



**HAL**  
open science

# Empirical phylogenies and species abundance distributions are consistent with preequilibrium dynamics of neutral community models with gene flow

Anne-Sophie Bonnet-Lebrun, Andrea Manica, Anders Eriksson, Ana S.L. Rodrigues

## ► To cite this version:

Anne-Sophie Bonnet-Lebrun, Andrea Manica, Anders Eriksson, Ana S.L. Rodrigues. Empirical phylogenies and species abundance distributions are consistent with preequilibrium dynamics of neutral community models with gene flow. *Evolution - International Journal of Organic Evolution*, 2017, 71 (5), pp.1149-1163. 10.1111/evo.13228 . hal-03015387

**HAL Id: hal-03015387**

**<https://hal.science/hal-03015387>**

Submitted on 19 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Post-print of: Bonnet-Lebrun, A.-S., Manica, A., Eriksson, A. & Rodrigues, A.S.L. (2017). Empirical phylogenies and species abundance distributions are consistent with preequilibrium dynamics of neutral community models with gene flow. *Evolution*, 71, 1149–1163. <https://doi.org/10.1111/evo.13228>

# Neutral community models with gene flow generate realistic communities under non-equilibrium conditions

Anne-Sophie Bonnet-Lebrun<sup>[1,2,3]</sup>, Andrea Manica<sup>[2]</sup>, Anders Eriksson<sup>[2,4]</sup>, Ana S.L. Rodrigues<sup>[1]</sup>

<sup>[1]</sup>Centre d'Ecologie Fonctionnelle et Evolutive CEFE UMR 5175, CNRS - Université de Montpellier - Université Paul-Valéry Montpellier – EPHE, 1919 route de Mende, 34293 Montpellier cedex 5, France

<sup>[2]</sup>Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom

<sup>[3]</sup>École Normale Supérieure, 45 rue d'Ulm, F-75230 Paris, France

<sup>[4]</sup>Integrative Systems Biology Laboratory, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia

**e-mail addresses:** Anne-Sophie Bonnet-Lebrun ([anne-sophie.bonnet-lebrun@cefe.cnrs.fr](mailto:anne-sophie.bonnet-lebrun@cefe.cnrs.fr)), Andrea Manica ([am315@cam.ac.uk](mailto:am315@cam.ac.uk)), Anders Eriksson ([aje44@cam.ac.uk](mailto:aje44@cam.ac.uk)) and Ana S. L. Rodrigues ([ana.rodrigues@cefe.cnrs.fr](mailto:ana.rodrigues@cefe.cnrs.fr)).

**Correspondence author:** Anne-Sophie Bonnet-Lebrun, UMR 5175, CEFE CNRS, 1919 route de Mende, 34293 Montpellier Cedex 5, France; E-mail: [anne-sophie.bonnet-lebrun@cefe.cnrs.fr](mailto:anne-sophie.bonnet-lebrun@cefe.cnrs.fr), Telephone: +33467613205.

**Key words:** neutral community models, protracted speciation, point mutation, gene flow, macroevolution, non-equilibrium, phylogenetic tree shape.

**Statement of authorship:** ASLR, AM and ASBL designed the project; ASBL wrote the models (with support from AM and AE), performed the modelling work and analysed the output data; all authors participated in the discussion of the results; ASBL wrote the first draft of the manuscript; all authors commented on the manuscript.

**Abstract:**

Community characteristics reflect past ecological and evolutionary dynamics. Here, we investigate whether it is possible to reproduce the shape of empirical phylogenetic trees of birds and mammals and the shape of the distribution of species' abundances from neutral community models. To test the effect of gene flow, we contrasted two spatially explicit individual-based neutral models: one with protracted speciation, delayed by gene flow, and one with point mutation speciation, unaffected by gene flow. The former produced more realistic communities (shape of phylogenetic tree and species-abundance distribution), suggesting that gene flow is a key process in macro-evolutionary dynamics. Earlier models struggled to capture the empirically observed branching tempo in phylogenetic trees, as measured by the gamma statistic. We show that the low gamma values typical of empirical trees can be obtained in modelled communities with protracted speciation at non-equilibrium, particularly if they are sampled incompletely (i.e., not all species are known, particularly if the unknown species are the youngest). Overall, our results demonstrate that the characteristics of empirical communities that we have studied can, to a large extent, be explained through a purely neutral model.

## INTRODUCTION

Evolutionary studies aim at explaining the diversity of life. However, evolutionary processes are difficult to observe in real time, and the fossil record is extremely incomplete for many groups. Investigating evolution's results, such as the diversity of species and the phylogenetic relationships among them, can give insights into the underlying processes. Over recent decades, the development of molecular and statistical tools, combined with the systematic sharing of genetic sequencing data (e.g. GenBank; Benson et al. 2005), have fostered an explosion in the number and detail of available phylogenetic trees, including the creation of super-trees covering entire taxonomic groups (e.g., mammals: Bininda-Emonds et al. 2007; and birds: Jetz et al. 2012). Patterns in the shape of phylogenetic trees reflect the macroevolutionary processes through which the diversity of life has been generated (signal), but also purely stochastic effects (noise). Because phylogenetic trees are reconstructed rather than directly observed, they are also influenced by the choice of model or decisions by systematists (bias). The quest of disentangling these three signals has motivated a recent emphasis on the study of the characteristics of phylogenetic trees produced under different types of speciation models (e.g. Nee et al. 1994a; Nee 2006a; Pigot et al. 2010; Davies et al. 2011a; Manceau et al. 2015; Missa et al. 2016). Models allow the formal exploration of processes affecting speciation while controlling for confounding factors (e.g., testing the effect of speciation rates while controlling for community size), thus generating predictions that can be tested by comparison with empirical phylogenetic trees.

Comparing phylogenetic trees generated by different models requires first summarising their characteristics into meaningful measures. Accordingly, a number of statistics describing tree shape have been proposed (Mooers and Heard 1997a; Pybus and Harvey 2000a; Blum and François 2005, 2006a; Mooers et al. 2007). Two aspects of tree shape have attracted particular attention in previous studies: tree balance and branching tempo. Measures of tree balance (e.g. Shao 1990; Pybus and Harvey 2000a; Blum and François 2006b) reflect the evenness of distribution in species richness among clades. Empirical trees tend to be imbalanced (Heard 1992; Purvis et al. 2011), reflecting differences among lineages in their diversification rates. Branching tempo, the distribution of branching events (i.e. the spacing of internal nodes) over time, reflects variation in differentiation rates (Pybus and Harvey 2000b). Empirical trees tend to have internal nodes relatively close to their root (McPeck 2008; Phillimore and Price 2008), a pattern that has been interpreted as a slowdown in diversification rates towards the present (Morlon et al. 2010; Moen and Morlon 2014; Morlon 2014). By attempting to reproduce these characteristics of empirical phylogenetic trees (and either succeeding or failing to do so), models can provide insights into how speciation works.

The simplest models of clade diversification are lineage-based pure-birth (Yule 1925) or birth-death (Kendall 1948) models, whereby a phylogenetic tree is generated as lineages stochastically split (speciation) and (in the case of birth-death models) disappear (extinction) with given homogeneous rates across lineages and across time (e.g., Nee et al. 1992, 1994b,c, Morlon et al. 2010, 2011). But these models fail to produce the imbalance and slowdown in diversification rates typical of empirical trees (Table 1), producing instead perfectly balanced trees with either constant diversification rates (in pure-birth models) or with an artefactual increase in speciation rate (the “pull of the present”, in pure birth-death models; Nee 2006b; but see Etienne and Rosindell 2012).

These deviations between empirical trees and the predictions of simple speciation models may reflect ecological differences between lineages. For example, tree imbalance may result from trait variations across species that affect the probability of speciation or extinction (e.g., floral nectar spurs: Hodges and Arnold 1995), whereas a slowdown in diversification has often been interpreted as evidence of ecological niches being filled after an adaptive radiation (Phillimore and Price 2008; McInerney and Etienne 2012; Wennekes et al. 2012). But previous studies have shown that in some conditions it is

also possible to obtain realistic phylogenetic trees from individual-based neutral models (Pigot et al. 2010; Manceau et al. 2015; Missa et al. 2016), i.e. without the need to invoke ecological differences between species.

Reviewing the contrasting results of recent studies that have investigated the shape of trees obtained from neutral evolutionary models (Pigot et al. 2010; Davis et al. 2011; Manceau et al. 2015; Missa et al. 2016; Table 1) shed light into the conditions in which realistic phylogenetic trees can (or cannot) be generated. Whereas realistically imbalanced trees were frequently obtained, trees with realistic branching tempo were only generated in out-of-equilibrium scenarios. In Pigot et al. (2010) and Manceau et al. (2015), this out-of-equilibrium dynamics was intrinsic to the models, simulated by considering unbounded communities whose diversity and/or community size could expand indefinitely. In Missa et al. (2016), out-of-equilibrium was a transition phase between the founding event and phylogenetic equilibrium. These results support the hypothesis that the pattern of branching tempo observed in empirical phylogenies reflects out-of-equilibrium dynamics (Manceau et al. 2015; Missa et al. 2016). Both in Manceau et al. (2015) and in Missa et al. (2016), models predicted that as communities approached equilibrium the trees obtained were no longer realistic in terms of branching tempo (and see also Gascuel et al. 2015 for similar results in a non-neutral spatially-explicit community model). If the same is true for empirical communities, then the slowdown in diversification rates typical of real phylogenetic trees may indicate that these clades are still relatively young (still highly affected by the initial conditions that originated the new clade) rather than being relatively old (approaching saturation in ecological niche space) (Gascuel et al. 2015). And yet, these studies also found that an out-of-equilibrium dynamics did not always result in realistic trees (Table 1). For example, in the individual-based neutral community model of Missa et al. (2016), realistic trees were only obtained when speciation was modelled as fixed fission (whereby 5% of individuals become a new species), but not with either point mutation (a randomly selected individual becomes a new species) or random fission mutation (a random fraction – 0 to 50% – of all individuals become a new species).

These results suggest that in order to obtain trees with realistic patterns of branching tempo, neutral models need to generate an initial burst of diversification followed by a slowdown, whereby species that appear earlier in the phylogeny have on average higher probabilities of speciation, or lower probabilities of extinction, or both, in relation to the species that appear later on. For example, in the fixed fission model of Missa et al. (2016), the species originating from the founding species have an initial population size that is much smaller (5%) than that of the latter, and are therefore less likely to further diversify (their own descendants being even rarer and likely to go extinct). We deduce that, as a result, the trees obtained by this model are initially dominated by lineages that descend directly from – and have lower diversification rates than – the founding species, explaining the slowdown in diversification rates in pre-equilibrium trees (the signal of this basal radiation being then progressively eroded as demographic stochasticity generates a wider range of abundances among the species in the community). We expect an even more pronounced radiation with the model of cladogenesis through peripatry of Pigot et al. (2012) with stable range sizes. Indeed, in this case the founding species (occupying 9% of the overall area) remains throughout the simulations much more widespread than its descendants (on average 0.25% of the overall area), and is thus at the origin of most speciation events.

Whereas these results demonstrate that realistic trees can be obtained through neutral models, it is difficult to find real life parallels to the particular assumptions that made these results possible. Pigot et al. (2010) and Manceau et al. (2015) assumed communities to be unbounded in diversity and/or community size, which is unlikely to reflect the conditions in which most real phylogenies diversified. Missa et al. (2016) demonstrate that such assumption is not necessary, but obtained realistic phylogenies only with a particular type of fixed fission speciation (whereby 5% of the individuals of

the parent species are assigned to the new species) and did not propose a realistic mechanism that would result in a comparable speciation pattern.

The results by Pigot et al. (2010; for peripatric speciation) and by Missa et al. (2016; for fixed fission) both suggest the need for a speciation mechanism allowing a founding widespread species to radiate into a diversity of species with small-but-not-too-small populations – sufficiently numerous to have a reasonable likelihood of persisting, but not too abundant so that they do not become themselves the basis of a highly diverse lineage. Here, we propose that local speciation may be such a mechanism, and that this may be obtained as an emergent property in neutral community models if these incorporate gene flow under the realistic scenario of a spatially structured environment.

To test this hypothesis, we develop a neutral, individual-based community model with gene flow, building from the work of Rosindell et al. (2010), Rosindell and Phillimore (2011) and Gascuel et al. (in review) on protracted speciation. Rosindell et al. 2010 first introduced the concept of protracted speciation to take this into account the fact that speciation takes time. This was originally implemented simply by adding a delay to point mutation, such that it takes time between an individual beginning to speciate and it becoming a full species. Applied to lineage-based models, it avoids the “pull of the present” artefact and results in phylogenies with a realistic slowdown in diversification towards the present (Etienne and Rosindell 2012). However, speciation takes time not only because of the time needed to accumulate significant genetic differences between populations (to allow for the evolution of reproductive barriers) but because of the homogenising effects of gene flow between populations. To take this into account, Rosindell & Phillimore (2011) subsequently proposed a spatially-explicit individual-based neutral model (based on a mainland-single island system) with protracted speciation delayed by gene flow. This model allowed local speciation to occur in islands sufficiently close to the mainland to be colonised from the mainland, but not so close that gene flow prevents speciation. Gascuel et al. (in review) extended this model to a mainland-archipelago system, showing how the spatial structure caused by gene flow substantially increased levels of island speciation in relation to a similarly sized but unstructured community.

Here, we consider a chain of demes as an example of a spatially-structured community, initially colonised by a single species. We then investigate the characteristics of neutral communities structured by gene flow as they diversify, analysing in particular if the phylogenetic trees obtained are realistic in terms of balance as well as branching tempo. As a further test to the realism of these communities, we also investigate their composition, as assessed by species-abundance distributions (SADs; histograms representing the number of species in each class of abundance) (McGill 2003). We characterise the effects of gene flow by comparing two speciation modes: protracted speciation with gene flow vs. and point mutation (in which gene flow has no effect). Given the results from previous model studies, we expect that, even in the presence of spatial structure with gene flow, phylogenies will have a realistic branching tempo only during the non-equilibrium dynamics: thus, we analyse the results of the simulations over time, from the single-species foundation events until the modelled communities reach ecological and evolutionary equilibrium. Furthermore, we expect that the initial range of the founder species – affecting the likelihood that it will generate a rapid initial radiation – should play an important role in shaping the resulting phylogenies. We test this hypothesis by contrasting our main scenario (of a three different scenarios: a founder species that has invaded a virgin environment and whose range covers all available demes, i.e., (a first-wave colonisation) with two others:; with a founder species that has colonised an environment where its competitors already occupy most of the available habitat (second-wave colonisation); and with a founder species that has colonised an environment in which its competitors already occupy half of the available space (mixed colonisation). We investigate the sensitivity of predicted community characteristics to the key model parameters that govern community size, speciation and dispersal rates. Finally, given that the shape of

empirical phylogenetic trees, to which our model outputs are compared, may be affected by methodological biases (Davies et al. 2011; Moen & Morlon 2014) mostly due to tree incompleteness (Cusimano and Renner 2010; Brock et al. 2011; Höhna et al. 2011), we tested the effect of tree incompleteness on the properties of the phylogenies produced by the models.

## METHODS

### Neutral community models

We simulated communities using individual-based neutral models (*sensu* Hubbell 2001), where all individuals are equivalent in their probabilities of dying, reproducing and dispersing. We used a spatially explicit framework, with a one-dimensional set of demes (e.g. islands) connected by migration between adjacent demes (i.e. a one-dimensional stepping stone). The total population size was finite and constant, with a fixed number of individuals,  $K$ , per deme.

We departed from Hubbell (2001) by modelling communities with non-overlapping generations: at each time step, all the individuals died and were replaced. Each individual in a new generation descended, with probability  $1-m$ , from a randomly sampled individual from the same deme, and with probability  $m$  (migration) from a randomly sampled individual in an adjacent deme.

We integrated the effects of gene flow by modelling speciation as protracted, (Rosindell et al., 2010; 2011), whereby new species form after populations are isolated long enough for genetic differentiation to occur, with gene flow slowing this process down. In order to model speciation as a process that takes time, we assigned to each population (i.e., individuals of the same species in a given deme) a speciation timer. Whenever a population first became isolated (i.e., it colonised a new deme), the timer was set to  $T_{sp}$ , the time to speciation. It then decreased a unit step with each generation, with the population becoming a new species when the timer reached zero. This process was however delayed by gene flow due to migration from other demes. Hence, whenever one or more migrants of a given species arrived to a deme with an already existing population of the same species, we added to the speciation timer of this resident population a (non-integer) value calculated as  $T_{delay} \times N_m/N_i$ , where  $T_{delay}$  is the time delay induced by migration,  $N_m$  is the number of migrants and  $N_i$  is the size of the resident population. We were able to apply this method because we simulated communities forward in time, and thus knew each species' abundance at any point in time (such approach would not have been possible if we had modelled communities through coalescence, as in Rosindell et al. 2011 Gascuel et al. in review). Speciation was irreversible: once a population's timer reached zero, it became a new species and no longer connected by gene flow to other populations that were previously of the same species.

We contrasted the results of our model with those obtained through a similar neutral model without gene flow, where speciation was instead modelled through point mutation (e.g., Hubbell 2001; Ricklefs 2003; Chave 2004; Etienne 2007; Davies et al. 2011b). In this model, each individual in a new generation had a given probability  $\nu$  of becoming a new species, and so the number of founding individuals in each species always equalled one.

All simulations were run in Julia 0.3.10 (Bezanson et al. 2014). We ran ten stochastic simulations for each set of parameters. For each simulation, at any given time, we recorded the state of the modelled community including, for each species: its ID; the ID of its mother species; its date of birth; and, if the species went extinct during the previous 20 generations, its date of death. We used this information to build phylogenetic trees in Matlab (The MathWorks, Inc., 2010) and R 3.2.2 (R Development Core Team 2015) with the extant species at each time step, from which we could quantify tree shape ( $\beta$  and

$\gamma$  metrics, see below for details) and SADs (see below for details). The Julia code is provided in Supplementary Materials.

### Main model parameters

We started our model exploration using a spatially-structured system of  $N=40$  demes, each with a carrying capacity  $K=250$  individuals, and migration rate  $m=0.05$ . After an initial exploration of parameters, we set the time to speciation  $T_{sp}=150$  and the delay penalty for immigration  $T_{delay}=14$  for protracted speciation, and selected a speciation rate  $\nu=0.0002$  for point mutation. This set of speciation parameters was selected in order to make results comparable across the two speciation models, as it generated in both cases communities with similar numbers of species ( $\approx 33$ ). We call this the “main” model, to distinguish it from various modifications outlined below. Computational limitations prevented us from focusing on communities with many more individuals/species and/or much lower speciation rates.

In this version of the model, all demes were initially filled with the same species, and the community then diversified through speciation (subsequently called ‘first wave colonisation’). The ecological analogy to this setting would be the first colonization of an empty archipelago by a single species, from a very remote mainland, in which the species initially colonizes the entire archipelago before going on to speciate.

### Metrics of tree shape

Tree balance was estimated using  $\beta$  (Blum & François 2006) from Aldous’  $\beta$ -split model of clade growth (Aldous 2001). The  $\beta$ -splitting family is a single parameter family of branching processes.  $\beta$  is estimated by maximum likelihood methods based on the cladogram. Its expected value under the constant-rate pure-birth model (Yule model) is zero, with negative values for trees that are more imbalanced than the average prediction of the Yule model (Fig. S1A), and positive values for trees that are more balanced (Fig. S1B). Compared to other measures of tree balance (Shao 1990; Mooers and Heard 1997b),  $\beta$  is unbiased with respect to species richness, allowing for the comparison of trees of different sizes. Accordingly, it is a commonly used measure in studies of phylogenetic tree shape (e.g., McPeck 2008; Phillimore and Price 2008).

We assessed the branching tempo through Pybus and Harvey’s  $\gamma$  index (Pybus and Harvey 2000b), measuring the distribution of internal nodes along the time axis of the phylogenetic trees. Under the Yule model,  $\gamma$  follows a standard Gaussian distribution (Yule 1925). Positive values are obtained when the internal nodes are closer to the leaves than under the Yule model, interpreted as an acceleration in diversification rates towards the present. Negative  $\gamma$  are obtained when internal nodes are closer to the roots, and are often interpreted as sign of a decrease in diversification rates with time, either an early burst of diversification or a slowdown near the present (Fordyce 2010; Morlon et al. 2010).

All analyses of the trees (stored in parenthetical format) were made in R, using the “apTreeshape” package (Bortolussi *et al.* 2006).

### Species-abundance distributions

SADs for real communities tend to be approximately lognormal (Preston 1962; Hubbell 2001). In order to check if the communities produced by the models were realistic, we built SADs by combining results across a given time period (e.g., for a given simulation, with a given set of parameters, we obtained a SAD from  $t=1020$  to  $t=2000$  by pooling all 40 communities obtained every 20 time steps within this interval). We present SADs averaged across the ten simulation runs (replicates) for each set of parameters values, using a Preston plot (a semi-log<sub>2</sub> representation following Preston 1962).

## Empirical phylogenetic trees

For comparative purposes, we calculated values of  $\beta$  and  $\gamma$  for two sets of empirical trees containing between 10 and 40 species – one for birds (29 trees), and one for mammals (112 trees). The bird trees were a subset of the 129 trees in Jetz et al. (2012; their supplementary data; number of species: median=24.00, Q1=16.00, Q3=33.00; tree age: median=0.24 million years [My], Q1=0.17 My, Q3=0.47 My). The mammal trees were all subclades with 10 to 40 species from a mammalian super tree developed by Bininda-Emonds et al. (2007; first extended by Fritz, Bininda-Emonds, & Purvis 2009 and then by Rodrigues et al. 2011 to include all living mammals) at the level of genus (number of species: median=14.50, Q1=12.00, Q3=19.75; tree age: median =14.35 My, Q1=9.07 My, Q3=20.40 My) or family (number of species: median=17.50, Q1=14.25, Q3=26.00; tree age: median=25.40 My, Q1 =18.42 My, Q3=31.07 My). Although these empirical trees were obtained from what are likely the most complete phylogenies available, they are nonetheless incomplete, particularly for mammals (Reeder et al. 2007; Fjeldsa 2013).

## Effects of mode of speciation on community characteristics

We contrasted communities simulated with the main model parameters and first wave starting conditions under the two modes of speciation described above – protracted *vs.* point mutation. Each simulation ran for 80,000 generations, long enough to reach both ecological equilibrium (stable number of species) and phylogenetic equilibrium (stable tree shape as measured by both  $\beta$  and  $\gamma$ ). We recorded the state of the modelled communities at intervals of 20 generations, allowing us to investigate how tree shape and SADs varied over time, from the founding event (i.e., when the set of demes is first colonised), to equilibrium. We then averaged results across 10 replicates. In the figures, we represent average values over 400 generations.

## Sensitivity to range of the founder species

To test the extent to which results were affected by the range of the founder species considered in the main model (first wave colonisation), we have also simulated communities obtained through protracted speciation starting from under two very different starting conditions: second-wave and mixed colonisation.

In the second wave setting, the archipelago was previously filled with individuals whose species was not recorded. In two adjacent demes, we then replaced all individuals from the initial community with individuals of a new founder species of an unrelated lineage, and tracked its fate. This new lineage often went quickly extinct, but sometimes it expanded and diversified, producing a new community with its own phylogeny and SAD. We analysed successfully established second wave lineages, defined as the daughter lineages of the newly introduced lineage. The ecological analogy of this model is the colonisation of an already saturated archipelago by a species taxonomically unrelated to those already in place.

In the mixed starting conditions, we simulated a situation intermediate between first and second wave colonisations. As in the second wave setting, we started with a saturated system, but the newly introduced lineage was allowed to occupy immediately twenty adjacent demes out of the forty demes (i.e. half of the available space).

## Sensitivity to model parameters

We found the first wave colonisation scenario to generate the most realistic phylogenies, so we used this scenario to thoroughly investigate the effect of different parameters on the shape of the resulting phylogenies. We explored the following ranges of values for each parameter independently, while

keeping all other parameters fixed to the values for the main model (i.e.,  $N=40$ ,  $K=250$ ,  $m=0.05$ ,  $v=0.0002$ ,  $T_{sp}=150$ ,  $T_{delay}=14$ ): {0.01, 0.03, 0.05, 0.07} for  $m$ ; {100, 150, 200} for  $T_{sp}$ ; {10, 12, 14, 15, 16} for  $T_{delay}$ ; and {0.0001, 0.0002, 0.0003, 0.0004, 0.0005} for  $v$ .

We also tested our results held with higher values of carrying capacity ({250, 500, 750, 1000} for  $K$ ), but, due to computational limitations, we had to reduce the number of demes in these simulations to  $N=20$  (all other parameters as in the main model).

### Effect of incomplete trees and sampling

The shape of empirical phylogenetic trees may to an extent reflect methodological biases (Davies et al. 2011; Moen & Morlon 2014). In particular, real trees are often incomplete, even among well-known taxa, as substantial fractions of species remain undescribed. Tree incompleteness has been shown to influence tree shape (Cusimano and Renner 2010; Brock et al. 2011), particularly when the sampling of nodes is not random (e.g., impact of oversampling older/younger nodes on the distribution of branch lengths, Höhna et al. 2011). We therefore tested the effect of tree incompleteness on the properties of the models produced by the models.

We simulated incomplete trees by removing one third of the species from those trees that had been obtained from the first wave model with protracted speciation. We tested two ways in which empirical trees are likely to be incomplete: by removing the rarest species (with the smallest number of individuals; less likely to be known to science and/or to be represented in the molecular databases used to build the empirical trees); and by removing the youngest species (more likely to be cryptic species, not yet differentiated from currently described species). Even though the youngest species tend to be rare (their initial abundance cannot exceed the deme size), these two sets of species are not necessarily the same, as the rarest species are not necessarily the youngest.

As controls, we also investigated the effects of removing one third of all species, uniformly selected among all extant species.

## RESULTS

### Effect of mode of speciation on community characteristics

Given the same parameters for community size ( $N$ ,  $K$ ) and migration rate ( $m$ ), the same initial conditions (first wave), and a careful selection of speciation parameters ( $v$ ,  $T_{sp}$ ,  $T_{delay}$ ) to ensure that a similar number of species was obtained in both cases ( $\approx 33$ ), speciation type had little effect on the equilibrium shape of the phylogenetic trees produced by our spatially-explicit individual-based neutral community models. Indeed, with the protracted speciation model, the values of  $\gamma$  and  $\beta$  tended, respectively, towards  $\approx 6$  and  $\approx -0.5$  (Fig. 1A, C), and with the point speciation model, they tended towards  $\approx 7.5$  and  $\approx -1$  (Fig. 1B, D).

There were however important differences in the shape of phylogenetic trees during pre-equilibrium phrase. In particular, whereas in point mutation communities  $\gamma$  values were always positive (Fig. 1B), protracted speciation led to a short initial phase (ca. 2000 generations) during which phylogenetic trees spanned the full range of the (mostly negative) empirical  $\gamma$  values (Fig. 1A). Protracted speciation simulations also took substantially longer to reach phylogenetic equilibrium (ca. 30,000 generations before  $\gamma$  stabilised) than point mutation models (ca. 20,000 generations), but reached ecological equilibrium faster (at less than 2000 generations, compared to ca. 5000 generations for point mutation).

The SADs of communities differed markedly between speciation types, being highly skewed to the left in communities produced by point mutation (a large number of singletons and rare species) and approximately log-normal in communities obtained through protracted speciation, both in the pre-equilibrium phase and after the system became stable (Fig. 1G-H).

### **Sensitivity to range of the founder species**

In protracted speciation models with second wave starting conditions, the large majority of simulations resulted in the rapid extinction of the colonising lineage; lineage survival was higher for mixed initial conditions (Fig. 2). In both cases, the long-lived lineages (that survived past 70,000 generations) tended to produce communities with similar properties to those obtained for first wave conditions at equilibrium, in terms of tree shape ( $\gamma \approx 6$  Fig. 2A-B;  $\beta$  generally slightly negative, Fig. 2C-D), community size ( $\approx 30$  species, Fig. 2E-F), and realistic SADs (Fig. 2G-H). Communities during the pre-equilibrium phase also had realistic SADs (Fig. 2G-H) but values of  $\gamma$  (Fig. 2 A-B) and  $\beta$  (Fig. 2 C-D) were less realistic than those obtained for the first wave colonisation conditions (Fig. 1 A-B), with  $\gamma$  still reaching negative values during the early part of the simulations but for an ever shorter period (particularly for the second wave colonisation scenario) and  $\beta$  behaving more erratically. Overall, then, the larger the range of the initial founder species (in decreasing order first wave, mixed and second wave colonisation) the more realistic the phylogenies obtained.

### **Sensitivity to model parameters**

The shape of communities simulated through individual-based neutral models with protracted speciation (first wave initial conditions) was robust to variation in the model's parameters. Indeed, for all values tested, we continued to obtain realistic (slightly negative) values of  $\beta$ , realistic (log-normal-like) SADs, and values of  $\gamma$  that span realistic (negative) empirical values but only during the initial, pre-equilibrium, phase (Fig. 3). Variation in model parameters had nonetheless some quantitative effects: higher values of the migration rate  $m$  (Fig. 3A) and, to a lesser extent, lower values of time delay  $T_{\text{delay}}$  (Fig. 3E) prolonged the time needed for reaching ecological and phylogenetic stability, and so the duration of the period during which negative  $\gamma$  values were obtained. An increase in the number of demes  $N$  resulted in the initial values of  $\gamma$  to be even more negative, but it did not affect the duration of the phase during which  $\gamma$  values remain negative (Fig. 3C).

### **Effect of incomplete trees and sampling**

Pruning the tree to mimic incomplete sampling increased substantially the duration of the phase during which communities obtained through protracted speciation (first wave initial conditions) had realistically negative values of  $\gamma$  (Fig. 4A). This was true for all types of pruning, but particularly noticeable for pruning of the youngest species (where negative  $\gamma$  values were obtained for about 5000 generations, compared with about 2000 for the full tree). Pruning slightly increased  $\beta$  values, but these remained generally slightly negative on average, and within the range observed in empirical trees (Fig. 4b). The shape of SADs was only noticeably affected by the rarest species pruning scheme, which, by definition, led to a lack of rare species (Fig. 4D).

## **DISCUSSION**

Our results support previous studies in showing that it is possible to obtain realistic phylogenetic trees from individuals-based neutral models (Pigot et al. 2010; Manceau et al. 2015; Missa et al. 2016). As in these previous studies, it was only in non-equilibrium circumstances that we were able to do obtain realistic trees, in particular when it came to obtaining realistic branching tempo ( $\gamma$ ), thus lending further weight to the hypothesis that the shape of real trees may at least in some cases reflect non-equilibrium

dynamics (Manceau et al. 2015; Missa et al. 2016). There is a growing body of evidence that real communities are often not at equilibrium, with lineage-through-time plots for empirical trees typically not showing evidence of equilibrium dynamics (Morlon 2014). This may reflect clades that are still young (at pre-equilibrium, or having just reached equilibrium recently; Rabosky & Glor, 2010) or the effects of carrying capacity changing as a response to changing environmental conditions (Quental & Marshall 2013).

We also found that non-equilibrium conditions, *per se*, do not guarantee realistic tree shapes (Missa et al. 2016, Davies et al. 2011). Like Pigot et al. (2010), Manceau et al. (2015) and Missa et al. (2016), we were only able to obtain realistic trees for particular speciation modes, but our model is more realistic than those of previous studies. Indeed, unlike in Pigot et al. (2010) and Manceau et al. (2015) we were able to obtain realistic tree shapes without needing to allow for ongoing community expansion, instead retaining the more realistic zero-sum assumption of neutral community models. In our simulations, like in Missa et al. (2016), the total number of individuals does not vary, showing that an expanding community size is not a necessary condition for obtaining realistic trees. But unlike in the fixed fission model of Missa et al. (2016), where the rate and form of cladogenesis is an input, in our model cladogenesis is an emergent property of the system, with local speciation occurring whenever populations are sufficiently separated in space that the effects of genetic drift (as modelled by the parameter  $T_{sp}$ ) overcome the effects of gene flow (controlled by  $m$  and  $T_{delay}$ ). Even though we were only able to obtain realistic trees (in terms of branching tempo) during a relatively short transition phase soon after the founding effect, we were able to reproduce these results for a wide variety of model parameters (Fig. 3). We found this transition phase is particularly extended in communities that are more interconnected (with higher levels of migration rates,  $m$ ) and, to a lower degree, for which it takes longer for species to separate (higher time delay,  $T_{sp}$ ), although if any of these parameters becomes too extreme speciation ceases completely. Furthermore, we show that the communities produced by our neutral model have realistically-shaped species-abundance distributions, including during the non-equilibrium phase (Fig. 1H).

Previous studies on lineage-based models had found that incomplete trees have lower  $\gamma$  values (Cusimano and Renner 2010; Brock et al. 2011). Our analysis extends this to trees generated by individual-based neutral community models (Fig. 4A). We also found that this effect was particularly noticeable when the missing species were the rarest or the youngest in the tree. It is plausible that empirical phylogenetic trees are biased in this way: rare species are more likely both to have not yet been described and to have been more affected by premature anthropogenic extinction, and young species are more likely to be difficult to differentiate from others and so to have been missed by taxonomists. Coherent with this hypotheses, in our sample of empirical trees the mammal phylogenies have on average more negative values of  $\gamma$  than bird phylogenies (respectively,  $\gamma = -3.1$  and  $\gamma = -1.4$ ), as expected given that the former are known to be more incomplete than the latter (Reeder et al. 2007; Fjeldså 2013). In our models, tree incompleteness on its own was not sufficient to drive  $\gamma$  below zero in communities at equilibrium, but in protracted speciation models tree pruning increased substantially the time during which realistically negative  $\gamma$  values were obtained (Fig. 4A).

Our results were highly dependent of the initial conditions (Fig. 2). As in Missa et al. (2016), our main model considers a scenario where the initial conditions consist of a single species occupying the entire territory and then differentiating. As these populations all start to speciate at the same time, an initial radiation is generated at the base of the phylogenetic tree. Given that each of these new species has a lower probability of speciation than the original one, the subsequent slowdown in speciation rates generates the negative values of  $\gamma$  we observed. A similar process likely occurs in Missa et al. (2016)'s fixed fission model as well as in the vicariance model by Pigot et al. (2010). Accordingly, we found it very difficult to produce trees that were realistic in terms of branching tempo in our second wave initial

conditions, where the founding species occupies a single deme (Fig. 2A), but we were still able to do so in our mixed colonisation conditions, where the founding species occupies half of the demes (Fig. 2B). In summary, we found that a necessary condition for obtaining realistic trees with our model was that an initially abundant and widespread species diversifies through local speciation into a set of rarer species. Such initial conditions may correspond to an ecological scenario in which a virgin territory is colonised from a distant source, with the initial species first dispersing across the entire territory before starting to speciate. An alternative scenario is the colonisation of a new niche after the evolution of an innovation, if this results in a new species that is very successful and thus abundant and widespread. Whilst this setting invokes an adaptive (i.e. non-neutral) explanation for the formation of a new guild, the following radiation dynamics within that guild can be neutral. Indeed, our results suggest that if the original widespread species diversifies into rarer (less likely to diversify) species, a slowdown in diversification rates emerges even if these new species are all functionally equivalent, and thus does not necessarily reflect the saturation of ecological niche space.

## ACKNOWLEDGEMENTS

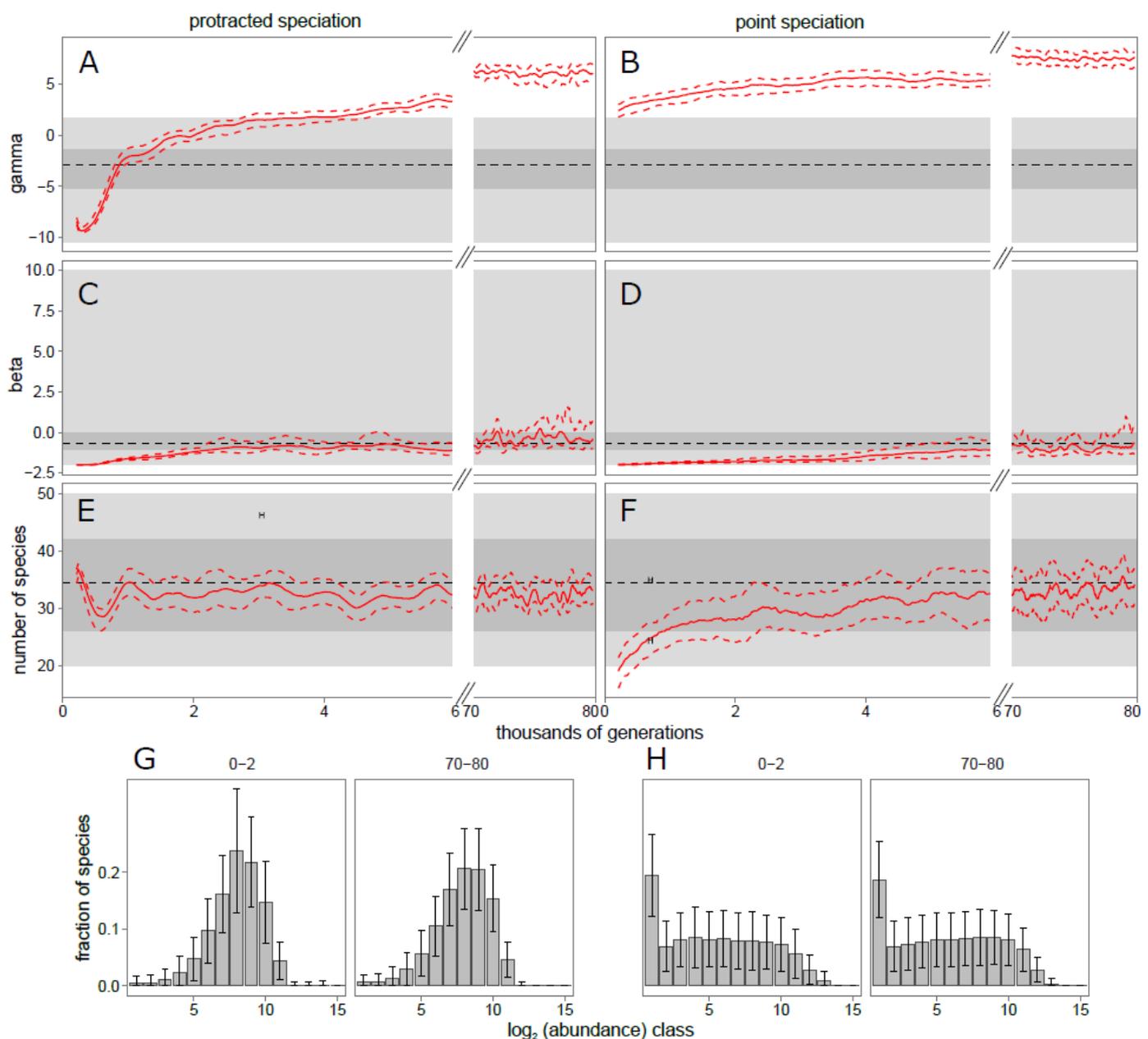
We are grateful to F. Laroche and F. Jabot for very constructive discussions and to O. Gimenez for technical support. We also thank James Rosindell and an anonymous referee for helping improve the manuscript.

## REFERENCES

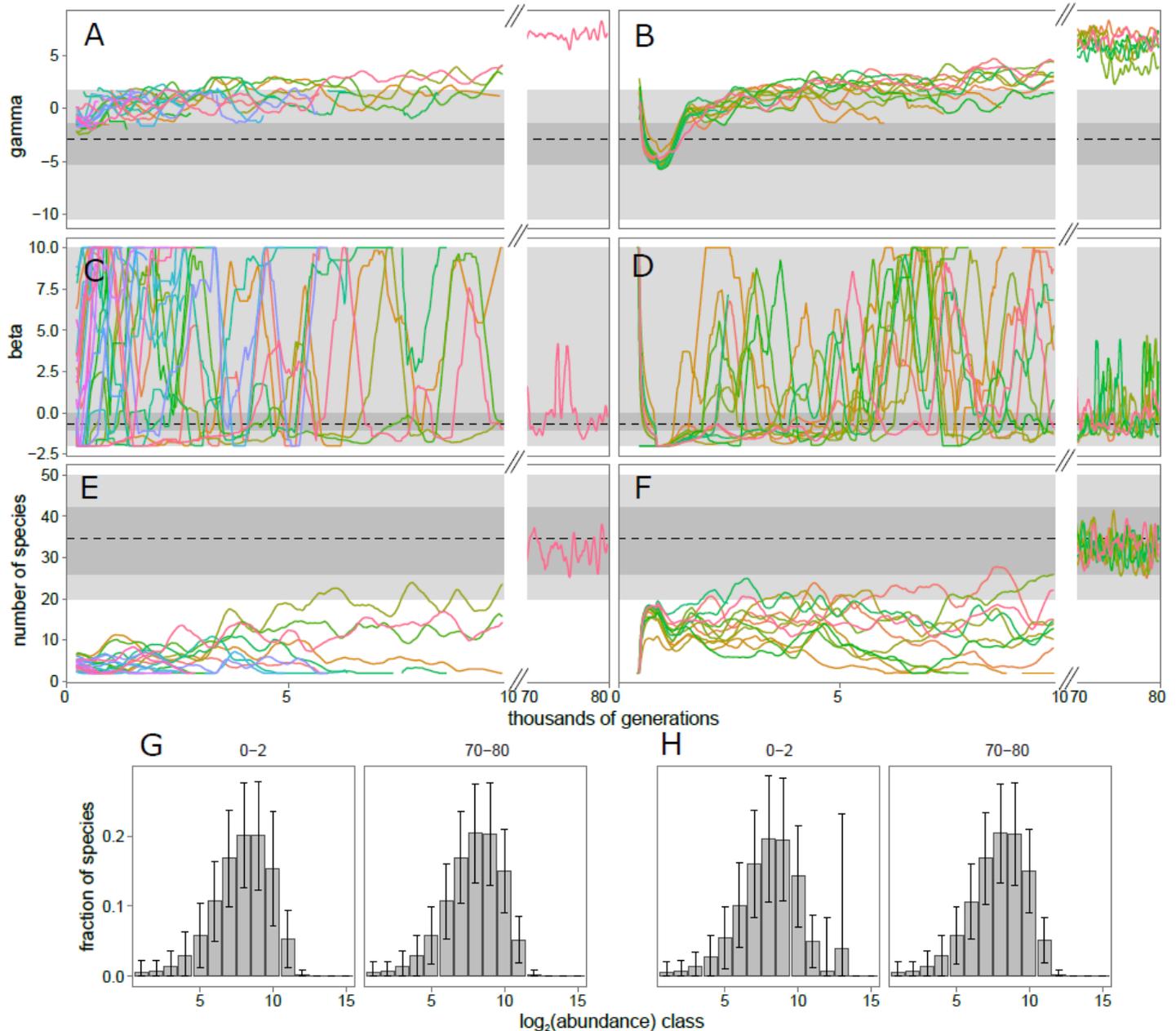
- Aldous, D. J. 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Stat. Sci* 16:23–34.
- Benson, D. A., Karsch-Mizrachi, Ilene, Lipman, David J., Ostell, James, and Wheeler, David L. 2005. GenBank. *Nucleic Acids Res.* 33:D34–D38.
- Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah. 2014. Julia: A Fresh Approach to Numerical Computing. ArXiv14111607 Cs.
- Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. *Nature* 446:507–512.
- Blum, M. G. B., and O. François. 2005. On statistical tests of phylogenetic tree imbalance: The Sackin and other indices revisited. *Math. Biosci.* 195:141–153.
- Blum, M. G. B., and O. François. 2006b. Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Syst. Biol.* 55:685–691.
- Brock, C. D., L. J. Harmon, and M. E. Alfaro. 2011. Testing for temporal variation in diversification rates when sampling is incomplete and nonrandom. *Syst. Biol.* 60:410–419.
- Chave, J. 2004. Neutral theory and community ecology. *Ecol. Lett.* 7:241–253.
- Cusimano, N., and S. S. Renner. 2010. Slowdowns in diversification rates from real phylogenies may not be real. *Syst. Biol.* 59:458–464.
- Davies, T. J., A. P. Allen, L. Borda-de-Água, J. Regetz, and C. J. Melián. 2011a. Neutral biodiversity theory can explain the imbalance of phylogenetic trees but not the tempo of their diversification. *Evolution* 65:1841–1850.

- 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., and J. Rosindell. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Syst. Biol.* 61:204–213.
- Fjeldså, J. 2013. The discovery of new bird species. Pp. 147–186 in *Handbook of the Birds of the World*. del Hoto, J., Elliot, A., Sargatal, J. & Christie, D., Barcelona.
- Fordyce, J. A. 2010. Interpreting the  $\gamma$  statistic in phylogenetic diversification rate studies: a rate decrease does not necessarily indicate an early burst. *PLoS ONE* 5:e11781.
- Gascuel, F., R. Ferrière, R. Aguilée, and A. Lambert. 2015. How Ecology and Landscape Dynamics Shape Phylogenetic Trees. *Syst. Biol.* 64:590–607.
- Heard, S. B. 1992. Patterns in tree balance among cladistic, phenetic and randomly generated phylogenetic trees. *Evolution* 46:1818–1826.
- Hodges, S. A., and M. L. Arnold. 1995. Spurring Plant Diversification: Are Floral Nectar Spurs a Key Innovation? *Proc. R. Soc. Lond. B Biol. Sci.* 262:343–348.
- Höhna, S., T. Stadler, F. Ronquist, and T. Britton. 2011. Inferring speciation and extinction rates under different species sampling schemes Inferring speciation and extinction rates. *Mol. Biol. Evol.*, doi: 10.1093/molbev/msr095.
- Hubbell, S. P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. *Nature* 491:444–448.
- Kendall, D. G. 1948. On the Generalized “Birth-and-Death” Process. *Ann. Math. Stat.* 19:1–15.
- Manceau, M., A. Lambert, and H. Morlon. 2015. Phylogenies support out-of-equilibrium models of biodiversity. *Ecol. Lett.* 18:347–356.
- McGill, B. J. 2003. A test of the unified neutral theory of biodiversity. *Nature* 422:881–885.
- McInerney, G. J., and R. S. Etienne. 2012. Ditch the niche – is the niche a useful concept in ecology or species distribution modelling? *J. Biogeogr.* 39:2096–2102.
- McPeck, M. A. 2008. The ecological dynamics of clade diversification and community assembly. *Am. Nat.* 172:270–284.
- Missa, O., C. Dytham, and H. Morlon. 2016. Understanding how biodiversity unfolds through time under neutral theory. *Phil Trans R Soc B* 371:20150226.
- Moen, D., and H. Morlon. 2014. Why does diversification slow down? *Trends Ecol. Evol.* 29:190–197.
- Mooers, A., and S. Heard. 1997a. Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* 72:31–54.
- Mooers, A. O., L. J. Harmon, M. G. B. Blum, D. H. J. Wong, and S. B. Heard. 2007. Some models of phylogenetic tree shape. Pp. 149–170 in O. Gascuel and M. Steel, eds. *Reconstructing evolution: new mathematical and computational advances*. Oxford University Press, Oxford.
- Morlon, H. 2014. Phylogenetic approaches for studying diversification. *Ecol. Lett.* 17:508–525.

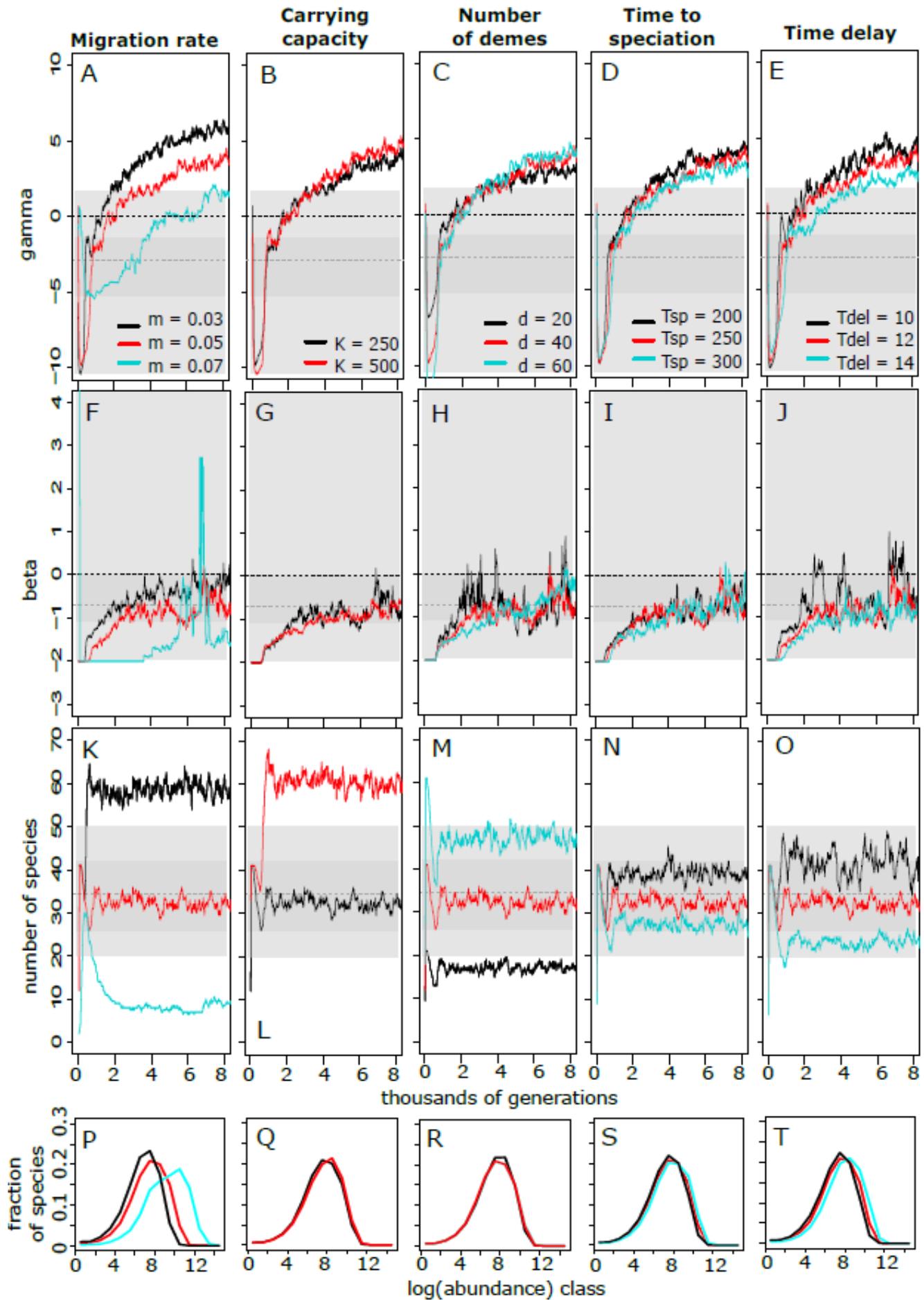
- Morlon, H., T. L. Parsons, and J. B. Plotkin. 2011. Reconciling molecular phylogenies with the fossil record. *Proc. Natl. Acad. Sci.* 108:16327–16332.
- Morlon, H., M. D. Potts, and J. B. Plotkin. 2010. Inferring the dynamics of diversification: a coalescent approach. *PLoS Biol* 8:e1000493.
- Nee, S. 2006a. Birth-death models in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* 1–17.
- Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994a. Extinction rates can be estimated from molecular phylogenies. *Philos. Trans. R. Soc. B Biol. Sci.* 344:77–82.
- Nee, S., R. M. May, and P. H. Harvey. 1994c. The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 344:305–311.
- Nee, S., A. O. Mooers, and P. H. Harvey. 1992. Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci.* 89:8322–8326.
- Phillimore, A. B., and T. D. Price. 2008. Density-dependent cladogenesis in birds. *PLoS Biol* 6:e71.
- Pigot, A. L., A. B. Phillimore, I. P. F. Owens, and C. D. L. Orme. 2010. The Shape and Temporal Dynamics of Phylogenetic Trees Arising from Geographic Speciation. *Syst. Biol.* 59:660–673.
- Purvis, A., S. A. Fritz, J. Rodríguez, P. H. Harvey, and R. Grenyer. 2011. The shape of mammalian phylogeny: patterns, processes and scales. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366:2462–2477.
- Pybus, O. G., and P. H. Harvey. 2000a. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. B-Biol. Sci.* 267:2267–2272.
- Reeder, D. M., K. M. Helgen, and D. E. Wilson. 2007. Global trends and biases in new mammal species discoveries. *Occas. Pap. Mus. Tex. Tech Univ.* 269:1–35.
- Ricklefs, R. E. 2003. A comment on Hubbell’s zero-sum ecological drift model. *Oikos* 100:185–192.
- Rosindell, J., S. J. Cornell, S. P. Hubbell, and R. S. Etienne. 2010. Protracted speciation revitalizes the neutral theory of biodiversity. *Ecol. Lett.* 13:716–727.
- Rosindell, J., S. P. Hubbell, and R. S. Etienne. 2011. The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol. Evol.* 26:340–348.
- Rosindell, J., and A. B. Phillimore. 2011. A unified model of island biogeography sheds light on the zone of radiation. *Ecol. Lett.* 14:552–560.
- Shao, K.-T. 1990. Tree Balance. *Syst. Biol.* 39:266–276.
- Wennekes, P. L., J. Rosindell, and R. S. Etienne. 2012. The neutral—niche debate: a philosophical perspective. *Acta Biotheor.* 60:257–271.
- Yule, G. U. 1925. A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 213:21–87.



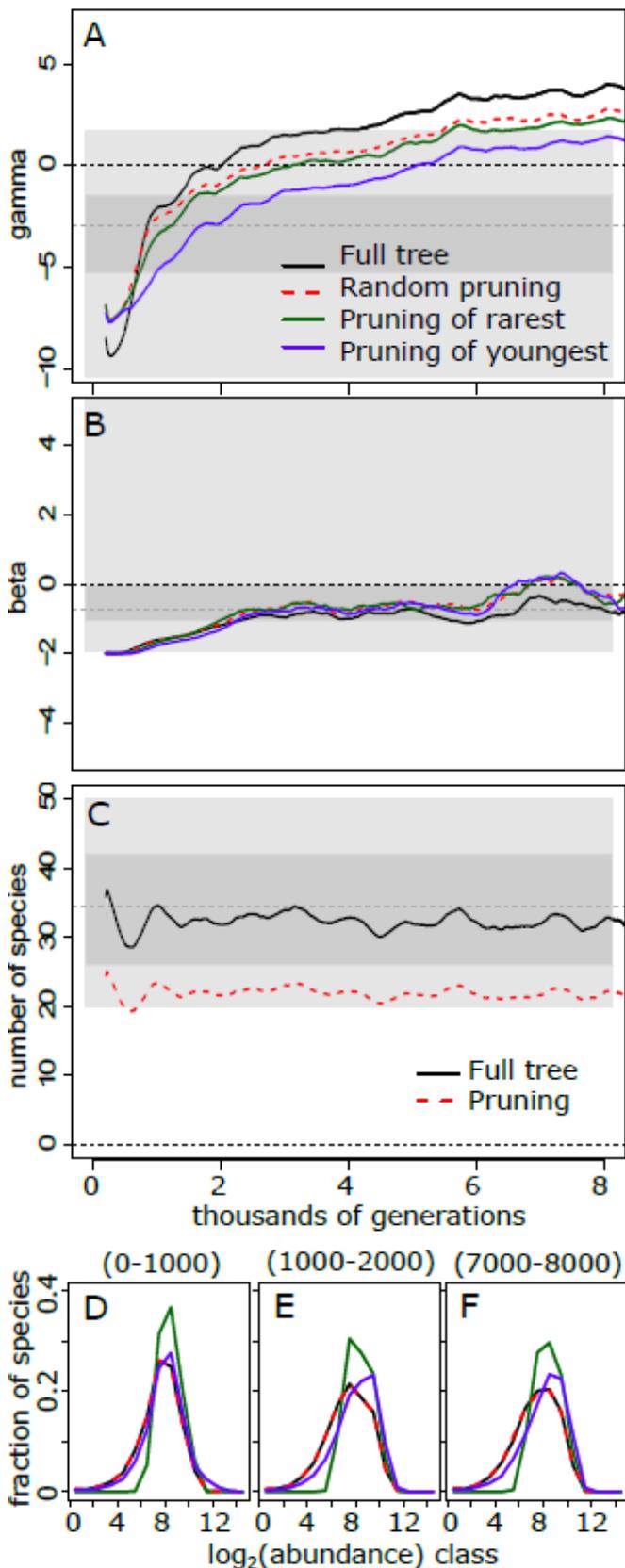
**Figure 1:** Effect of speciation type – protracted (left column) vs. point mutation (right column) – on tree shape (gamma  $\gamma$  statistic, characterising the distribution of branch length across time, A-B; and beta  $\beta$  statistic, characterising balance; C-D); community size (number of species; E-F) and community composition (species-abundance distributions, SAD; G, H), in communities obtained from individual-based neutral models, compared with empirical values. Model parameters are: number of demes  $N=40$ ; carrying capacity of each deme  $K=250$ ; migration rate  $m=0.05$ ; in protracted speciation models:  $T_{sp}=150$ , time delay  $T_{delay}=14$ ; in point mutation models, speciation rate  $\nu=0.0002$ ; initial conditions: all demes occupied by a single species (first wave colonisation). In A-F: dark lines represent the median of the statistics, with the first and third quartiles represented by the dashed line, smoothed by calculating the mean value in a 400 time steps interval; grey bands represent the whole range (lighter) and the 50% interval of empirical values (darker), and the horizontal dashed line the median, for empirical values obtained from 29 bird phylogenies and from 112 mammal phylogenies. In G-H, each histogram represents the SAD for the time interval between brackets, obtained by pooling all communities obtained every 20 time steps within this interval, and averaging across 10 simulations; error bars for each class represent the standard deviation around the mean.



**Figure 2:** Effect of initial conditions – second wave (left column) vs. mixed (right column) – on tree shape (gamma  $\gamma$  statistic A-B; and beta  $\beta$  statistic C-D); community size (number of species, E-F) and community composition (species-abundance distributions, SAD, G-H), in communities obtained from individual-based neutral models with protracted speciation, compared with empirical values. Model parameters as in Figure 1, except for the initial conditions (second wave: two demes filled with new lineage; mixed: 20 demes filled with new lineage). Each coloured line shows the results of a single simulation; some become extinct before 80,000 generations. Representation of empirical data (gray bands, dashed line) as in Figure 1. SADs are only represented for the lineages that are still extant at the end of the simulation.



**Figure 3:** Effect of model parameters on tree shape (gamma  $\gamma$  statistic A-E; and beta  $\beta$  statistic F-J); community size (K-O) and community composition (species-abundance distributions, SAD; P-T), in communities obtained from individual-based neutral models with protracted speciation, in pre-equilibrium conditions ( $\leq 8,000$  generations). The parameters tested were: migration rate ( $m$ ), carrying capacity ( $K$ ), number of demes ( $d$ ), time to speciation ( $T_{sp}$ ) and time delay ( $T_{delay}$ ). Coloured lines correspond to median values of the 10 replicates. In each column, we varied the specific parameter being tested while keeping all other parameters as in Figure 1. SADs obtained as histograms with the same class intervals as in Fig. 1, represented here as lines to facilitate comparisons.



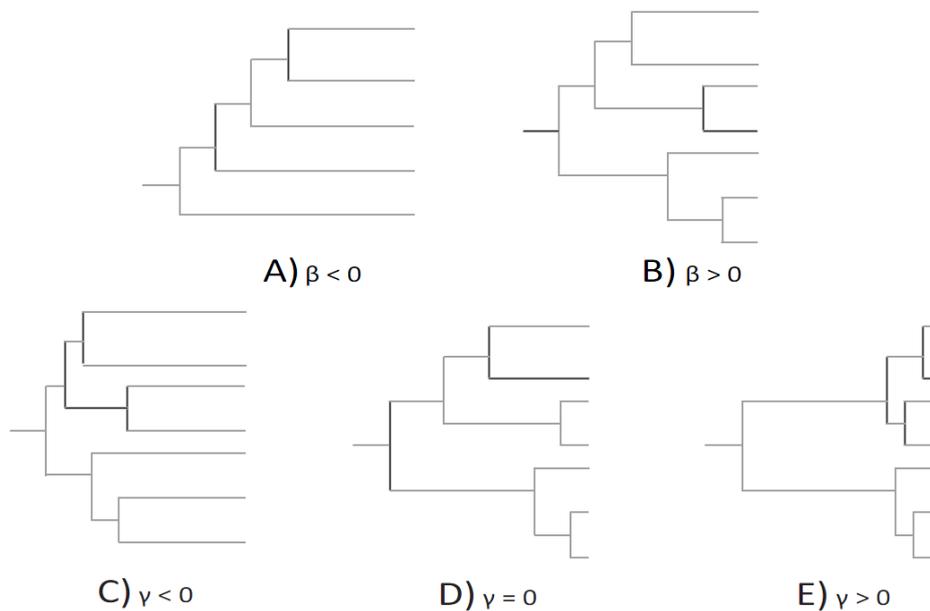
**Figure 4:** Effect of incomplete knowledge on: on tree shape (gamma  $\gamma$  statistic A; and beta  $\beta$  statistic B); community size (C) and community composition (species-abundance distributions, SAD; D-F) in pre-equilibrium conditions ( $\leq 8,000$  generations). Incomplete knowledge was simulated by pruning the phylogenetic trees obtained from individual-based neutral models with protracted speciation (full tree results are those represented in Figure 1, left column). Three pruning schemes were tested – random, rarest, and youngest – whereby one third of the species were deleted. Model parameters and empirical data sets as in Figure 1. SAD representations as in Figure 3.

**Table 1.** Summary of analyses of the shape of phylogenetic trees obtained from neutral models. In each case, the realism of the modelled trees is compared with that of empirical trees, in terms of: balance (empirical trees are generally imbalanced, with  $\beta < 0$ ); diversification rates over time (empirical trees generally have  $\gamma < 0$  or  $\rho < 0$ , a slowdown in diversification rates).

Model	Speciation mode	Tree balance	Branching tempo
<b>Pure birth model</b> (Yule, 1925; Pigot et al. 2010; Morlon 2014). Lineage-based, starts with a single lineage that splits (speciation) with a given probability (constant over time). Zero extinction.	Cladogenesis happens as a lineage splits in two.	Unrealistic $\beta = 0$	Unrealistic $\rho = 0$
<b>Pure birth-death model</b> (Kendall, 1948; Pigot et al. 2010; Morlon 2014). Lineage-based, starts with a single lineage that splits (speciation) or disappears (extinction) with given probabilities (constant over time).	Cladogenesis happens as a lineage splits in two.	Unrealistic $\beta = 0$	Unrealistic $\rho > 0$
<b>Geographical models of cladogenesis</b> (Pigot et al. 2010). Species-based, spatially-explicit. Starts with a single founding species', its range randomly placed into a (finite) landscape, which diversifies into a clade. Species' range borders drift at random according to a normal distribution of mean $\mu$ (controlling change in average range size: stable for $\mu=0$ , increasing for $\mu>0$ ) and variance $\sigma^2$ (controlling stochasticity in border position: stable borders for low $\sigma^2$ ; highly variable borders for high $\sigma^2$ ). Once formed, species' dynamics are independent. Hence, the system does not necessarily reach an equilibrium and the number of species can increase indefinitely.	Geographic speciation through vicariance: pairs of sister species obtained through random bisection of the parent species' distribution, caused by barriers of random length, randomly placed. Species of intermediate size more likely to speciate (small ranges encounter fewer barriers, larger ranges may not be completely bisected). Widespread species less likely to go extinct (through stochastic range border changes).	Generally unrealistic $\beta > 0$ but realistic $\beta < 0$ obtained for high rates of range expansion	Generally unrealistic $\rho \geq 0$ , but realistic $\rho < 0$ obtained for high rates of range expansion
	Geographic speciation through peripatry: a new species forms from through dispersal, as a peripheral isolate outside the ancestor's range. Probability of speciation larger for species with larger range perimeter. New species have small ranges (on average, 0.25% of total area; compared to 9% for the founding species).	Generally unrealistic $\beta > 0$ but realistically $\beta < 0$ when ranges are very stable ( $\mu=0$ , low $\sigma^2$ ).	Generally unrealistic $\rho > 0$ , but realistic $\rho < 0$ , when ranges are very stable ( $\mu=0$ , low $\sigma^2$ ).
<b>Individual-based neutral community models, at equilibrium</b> (Davies et al. 2011). All individuals are equivalent in their rates of birth, death, dispersal and speciation. Communities have a limited (fixed) size so individuals compete for space/resources. Both spatially-implicit and spatially-explicit models tested (with similar results). Starts from a monospecific community. Results are for equilibrium conditions.	Point mutation: one randomly selected individual becomes a new species.	Realistic $\beta < 0$ for a range of speciation and migration rates	Unrealistic $\gamma > 0$
	Equal splits fission mutation: half of the individuals (randomly selected) in parent species assigned to a new species.	Realistic $\beta < 0$ only for large mutation rates	Unrealistic $\gamma > 0$
	Random fission mutation: a random percentage (0 to 50%) of randomly selected individuals from the parent species is assigned to new species.	Unrealistic $\beta > 0$	Unrealistic $\gamma > 0$
	Fixed incipient species abundance: 10 individuals become a new species.	Realistic $\beta < 0$	Unrealistic $\gamma > 0$

<p><b>Individual-based neutral community model with speciation by genetic differentiation</b> (Manceau et al. 2015). Spatially implicit. All individuals are equivalent (in their rates of birth <math>b</math>, death <math>d</math>, and mutation <math>\nu</math>). Community size is unbounded, and thus does not necessarily reach an equilibrium; positive growth rates if <math>b &gt; d</math>.</p>	<p>Speciation by genetic differentiation: genetic mutations with rate <math>\nu</math> give rise to a new genetic type. Speciation happens when two system populations no longer contain individuals of the same genetic type. Speciation thus takes time. Speciation rates are an emergent property of the model dynamics.</p>	<p>Wide range of <math>\beta</math> values, including realistic <math>\beta &lt; 0</math> when population expanding (out-of-equilibrium dynamics)</p>	<p>Wide range of <math>\gamma</math> values, including realistic <math>\gamma &lt; 0</math> when population expanding (out-of-equilibrium dynamics)</p>
<p><b>Individual-based neutral community models, at pre-equilibrium</b> (Missa et al. 2016). All individuals are equivalent in their rates of birth, death, dispersal and speciation. Spatially-explicit model. Communities have a fixed size so individuals compete for space/resources. Results are from the foundation event (monospecific community) to equilibrium conditions.</p>	<p>Point mutation: one randomly selected individual becomes a new species</p>	<p>Realistic <math>\beta &lt; 0</math></p>	<p>Unrealistic <math>\gamma &gt; 0</math></p>
	<p>Random fission mutation: a random percentage (0 to 50%) of randomly selected individuals from the parent species assigned to new species</p>	<p>Unrealistic <math>\beta \geq 0</math></p>	<p>Realistic <math>\gamma &lt; 0</math> during an initial, pre-equilibrium, phase, then increasing to unrealistic <math>\gamma &lt; 0</math></p>
	<p>Fixed fission: 5% of individuals (randomly selected) from the parent species assigned to new species</p>	<p>Realistic <math>\beta &lt; 0</math></p>	<p>Realistic <math>\gamma &lt; 0</math> during an initial, pre-equilibrium phase, then increasing to unrealistic <math>\gamma &lt; 0</math></p>

## SUPPLEMENTARY MATERIAL



**Figure S1:** Examples of trees with variation in balance (distribution of species richness across clades, characterised by the beta  $\beta$  statistic) and in the distribution of branch length across time (characterised by the gamma  $\gamma$  statistic). (A) and (B) show trees with constant speciation rates, but (A) is imbalanced (with species disproportionately concentrated in particular clades) while (B) is balanced (with species distributed evenly among clades). Irrespective of balance, in (C) trees have internal nodes close to the root of the tree (reflecting a slow-down in diversification), in (D) the internal nodes are evenly distributed, and in (E) internal nodes are close to the tips of the tree (reflecting an acceleration in diversification).