Version dated: June 12, 2015

How Ecology and Landscape Dynamics Shape Phylogenetic Trees

Fanny Gascuel[1,2,3,*], Régis Ferrière[3,4,+], Robin Aguilée[5], and Amaury Lambert[1,6,+]

[1] *Center for Interdisciplinary Research in Biology, CNRS UMR 7241, Collège de France, Paris, France;*

[2] *Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7625, Laboratoire Écologie et Évolution, Paris, France;*

[3] *Institut de Biologie de l'Ecole Normale Supérieure, CNRS UMR 8197, École Normale Supérieure, Paris, France;*

[4] *Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, USA;*

[5] *Laboratoire Évolution et Diversité Biologique, CNRS UMR 5174, Université Paul Sabatier, Toulouse, France;*

[6] *Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7599, Laboratoire Probabilités et Modèles Aléatoires, Paris, France*

[*] **Corresponding author:** Fanny Gascuel, Center for Interdisciplinary Research in Biology, Equipe Stochastic Models for the Inference of Life Evolution, Collège de France, 11 place Marcelin Berthelot, 75005 Paris, France; E-mail: fanny.gascuel@college-de-france.fr.

MODEL VERSION 2: Refining the criterion for species delineation

In Gascuel et al. (2015), our method to delineate species and follow their evolutionary history was based on the probabilities of reproductive isolation between individuals, calculated from their phenotypic and assortative mating traits. This method did not account for geography: populations living in different geographical sites were grouped into the same species if their individuals had similar phenotypes. In an additional version of our model (presented in the online Appendix 6 of Gascuel et al. 2015, available at http://dx.doi.org/10.5061/dryad.3bp51), we wanted to account for the additional assumption that if species have diverged for long time in separated geographical sites, the accumulation of genetic incompatibilities may prevent them from hybridizing, even if their phenotypic traits are identical. We therefore introduced a mechanism of long-term

irreversible isolation resulting from the accumulation of genetic incompatibilities. This mechanism prevented mating between individuals that diverged a long time before.

In the new version on the model (Version 2, introduced herein), we also account for genetic incompatibilities to delineate species and follow their evolutionary history. When genetic incompatibilities influence mating between individuals but not species delineation (as in Appendix 6 in the Dryad repository 10.5061/dryad.3bp51), the model would assume that two genetically incompatible species (either in the same or in different sets of connected sites) that converge phenotypically would eventually hybridize, with one of them going extinct. It would also assume that if the species' traits diverged again after some time, a new speciation event would occur on the phylogeny. But in fact neither hybridization leading to extinction nor speciation actually occurred. This may generate discrepancies between predicted reproductive isolation between populations and their actual occurrence. In conclusion, accounting for genetic incompatibilities only at the individual level might lead to overestimate species turnover, underestimate the number of species, and distort descent relationships between species.

In Gascuel et al. (2015), Appendix 6 was designed to investigate the dynamics of species trees when hybrid collapses are prevented (by the accumulation of genetic incompatibilities at the individual level). With the new version of the model, we account for the effect of genetic incompatibilities on species delineation to revisit how hybridization influences species turnover at steady state. We further investigate how this refined criterion for species delineation impacts the structure and dynamics of species trees.

How the model can be modified to account for genetic incompatibilities in species delineation

In this new version of the model, we delineate species and follow their evolutionary history during diversification as in Gascuel et al. (2015, with methodological details in the online Appendix 4, at http://dx.doi.org/10.5061/dryad.3bp51), but with the following refinements:

- We extract the distribution of individuals in the phenotypic space (x_1, x_2) and group individuals within adjacent high density cells into populations *independently for each set of connected sites*;

- Populations are grouped into the same species if (i) they can interbreed (as defined Gascuel et al. 2015, if the probability of reproductive isolation is below the threshold thr_{ri}=99%), and (ii) *if the two populations are also genetically compatible*. We assess genetic incompatibility between populations using the same function of genetic distance as for individuals (see Appendix 6). The genetic distance between two populations is the number of different genetic loci harbouring incompatible alleles carried by their individuals, weighted by their frequency in each population.

- We determine descent relationships between species at time t and species at time $t-1$ based on *the minimum number of genetic incompatibilities between their populations.* We only use minimum Euclidian distances between their average phenotypic traits to

discriminate the line of descent if populations of a species at time t have an equal (minimum) number of genetic incompatibilities with several populations from different species at time $t-1$.

Figures 1 and 2 illustrate this new method to delineate species and follow their evolutionary history (update of Figures A4 and A5 in Appendix 4).

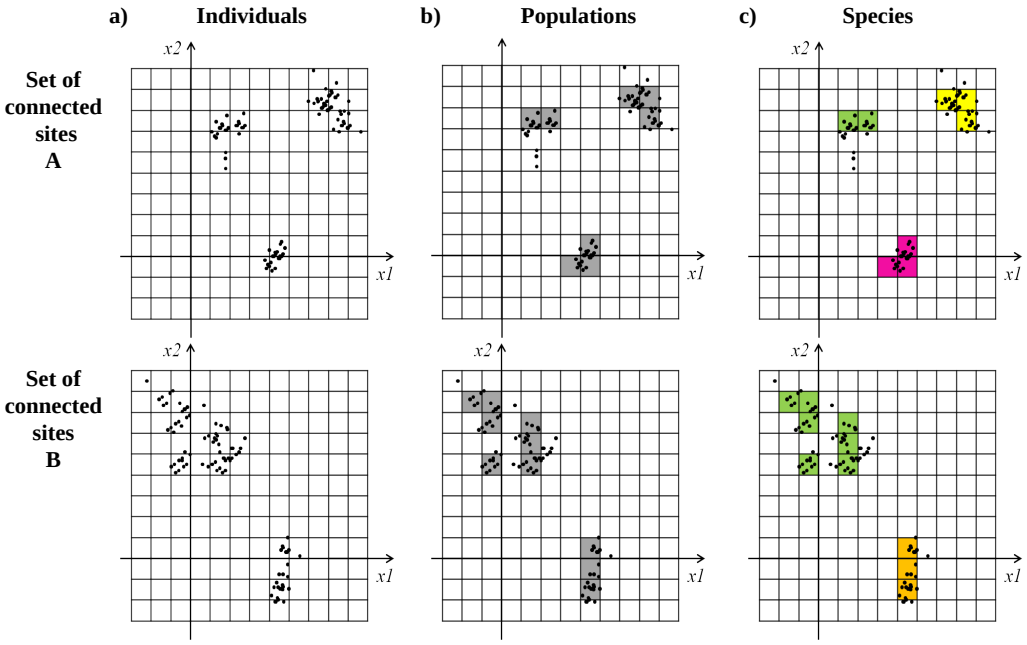

FIGURE 1: **How to delineate species when accounting for genetic incompatibilities.** In this example, we consider two sets of connected geographical sites (A and B), and represent individuals (black dots) in the phenotypic space. We first group individuals into populations by (a) extracting, for each set of connected sites, the distribution of individuals in the phenotypic space (x_1, x_2), according to cells of width $\sigma_{\mu_{x_1}} \times \sigma_{\mu_{x_2}}$; and (b) determining high density phenotypic cells (in grey), and group individuals within adjacent high density cells into populations. Then, (c) we consider populations of all sets of connected sites (A, B,...) and group into species (same color) those that (i) can interbreed (probability of reproductive isolation below the threshold thr_{ri}=99%) and (ii) are not genetically incompatible. Here, the green populations in the sets of connected sites A and B can interbreed and are not genetically incompatible, the orange and pink populations could interbreed based on their phenotypic and assortative mating traits but are genetically incompatible, and the yellow population albeit not genetically incompatible with the green ones (due to recent divergence) cannot interbreed with them due to quite different phenotypic traits.

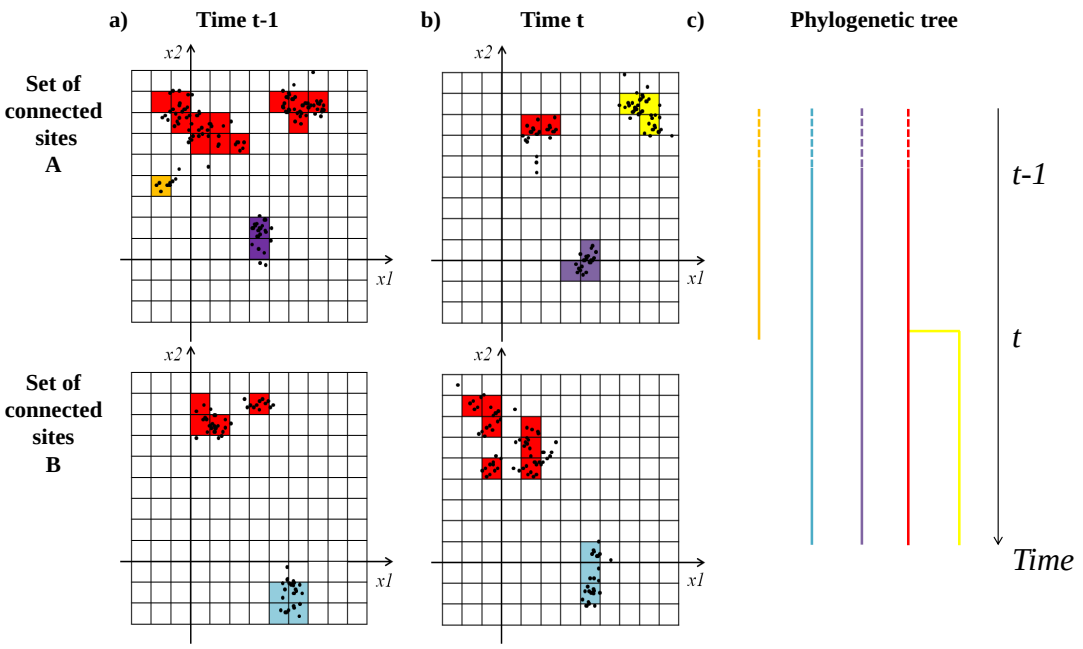

FIGURE 2: **How to determine species ancestry when accounting for genetic incompatibilities.** To determine the ancestry of the species $[S_t^{(i)}]_{i\in\{1,\dots,N_t\}}$ delineated at time step t (as presented in Fig. 1), and thus the changes in the phylogenetic tree between time step $t-1$ and time step t (c), we compare the loci carrying genetic incompatibilities and the phenotypic traits of species and populations at time t (b) to those at time $t-1$ (a). Descent relationships are determined by minimum number of genetic incompatibilities between entities at time t and at time $t-1$ and, if the latter are equal for multiple species at time $t-1$, by minimum Euclidian distances between average phenotypic traits. Speciation occurs if different species $S_t^{(i)}$ descend from the same species $S_{t-1}^{(j)}$; in that case (e.g. the red and yellow species at time t), the species $S_t^{(i)}$ most similar to $S_{t-1}^{(j)}$ is $S_{t-1}^{(j)}$ (here the red species), whereas other species $S_t^{(i)}$ (here the yellow species) are new ones, descending from $S_{t-1}^{(j)}$. Hybridization occurs if a species $S_t^{(i)}$ includes populations descending from populations of several different species $S_{t-1}^{(j)}$; in that case $S_t^{(i)}$ (e.g. the red species at time t), has evolved from the most similar of its parental species $S_{t-1}^{(j)}$ (here the red species), partly due to hybridization, whereas other species $S_{t-1}^{(j)}$ (here the orange species) might go extinct.

New results

In the following section, we first explore the influence of the rate of accumulation of genetic incompatibilities on species turnover at steady state. By doing so, we want to revisit our previous conclusions (presented in Appendix 6 of Gascuel et al. 2015) on the importance of hybridization in the species turnover in Gascuel et al. (2015). Then, we use these results to calibrate the rate of accumulation of genetic incompatibilities in order to prevent hybridization between distantly related species. This allows us to investigate how accounting for genetic incompatibilities to delineate species impacts the structure and dynamics of species trees, by comparing the shapes of species trees generated by this new version of model to those presented in Gascuel et al. (2015).

Figure 3 shows that about fifty percent of the species turnover observed at steady state in Gascuel et al. (2015) was due to hybridization events that are predicted but do not actually occur (i.e. due to trait convergence of distantly related species in different sets of connected sites).

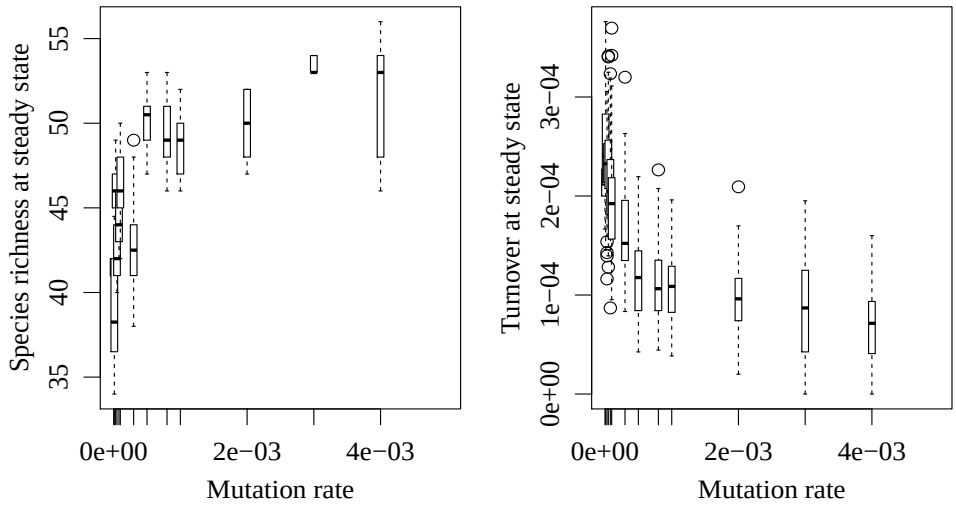

FIGURE 3: **Species richness and turnover at steady state as a function of the mutation rate of loci harbouring incompatible alleles.** Simulations were replicated 5 times. Parameter values: C=0.1, q=0.1, $\mu = 1.10^{-3}$, σ_C=0.3, and σ_K=0.8. See Table 1 in Gascuel et al. (2015) for other parameter values.

To address the effect of these observed hybridization events on phylogenetic tree shape, we ran again all simulations from Gascuel et al. (2015) with this new version of the model. Below, we provide the equivalent of figures 1 to 3 in Gascuel et al. (2015, using the same parameter values). Parameter C quantifies the threshold in number of incompatible loci needed to generate reproductive isolation; it was set to C=0.1. Parameter q gives the probability of incompatibility between two genetic loci; it was set to q=0.1. The mutation rate for new alleles was set to $\mu = 1.10^{-3}$. The probability of genetic incompatibilities was

computed as explained in the online Appendix 6. The above parameter values ensure that genetic incompatibilities accumulate fast enough to prevent hybridization between distantly related species (reaching a plateau in species turnover, see Fig.3), but slow enough to remain computationally tractable and not be the cause of speciation (so that the drivers of speciation are kept ecological, as opposed to genetic).

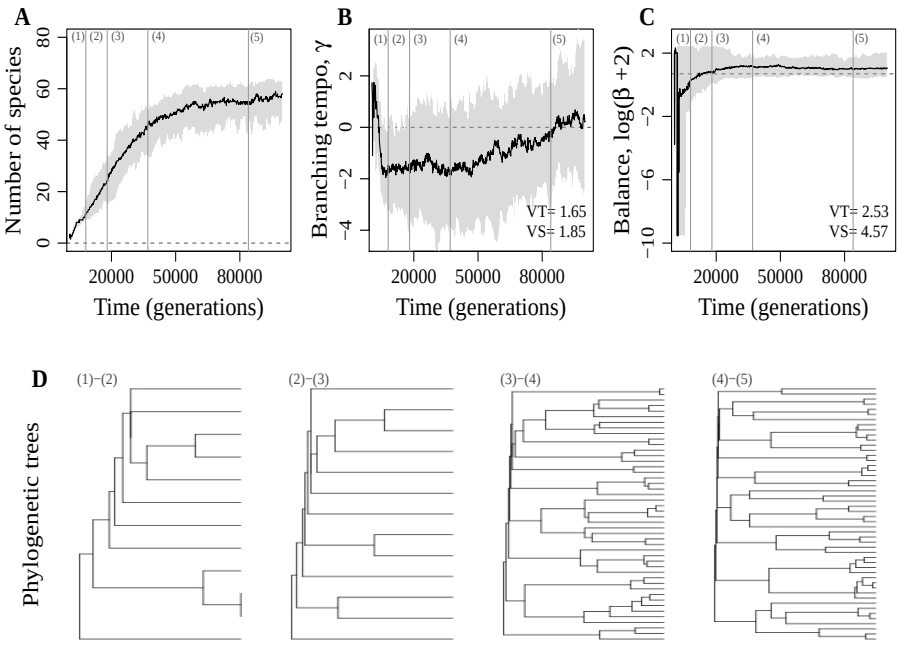

FIGURE 4: **Species diversity and phylogenetic tree shape (branching tempo γ and balance β) as a function of time.** Time is measured from the common ancestor's introduction in the landscape. Simulations were replicated 50 times. Top row: Black lines give median values; 95% confidence interval shown in grey. Dashed horizontal lines indicate γ=0 and β=0. VT and VS give the average variance, respectively over time (corrected according to the median among simulation replicates) and among simulation replicates. Vertical grey lines and associated labels highlight five different stages of γ and β variation across time. Bottom row: examples of species trees built from the end of each of the four stages of diversification. Competition parameters: σ_C=0.4, σ_K=1.1. See Table 1 in Gascuel et al. (2015) for other parameter values.

FIGURE 5: **Effect of the scaled width of competition and interaction with abiotic factors on the evolution of species diversity, branching tempo γ and balance β of phylogenetic trees as a function of time.** Time is measured from the common ancestor's introduction in the landscape. Simulations were replicated 50 times. Black lines give the median (with 95% confidence interval in dark grey) under wide competition (σ_C/σ_K=0.75); white lines give the median (with 95% confidence interval in light grey) under narrow competition (σ_C/σ_K=0.25). Stars indicate statistically significant differences ($p - value$ <0.05) between wide and narrow competition (stars near black lines for differences in median values, t-test; stars on top of dark grey areas, for differences in variance, F-test), either at the beginning or at the end of the simulations (respectively first and last 20,000 generations). Dashed horizontal lines indicate γ=0 and β=0. Rows B1 and B2 test for the effect of the pace of landscape dynamics (default A $\approx 4.76.10^{-5}$, B1 $\approx 9.9.10^{-6}$ and B2 $\approx 9.8.10^{-5}$). Rows C1 and C2 test for the effect of time in isolation (default A = $\approx$ 0.95, C1 $\approx$ 0.52, and C2 $\approx$ 0.99). Row D tests for the effect of local catastrophes (default A = 0, and D = 2.10^{-4}). See Table 1 in Gascuel et al. (2015) for other parameter values.

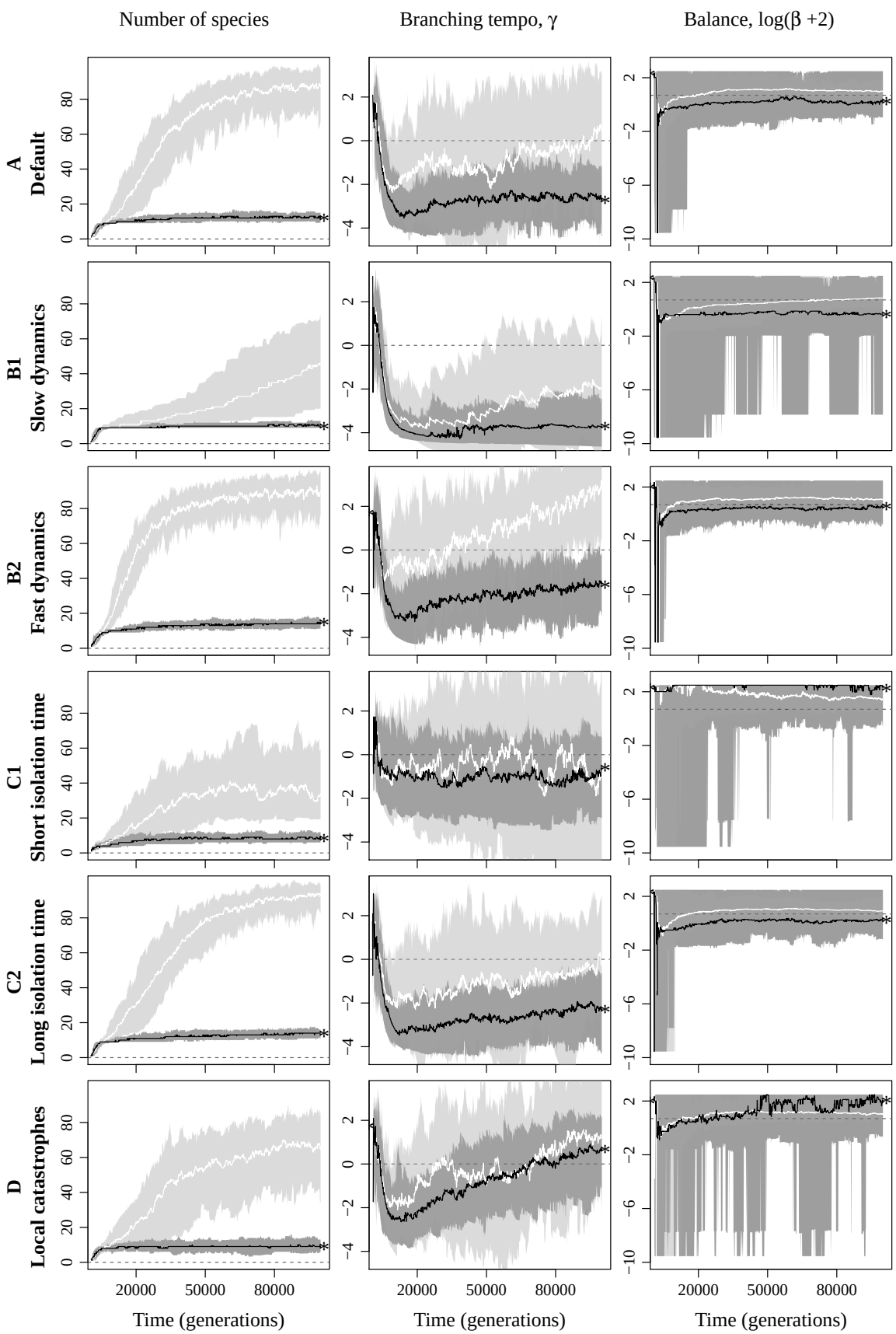
Number of species
Branching tempo, γ
Balance, log(β +2)
A
Default
B1
Slow dynamics
B2
Fast dynamics
C1
Short isolation time
C2
Long isolation time
D
Local catastrophes
0
20
40
60
80
−4
−2
0
2
−10
−6
−2
2
20000
50000
80000
Time (generations)

FIGURE 6: **Effect of spatial heterogeneity in the scaled width of competition and interaction with abiotic factors on the evolution of species diversity, branching tempo γ and balance β of phylogenetic trees as a function of time.** Time is measured from the common ancestor's introduction in the landscape. Simulations were replicated 50 times. Black lines give the median (with 95% confidence intervals in dark grey) with heterogeneity; white lines give the median (with 95% confidence intervals in light grey) without heterogeneity. Stars indicate statistically significant differences ($p - value$ <0.05) with versus without heterogeneity (stars near black lines for differences in median, t-test; stars on top of dark grey areas for differences in variance, F-test), either at the beginning or at the end of the simulations (respectively first and last 20,000 generations). Dashed horizontal lines indicate γ=0 and β=0. Rows B1 and B2 test for the effect of the pace of landscape dynamics (default A $\approx 4.76.10^{-5}$, B1 $\approx 9.9.10^{-6}$ and B2 $\approx 9.8.10^{-5}$). Rows C1 and C2 test for the effect of time in isolation (default A = $\approx$ 0.95, C1 $\approx$ 0.52, and C2 $\approx$ 0.99). Row D tests for the effect of local catastrophes (default A = 0, and D = 2.10^{-4}). Row E shows results when spatial heterogeneity impacts landscape dynamics rather than competition width (heterogeneity intensity $\approx$ 0.46). See Table 1 in Gascuel et al. (2015) for other parameter values.

Number of species
Branching tempo, γ
Balance, log(β +2)
A
Default
B1
Slow dynamics
B2
Fast dynamics
C1
Short isolation time
C2
Long isolation time
D
Local catastrophes
E
Heterogeneous landscape dynamics
Time (generations)
Time (generations)
Time (generations)

Conclusions

Accounting for long-term irreversible reproductive isolation between species increases the species richness and decreases the species turnover at steady state, but does not change phylogenetic tree shape under any of all biotic or abiotic conditions we tested (compare above Figures 4, 5 and 6 to Figures 1, 2 and 3 in Gascuel et al. 2015). Therefore, all results on the role of ecology and landscape dynamics on phylogenetic tree imbalance and branching tempo from Gascuel et al. (2015) extend to this refined version of the model.

Future work will use this new version of the model, in which species delineation and evolutionary history are determined by accounting for both populations' evolutionary history and individuals' reproductive isolation. This new version allows trait convergence to occur without hybridization between species evolving in disconnected geographical sites.

*

References

Gascuel, F., R. Ferriere, R. Aguilee, and A. Lambert. 2015. How ecology and landscape dynamics shape phylogenetic trees. Syst. Biol. 0:1–18.