation-initiation rate μ, and not by the speciation rate ν. This means that the protracted speciation model makes exactly the same predictions (except for the rarest species) as a point mutation speciation model with speciation rate μ. Neutral models have been criticised for requiring unrealistically high speciation rates (Ricklefs, 2003; Rosindell & Cornell, 2009) when fitted to data, but we can now interpret such studies as estimating μ rather than the true speciation rate ν, which can be several orders of magnitude smaller than μ (see equation 1). We will show below that, likewise, it is μ and not ν that controls the SAD in the local community. We stress that the transition time τ is an important component of the speciation process that influences the model's predictions for rare species, but not significantly for common species.

The DLS differs from the log-series at small j (figure 3). The degree to which this can be detected in empirical data will depend on the size of the metacommunity, the value of τ, how many individuals are sampled and the randomness of the sampling process (see appendix B). In the case of tropical forest trees, even the most extensive empirical studies of pooled abundance data only sample one in a million individuals from the metacommunity which can contain as many as 10^{11} individuals (Hubbell *et al.*, 2008). As we show in figure 3 even a very significant effect of protracted speciation cannot be detected in such a sparse sample. Log-series data are thus consistent with protracted speciation; they arise in empirical data as a result of restricted sampling or restricted spatial scale.

Species longevity

Neutral theory's predictions for species longevity have been criticized for two reasons. First, under point mutation speciation the average species longevity L is proportional to $|\log \mu|$ (Ricklefs, 2003). This entails unreasonably short-lived species even for small speciation rates. In our new model, L is given by

$$L = \left(\frac{1+\tau}{1-\mu}\right) \ln\left(\frac{1+\tau\mu}{\mu+\tau\mu}\right) \approx -\tau \ln \tau\mu \tag{3}$$

(see box 1). The factor of τ allows the species lifetimes to be much larger than in the point mutation speciation model.

Second, the classic neutral model predicts that the *average* age of a common species is incredibly long, being of the order of the population size (in generations) (Nee, 2005; Ricklefs, 2006). However, the average is not the most informative statistic, because it does not sufficiently capture the possibility that at least some species may attain high abundance within reasonable time (which is all that is needed). A percentile (e.g. the median) is a much more relevant statistic. In appendix C we show that the median lifetime of a species with high abundance $j \gg 1/\mu$ is approximately $\frac{1}{\mu}\log(j\mu)$. This need not be excessively long because of the weak logarithmic behavior on j, and because the speciation-initiation rate μ is not required to be small under protracted speciation. Although this makes parameters more realistic, we suggest that further refinements of the model are probably necessary to completely resolve the problem of the age of common species (Allen & Savage, 2007; Zhou & Zhang, 2008).

Metacommunity data comparison

In some studies, species abundances are not supportive of the log-series and are instead fitted well by a log-normal (Connolly *et al.*, 2005; Gregory, 2000; McGill, 2003a). This suggests that a protracted speciation model—which can produce log-normal-like metacommunity distributions—would be an improvement on the classic neutral model, which predicts too many rare species. We note, however, that in (Connolly *et al.*, 2005) the convincing log-normal abundance distributions are calculated using 'cover' and 'biomass' rather than individuals (or colonies in the case of coral). The appendix of the same work (Connolly *et al.*, 2005) shows abundance plots using individuals which do not display a clear shortage of rare species.

Here we fit neutral models with both protracted and point mutation speciation to metacommunity data for corals(Dornelas & Connolly, 2008) and for reef fish (Connolly *et al.*, 2005). We

expect the protracted speciation model to improve the predictions for species lifetimes and speciation rates but it will only improve on the fit to abundance data if the data represents a sufficiently large sample from the metacommunity. We sampled from a metacommunity of size J_M, with speciation-initiation rate μ and transition period of length τ. Although there are three parameters, it is already known that for $\tau = 0$ the resulting sampled distribution depends only on the compound parameter $\theta = \frac{\mu J_M}{1-\mu}$. We show in Box 1, that the SAD for a random sample from a finite metacommunity is given by

$$\mathbb{E}[S(j)|J] = \frac{J!}{(J-j)!}\frac{\vartheta}{j}\left[\frac{\Gamma\left(J+\frac{\vartheta\beta}{\theta+\beta}-j\right)}{\Gamma\left(J+\frac{\vartheta\beta}{\theta+\beta}\right)} - \frac{\Gamma\left(J+\beta-j\right)}{\Gamma\left(J+\beta\right)}\right]$$

and thus depends only on two parameters, θ and the rescaled transition time$\tau' = \frac{1}{\beta} \approx \frac{\tau}{J_M}$. Because the exact sampling formula Etienne & Alonso (2005) is computationally demanding, we use the approximate likelihood formula of Alonso & McKane (2004); Volkov *et al.* (2005) to obtain the parameters that best fit the model. This formula is given by

$$\mathcal{L} = \frac{S!}{\prod_i S_i!}\prod_i\left(\frac{E[S(i)|J]}{E[S|J]}\right)^{S_i} \tag{4}$$

(Alonso & McKane, 2004). We used simplex optimisation to find the parameters that maximise this likelihood $\mathcal{L}$.

For both fish and corals, we find that almost identical fits to the species abundance distributions can be obtained for any τ' between 0 and 10^{-5}, but the true global optima are at $\tau' = 2.23 \times 10^{-6}$ for corals and $\tau' = 3.77 \times 10^{-5}$ for fish. In both data sets, when just considering species abundances, no significant difference can be seen between the global optimum fit and the $\tau' = 0$ fit (figure 4). This indicates that, although in theory the $\tau' \neq 0$ fit must be better, in practice the species abundance data, despite an enormous sampling effort, do not have sufficient resolving power to distinguish between these possible values for τ'. The good fit of the difference log-series over this range of values of τ' is the result of strong sampling effects as demonstrated in figure 3. We note that the fit to reef fish data is not so good for the most common species, this

could perhaps be due to the pooling of fish from different guilds, or due to the spatial structure of reef fish communities not being well captured by a spatially implicit model. Because the value of τ and hence the other predictions of the model will depend on metacommunity size, we show predictions for species lifetime (in generations) and speciation rate (per species per generation) as a function of metacommunity size in figure 4. For example, the point mutation case yields clearly unrealistic lifetimes of around 10 coral generations for all values of J_M Ricklefs (2003), but the global optimum of $\tau' = 2.23 \times 10^{-6}$ predicts coral species lifetimes of approximately $200,000$ generations and speciation rates of 4.5 species per species per million generations when the metacommunity is of size 10^{10}. Reef fish lifetimes and speciation rates take similar values to those of corals (figure 4). The new model consequently marginally improves on the fit of the existing model for species abundances, but dramatically improves on its predictions of species lifetimes and speciation rates.

Local Community Dynamics

The original neutral model describes the dynamics of a 'local community' consisting of a dispersal-limited sample from the metacommunity source pool. In each time step, one individual in the local community dies and is then replaced with offspring from another individual in the local community with probability $(1 - m)$ and with offspring from the metacommunity (immigration) with probability m. The local community is sampled in a state of equilibrium between immigration and local extinction. Our model follows exactly the same convention except that the metacommunity abundance distribution is given by the new protracted speciation model. We demonstrate the distinction between the local community and metacommunity with simulation results (figure 5). The results show that very different metacommunities can produce almost identical local community distributions, again because the local community represents a very small sample from the metacommunity.

We compared the local community predicted by protracted speciation to empirical data using methods similar to those that were used for the metacommunity data. The approximate likelihood formula must be expanded to encompass dispersal limitation and the parameter m (see

box 1) . There are now three parameters to be fitted: m , θ and τ'. We fitted SAD data for tropical forest trees on Barro Colorado Island (BCI) (Condit, 1998; Hubbell *et al.*, 1999, 2005) (figure 5) which contains 21457 individuals. As for the metacommunity data, we again find almost exactly the same fit to the data for a range of values for τ', but the global optimum is at $\tau' = 8.15 \times 10^{-8}$. Although we do get a very similar value of θ as under point mutation, we now have $\theta = \frac{J_M \mu}{1-\mu} \approx J_M \mu$ and not $\theta \approx J_M \nu$, so the parameter θ should be reinterpreted in terms of speciation-initiation rate rather than speciation rate in all previous studies that used point mutation speciation. We show in figure 5 that values of τ' consistent with the SAD data can provide dramatically improved predictions for species lifetimes and speciation rates, whereas the original point mutation model's predictions cannot be considered as being even remotely credible (Ricklefs, 2003) (figure 5).

'Good' species and 'incipient' species

During the transition period of a lineage undergoing protracted speciation, the individuals of this lineage are interpreted as an 'incipient species' and appear conspecific with, but are slowly diverging from, their parent species. If 'incipient species' are observed during the transition time, they might be regarded as simply a natural variation in the population. The lineage forms a novel 'good species' only if it survives after the transition period has passed. We can calculate the expected richness of incipient species $\mathbb{E}[[S_c]$, by identifying any species as incipient when their source mutation occurred within the last τ generations. The number of incipient species is therefore the difference between the species richness of the old model and the new model (see box 1):

$$\mathbb{E}[S_c] = \theta \ln\left(\frac{1}{\mu}\right) - \theta \ln\left(\frac{1+\tau\mu}{\mu+\tau\mu}\right) \approx \theta \ln(\tau) \tag{5}$$

This will be of comparable size to the expected number of 'good species', given by $\mathbb{E}[S_g] = \theta \ln\left(\frac{1+\tau\mu}{\mu+\tau\mu}\right) \approx \theta \ln\left(\frac{1}{\tau\mu}\right)$, if $\tau^2 \approx \mu^{-1}$. A more germane measure is the probability that a randomly sampled individual belongs to an incipient species, which is approximately equal to $\tau\mu$

(see appendix D). The study of incipient species provides a way to estimate $\tau\mu$ and further test the new predictions of our model. This suggests that any data on individuals whose species identity is uncertain would be extremely valuable for understanding the properties of an ecological community.

A further notable prediction of our model is that for every incipient species that becomes a good species, there are τ incipient species that become extinct before the transition time has elapsed. This is analogous to the process of maturation for individuals: there may be many juveniles, but only very few survive to adulthood.

Some taxonomists use the time since divergence of pairs of lineages as a criterion for whether certain variants should classed as 'good' species (Hubbell, 2003). By setting τ according to this time, our model adopts this species concept in contrast to the original neutral model where diversity is fractal and infinitely divisible making any concept of species arbitrary (Hubbell, 2001).

Discussion

Introducing protracted speciation resolves serious difficulties with earlier neutral models, implying much more realistic numbers of rare species, speciation rates and species lifetimes (Ricklefs, 2003) and helping to explain how abundant species can be young (Nee, 2005; Ricklefs, 2006). The 'difference log-series' (DLS) model for metacommunity abundances is a simple and testable metacommunity model that contains a combination of favorable properties from both the log-series and log-normal distributions. In contrast to previous attempts to address these problems (Hubbell, 2003; Allen & Savage, 2007), our model remains tractable and retains the good fit to abundance data. Our model maintains the lineage structure of the original model, but makes new testable predictions regarding incipient species and rare species. We argue that many of the criticisms previously wielded against neutral theory should have been aimed more specifically at the point mutation mode of speciation rather than the neutrality assumption. Our results therefore support the idea that, before we can fairly evaluate the utility of the neutrality assumption, we must first experiment with relaxing the auxiliary assumptions associated with

the original neutral theory (Leigh, 2007; Etienne, 2007; Rosindell & Cornell, 2009). Furthermore, our results suggest that other models involving 'instantaneous' speciation should be replaced with a form of protracted speciation to avoid unrealistic predictions. For example, the random fission mode of speciation can be made protracted in the same way as we did for the point mutation model.

For samples from both local communities and metacommunities, our results suggest that a spectrum of possible values for τ' are consistent with the same fit to SAD data. We have shown that, in some cases, the parameter τ' plays no significant role for species abundance distributions, which permits using existing techniques for parameter estimation, but requires reinterpreting the parameters. If there is a very large sample of data, or if more than just species abundances are being considered, the protracted speciation model should be used. Protracted speciation always makes reasonable predictions for species lifetimes and speciation rates. To the best of our knowledge, protracted speciation is the first alternative mode of speciation that has actually been shown to match the performance of point mutation for fitting empirical SADs, because all other alternatives either remained untested or have performed worse. One of the advantages of neutral theory is that it can make testable predictions beyond those of SADs. We have shown in figure 5 that protracted speciation does make a dramatic improvement to the realism of such predictions. For further predictions of the theory, such as predictions of phylogeny, we expect that protracted speciation will once again have an important role to play.

A thorough investigation of the neutral model and its predictions really requires a fully spatially explicit model which can make spatially explicit predictions. Although this is beyond the scope of this paper, we note that protracted speciation is equally amenable to spatially explicit simulations as the original model. Coalescence techniques (such as those recently developed for simulating infinitely large and complex spatial structures (Rosindell *et al.*, 2008)) can be straightforwardly adapted to the new framework, but would be extremely complex or impossible to apply to the other alternatives to point mutation speciation (Hubbell, 2001, 2003; Allen & Savage, 2007; Etienne *et al.*, 2007). Furthermore, our expressions for species lifetime, speciation rate and metacommunity SADs will apply as approximations in any large spatially explicit

model where all individuals have equal intrinsic chances of reproduction and death.

There are many possibilities for further research: 1) One could study the changes to the model expected from adopting a Wright-Fisher model rather than a Moran model (Blythe & McKane, 2007) 2) While the DLS formula applies under the assumption of a large metacommunity, it would be very useful to have a formula that does not rely on this assumption, especially for analysing cases where the metacommunity may be relatively small in size. 3) A small point mutation speciation rate could be added to the model to represent the rare speciation events that can occasionally occur in a single generation, e.g. through polyploidy (Coyne & Orr, 2004) or these events could be interpreted as long distance dispersal events (Rosindell & Cornell, 2009). 4) Predictions about sub-species could be made by introducing a second shorter transition time describing the time required to become recognised as a distinct sub-species rather than just an incipient species. 5) Although our equations go some way to explain how really common species can come to exist in a relatively short time, we do not believe that this particular problem is entirely solved, but requires explicit incorporation of biogeography. For example, because the common ancestor of all trees found on BCI dates back to 150 Mya Kembel & Hubbell (2006), there are many bio-geographical influences, such as dramatic fluctuations in the metacommunity size, that could affect all individuals in this guild equally but which have thus far been ignored in all neutral models. An interesting and novel future study is therefore to incorporate these fluctuations which may explain how species can quickly become common (Alonso *et al.*, 2007).

Further implications of protracted speciation follow when considering the phylogeny. For example, some studies observe a slow-down in diversification near the present day, which is often attributed to the saturation of available niches (Barraclough & Nee, 2001; Phillimore & Price, 2008). Protracted speciation presents an alternative and entirely neutral explanation for the apparent slowdown: speciation processes having started but not yet finished. A stochastic distribution for τ would cause a gradual slow-down in diversification (rather than the sudden jump to zero diversification rate caused by a fixed τ). An instantaneous mode of speciation in a neutral model cannot explain the apparent slowdown in diversification. The phylogeny

could also be used as a valuable additional resource, for example pairs of sister species share a common ancestor 10 Mya on BCI (Kembel & Hubbell, 2006); consequently speciation cannot possibly take longer than 10 Myr to complete, providing us with an (albeit generous) upper bound for τ.

In this paper we have presented a candidate for the successor to the unified neutral theory of biodiversity and biogeography—one that is tractable and straightforwardly absorbed into existing analytical and computational research, but can also provide new and much more reasonable predictions and parameter estimates compared to the classic neutral model. The empirical data overwhelmingly support the use of protracted speciation over the point mutation speciation alternative. It is generally accepted that SADs alone can tell us little about the truth of a model, and thus neutral models should try to predict more than just species abundances in the future. In these cases, protracted speciation should certainly be utilised as it is essential for obtaining credible results. Protracted speciation can straightforwardly be implemented as it can be fitted using similar methods to those developed for the original theory. We hope that models incorporating protracted speciation will act as stepping stones to improved models of biodiversity and closer links with population genetics.

References

Allen, A.P. & Savage, V.M. (2007). Setting the absolute tempo of biodiversity dynamics. *Ecol. Lett.*, 10, 637–646.

Alonso, D., Etienne, R.S. & McKane, A.J. (2006). The merits of neutral theory. *Trends Ecol. Evol.*, 21, 451–457.

Alonso, D., Etienne, R.S. & McKane, A.J. (2007). Response to benedetti-cecchi: Neutrality and environmental fluctuations. *Trends Ecol. Evol.*, 22, 232–232.

Alonso, D. & McKane, A.J. (2004). Sampling hubbell's neutral theory of biodiversity. *Ecol. Lett.*, 7, 901–910.

Barraclough, T.G. & Nee, S. (2001). Phylogenetics and speciation. *Trends Ecol. Evol.*, 16, 391–399.

Blythe, R.A. & McKane, A.J. (2007). Stochastic models of evolution in genetics, ecology and linguistics. *J. Stat. Mech.-Theory Exp.*, 07018.

Condit, R. (1998). *Tropical Forest Census Plots*. Springer-Verlag and R. G. Landes company, Berlin, Germany.

Connolly, S.R., Hughes, T.P., Bellwood, D.R. & Karlson, R.H. (2005). Community structure of corals and reef fishes at multiple scales. *Science*, 309, 1363–1365.

Coyne, J. & Orr, H. (2004). *Speciation*. Sinauer Associates.

Dornelas, M. & Connolly, S.R. (2008). Multiple modes in a coral species abundance distribution. *Ecol. Lett.*, 11, 1008–1016.

Etienne, R.S. (2005). A new sampling formula for neutral biodiversity. *Ecol. Lett.*, 8, 253–260.

Etienne, R.S. (2007). A neutral sampling formula for multiple samples and an 'exact' test of neutrality. *Ecol. Lett.*, 10, 608–618.

Etienne, R.S. & Alonso, D. (2005). A dispersal-limited sampling theory for species and alleles. *Ecol. Lett.*, 8, 1147–1156.

Etienne, R.S., Apol, M.E.F., Olff, H. & Weissing, F.J. (2007). Modes of speciation and the neutral theory of biodiversity. *Oikos*, 116, 241–258.

Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.*, 12, 42–58.

Friedman, M. (1966). *The Methodology of Positive Economics*. University of Chicago press.

Gavrilets, S. (2004). *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, NJ/Oxford.

Gregory, R.D. (1994). Species abundance patterns of british birds. *Proc. R. Soc. B-Biol. Sci.*, 257, 299–301.

Gregory, R.D. (2000). Abundance patterns of european breeding birds. *Ecography*, 23, 201–208.

Holt, R.D. (2006). Emergent neutrality. *Trends Ecol. Evol.*, 21, 531–533.

Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ/Oxford.

Hubbell, S.P. (2003). Modes of speciation and the lifespans of species under neutrality: A response to the comment of robert e. ricklefs. *Oikos*, 100, 193–199.

Hubbell, S.P. (2006). Neutral theory and the evolution of ecological equivalence. *Ecology*, 87, 1387–1398.

Hubbell, S.P., Condit, R., & Foster, R.B. (2005). Barro colorado forest census plot data. http://ctfs.si.edu/datasets/bci.

Hubbell, S.P., Foster, R.B., O'Brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J. & de Lao, S.L. (1999). Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science*, 283, 554–557.

Hubbell, S.P., He, F.L., Condit, R., Borda-de Agua, L., Kellner, J. & ter Steege, H. (2008). How may tree species are there in the amazon and how many of them will go extinct? *Proc. Natl. Acad. Sci. U. S. A.*, 105, 11498–11504.

Hubbell, S.P. & Lake, J. (2002). *The neutral theory of biodiversity and biogeography, and beyond - In: Blackburn, T. and Gaston, K. (eds). Macroecology: patterns and process.* Blackwell.

Jabot, F. & Chave, J. (2009). Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecol. Lett.*, 12, 239–248.

Kembel, S.W. & Hubbell, S.P. (2006). The phylogenetic structure of a neotropical forest tree community. *Ecology*, 87, S86–S99.

Kurka, P., Sizling, A.L. & Rosindell, J. (2010). Analytical evidence for scale-invariance in the shape of species abundance distributions. *Math Biosci*, 223.

Leigh, E.G. (2007). Neutral theory: a historical perspective. *J. Evol. Biol*, 20, 2075–2091.

McGill, B.J. (2003a). Does mother nature really prefer rare species or are log-left-skewed sads a sampling artefact? *Ecol. Lett.*, 6, 766–773.

McGill, B.J. (2003b). A test of the unified neutral theory of biodiversity. *Nature*, 422, 881–885.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F.L., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.*, 10, 995–1015.

Muneepeerakul, R., Bertuzzo, E., Lynch, H.J., Fagan, W.F., Rinaldo, A. & Rodriguez-Iturbe, I. (2008). Neutral metacommunity models predict fish diversity patterns in mississippi-missouri basin. *Nature*, 453, 220–222.

Nee, S. (2005). The neutral theory of biodiversity: do the numbers add up? *Funct. Ecol.*, 19, 173–176.

Nee, S., Harvey, P.H. & May, R.M. (1991). Lifting the veil on abundance patterns. *Proc. R. Soc. B-Biol. Sci.*, 243, 161–163.

O'Dwyer, J.P. & Green., J.L. (2009). Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity. *Ecol. Lett.*

Phillimore, A.B. & Price, T.D. (2008). Density-dependent cladogenesis in birds. *PLoS. Biol.*, 6, 483–489.

Preston, F.W. (1981). Pseudo-lognormal distributions. *Ecology*, 62, 355–364.

Ricklefs, R.E. (2003). A comment on hubbell's zero-sum ecological drift model. *Oikos*, 100, 185–192.

Ricklefs, R.E. (2006). The unified neutral theory of biodiversity: Do the numbers add up? *Ecology*, 87, 1424–1431.

Rosindell, J. & Cornell, S.J. (2007). Species-area relationships from a spatially explicit neutral model in an infinite landscape. *Ecol. Lett.*, 10, 586–595.

Rosindell, J. & Cornell, S.J. (2009). Species-area curves, neutral models and long distance dispersal. *Ecology*, 90, 1743–1750.

Rosindell, J., Wong, Y. & Etienne, R.S. (2008). Coalescence methods for spatial neutral ecology. *Ecol. Inform.*, 3, 259–271.

Scheffer, M. & van Nes, E.H. (2006). Self-organized similarity, the evolutionary emergence of groups of similar species. *Proc. Natl. Acad. Sci. U. S. A.*, 103, 6230–6235.

Schluter, D. (2009). Evidence for ecological speciation and its alternative. *SCIENCE*, 323, 737–741.

Schluter, D. & Weir, J. (2007). Explaining latitudinal diversity gradients - response. *Science*, 317, 452–453.

Sizling, A.L., Storch, D., Sizlingova, E., Reif, J. & Gaston, K.J. (2009). Species abundance distribution results from a spatial analogy of central limit theorem. *Proc. Natl. Acad. Sci. U. S. A.*, 106, 6691–6695.

Volkov, I., Banavar, J.R., He, F.L., Hubbell, S.P. & Maritan, A. (2005). Density dependence explains tree species abundance and diversity in tropical forests. *Nature*, 438, 658–661.

Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, 424, 1035–1037.

Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2007). Patterns of relative species abundance in rainforests and coral reefs. *Nature*, 450, 45–49.

Williamson, M. & Gaston, K.J. (2005). The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. *J. Anim. Ecol.*, 74, 409–422.

Zhou, S.R. & Zhang, D.Y. (2008). A nearly neutral model of biodiversity. *Ecology*, 89, 248–258.

Acknowledgements

To the best of our knowledge we have offset carbon emissions for this project arising from flights, high performance computer use and lead author office space heating and lighting, we thank 'The Clean Planet Trust' registered charity 1112249 for their assistance with this. We are especially grateful to Bart Haegeman for crucial discussions about the derivation of the sampling formula. We thank David Alonso, Tim Benton, John Grahame, Luke Harmon, Patrick Jansen, Michael Kopp, Bill Kunin, Bert Leigh, Ben Miller, Han Olff, George Perry, Albert Phillimore, Robert Ricklefs and Yan Wong for useful comments and discussion, also Bill Allombert, Rick Condit, Xavier Didelot, Frank Hindriks, Theo Kuipers and Jan-Willem Romeijn for their assistance and David Bellwood for the use of his reef fish abundance data. We thank the EPSRC (grant EP/F043112/1) for funding JR, the NWO for funding RSE and The University of Leeds for hosting RSE as a visiting research fellow. The BCI forest dynamics research project was made possible by National Science Foundation grants to SPH, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Celera Foundation, and through the hard work of over 100 people from 10 countries over the past two decades.

Figure Legends

Figure 1

A simple example genealogy for various possible modes of speciation with different colours denoting different species. The left panel shows the point mutation speciation case where new species arrive as singletons. The centre panel shows protracted speciation where instead of appearing instantaneously, new species gradually evolve over a period of time τ shown with the colour of one species gradually fading into the colour of the next. Individuals belonging to an 'incipient' species (e.g. shades of dark green) would not be recognised as being a novel species and are classified as natural variants of the parent species shown in solid black, after the transition time a new species (solid green) is clearly recognised. The far right panel shows peripheral isolate speciation, random fission speciation and the case where initial abundances are fixed, but greater than one. In each case we show the phylogeny of good species just to the right of the genealogy. Splits in the phylogeny correspond to speciation events that have had time to complete.

Figure 2

The probability distributions of initial abundances of new species under various speciation mechanisms. In point mutation speciation and the alternative of fixed initial abundances proposed by Allen & Savage (2007), each new species begins with a fixed abundance. In random fission a population is randomly cleaved in two with the smaller part forming a new species leading to a constant probability for all possible initial abundances up to a size half that of the parent population. Peripheral isolate speciation yields a normal distribution with both mean and variance as model parameters and so it reduces to the case of fixed initial abundance (Allen & Savage, 2007) when the variance is set to zero. The only mechanistically derived distribution here is that of protracted speciation where initial abundances follow the distribution given in equation (8).

Figure 3

Species abundance distributions (SADs) for the log-series and difference log-series (DLS). Left panel: histogram of numbers of species in each abundance class for $J_M = 10^9$ and $\mu = 0.000001$ predicted by the log-series distribution (red line) and the Difference Log-Series (DLS) with a variety of different transition times $\tau = 0$, 1 , 10 , 100 , 1000 , 10,000 (bars with darker colours corresponding to larger τ and τ =10,000 in green). The j^{th} abundance class is defined as containing species with $j \leq$ abundances $< 2j$. As τ increases, there are dramatically fewer rare species, but the more abundant species follow the log-series distribution very closely. The remaining panels show the same data for the log-series (red line) and DLS with $\tau = 10,000$ (green line) as a plot of log abundance vs. rank in abundance. The whole metacommunity is shown in the centre, where the DLS is clearly distinct from the log-series. A random sample of 0.01% of individuals is plotted on the right showing that after sampling it is nearly impossible to observe the difference between the DLS and log-series.

Figure 4

Top left panel: a fit to coral reef metacommunity data (shown in blue) using the standard point mutation model with $\theta = 19.85$ (red) and protracted speciation model with $\theta = 20.13$ and $\tau' = 2.23 \times 10^{-6}$ (green). These parameters represent the global optimum in parameter space calculated with equation 4. The two black lines represent one standard deviation either side of the mean prediction under protracted speciation, produced by analysing 4,000,000 simulations of the model (using algorithms described in appendix A). It can clearly be seen that both models fit the abundance data equally well: we increased the thickness of the red line otherwise it would be completely hidden behind the green line. The top centre and top right panels show the species lifetimes (in generations) and speciation rates (per species per generation) respectively and were plotted using the same parameter values as those used for the top left panel. The lower set of panels show the same results for a reef fish metacommunity sample for which the best fitting θ was 15.82 under point mutation (red) and $\theta = 16.21$, $\tau' = 3.77 \times 10^{-5}$ under protracted speciation (green).

Figure 5

Local community species abundances compared with the regional species abundances and empirical data. Top left panel: a metacommunity abundance distribution with $J_M = 10^{10}$, $\theta = 100$ under different modes of speciation. Point mutation speciation (red bars), protracted speciation with $\tau = 100$ (green line) protracted speciation with $\tau = 10,000$ (black line). Top centre and right panels show dispersal limited local community samples $m = 0.1$ (top centre) and $m = 0.7$ (top right) for all three metacommunities. It is clear that in these local community samples we no longer see the very significant differences in the metacommunities from which they are drawn. A fit to locally sampled species local community abundance data for trees on BCI (Condit, 1998; Hubbell *et al.*, 1999, 2005) is shown in the bottom left panel. The parameters used were $\theta = 47.67$ and $m = 0.093$ for point mutation (red) and $\theta = 47.96$, $\tau' = 8.15 \times 10^{-8}$ and $m = 0.093$ for protracted speciation (green). Blue bars show empirical data and the two black lines represent the one standard deviation on either side of the mean of the protracted speciation result, calculated by simulation as in figure 4. The bottom centre and bottom right panels show speciation rate and species lifetime for the same optimal parameters as used to fit the species abundances in the bottom left panel.

Figures

Figure 1

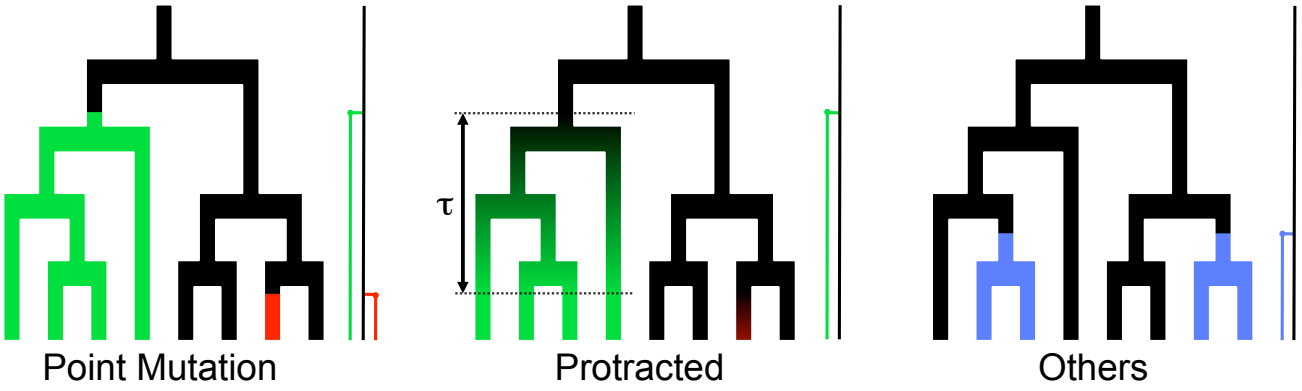


Figure 2

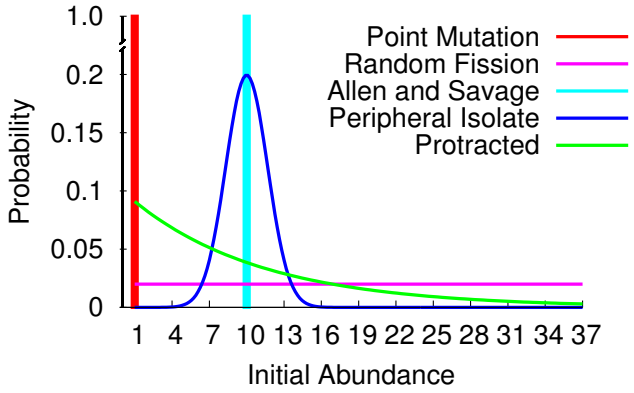


Figure 3

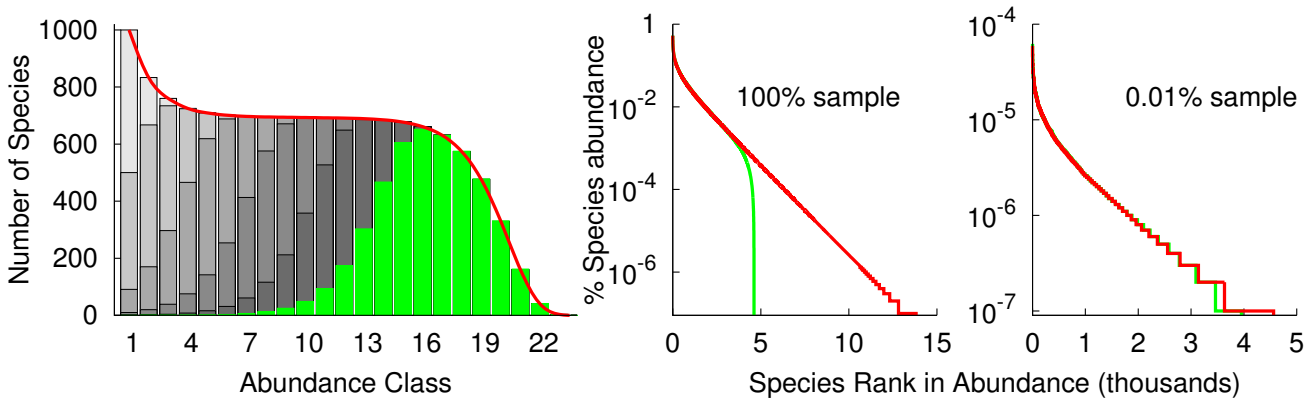


Figure 4

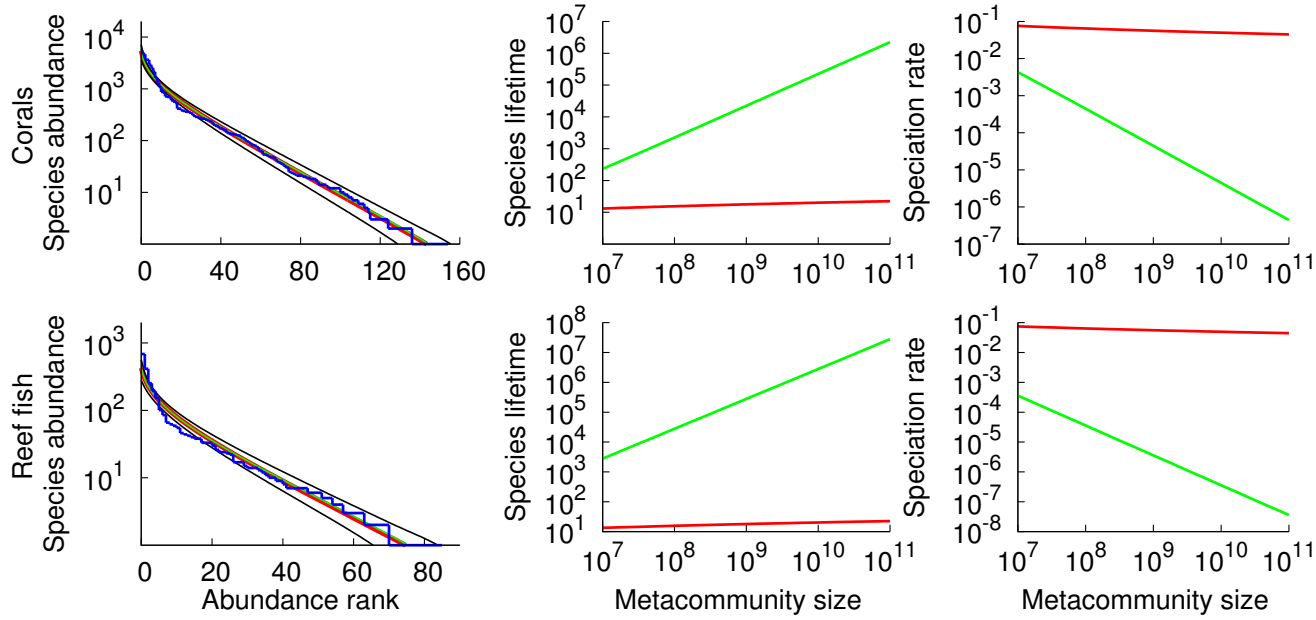
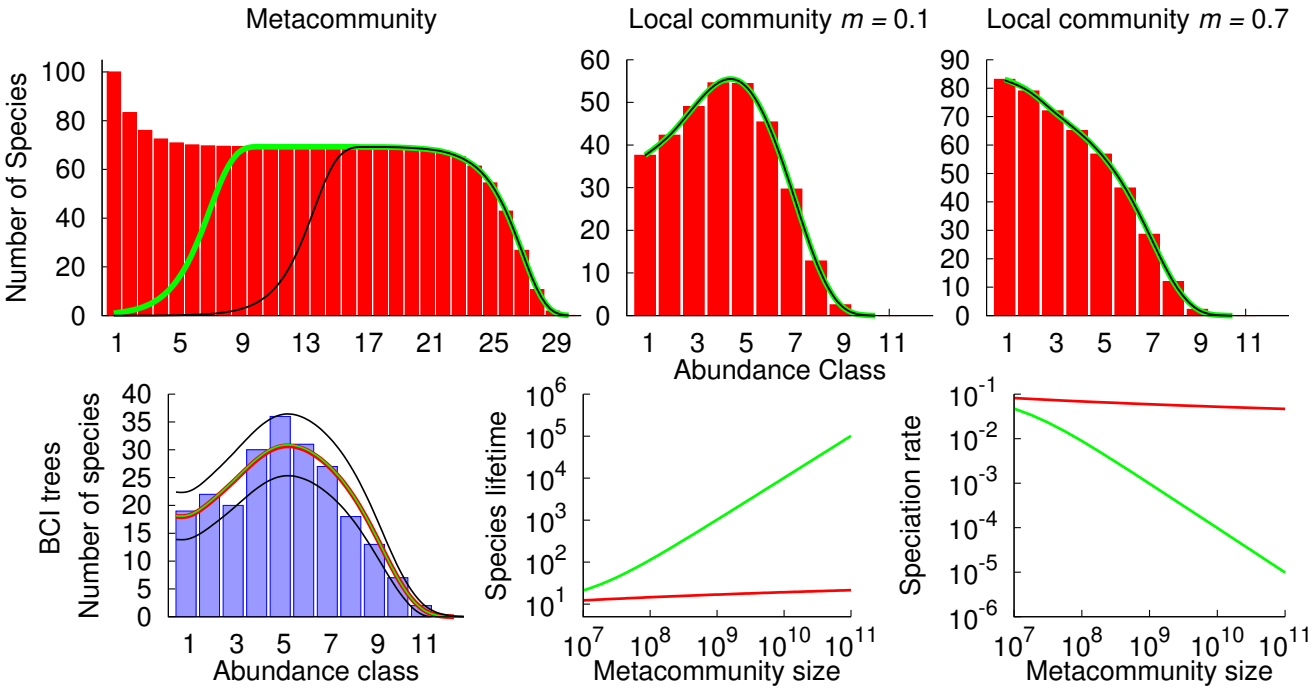


Figure 5


Metacommunity
Local community m = 0.1
Local community m = 0.7
Number of Species
Abundance Class
BCI trees
Number of species
Abundance class
Species lifetime
Metacommunity size
Speciation rate
Metacommunity size

Box 1

We use the formalism of probability generating functions to calculate the properties of the protracted speciation model. These can be derived by simply ignoring all of the mutations that have occurred in the most recent τ generations in original point mutation speciation model. We first consider the point mutation speciation model with general speciation-initiation rate μ. If the metacommunity size J_M is large, the probability that a species has abundance j at time t in the point mutation speciation model satisfies the equation

$$\frac{dP(j,t)}{dt} = (j+1)P(j+1,t) + (j-1)(1-\mu)P(j-1,t) - j(2-\mu)P(j,t) \tag{6}$$

(See equations (1), (8), (9) in reference (Volkov *et al.*, 2003) with $J_M \to \infty$ and $m = 0$), where the time unit is one generation. Note that the same equation would also apply if $P(j,t)$ denoted the number of species with abundance j, rather than the probability that a given species has abundance j. The probability generating function (PGF) $\Theta(z,t) \equiv \sum_{j=0}^{\infty} z^j P(j,t)$ satisfies the following equation, which can be obtained by multiplying equation (6) by z^j and summing over j:

$$\frac{\partial \Theta}{\partial t} = [1-z][1-(1-\mu)z]\frac{\partial \Theta}{\partial z}.$$

The general solution to this equation, starting from the initial condition $\Theta(z,0) = \Theta_0(z)$, is

$$\Theta(z,t;\mu) = \begin{cases} \Theta_0\left(\frac{1-z(1-\mu)-(1-z)e^{-\mu t}}{1-(1-\mu)(z+(1-z)e^{-\mu t})}\right) & \text{for } \mu > 0 \\ \Theta_0\left(\frac{t-z(t-1)}{t+1-zt}\right) & \text{for } \mu = 0 \end{cases} \tag{7}$$

Metacommunity abundance distribution

Denote by $\Theta^1(z,t;\mu) = \sum_{j=0}^{\infty} z^j P^1(j,t)$ the PGF of a species' abundance distribution at time t starting from a single individual at time 0, which is obtained by setting $\Theta_0(y) = y$ in equation (7). For the protracted speciation model, the PGF $\Theta_{\text{initial}}(z)$ of initial species abundance at the moment it becomes a 'good' species is determined by the descendants of a single individual after τ generations of neutral drift (i.e. no further mutations), and can be obtained from $\Theta^1(z,t;\mu)$

by setting $t = \tau$ and $\mu = 0$ in (7):

$$\Theta_{\text{initial}}(z) = \frac{\tau - z(\tau - 1)}{\tau + 1 - z\tau}$$

The probability that the lineage has survived to time τ is $\sum_{j=1}^{\infty} P^1(j, \tau) = 1 - \Theta_{\text{initial}}(0) = \frac{1}{\tau+1}$ (Leigh, 2007), while the initial abundance distribution conditional on the lineage surviving is obtained by expanding $\frac{\Theta_{\text{initial}}(z)}{1-\Theta_{\text{initial}}(0)}$ in powers of z:

$$P^{\text{initial}}(j|\text{survival}) = \frac{\tau^{j-1}}{(\tau+1)^j}, \tag{8}$$

The average initial abundance is $\sum_{j=1}^{\infty} jP^{\text{initial}}(j|\text{survival}) = (1 + \tau)$.

In the point mutation speciation model, new species arise at a rate of μJ_M so the equilibrium number of species $\psi_j^{\text{point}}$with abundance j is given by $\psi_j^{\text{point}} = J_M \mu \int_{-\infty}^{t} P^1(j, t - s)ds$. Define $\Theta_{\text{eq}}^{\text{point}}(z) = \sum_{j=1}^{\infty} z^j \psi_j^{\text{point}} = J_M \mu \int_0^{\infty} \left[\Theta^1(z, t; \mu) - \Theta^1(0, t; \mu)\right] dt$, which by setting $\Theta_0(y) = y$ in equation (7) and integrating yields

$$\Theta_{\text{eq}}^{\text{point}}(z) = J_M \frac{\mu}{1-\mu} \log\left[1 - (1-\mu)z\right] = \frac{J_M \mu}{1-\mu} \sum_{j=1}^{\infty} \frac{(1-\mu)^j}{j} \tag{9}$$

which is the PGF for Fisher's log-series.

The PGF $\Theta_{\text{eq}}^{\text{protract}}(z) = \sum_{j=1}^{\infty} z^j \psi_j^{\text{protract}}$ for the equilibrium distribution of abundances in the protracted speciation model is obtained by starting with the equilibrium distribution for the point mutation speciation model with speciation-initiation rate μ, then allowing τ generations of neutral drift ($\mu = 0$), i.e. by setting $\Theta_0(z) = \Theta_{\text{eq}}^{\text{point}}(z)$ and $t = \tau$ in equation (7):

$$\Theta_{\text{eq}}^{\text{protract}}(z) = -J_M \frac{\mu}{1-\mu} \log\left[\frac{1 - z\left(1 - \frac{\mu}{1+\mu\tau}\right)}{1 - z\left(1 - \frac{1}{1+\tau}\right)}\right] + J_M \frac{\mu}{1-\mu} \log\frac{1+\mu\tau}{1+\tau}$$

which is the PGF of the difference log series, equation (2). The term that is independent of z represents the species that have gone extinct in the period τ.

Species richness and lifetimes

The expected species richness is $\mathbb{E}[S] = \sum_{j=1}^{\infty} \psi_j^{\text{protract}} = \Theta_{\text{eq}}^{\text{protract}}(1) - \Theta_{\text{eq}}^{\text{protract}}(0) = J_{\text{M}} \frac{\mu}{1-\mu} \log\left[\frac{1+\tau\mu}{\mu+\tau\mu}\right]$. The average species lifetime L can be derived from the standard result Ricklefs (2003) $L = \frac{\text{species richness}}{\text{speciation rate}}$, yielding equation 3 .

Sample abundance distribution

The expected number of species with abundance j in a sample of size J is given by (Etienne & Alonso 2005):

$$\mathbb{E}\left[S(j)|J\right] = \frac{J!}{j!\,(J-j)!} \int_0^1 \frac{(Ip)_j \left(I\left(1-p\right)\right)_{J-j}}{(I)_J} \rho(p) dp \tag{10}$$

where $\rho(p)$ is the species density for an infinite metacommunity, $(x)_y = \frac{\Gamma(x+y)}{\Gamma(x)}$ denotes the Pochhammer notation, and I is the fundamental dispersal number which is related to the immigration probability by $I = \frac{m}{1-m}\left(J-1\right)$. The species density for an infinite metacommunity can be computed by taking the appropriate limit of the metacommunity abundance distribution:

$$\rho(p) = \lim_{J_{\text{M}} \to \infty} J_{\text{M}} \mathbb{E}\left[S(pJ_{\text{M}})\right] \tag{11}$$

For the protracted-speciation model $\mathbb{E}\left[S(pJ_{\text{M}})\right]$ is simply given by the difference log-series for abundance pJ_{M}. We only have to use the appropriate scaling of the parameters for $J_{\text{M}} \to \infty$. We already have the scaling for the fundamental biodiversity number θ:

$$\theta = \frac{\mu}{1-\mu}\left(J_{\text{M}} - 1\right) \Rightarrow \mu = \frac{\theta}{\theta + J_{\text{M}} - 1} \tag{12}$$

Now we also need a proper scaling for the transition time. When metacommunity becomes large, we expect the transition time to become large as well. Defining

$$\beta = \frac{1}{1+\tau}\left(J_{\text{M}} - 1\right) \Rightarrow \tau = \frac{J_{\text{M}} - 1}{\beta} - 1 \tag{13}$$

we therefore let J_M and τ go to infinity such that β is constant. This leads to

$$\rho(p) = \lim_{J_M \to \infty} J_M \frac{\theta}{p J_M} \left(\left(1 - \frac{\mu}{1+\tau\mu} \right)^{p J_M} - \left(1 - \frac{1}{1+\tau} \right)^{p J_M} \right) = \frac{\theta}{p} \left(e^{-\frac{\theta\beta}{\theta+\beta} p} - e^{-\beta p} \right) \quad (14)$$

Thus we have

$$\mathbb{E}\left[S(j) | J \right] = \frac{J!}{j!\,(J-j)!} \theta \int_0^1 \frac{(Ip)_j \left(I\left(1-p\right) \right)_{J-j}}{(I)_J} \frac{e^{-\frac{\theta\beta}{\theta+\beta} p} - e^{-\beta p}}{p} dp \quad (15)$$

When there is no dispersal limitation, this expression reduces to

$$\begin{aligned} \mathbb{E}\left[S(j) | J \right] &= \frac{J!}{j!\,(J-j)!} \theta \int_0^1 p^j \left(1-p\right)^{J-j} \frac{e^{-\frac{\theta\beta}{\theta+\beta} p} - e^{-\beta p}}{p} dp \\ &\approx \frac{J!}{j!\,(J-j)!} \theta \int_0^1 p^j \left(1-p\right)^{J-j} \frac{\left(1-p\right)^{\frac{\theta\beta}{\theta+\beta}-1} - \left(1-p\right)^{\beta-1}}{p} dp \\ &= \frac{J!}{(J-j)!} \frac{\theta}{j} \left(\frac{\Gamma\left(J + \frac{\theta\beta}{\theta+\beta} - j \right)}{\Gamma\left(J + \frac{\theta\beta}{\theta+\beta} \right)} - \frac{\Gamma\left(J + \beta - j \right)}{\Gamma\left(J + \beta \right)} \right) \end{aligned} \quad (16)$$

Equations (15) and (16) which are numerically computable, can be inserted in the approximate likelihood (4).