

UNIVERSIDADE ESTADUAL DE CAMPINAS
SISTEMA DE BIBLIOTECAS DA UNICAMP
REPOSITÓRIO DA PRODUÇÃO CIENTÍFICA E INTELLECTUAL DA UNICAMP

Versão do arquivo anexado / Version of attached file:

Versão do Editor / Published Version

Mais informações no site da editora / Further information on publisher's website:

<https://onlinelibrary.wiley.com/doi/10.1111/ecog.04937>

DOI: 10.1111/ecog.04937

Direitos autorais / Publisher's copyright statement:

©2020 by Wiley-Blackwell Publishing. All rights reserved.

DIRETORIA DE TRATAMENTO DA INFORMAÇÃO

Cidade Universitária Zeferino Vaz Barão Geraldo

CEP 13083-970 – Campinas SP

Fone: (19) 3521-6493

<http://www.repositorio.unicamp.br>

Research

Allopatry increases the balance of phylogenetic trees during radiation under neutral speciation

Flavia Maria Darcie Marquitti, Lucas D. Fernandes and Marcus Aloizio Martinez de Aguiar

F. M. D. Marquitti (<https://orcid.org/0000-0003-0510-3992>) ✉ (flamarquitti@gmail.com) and M. A. M. de Aguiar (<https://orcid.org/0000-0003-1379-7568>), Inst. de Física 'Gleb Wataghin', Univ. Estadual de Campinas, Campinas, SP, Brazil. – L. D. Fernandes (<https://orcid.org/0000-0001-6053-2277>), Dept of Life Sciences, Imperial College London, Silwood Park, Ascot, Berkshire, UK.

Ecography

43: 1487–1498, 2020

doi: 10.1111/ecog.04937

Subject Editor: Susanne Fritz

Editor-in-Chief: Miguel Araújo

Accepted 15 June 2020



The shape of a phylogenetic tree is defined by the sequence of speciation events, represented by its branching points, and extinctions, represented by branch interruptions. In a neutral scenario of parapatry and isolation by distance, species tend to branch off the original population one after the other, leading to highly unbalanced trees. In this case the degree of imbalance, measured by the normalized Sackin index, grows linearly with species richness. Here we claim that moderate values of imbalance for trees with large number of species can occur if the geographic distribution involves more than one deme (allopatry) and speciation is parapatric within demes. The combined values of balance (normalized Sackin index) and species richness provide an estimate of how many demes were involved in the process if it happened in such neutral scenario. We also show that the spatial division in demes moderately slows down the diversification process, portraying a neutral mechanism for structuring the branch length distribution of phylogenetic trees.

Keywords: neutral theory, phylogeography, Sackin index, speciation rate, Yule model

Introduction

Models of evolution have been instrumental to understand how the patterns observed in phylogenetic trees were shaped by ecological processes (Schluter 2001, Ricklefs 2010), geographical structure (Pigot et al. 2010, Aguilée et al. 2013) and demographic effects (Rosindell et al. 2010, 2015, Davies et al. 2011). Two important structural features of phylogenetic patterns are 1) the distribution of branching times, i.e. the intervals between consecutive speciation events, and 2) the balance of the tree, i.e. the number of species in each side of its branches. Most of the known phylogenetic trees show a slowdown in the net speciation rate, displaying long branches in recent times (Nee et al. 1992, McPeck 2008, Phillimore and Price 2008, Rabosky and Lovette 2008, Morlon et al. 2010) and are more unbalanced than predicted by pure birth process (Equal Rate Markov, Yule) (Mooers 1995, Mooers and Heard 1997, Blum and François 2006, Hagen et al. 2015).

Both the branching tempo and the phylogenetic imbalance (hereafter referred as imbalance only, as opposed to balance) are expected to be affected by the geographic

mode of speciation. Branching times depend on the rates of speciation and extinction and may result in different branch sizes in the early and late observable branches. Previous studies have argued that branching times may slowdown as a consequence of the geography, since as species are subdivided geographically by vicariance, per-species rates of speciation decline and extinction rates increase as population sizes become smaller (diversity-dependent hypothesis) (Phillimore and Price 2008, Rabosky and Lovette 2008). Other causes have also been pointed out as decelerators of diversification, such as competition for limited resources, which decreases diversification as resources become rare (Nee et al. 1992, Ricklefs 2006, 2010). The inability to keep pace with the changing environment can also make non-adapted species disappear, resulting in fewer and fewer species as time passes (Quental and Marshall 2013). Failure to detect species in recent times as well as incipient events of speciation that occurred and left no signals (protracted speciation) may also indicate a slowdown in the diversification process (Rosindell et al. 2010, Etienne and Rosindell 2012).

The imbalance of phylogenetic trees, on the other hand, is mainly explained as a result of species evolving locally in space as a consequence of ecological niches and/or spatial heterogeneity. Key innovations allow species to occupy new niche spaces, leading to a higher diversification in a given clade compared to others, resulting in disequilibrium of species numbers in each branch, i.e. in more unbalanced trees (Purvis et al. 2011). If space is heterogeneous, populations split into sub-populations that adapt locally, also leading to unbalanced trees (Gascuel et al. 2015). In a neutral scenario, i.e. when fitness, differential reproduction/survival and selection are not important, isolation by distance may cause the splitting of a population into different species over space (Davies et al. 2011). The sequential branching of new species from the original population in conjunction with partition of geographic space leads to highly unbalanced trees. As a result, the Sackin index, a common measure of imbalance, grows with the square of the number of species, and not linearly as in neutral models such as the Yule model. Other processes have also been suggested as causes of tree imbalance, such as the allopatric (populations separated by the presence of barriers in the spatial structure) and peripatric (speciation occurring in an isolated peripheral population) geographic modes of speciation (Losos and Glor 2003, Pigot et al. 2010) and the emergence of traits that promote speciation (Mooers and Heard 1997, Magnuson-Ford and Otto 2012). Taxonomic level is also an important ingredient in tree balance, as the higher the taxonomic level, the higher the balance of the phylogenetic tree (Mooers 1995).

In this paper we study tree balance and branch time properties in evolutionary radiations using a neutral model in which individuals are defined by a binary genome (de Aguiar et al. 2009). Individuals reproduce sexually in a mating range and the species definition is based on reproductive isolation. Genetic diversity, generated by recombination and mutation, correlates with the spatial dimension (of mating

range) (Baptistini et al. 2013a, b, Martins et al. 2013). As a result, individuals that are far apart tend to be genetically different, forming species in spatial clusters of genetically similar individuals (de Aguiar et al. 2009, Baptistini et al. 2013a). Previous results using the same modelling framework (Costa et al. 2019) have shown that the mating range significantly affects the branching tempo: larger mating ranges (when individuals use the same geographic region, corresponding to sympatry) lead to an increase in the rate of speciation throughout the radiation process. These results also indicate that tree imbalance, as measured by the normalized Sackin index, increases linearly with species richness as a result of parapatry (i.e. populations are not separated by geographical barriers but have limited usage of the available space), implying that trees with a large number of species become unrealistically unbalanced.

Here, we tested if allopatry, i.e. the presence of barriers in the spatial structure, also affects the topology of the phylogenetic trees resulting from the neutral model. Since increasing the geographic complexity from sympatry to parapatry has been shown not to affect the balance of phylogenetic trees (Costa et al. 2019), we would expect a similar result when altering the geographic structure by dividing space into demes. Deme, in our definition, is a spatially segregated area where individuals live, reproduce and evolve. Demes can be connected through a connection where interchange of individuals between demes is possible. The arrangement is similar to islands, but the connection and barriers are juxtaposed, as described in detail below. Following the previous results (Costa et al. 2019), our expectation is that the branching tempo might speed up when increasing the geographic complexity by adding the barriers. We compare the simulations under a neutral model to 10 empirical phylogenetic trees related to adaptive and non-adaptive radiations available in the literature. We look for understanding if the geographic mode of speciation attributed to these empirical trees can be detected on the structural properties of simulated trees under different geographic scenarios. We would expect the non-adaptive radiation trees to present a topological structure similar to that of simulated trees with the respective geographic scenario. For adaptive radiation trees, we also expect the same result if the adaptive response is also linked to the environmental structure.

Material and methods

We performed simulations of population evolution using an individual based model (IBM), described in the next subsection. The population was arranged into weakly connected demes and we simulated spatial structures consisting in 1, 2, 3 or 4 demes (see the second subsection Geographic scenarios), focusing on the properties of the corresponding phylogenetic trees in each scenario. Next we describe the radiation of 10 groups of species that form the empirical set of phylogenetic trees used to compare with our simulations (subsection

Empirical phylogenetic radiations). Finally we describe the metrics used to characterize the topology of trees (empirical and arising from the simulations) in subsection Structural characterization of simulated and empirical phylogenies. More information about the methods we used in this study can be found in the Supplementary material Appendix 1.

Individual based model

Here we use the IBM of speciation proposed in Costa et al. (2019), which in turn is a modified version of the model by de Aguiar et al. (2009). We included this brief explanation about the model to make the present study easier to understand on its own. We consider a population of M haploid individuals randomly distributed in a spatial area consisting of discrete lattice sites. Because we are interested in simulating more than one deme we use reflective boundary conditions, instead of the periodic conditions used in Costa et al. (2019). No significant differences in the phylogenetic tree metrics were found between periodic and non-periodic boundary conditions (Supplementary material Appendix 1 Fig. A1). The overall shape of the living area varies from a square lattice, representing a single deme, to divisions of the square lattice by barriers to simulate two or more demes. Specifically, we consider four geographic scenarios, representing one, two, three and four demes, which are weakly connected as shown in Fig. 1 (see more in the Geographic configurations subsection below). The size of the gap where demes are connected is 0.9 of the diameter size of each individual neighborhood (parameter S , see below). Time is counted in number of generations, which are non-overlapping.

The genome of individual i is represented by a binary string of size B , $(\sigma_1^i, \sigma_2^i, \dots, \sigma_B^i)$, where σ_k^i can assume the allele values 0 or 1. The genetic distance $d^{(i,j)}$ between two individuals i and j is the Hamming distance between the corresponding sequences and measures the number of genes bearing different alleles:

$$d^{(i,j)} = \sum_{k=1}^B |\sigma_k^i - \sigma_k^j|$$

Individual i can only choose as mating partner those inside a circular neighborhood of radius S centered in its spatial location (the mating range), and whose genetic distance satisfies $d^{(i,j)} \leq G$ (the genetic threshold of reproductive isolation). Focal individual i mates with one of the compatible individuals within the mating range, chosen randomly with the same probability. For the simulations, we used $S=5$, $B=150$ and $G=0.05B$. Each focal individual has a chance of reproducing, but there is a probability Q that it will not do so. If this is the case, another individual inside the mating range of this non-successful mating individual is randomly chosen to reproduce in its place, keeping the population size constant. In our simulations, we set $Q=0.37 \approx e^{-1}$, which corresponds approximately to the probability of an individual

not to be selected in M trials with replacement, in accordance with the Derrida–Higgs model (Higgs and Derrida 1991, de Aguiar 2017).

During mate choice, the number of individuals in the mating range S may be very low as a consequence of spatial distribution. To avoid this situation, we follow the procedure introduced in de Aguiar et al. (2009): if the number of possible mates (genetically compatible) in the range is smaller than a set value P ($P=3$ in our simulations), the individual searches in a radius of $S+1$. If the number of compatible mates is still smaller than P , the process is repeated once more to $S+2$, and if there is still less than P potential mates, another neighbor is randomly selected to reproduce in its place. If the neighborhood is empty, which might happen for small values of S , reproduction fails and the population decreases by 1. In this case random individuals of the next generation will have the opportunity to have a second offspring in order to increase the population back to the specified value M . Therefore, for small mating neighborhoods the population size might fluctuate slightly around M .

Reproduction is sexual and the offspring inherits, locus by locus, the allele of either parent with equal probability. For simplicity, we consider the individuals hermaphrodite, so that there are no sex incompatibilities between parents. The reproductive process is repeated until all M individuals had a chance to reproduce. After reproduction, the offspring is also subjected to mutations with probability μ per locus ($\mu=0.00025$ in our simulations).

The offspring generated is either positioned at the location of the focal parental or it disperses with probability D (here we set $D=0.02$) to one of the 20 nearest sites. Because we set $S=5$ in our simulations, the mating range of the focal parent has approximately 78 sites (πS^2). Therefore when dispersal happens, the offspring is positioned inside the mating range of the focal parent. This guarantees that close to the location of every individual of the previous generation there will be an individual in the present generation, avoiding to form clusters of individuals in the space.

We identify species as groups of individuals reproductively isolated from all others by the genetic threshold on mating defined by parameter G . Notice that not all members of the group have to be able to mate with each other, but could maintain an indirect gene flow through an intermediary individual (de Aguiar et al. 2009, Costa et al. 2019). No other conditions but the genetic are imposed on the members of a species. Individuals are genetically identical at the beginning of the simulation, but mutations and recombination create diversity that might lead to the formation of species. After a transient time, speciation and extinction events reach an equilibrium and the number of species remains approximately constant. In all cases studied here the populations were evaluated at the equilibration time, so as to describe the process of radiation from a given population inhabiting a single area.

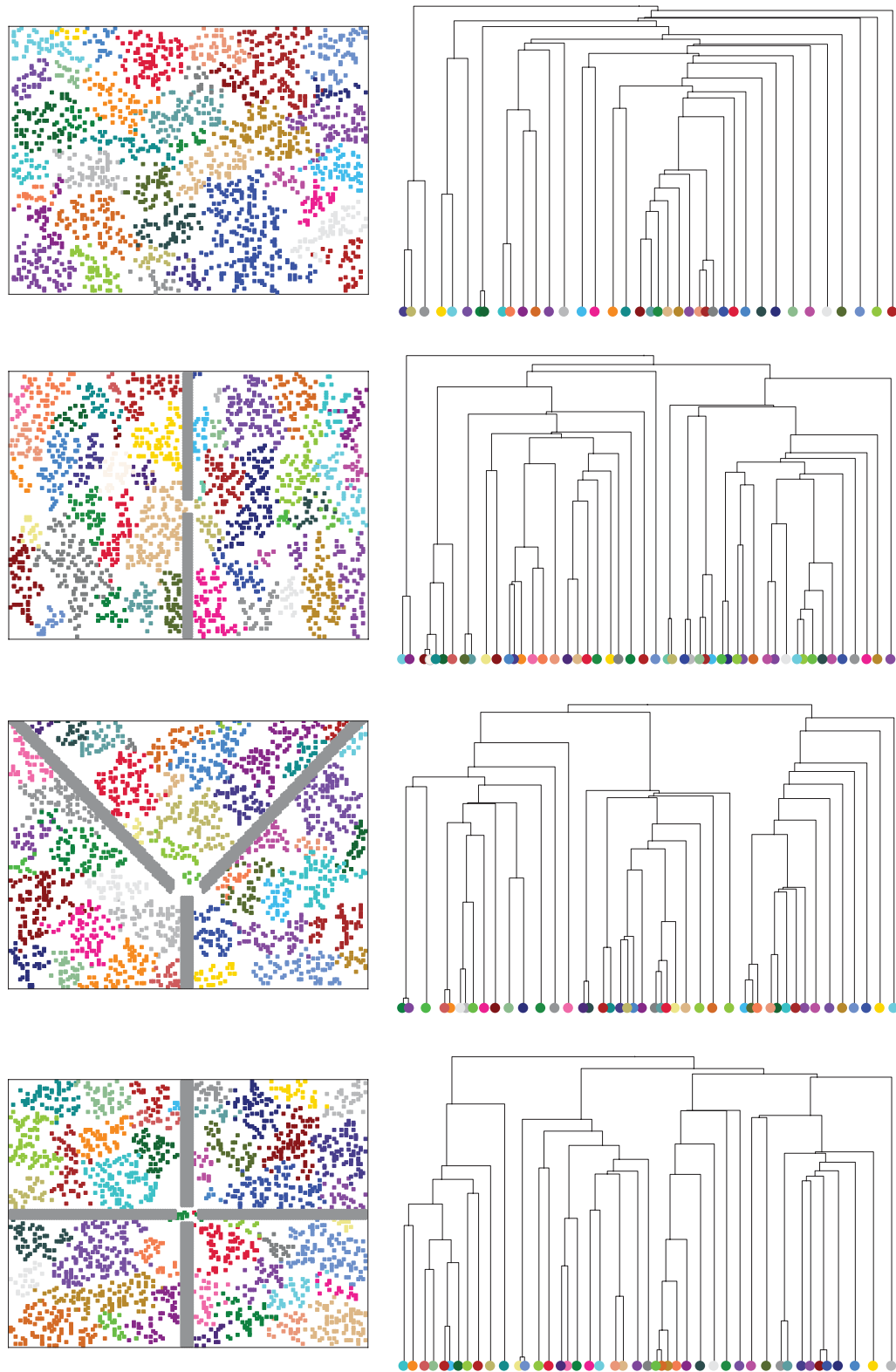


Figure 1. Geographical configurations. Scenarios with 1, 2, 3 and 4 demes divided by barriers. Demes are connected at the barriers gap, so that the initial population is a single species throughout the landscape. Species are drawn in different colors for one simulation and the corresponding tree is displayed on the right.

Geographic scenarios

Figure 1 shows the four geographical configurations considered here with an example of the final population configuration for $M=2000$ individuals, along with the corresponding phylogenetic tree (see more about how we build the phylogenetic trees from the simulation in the Supplementary material Appendix 1). In each case all demes have the same area and are separated by barriers that cannot be transposed by individuals (barrier width is larger than S). The number of large clades on the phylogenetic trees is equal to the number of demes. The size of the area connecting the demes was kept constant for all cases. Increasing the size of the gap, and thus increasing the connectivity between each deme, the resulting phylogenetic trees are more unbalanced (Supplementary material Appendix 1 Fig. A2).

We considered different population sizes ($M=\{250, 500, 1000, 2000, 3000, 4000\}$) for each geographical configuration, adjusting the total area to keep the average density of individuals constant. Because of the geographic structure limitations, the populations size $M=250$ was simulated only for the 1 deme scenario and $M=500$ was performed for both 1 and 2 demes scenarios. We ran 100 replicas for each geographic scenario (1, 2, 3 and 4 demes) and each value of population size (M), up to equilibration time. The equilibration time is defined as the time when the number of species stops to increase. The values of equilibration time used in our simulations are presented in Supplementary material Appendix 1 Table A1. For each simulation we computed the phylogenetic tree and its corresponding metrics of tree topology (Sackin index and α -value, see Supplementary material Appendix 1 for more details).

It is expected that after the equilibration on species richness, i.e. when average extinction and speciation rates are equal, the balance will also reach an equilibrium, resulting in trees as balanced as those predicted by the Yule model after enough time is given for the process. To show the effect of the geographic structure after equilibration time, we ran 50 replicas of each deme configuration for $M=2000$ individuals and 10 000 generations.

Empirical phylogenetic radiations

We compared our simulations with 10 groups of species (Table 1) whose phylogenetic trees were classified in the literature according to their corresponding process of radiation: adaptive (points number 1–7) and non-adaptive (points number 8–10) radiations. We considered only trees with low or intermediary gene flow, which are expected for allopatric diversification. We used the same qualitative level of gene flow and adaptiveness classifications as defined in the study (Costa et al. 2019).

Some of the empirical phylogenetic trees have a strong correlation with the spatial structure. For instance, the Silverwood alliance (data point number 1) presents four main clades (*Wilkisia* + *Dubatia*, *Dubatia* (K+), *Dubatia* (OMH) and *Argyroxiphium*) which can be associated to the hawaiian islands: *Wilkisia* + *Dubatia* exclusively present at Kaua'i, the *Dubatia* (K+) clade predominantly present on the Kaua'i island, and finally *Argyroxiphium* and *Dubatia* (OMH) inhabiting the islands of Hawaii and the smaller islands Maui, O'ahu and Lana'i (Blonder et al. 2016, Landis et al. 2018).

The *Anolis* lizards (data point number 2) radiation has undergone two important colonization processes, one by southern part of the Lesser Antilles and the second to the rest of the Caribbean islands, where they initially diversified an adaptive radiation on the two larger Greater Antillean islands, coming across a secondary radiation to all islands, finally reaching the continent in a subsequent event (Alföldi et al. 2011). This process resulted into two main clades linked to the geographic space: one clade associated to the islands and another one to the continent.

A relationship between clade and islands is also remarkable among the *Tretagnatha* spider (data point number 3). The clades are related to the four main islands of Hawaii: Hawaii Island, Maui, Kaua'i and O'ahu (Gillespie 2004). Similar ecomorphs are often distributed among islands, with independent origins on different islands, characterizing this radiation (Gillespie 2004).

Secondary contact and hybridization among Darwin's finches (data point number 4) may have erased the signal of any possible allopatric speciation, while some degree of

Table 1. Empirical radiations. The point number identifying the tree in this table is used in Fig. 3 (Results section). N_s is the number of species (tips) in each empirical tree.

Point number	N_s	Species	Gene flow during diversification	Radiation	Reference
1	35	Hawaiian silversword alliance	intermediate	Adaptive	Blonder et al. 2016
2	71	Caribbean <i>Anolis</i> lizards	intermediate	Adaptive	Alföldi et al. 2011
3	25	<i>Tretagnatha</i> spiders	intermediate	Adaptive	Gillespie et al. 2004*
4	14	Darwin's finches	intermediate	Adaptive	Lamichhaney et al. 2015 Clarke et al. 2017
5	44	Tanganyika cichlids – 1	low	Adaptive	Meyer et al. 2015
6	40	Tanganyika cichlids – 2	low	Adaptive	McGee et al. 2016
7	16	Malawi cichlids	low	Adaptive	McGee et al. 2016
8	68	Lichen <i>Sticta</i>	intermediate	Non-adaptive	Simon et al. 2018
9	46	Caviomorph rodents <i>Ctenomys</i>	low	Non-adaptive	Álvarez et al. 2017
10	53	New World titi monkeys	low	Non-adaptive	Byrne et al. 2016

* Personal communication.

allopatry cannot be detected in phylogenies because individuals presenting some degree of allopatry are still classified as the same species (Grant et al. 2000, Lamichhaney et al. 2015, Clarke et al. 2017).

Malawi cichlids, Tanganyika cichlids-1 and 2 (data point number 5, 6 and 7, respectively) are phylogenetically grouped by the feeding mode these fish species use: suction and biting (McGee et al. 2016). These modes are also related to the use of the space in aquatic systems. The suction mode of feeding is typical of pelagic fish, which live and feed in the water column. The biting mode is typical of shallow water areas, where molluscs, algae or other fish fins are available (McGee et al. 2016).

The evolutionary history of lichen *Sticta* (data point number 8) strongly supports a local radiation starting concomitantly with the emergence of the Mascarene archipelago (Simon et al. 2018). A likely single dispersal event and subsequent migrations between islands (Madagascar, Mauritius and Réunion) resulted in the observed complex of small range endemic species. Five main lineages can be distinguished and are distributed in three or more geographic regions (Simon et al. 2018). Two of them are endemic to the Mascarene archipelago. One lineage is endemic to Madagascar. The other two lineages, in addition to being more widespread throughout the island complex, contain species that are endemic to only one or two islands.

Clades of *Ctenomys* rodents (data point number 9) are related to major regions in South America (Álvarez et al. 2017, Caraballo and Rossi 2018). The greater region, which comprises the group of *Ctenomys torquatus*, was likely invaded by *Ctenomys* spp of other regions on recent years. After this secondary contact, at least three main regions still persist and they are related to the clades: *C. perasoni* clade, *C. minutus* clade and the Corrientes group clade (Caraballo and Rossi 2018).

Structural characterization of simulated and empirical phylogenies

Phylogenetic trees for the evolutionary model described in the Individual based model subsection can be constructed using different techniques (Costa et al. 2018a). In the Supplementary material Appendix 1 we detail how to build the corresponding phylogenetic trees. To characterize the tree topology of the trees resulting from the simulations and also the empirical phylogenetic trees (see in the previous subsection), we used two metrics. These metrics give information about the balance and the branching tempo. For the branching tempo, we estimated the acceleration of the diversification process using the α -value, a metric derived from the γ -statistics. We detail in the Supplementary material Appendix 1 how γ -statistics and α -value are related and computed. Positive values of α -value correspond to tippy trees, with speciation events accumulating close to the leaves and corresponding to a speed up of the diversification process. On the other hand, negative values indicate steammy trees, with most speciation occurring close to the root, corresponding

to slowdown of the diversification process. The α -value, in opposition to γ -statistics (McPeck 2008, Phillimore and Price 2008), has no dependency on the number of species of the tree, allowing direct comparisons between trees of different sizes.

For the balance metric we used the Sackin index I . The average value of the Sackin index for the stochastic Yule model, $E(I)$, is proportional to the number R of species (tips) of the tree and the variance σ_R^2 to R^2 (Cardona et al. 2013). Therefore, to compare trees with different number of tips, we work with the normalized Sackin index,

$$I_n(R) = \frac{I(R) - E[I(R)]}{\sqrt{\sigma_R^2}}$$

Although $I_n(R)$ would be close to zero for trees generated with the Yule model, independent of the species richness R , in the parapatric model $I(R)$ increases with R^2 because of the sequential branching of species from the spatially distributed population. In this case I_n itself grows linearly with R , producing unrealistic values of imbalance for large species richness (see details in the Supplementary material Appendix 1).

Results

The results show significant increase of the normalized Sackin index with population size (and consequently with number of species – Supplementary material Appendix 1 Fig. A3) at the equilibration time. In contrast, the average α -value changed very little for all deme configurations (Fig. 2). In general, for all tested populations sizes, the number of demes had a significant impact on both average metrics – $\langle I_n \rangle$ and $\langle \alpha \text{-value} \rangle$ (ANOVA and Tukey's tests for both metrics against the number of demes, respectively run for each population size $M = \{1000, 2000, 3000, 4000\}$, Supplementary material Appendix 1 Table A2, A3, Fig. A4, A5). Figure 3 shows how the average values of tree metrics depend on the number of demes, which can be seen to have little effect on the diversification acceleration (α -value).

We found a linear relationship between the normalized Sackin index (I_n) and the number of species at the equilibration time (Fig. 2 and Supplementary material Appendix 1 Table A4 for linear regression results), following the behavior predicted by Eq. A8 in Supplementary material Appendix 1 for highly unbalanced trees. We also used the β value as a test of tree balance (Aldous 1996, Blum and François 2006) as it is known to be invariant under the number of tips in the tree. Indeed, the balance of the trees calculated via β value also increased with the number of demes (Supplementary material Appendix 1 Fig. A6, Table A5–A6 for ANOVA and Tukey's test). For the normalized Sackin index (I_n), we show the positive relationship through the linear regression by the number of species. However, the curves differ in their slope for different number of demes (Fig. 2, check R^2 values at Supplementary material Appendix 1 Table A4). For a

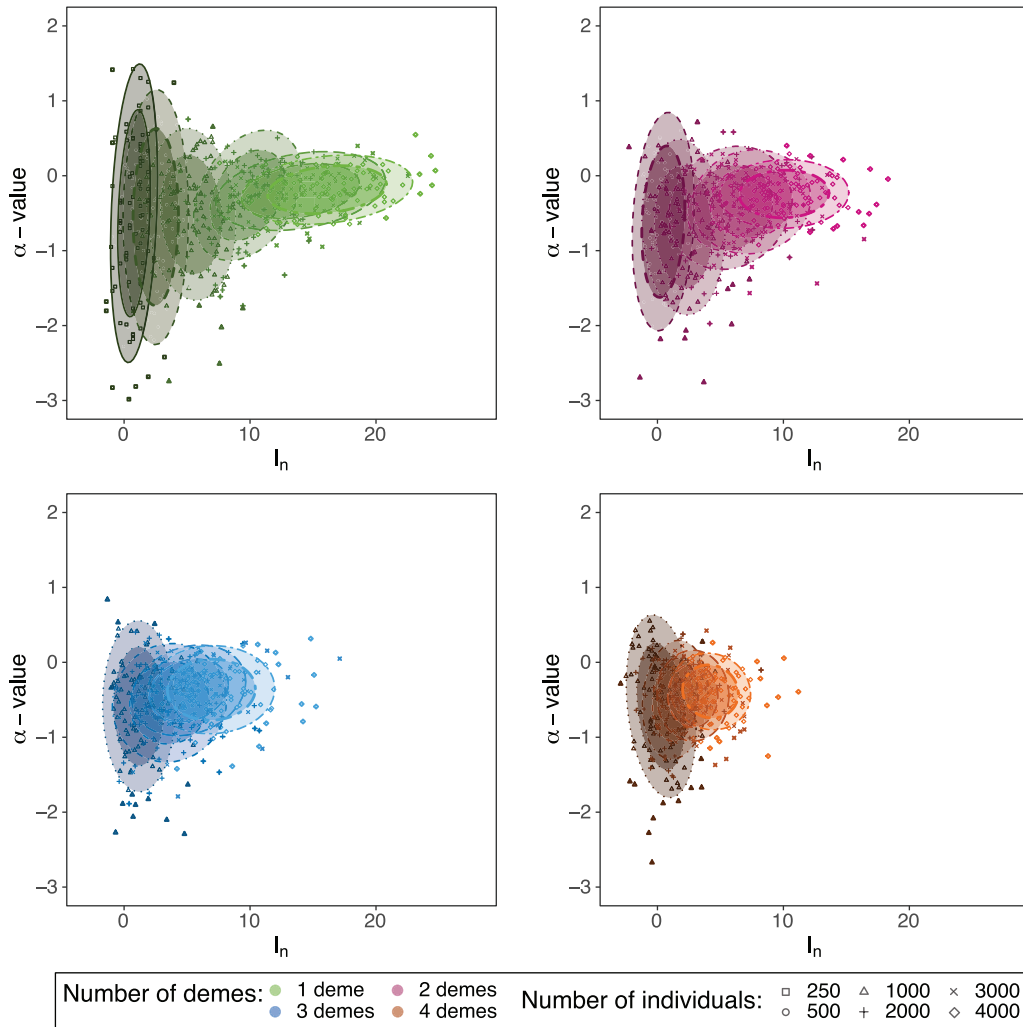


Figure 2. The relationship between phylogenetic tree metrics for the four deme configurations: the tree balance, measured by the normalized Sackin index (I_n) on the x-axis, and the branch length metric, measured by the α -value on the y-axis. A total of 100 runs were performed for each geographic configuration (number of demes) and number of individuals. We present the results for different population sizes (different symbols, see legend) and different number of demes (different plots: up-left (green): 1 deme, up-right (pink): 2 demes, down-left (blue): 3 demes, and down-right (orange): 4 demes). To help distinguishing the different populations sizes (different symbols), we make colors darker for smaller population sizes. The ellipses represent 63% (inner) and 87% (outer) of the variation in both tree metrics.

single deme, for example, clades with $N=50$ species should have $I_n \approx 10$, which seems unrealistic for empirical phylogenetic trees (Blum and François 2005). Empirically reasonable values of I_n for $N=50$ in our model are obtained for geographical configurations involving more than one deme. The number of species had weak effect on the α -value for different spatial configurations (see results of linear regression in Supplementary material Appendix 1 Table A4), with an average α -value between 0 and -1 .

To compare our results with empirical data, we selected 10 empirical radiations (see Table 1 for further information). The empirical phylogenetic trees are represented in the plots with different symbols according to the corresponding radiation processes being adaptive or non-adaptive and to the gene flow level (low and medium) during radiation (color of the symbol). All the empirical trees' metrics are below the curve

predicted for one deme in their Sackin index (and only two of them are above 2 demes).

In the long term, i.e. many generations after the radiation process ended, the α -value increases with an apparent attenuation after 5000 generations for every configuration of demes with $M=2000$ individuals (Fig. 4, first plot). We found that the signature left by the number of demes on the normalized Sackin index diminishes in the long term, tending towards zero (balanced) for all configurations of demes for $M=2000$ (Fig. 4, second plot). The number of extant species shows an overshooting in the number of species close to the equilibration time and stabilizes around 35 species in this case (Fig. 4, third plot). The number of species and Sackin index in the 10 000th generation (the last one in our simulations) show significant difference among demes in the 10 000th generation, although the simulated trees under 2, 3 and 4 demes can not

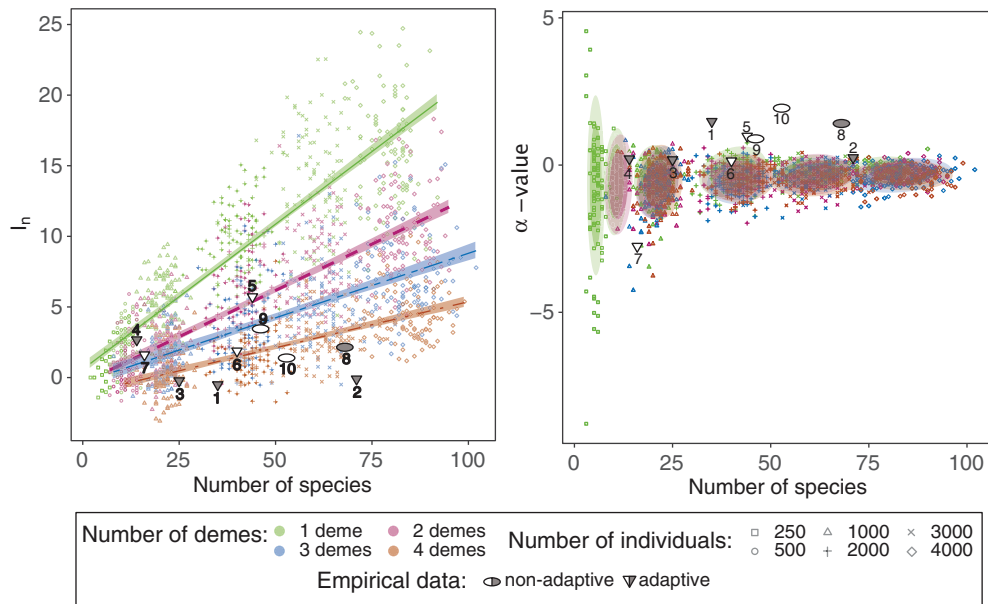


Figure 3. Normalized Sackin index (I_n) and α -value for different geographic configurations (colors) and species richness (R). The slope represent the linear regression of the I_n by number of species. On the right, the ellipses represent 63% (inner) and 87% (outer) of the variation on α -value and number of species. The different symbol shapes represent the differing number of individuals. A total of 100 runs were performed for each geographic configuration (number of demes) and number of individuals. The empirical phylogenetic trees are classified according to the corresponding radiation process: adaptive and non-adaptive radiation. Symbols are filled by gray and white according to the genetic flow: white is low and gray is medium genetic flow. Labels for the empirical trees (numbers) can be found at Table 1.

be differentiated by their normalized Sackin index and number of species formed (Supplementary material Appendix 1 Table A7–A8).

Discussion

Our simulations show that speciation in an allopatric scenario depends on the spatial structure of the population. Here we have considered a scheme where the population is divided into equally large demes which are weakly connected. Our three main results are 1) the branch lengths distribution is independent of the geographic structure (α -values change marginally according to the number of demes), 2) the trees become more balanced as the number of demes increases (the normalized Sackin index decreases with higher number of demes), displaying one large clade per deme and 3) in the long term the differences between the differing geographic configurations vanish and the balance of trees reach the value expected by the Yule model (or the Equal Rate Markov process). In opposition to the expectations based on augment of geographic complexity, points 1) and 2) show that increasing the number of demes (this work) is different from the complexity of the sympatry–parapatry dualism (as discussed in Costa et al. 2019).

Studies that evaluate the diversification tempo in empirical phylogenies have found evidence of slowdowns in diversification rates for several different groups (Phillimore and Price 2008, Rabosky and Lovette 2008). This slowdown of diversification rates has been associated to geographic

structuring and niche filling, leading to the saturation of the accumulation curves for species richness as the available niches are taken (Phillimore and Price 2008, Rabosky and Lovette 2008). We observed a weak dependency of the average α -values according to the number of demes and individuals of the population, and the variance of the distribution decreased with the number of individuals (notice that there is no trend on the y-axis on Fig. 2 and the ellipses overlap on Fig. 2, right panel). Thus our results indicate that, under a neutral assumption, diversification tempo may be mainly determined by microevolutionary forces during the radiation, such as mutation rate. However, these small effects on the α -values point to a slowdown of the diversification rate under a neutral speciation process (Supplementary material Appendix 1 Fig. A5 with smaller α -values for many demes (3 and 4) when compared with few demes (1 and 2)). We believe further analysis under higher spatial complexity might result in more notable diversification slowdown, indicating that ecological causes would not be the only mechanism of deceleration in phylogenetic trees. In Costa et al. (2019), the authors discuss the dependency of this metric (α -value) on genome size and mating range, which are directly linked to microevolutionary forces. The decrease in variance of α -value, which can be noted by smaller ellipses in Fig. 2 is then a direct effect of the increase in the number of individuals (and consequently as a number of species), and should not depend on the number of demes in space.

Contradicting previous models of peripatric and allopatric speciation, showing that trees tend to become more

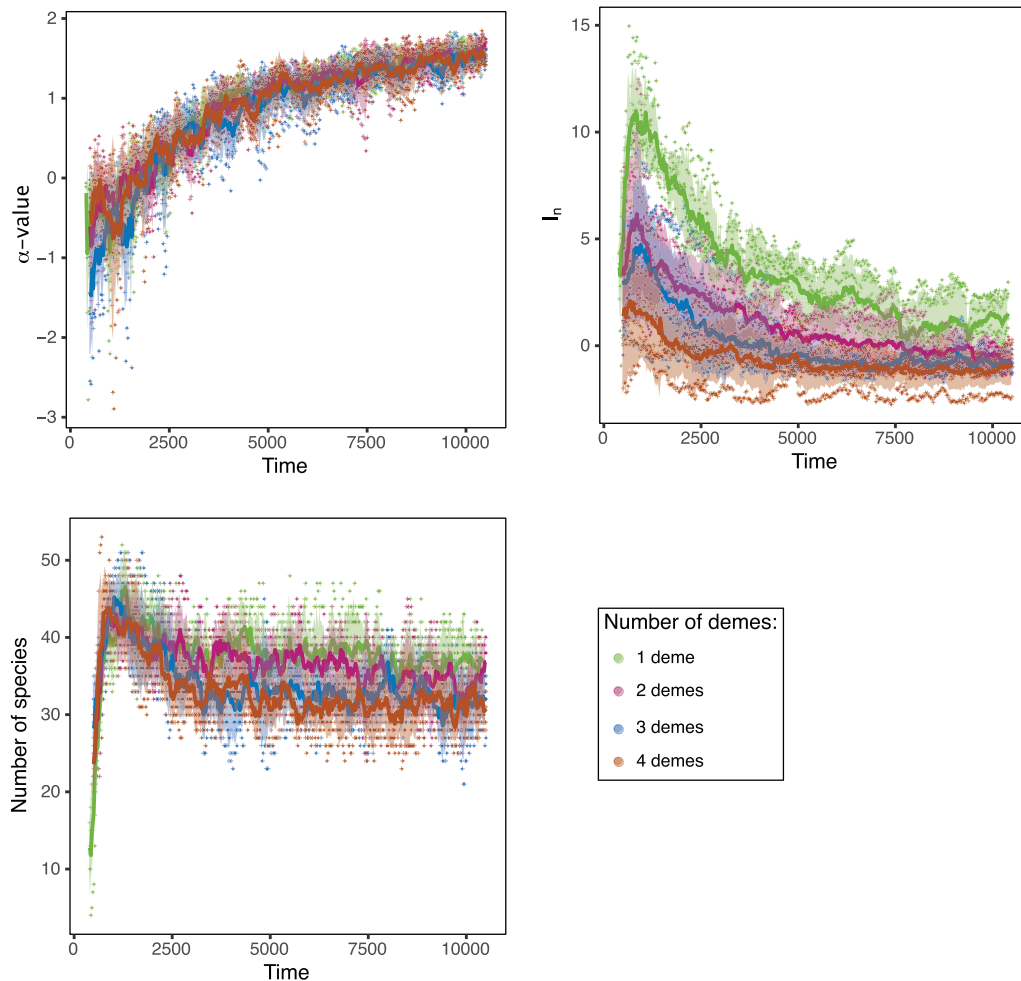


Figure 4. Tree indices and species richness evolution in long term. Normalized Sackin index (I_n), α -values and species number in this figure correspond to 50 runs made for populations of 2000 individuals in the four different geographic scenarios.

unbalanced according to geographic structure (Losos and Glor 2003, Pigot et al. 2010, Davies et al. 2011), our results show phylogenetic trees becoming more balanced as space is structured with more demes (Fig. 2, left panel). By analysing balance as a function of the number of species (tips) in the trees, we show that empirical phylogenetic radiation trees are more balanced than expected by the one deme configuration with the same number of species in equilibrium. Trees with large number of species, in particular, are far from the predicted normalized Sackin index according to non-structured populations. These results are coherent when taking into account a direct relationship between the main clades in the phylogenetic tree and the structure of the space (here, demes). We argue that if deme sizes are similar and the same parapatric process happens in each deme, we should expect more balanced trees than in a single deme. We first note that the phylogenetic trees in k demes display k similar sub-trees corresponding to the clades in each location. Since these sub-trees are connected by the root, they confer the entire tree a symmetry at this first level. Because allopatry usually occurs on large regions, if the regions are under the same

evolutionary conditions in all different parts of the space, we should expect more balanced trees with the increase in the structuring of the space. In contrast, a greater imbalance should be expected as a result of the local usage of the space by individuals in only one deme, i.e. a more parapatric mode of speciation. This might happen because there are no guarantees that sister species will give rise to the same number of descendant lineages.

Although we have not simulated scenarios with demes of different sizes, we may infer the consequences of such outline based on our results: for instance, considering a scenario with two demes, with one large and one small, the balance of the resulting phylogenetic tree would be between the one deme and the two demes configurations for that number of species. In our Fig. 2 (left panel), this means that the expected balance would be between the green (continuous line) and the pink (simple dashed line) regressions for a fixed value of number of species (x-axis). In a three demes scenario with two large demes and one smaller, the expected balance would be between the pink (simple dashed line) and the blue (two-dashed line). In other words, the largest deme (or the prevalent

demes scenario) can inform us about the expected balance under a neutral process. Nonetheless, apart from variation in deme sizes, we can not draw strong conclusions about the consequences of different degrees of connectivity between the different demes at this point. Our model, however, provides means to analyse the signatures left in the phylogenetic trees by different migration rates, under a source–sink islands scenario or under continent–islands geographic configurations. Moreover, the transition between local parapatry (this study) and local sympatry, in a regional allopatric scenario can be evaluated in a future work.

Tree imbalance indeed varied considerably in our model, with trees becoming more balanced under increasing spatial compartmentalization. Previous results by Gascuel et al. (2015) show an increase in tree balance as the landscape dynamics increase in pace, similar to our results for when space is divided into demes. However, these authors predict unbalanced trees in the end of the radiation process only for the scenario of wide competition, which is the most non-neutral of their scenarios. Here we show that unbalanced trees may be obtained even under a neutral scenario, reinforcing that ecological processes are not strictly necessary to explain tree imbalance, as Davies and collaborators have shown previously (Davies et al. 2011). In fact, the differences in branching probabilities that appear between leaves in the tree, which ultimately leads to tree imbalance, is a result of different speciation rates associated to differences in the number of individuals of each species, which emerges in our model but is directly modeled by Davies et al. (2011). During the parapatric process, speciation proceeds as the segregation of a group of individuals (due to spatial and genetic constraints) from the initial founder population. The segregated group, usually smaller in size, will most likely not speciate again before the end of the radiation process.

Additionally, we show that theoretical phylogenetic trees become more balanced a long time after the equilibrium number of species number is reached (see also de Aguiar 2017). Although it is still possible to notice small differences on the normalized Sackin index for different numbers of demes (phylogenetic trees in one deme have a higher Sackin index than in four demes), the trend seems to be balanced trees, according to the Yule model ($I_n = 0$), for very long times. This can be explained again by purely demographic effects. After reaching equilibrium species richness, the distribution of population sizes for different species tends to be more equitable, leading to fairly similar speciation rates for each species (and, as such, equal branching probabilities for different tree tips). Thus, if enough time after radiation is considered, the balance of the tree can be reasonably described by the Yule model. This similar trend in long time simulations for all geographic configurations shows how signatures of the radiation process can vanish (no long time differences appear among 2, 3 and 4 demes) after the equilibration of the rates of speciation and extinction. Thus, phylogenetic trees which had reached equilibrium species richness for a long time may

be uninformative about spatial compartmentalization in the process of speciation.

We have also compared our results to the tree statistics of ten empirical radiations, relating structure of these trees to the spatial configuration experimented during the evolutionary process according to the literature. For some of the empirical data, our model points to the right number of demes (islands or main regions) according to the balancing and number of species found. For instance, clades of *Ctenomys* rodents (data point number 9) are present in three main regions of South America (Caraballo and Rossi 2018) and this number of main regions coincides with the number of demes predicted by our model. The evolutionary history of lichen *Sticta* (data point number 8) has five main lineages which are distributed in three or more geographic regions (Simon et al. 2018), and this geographic configuration agrees with the results of our model, where we predict lichen evolution has likely happened under three or more demes. Hawaiian species are also an example of how well the spatial evolutionary history could be predicted by our model. Both the Silverwood alliance (data point number 1) and the *Tretagnatha* spider (data point number 3) present four main clades (associated with hawaiian islands), which coincides with the four demes predicted by our model for both of them.

In some cases, our model is a bad predictor of the number of demes, as islands or regions upon which speciation has occurred. This is the case for the finches, whose clades can not be related to any geographic regions today (Lamichhaney et al. 2015, Clarke et al. 2017), but according to our model, finches were predicted to be the result of an evolutionary process which happened under two demes. It is important to note, however, that the number of finch species is small and our model is less accurate for smaller numbers of species. Our model also seems to be a poor predictor of the number of regions under which *Anolis* lizards (data point number 2) evolved. These lizards have two main clades linked to the geographic space (Alföldi et al. 2011), while we observe an imbalance expected for four demes in our model. Finally, our model also predicts the corrected number of demes that would be expected for cichlids in the Malawi tree (data point number 7) and also for the Tanganyika cichlids-1 (data point number 5), which is two clades according to the use of space for feeding and the non-expected number of demes for Tanganyika cichlids-2 – four demes. However, the Tanganyika cichlids are not grouped in their respective phylogeny according to their feeding mode or use of space (McGee et al. 2016).

To conclude, we believe our comparisons of empirical data and theoretical models can guide researchers to look for the geographic mode of speciation a certain group of related species has experienced. Here we essentially analysed the evolutionary process until equilibration time with an initial speciation rate higher than the extinction rate (radiation process), and therefore the signatures found describe only this process under a neutral model. We hypothesize that the

signatures left when the extinction rate is higher than the speciation rate are different, and this could inform further studies about biotic crises (Heard and Mooers 2002, Jablonski 2005, Crisp and Cook 2009). The balance of these trees is expected to decrease with the extinction of endangered species, specially those that are closely related and have emerged in the same deme. Future research could predict how strong local extinctions are based on the expected balance decrement according to the number of species and demes.

Data availability statement

Data available from the Dryad Digital Repository: <<https://datadryad.org/stash/dataset/doi:10.5061/dryad.4kn5j5d>> (Costa et al. 2018b).

Acknowledgements – The authors thank Espaço da Escrita – Pró-Reitoria de Pesquisa – UNICAMP – for the language services provided.

Funding – This work was partly funded by the Brazilian agencies: Capes, grant no. 301082/2019-7 (CNPq), no. 2016/01343-7 (ICTP-SAI FR FAPESP) and no. 2019/20271-5 (FAPESP). FMDM is supported by grant no. 2015/11985-3 (FAPESP).

References

- Aguilée, R. et al. 2013. Adaptive radiation driven by the interplay of eco-evolutionary and land-scape dynamics. – *Evolution* 67: 1291–1306.
- Aldous, D. 1996. Probability distributions on cladograms. – In: Aldous, D. and Pemantle, R. (eds), *Random discrete structures*. Springer, pp. 1–18.
- Alföldi, J. et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. – *Nature* 477: 587.
- Álvarez, A. et al. 2017. Diversification patterns and size evolution in caviomorph rodents. – *Biol. J. Linn. Soc.* 121: 907–922.
- Baptistini, E. M. et al. 2013a. Conditions for neutral speciation via isolation by distance. – *J. Theor. Biol.* 335: 51–56.
- Baptistini, E. M. et al. 2013b. The role of sex separation in neutral speciation. – *Theor. Ecol.* 6: 213–223.
- Blonder, B. et al. 2016. Variation and macroevolution in leaf functional traits in the Hawaiian silversword alliance (Asteraceae). – *J. Ecol.* 104: 219–228.
- Blum, M. G. and François, O. 2005. On statistical tests of phylogenetic tree imbalance: the Sackin and other indices revisited. – *Math. Biosci.* 195: 141–153.
- Blum, M. G. and François, O. 2006. Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. – *Syst. Biol.* 55: 685–691.
- Byrne, H. et al. 2016. Phylogenetic relationships of the New World titi monkeys (*Callicebus*): first appraisal of taxonomy based on molecular evidence. – *Front. Zool.* 13: 10.
- Caraballo, D. and Rossi, M. 2018. Spatial and temporal divergence of the torquatus species group of the subterranean rodent *Ctenomys*. – *Contrib. Zool.* 87: 11–24.
- Cardona, G. et al. 2013. Exact formulas for the variance of several balance indices under the Yule model. – *J. Math. Biol.* 67: 1833–1846.
- Clarke, M. et al. 2017. Trait evolution in adaptive radiations: modeling and measuring interspecific competition on phylogenies. – *Am. Nat.* 189: 121–137.
- Costa, C. L. N. et al. 2018a. Registering the evolutionary history in individual-based models of speciation. – *Phys. A Stat. Mech. Appl.* 510: 1–14.
- Costa, C. L. N. et al. 2018b. Data from: Signatures of microevolutionary processes in phylogenetic patterns, v2. – *Dryad Digital Repository*, <<https://datadryad.org/stash/dataset/doi:10.5061/dryad.4kn5j5d>>.
- Costa, C. L. et al. 2019. Signatures of microevolutionary processes in phylogenetic patterns. – *Syst. Biol.* 68: 131–144.
- Crisp, M. D. and Cook, L. G. 2009. Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. – *Evol. Int. J. Organic Evol.* 63: 2257–2265.
- Davies, T. J. et al. 2011. Neutral biodiversity theory can explain the imbalance of phylogenetic trees but not the tempo of their diversification. – *Evolution* 65: 1841–1850.
- de Aguiar, M. A. M. 2017. Speciation in the Derrida–Higgs model with finite genomes and spatial populations. – *J. Phys. A Math. Theor.* 50: 085602.
- de Aguiar, M. A. M. et al. 2009. Global patterns of speciation and diversity. – *Nature* 460: 384–387.
- Etienne, R. S. and Rosindell, J. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. – *Syst. Biol.* 61: 204–204.
- Gascuel, F. et al. 2015. How ecology and landscape dynamics shape phylogenetic trees. – *Syst. Biol.* 64: 590–607.
- Gillespie, R. 2004. Community assembly through adaptive radiation in Hawaiian spiders. – *Science* 303: 356–359.
- Grant, P. R. et al. 2000. The allopatric phase of speciation: the sharp-beaked ground finch (*Geospiza difficilis*) on the Galápagos islands. – *Biol. J. Linn. Soc.* 69: 287–317.
- Hagen, O. et al. 2015. Age-dependent speciation can explain the shape of empirical phylogenies. – *Syst. Biol.* 64: 432–440.
- Heard, S. B. and Mooers, A. Ø. 2002. Signatures of random and selective mass extinctions in phylogenetic tree balance. – *Syst. Biol.* 51: 889–897.
- Higgs, P. G. and Derrida, B. 1991. Stochastic models for species formation in evolving populations. – *J. Phys. A Math. Gen.* 24: L985.
- Jablonski, D. 2005. Mass extinctions and macroevolution. – *Paleobiology* 31: 192–210.
- Lamichanay, S. et al. 2015. Evolution of Darwin’s finches and their beaks revealed by genome sequencing. – *Nature* 518: 371–375.
- Landis, M. J. et al. 2018. Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic and paleogeographic uncertainty. – *Evolution* 72: 2343–2359.
- Losos, J. B. and Glor, R. E. 2003. Phylogenetic comparative methods and the geography of speciation. – *Trends Ecol. Evol.* 18: 220–227.
- Magnuson-Ford, K. and Otto, S. P. 2012. Linking the investigations of character evolution and species diversification. – *Am. Nat.* 180: 225–245.
- Martins, A. B. et al. 2013. Evolution and stability of ring species. – *Proc. Natl Acad. Sci. USA* 110: 5080–5084.
- McGee, M. D. et al. 2016. Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. – *Proc. R. Soc. B* 283: 20151413.

- McPeck, M. A. 2008. The ecological dynamics of clade diversification and community assembly. – *Am. Nat.* 172: E270–E284.
- Meyer, B. S. et al. 2015. A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. – *Mol. Phylogenet. Evol.* 83: 56–71.
- Mooers, A. O. 1995. Tree balance and tree completeness. – *Evolution* 49: 379–384.
- Mooers, A. O. and Heard, S. B. 1997. Inferring evolutionary process from phylogenetic tree shape. – *Q. Rev. Biol.* 72: 31–54.
- Morlon, H. et al. 2010. Inferring the dynamics of diversification: a coalescent approach. – *PLoS Biol.* 8: e1000493.
- Nee, S. et al. 1992. Tempo and mode of evolution revealed from molecular phylogenies. – *Proc. Natl Acad. Sci. USA* 89: 8322–8326.
- Phillimore, A. B. and Price, T. D. 2008. Density-dependent cladogenesis in birds. – *PLoS Biol.* 6: e71.
- Pigot, A. L. et al. 2010. The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. – *Syst. Biol.* 59: 660–673.
- Purvis, A. et al. 2011. The shape of mammalian phylogeny: patterns, processes and scales. – *Phil. Trans. R. Soc. B* 366: 2462–2477.
- Quental, T. B. and Marshall, C. R. 2013. How the red queen drives terrestrial mammals to extinction. – *Science* 341: 290–292.
- Rabosky, D. L. and Lovette, I. J. 2008. Density-dependent diversification in North American wood warblers. – *Proc. R. Soc. B* 275: 2363–2371.
- Ricklefs, R. E. 2006. Global variation in the diversification rate of passerine birds. – *Ecology* 87: 2468–2478.
- Ricklefs, R. E. 2010. Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. – *Proc. Natl Acad. Sci. USA* 107: 1265–1272.
- Rosindell, J. et al. 2010. Protracted speciation revitalizes the neutral theory of biodiversity. – *Ecol. Lett.* 13: 716–727.
- Rosindell, J. et al. 2015. Unifying ecology and macroevolution with individual-based theory. – *Ecol. Lett.* 18: 472–482.
- Schluter, D. 2001. Ecology and the origin of species. – *Trends Ecol. Evol.* 16: 372–380.
- Simon, A. et al. 2018. High diversity, high insular endemism and recent origin in the lichen genus *Stricta* (lichenized Ascomycota, Peltigerales) in Madagascar and the Mascarenes. – *Mol. Phylogenet. Evol.* 122: 15–28.

Supplementary material (available online as Appendix ecog-04937 at <www.ecography.org/appendix/ecog-04937>). Appendix 1.