Rapid speciation in small populations challenges the dominance of ecological speciation

Pierre Veron [1,2,#], Anaïs Spire [1], Agathe Chave-Lucas [1], Tatiana Giraud [2], and Hélène Morlon [1]

[1] *Institut de Biologie de l'École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France*
[2] *Écologie Société et Évolution, CNRS, Université Paris-Saclay, AgroParisTech, Gif-sur-Yvette, France*
[#] *Corresponding author, pveron@bio.ens.psl.eu*

February 19, 2026

Keywords — speciation | micro-macro | holey adaptive landscape | plants.

Abstract

Speciation – the process by which two lineages become reproductively isolated – plays a key role in the emergence and maintenance of biodiversity. Yet, our understanding of the time it takes for speciation to occur, and of the microevolutionary processes that influence this tempo, remains limited. Here, we thoroughly characterize how population size, mutation rate, local adaptation and migration are expected to influence the duration of speciation, as well as the shape of the "grey zone" of speciation. We show that the relationship between population size and speciation time is indicative of the speciation mode, as faster speciation in smaller populations only occurs in the case of non-ecological speciation. Leveraging genomic estimates of population size and speciation duration across 196 pairs of plant species, we uncover a positive association between population size and speciation duration. Taken together, these results challenge the view that ecological speciation is the source of much of species diversity.

Significance statement

How new species arise, and how quickly they do so, plays a key role in shaping Earth's biodiversity. Yet, the links between how speciation occurs, how fast it proceeds, and basic properties of species such as population size, remain debated. We clarify these relationships using an integrative model of speciation. We show, in particular, that faster speciation in smaller populations is a signature of non-ecological speciation. Analyses of genomic data from plants reveal precisely this pattern, challenging the prevailing view that speciation in nature is predominantly ecological.

1 Introduction

Speciation – the process by which a new species emerge – is one of the most fundamental processes in biology. It is at the origin of species diversity. In spite of substantial research on speciation, many controversies remain on its mode, i.e, how it occurs, and tempo, i.e., at which pace it occurs (Coyne and Orr, 2004). There is however a general consensus that speciation takes time (Benton and Pearson, 2001; Etienne et al., 2014), except in exceptional cases, such as speciation by polyploidization in plants (Rieseberg and Willis, 2007) or speciation by host shifts in pathogens (Giraud et al., 2010). Intuitively, fast-speciating groups will experience speciation events more frequently than slow-speciating ones. Together with the frequency of extinction events, these differences can explain differences in species richness across groups, and together with the frequency of dispersal events, differences in species richness across geographic regions (Wiens and Donoghue, 2004; Schluter and Pennell, 2017).

Insights on the tempo of speciation can be gained by fitting diversification models to extant phylogenies (Nee, 2006; Morlon et al., 2024). Fitting birth-death models representing lineages birth (speciation) and death (extinction), assumed to occur instantaneously, provides estimates of speciation and extinction rates, i.e., estimates of the average number of events occurring per lineage in a given amount of time. Such studies have shown that speciation rates vary by several orders of magnitude across lineages (Rabosky, 2016; Maliet et al., 2019; Quintero et al., 2024). For example, estimates range from 0.01 to 5 $\text{spp} \cdot \text{Myr}^{-1}$ in birds (Maliet et al., 2019), and from 0.005 to 1.5 $\text{spp} \cdot \text{Myr}^{-1}$ in mammals (Quintero et al., 2024).

A central challenge lies in understanding if and how microevolutionary (intraspecific) processes, such as drift, selection and migration, can explain this speciation rate heterogeneity (Harvey et al., 2019; Rolland et al., 2023; Morlon et al., 2024). In comparison with the numerous studies that have investigated correlates of speciation rates with species traits, as well as the abiotic and biotic environment they experience (Benton and Pearson, 2001; Lagomarsino et al., 2016; Schluter and Pennell, 2017; Landis et al., 2022; Wiens, 2024), empirical studies correlating speciation rates to metrics reflecting microevolutionary processes remain rare. The few studies that have investigated the relationship between speciation rates and such parameters have found contrasting or counter-intuitive results, in particular in terms of correlations with the velocity of reproductive isolation (Rabosky and Matute, 2013), population divergence rates (Harvey et al., 2017, 2019), population structure (Singhal et al., 2018; Burbrink et al., 2023), substitution rates (Lanfear et al., 2010; Goldie et al., 2011), or genetic diversity (Huang et al., 2018; Perez-Lamarque et al., 2022; Afonso Silva et al., 2025).

Mathematical models have been central in our understanding of the dynamics of speciation, and of the microevolutionary processes that influence these dynamics (Gavrilets, 2003; Coyne and Orr, 2004). In particular, Gavrilets' holey adaptive landscape (HAL) provided a mathematically tractable extension of Wright's rugged adaptive landscape (Wright, 1932) to high dimensional genotypic spaces (Gavrilets et al., 1998). He showed that – simply as a result of the high dimensionality of adaptive landscapes – populations can become separated by regions of low fitness ("holes"), and thus become

reproductively isolated, without crossing deep valleys of low fitness by evolving along nearly neutral ridges (Gavrilets, 2003). This framework is therefore well suited for analyzing how drift and selection jointly influence the dynamics of speciation. In his 1999 paper, Gavrilets derived a series of mathematical results describing the dynamics of divergence within- and between-populations under strict allopatry, in the presence of migration, and with local adaptation. These provide the foundation for obtaining analytical results concerning the expected duration of speciation, although those results were not derived. A later review focused on this question (Gavrilets, 2003) but without taking into account within-population genetic variation. Yet, polymorphism can substantially alter the speed at which mutations causing reproductive incompatibilities accumulate, and therefore speciation duration (Gavrilets, 1999).

Population size has occupied a central position in debates about the relative role of drift and selection in the speciation process (Coyne and Orr, 2004). This idea traces back to Mayr's verbal "genetic revolution" hypothesis that population bottlenecks, for example during founder events, can trigger a shift in allele frequency over several linked loci leading to speciation (Mayr, 1963). Under this hypothesis, speciation should be faster in smaller populations. Opponents of this hypothesis, however, predicted that speciation should be faster in larger populations, where natural selection is more efficient (Orr and Orr, 1996). Recent developments in phylogenetic models of diversification, combined with the availability of large empirical datasets, have allowed for new tests of the relationship between proxies of population size, such as range size or nucleotide diversity, and speciation rates (Maya-Lastra and Eaton, 2021; Smyčka et al., 2023; Afonso Silva et al., 2025). These tests, however, treat speciation as an instantaneous event, which rate is only weakly influenced by the duration of speciation (Veron et al., 2025), and can be confounded by the reciprocal effect of speciation on population sizes (Smyčka et al., 2023). Inferences from speciation genomics, on the other hand, can provide estimates of speciation duration *per se*, along with estimates of population sizes before and after the speciation event (Fraïsse et al., 2021).

Here, we begin by deriving new analytical predictions for the duration of speciation under the joint influence of drift, selection and gene flow, focusing on the effect of population size. Our derivations build upon Gavrilet's holey adaptive landscape (HAL) model (Gavrilets, 1999). We then design a test based on our theoretical results on the relationship between speciation duration and population size to assess speciation mode based on genomic data, which we apply to 196 pairs of plant species (Monnet et al., 2025).

2 Results and discussion

We consider the HAL model of speciation (Gavrilets et al., 1998; Gavrilets, 1999; Materials and methods). Initially, mutations arise in a population made of N individuals, with a per-individual per-generation mutation rate ν. Any pair of individuals is thus characterized by a genetic distance d, defined as the number of nucleotide sites at which they differ. If d is above a given threshold K called "genetic incompatibility threshold", the individuals cannot interbreed. Hence K is the maximum number of loci

on which two individuals can be different and still be interfertile. This mathematical simplification is motivated by the decrease in the fitness of hybrids with the genetic distance between their parents (Price and Bouvier, 2002; Rabosky and Matute, 2013; Christie and Strauss, 2018; Christie et al., 2022). This framework provides a mathematically tractable way to represent high dimensional, correlated adaptive landscapes (Gavrilets, 1997, figure 1A), and to account for the multi-genic origin of incompatibilities without the complexity of a large set of Dobzhansky-Muller incompatibilities (DMIs; Maya-Lastra and Eaton, 2021). Even if mutations are neutral in themselves, they eventually lead to the reproductive incompatibility of genetically distant individuals, as in DMIs. Following Gavrilets (1999), we consider three versions of the model. In the "allopatric neutral" scenario, the initial population experiences a random split at time t_{split}, for example due to the emergence of a geographic barrier isolating two populations, after which strict isolation is maintained. The two populations then gradually accumulate divergence, until speciation occurs (figure 1B). In the "parapatric neutral" scenario, the two populations can exchange alleles through migration, with a migration rate m. In the "adaptive" scenario, mutations confer a selective advantage in the local population in which they arise, with coefficient s_{LA}.

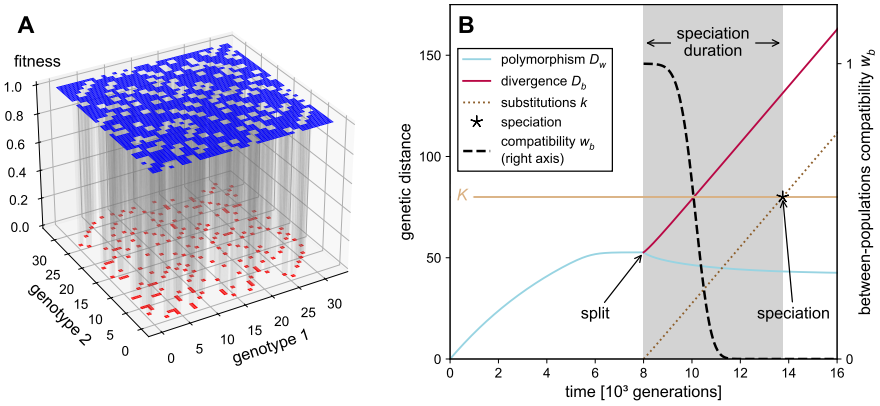


Figure 1. A. Illustration of the holey adaptive landscape (HAL) with 5 polymorphic sites (32 different genotypes) and a compatibility threshold $K = 3$. Populations can evolve along nearly neutral ridges (in blue) and end up being separated by fitness holes (in red). **B.** Example of temporal dynamics under the HAL model with neutral mutations and no migration. The ancestral population is split in two after 8000 generations. We track within-population polymorphism (D_w), between-population divergence (D_b), the number of different substitutions between the two populations (k), as well as between population compatibility (w_b, right axis). The star indicates speciation, characterized by the time when $w_b = 0$, or, alternatively, $k(t) = K$ with K the incompatibility threshold. Here $N_T = 6000$, $N_1 = N_2 = 3000$, $\nu = 0.007$, $K = 80$.

Following Gavrilets (1999), we use a deterministic approximation to investigate the temporal dynamics of both within-population polymorphism D_w, defined as the mean pairwise distance between individuals within populations (w for within), and between-population divergence D_b, defined as the mean pairwise distance between individuals

from the two different populations (b for between; figure 1, Materials and methods). We also track the probabilities w_w and w_b that two randomly chosen individuals from the same or different populations, respectively, can successfully interbreed. These probabilities are referred to as within- and between-population compatibility. Finally, we follow the number k of substitutions (fixed alleles) that differentiate the two populations. We consider speciation to occur if (and when) no individual from one population can successfully interbreed with any individual from the other population (i.e., $w_b = 0$; Mayr, 1942). This occurs if (and when) k reaches K (figure 1B); indeed, in this case, any two individuals picked in different populations are distant by at least K loci and are thus incompatible. We also investigate the shape of the grey zone of speciation, when the reproductive compatibility between the populations w_b is reduced or even close to zero, but does not reach zero. This corresponds to the often-encountered intermediate situation in nature where populations are not fully compatible, but speciation is not complete. We validate our theoretical predictions using intensive simulations (supplementary text S2 and figures S7 to S10).

In the absence of migration, populations ineluctably diverge and speciation occurs after enough time elapses. In the allopatric neutral scenario, we show that the speciation duration τ, defined as the expected time between t_{split} and the time when speciation occurs, is given by :

$$\tau = \frac{K}{\nu(R_1 + R_2)} \tag{1}$$

where R_1 and R_2 are, in each of the two sub-populations, coefficients of fixation which capture the efficiency of purifying selection to remove incompatibilities and depend on population size, mutation rate, and the incompatibility threshold (equation 3 in the Materials and methods). $R \approx 1$ represents the case when this effect is negligible. For simplicity, we assume in what follows that the ancestral population is split into two populations of equal sizes, but our equations can accommodate asymmetric population sizes. As K decreases, and N or ν increase, the effect of purifying selection against incompatibilities is stronger, and R decreases (see figure S1), reducing within-population polymorphism (figure 2, lower row) and affecting the duration of speciation (figure 2, upper row, blue curves). Equation 1 holds in the adaptive scenario, with an expression for R that depends on the local selective advantage of mutations s_{LA} (Materials and methods). As s_{LA} increases, the mutations in each environment confer a local adaptive advantage, and their fixation is faster. Hence the coefficient of fixation increases and can be larger than 1. As expected, local adaptation speeds up speciation (figure 2, upper row, green curves). In the presence of migration, we can determine when speciation occurs by solving a set of ordinary differential equations (Materials and methods). As expected, gene flow slows down speciation (figure 2). In fact, speciation is no longer ineluctable, and we designed an efficient way to predict whether it will occur or not (Materials and methods and supplementary text S1; figure 2 upper row, pink curves).

High mutation rates are generally thought to increase the rate at which populations acquire substitutions and reproductive isolation, resulting in higher speciation rates (Lanfear et al., 2010; Hua and Bromham, 2017). This expectation is part of the integrated evolutionary speed hypothesis, which stipulates that the short generation times and high

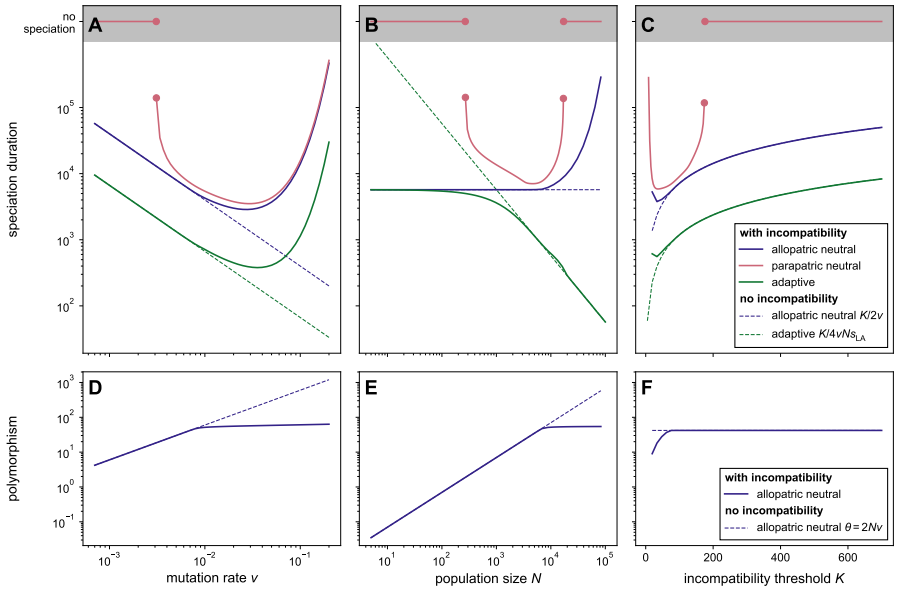


Figure 2. Expected speciation duration (upper row) and polymorphism (lower row) under the holey adaptive landscape model (HAL; solid lines), as a function of mutation rate (**A**, **D**), population size (**B**, **E**) and incompatibility threshold (**C**, **F**). Three versions of the model are considered: strict allopatry (in blue), with migration (parapatric, in red), and with local adaptation (in green). Expectations under a purely theoretical scenario without purifying selection against incompatibilities within populations ($R = 1$, $s = 0$) are shown with dashed lines for comparison. On the upper row, the pink line in the shaded region indicates cases where speciation does not occur. In each panel, one of the three parameters varies with the others kept constant at $\nu = 0.0007$, $N = 6000$, and $K = 80$. For the parapatric scenario, the migration rate is $m = 10^{-4}$; for the adaptive scenario, the coefficient of selection is $s_{\mathrm{LA}} = 0.001$.

mutation rates found in warm (tropical) environments result in fast genetic change and speciation (Rohde, 1992). In a purely theoretical scenario without within-population purifying selection against incompatibilities, increased mutation rates indeed always speed up speciation (figure 2A, dashed curves). When within-population incompatibilities are selected against, however, this expectation is true only under low-to-moderate mutation rates; when mutations become more frequent (> 0.05 per generation), purifying selection against incompatibilities limits their accumulation both within and between populations, and slows down speciation, contrary to the evolutionary speed hypothesis (figure 2A, solid curves). This speciation slowdown under a high mutation rate regime occurs in the three scenarios we considered. The non-monotonic dependency of speciation duration to mutation rates may explain why tests of the evolutionary speed hypothesis have found mixed evidence. A positive association was found in birds (Lanfear et al., 2010) and plants (Bromham et al., 2015), but a negative one was found in mammals (Afonso Silva et al., 2025). In the latter study, the authors attributed this *a priori* unexpected result to methodological artifacts; our results suggest

that it could in fact be real, due to a high purifying selection, and slow accumulation of incompatibilities, when mutations are frequent.

The dependency of speciation duration to population size strongly depends on whether speciation is ecological (adaptive scenario, in green) or not (neutral scenarios, in blue and red) (figure 2B). As expected, ecological speciation is faster in large populations where positive selection is more effective; this effect becomes visible when the population size is larger than the drift-selection barrier ($1/4s_{\mathrm{LA}} = 250$ here; Kimura and Ohta, 1971). This prediction would also hold in the presence of migration, as migrants, which are less fit than individuals from the local population, will be more rapidly selected against in large populations, reducing their chance of homogenizing diverging populations. In the allopatric neutral scenario, and in a purely theoretical situation without within-population purifying selection against incompatibilities, population size would not have an effect on the duration of speciation (figure 2B, dashed blue curve). When within-population incompatibilities are selected against, however, this purifying selection is more efficient in larger populations, slowing down both the accumulation of incompatibilities and speciation (figure 2B & E, solid blue curve). After a given size threshold, the relationship between population size and speciation duration is thus positive. The same holds with migration, except that, when populations decrease below a certain size threshold, speciation duration increases as populations become smaller, since the probability of a given migrant genotype fixing and homogenizing populations is higher in smaller populations (figure 2B, solid pink curve). Importantly, these results show that faster speciation in larger populations can occur under both ecological and non-ecological speciation, but faster speciation in smaller populations occurs only under non-ecological speciation (with or without gene flow). We additionally find that speciation duration is primarily influenced by the size of the smallest of the two separate populations in non-ecological speciation, whereas it is primarily influenced by the size of the largest one in ecological speciation (figure S4). This makes intuitive sense, as mutations causing reproductive incompatibilities accumulate faster in the smallest of the two populations if they are not adaptive, and in the largest population if they are adaptive.

The incompatibility threshold K, i.e., the number of genetic differences at which individuals can no longer interbreed, reflects the genomic architecture of speciation, ranging from few large-effect genes (or structural genomic changes) for small K to many small-effect genes spread across the genome for large K. In a purely theoretical scenario without within-population purifying selection against incompatibilities, the duration of speciation increases with K, as a larger number of fixed differences is required for speciation to occur (figure 2C, dashed curves). It is also largely the case when incompatibilities are selected against, except for very small values of K, where speciation duration increases (figure 2C, solid curves). This is due to the extreme purifying selection acting on genomic changes with large effects, which hampers their fixation (see the reduced coefficient of fixation R in figure S1 for small K) and slows down speciation, despite the small number of fixed differences required for speciation to occur; this effect is reflected by a reduced polymorphism in the population (figure 2F). Hence, the more polygenic speciation is, the more time it is expected to take, except in cases of very large effect genomic changes that have a hard time fixing.

Mutation rates, population sizes, and the genomic architecture of speciation influence not only speciation duration, but also the timing and shape of the grey zone, i.e., when and at which pace the between-population compatibility declines. We fitted a decreasing sigmoid to the between-population compatibility $w_b(t)$ (dashed line in figure 1) and analyzed its timing t_{GZ} (inflection point) and slope r_{GZ} (Materials and methods). The results mirror the dependencies found for speciation duration, with earlier/faster drops in between-population compatibility corresponding to shorter speciation durations, while later/slower drops correspond to longer speciation durations (see the concordance between figure 3A and figure 2A-C). These analyses further show that the only scenario with an early drop in between-population compatibility is the scenario with local adaptation when populations are large (figure 3A2, in green), and that migration enforces a slow compatibility drop (figure 3A2, in red). Hence, we expect – most often – a substantial lag between the time when populations split and the time when compatibility between them drops, and fuzzy species boundaries that extend for a long period of time as soon as there is between-population gene flow.

Examining the shape of the grey zone as a function of the net divergence (defined as the difference between divergence and polymorphism, $D_a := D_b - D_w$) instead of time, as is typically done in genomic-based empirical analyses of the grey zone of speciation (Roux et al., 2016; Monnet et al., 2025), reveals a completely different picture (figure 3B). In this case, the timing and slope of the grey zone are consistent across scenarios. Thus, local adaptation and gene flow have a minor influence on the shape of the genomic grey zone. Population size, mutation rate and genomic architecture all affect the timing of the drop, while its slope is primarily influenced by population size, with a more marginal effect of the other parameters. Earlier drops, reflecting decreases in compatibility at lower levels of divergence, occur under more frequent mutations, larger population sizes, and larger mutational effects. This is explained by the larger polymorphism of the populations at equal level of divergence under these conditions (see figure S2), which makes it more likely to find pairs of individuals that are not compatible between the two populations. Populations of small size or small mutation rate, on the contrary, maintain a high level of compatibility while accumulating net divergence, until the latter reaches K and compatibility drops sharply. Slower drops, reflecting a fuzzier level of divergence at which speciation occurs, take place in larger populations, and, with a smaller effect, under more frequent mutations and smaller mutational effects. The difference between the genomic and temporal grey zones (figure 3B versus figure 3A) show that viewing genomic divergence as a proxy for time since the separation of populations in empirical grey zone curves can lead to biased interpretations. Our theoretical results on the genomic grey zone should help interpreting the shape of the grey zone obtained from empirical data (Roux et al., 2016; Monnet et al., 2025). For example, Monnet et al. (2025) found that reproductive isolation is achieved at a smaller level of divergence in plants compared to animals. They hypothesized that this difference could be explained by reproductive isolation occurring with fewer reproductive barriers in plants. We show here that the number of loci required to achieve reproductive isolation (K) indeed has a large effect on the timing of the grey zone (figure 3B). Our results suggest that larger population sizes and/or higher mutation rates in plants compared to animals could also explain the pattern. Differences in

gene flow or local adaptation, to the contrary, are not expected to impact the shape of the genomic grey zone.

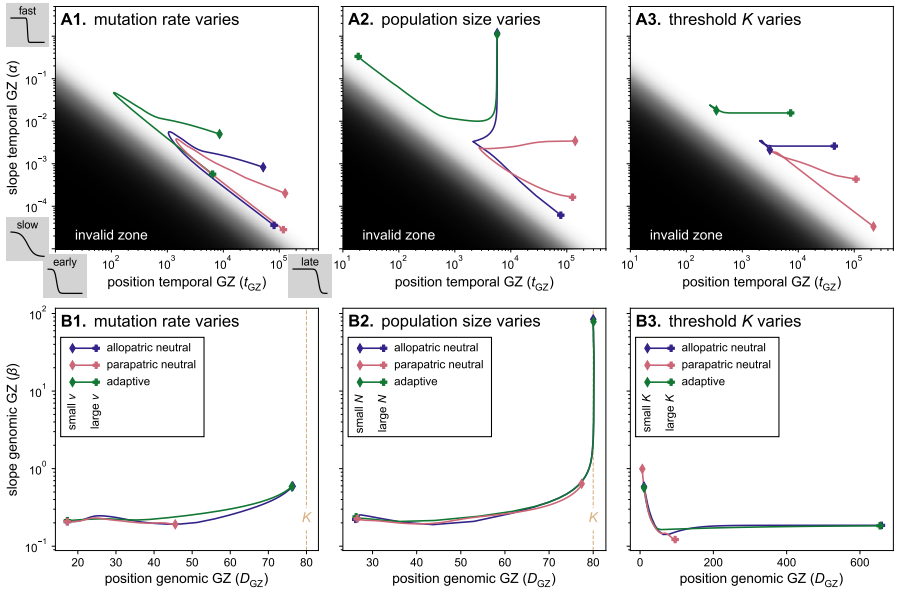


Figure 3. Characteristics of the temporal (**A**) and the genomic (**B**) grey zone (GZ) of speciation, with varying mutation rate (**1**), population size (**2**) and incompatibility threshold (**3**). The temporal GZ (resp. genomic GZ) are characterized by a slope α (resp. β) and a position t_{GZ} (resp. D_{GZ}), see Materials and methods. On each line, one parameter changes (increasing from ♦ to +), each line corresponds to one scenario of speciation. On the upper row, the dark zone of the graphics indicates mathematically unreachable regions; typically early speciation with small slope would require that initial compatibility w_b significantly different from 1 (the middle of the border correspond to $w_b = 0.75$).

Our theoretical predictions indicated that a positive correlation between population size and the duration of speciation would be indicative of non-ecological speciation and that, in this case, population duration should be primarily influenced by the size of the smallest population. We evaluated these correlations using genomic-based demographic inferences for 196 pairs of plant species for which speciation was inferred to be complete (i.e., no gene flow at present; Monnet et al., 2025). We used posterior distribution estimates of speciation duration and of the sizes of the ancestral and two descendant populations, and compared the correlations with a null model to account for the structure in pairs of the data (see subsection 4.4). The correlation between speciation duration and population size was significantly positive ($p < 0.05$) in 98.8% of the posterior samples when considering the small descendant population (25% in the null model), 59.5% of the samples when considering the larger descendant population (22.2% in the null model), and none of the samples when considering the ancestral population (22% in the null model; figure 4). These results suggest that non-ecological speciation dominates in these plant species pairs. The absence of correlation between

the ancestral population size and the duration of speciation may seem surprising at first sight, since population size typically influences the initial polymorphism within- and between-populations. However, the positive correlation between speciation duration and descendant population sizes suggests that we are in the large-population regime (figure 2B, right-most part) where polymorphism is expected to be independent of population size (figure 2E, right-most part). The empirical correlations are therefore strikingly consistent with expectations under non-ecological speciation and large population sizes.

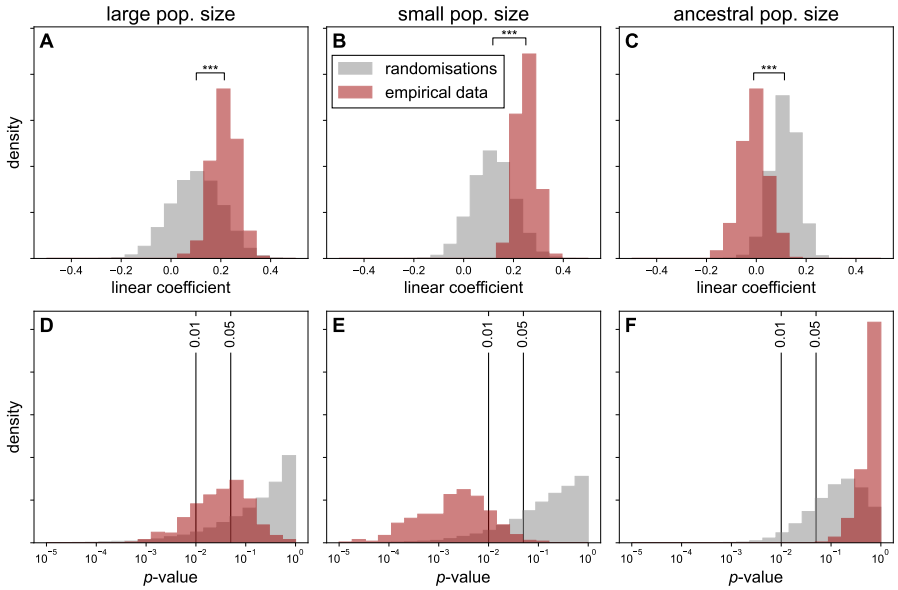


Figure 4. Distribution of linear coefficients (**A**, **B**, **C**) and p-values (**D**, **E**, **F**) for the correlation between speciation duration and population size across posterior samples (red) and randomisations (grey). We fitted the model: $\log(\text{spec. time}) \sim \log N_{\text{large}} + \log N_{\text{small}} + \log N_{\text{ancestral}}$ to 400 posterior samples from demographic parameters inferred on 196 pairs of plants, compared with 100 randomisations between pairs of species for each sample. On the upper row, the difference between the linear coefficient from the randomisations and the empirical data is assessed with a paired t-test (see detailed results in table S1).

3 Conclusion

We investigated how mutation rates, population sizes, the genomic architecture of speciation, selection, and migration shape the tempo of speciation in a multigenic, correlated adaptive landscape. As expected, higher mutation rates accelerated speciation, consistent with the evolutionary speed hypothesis, whereas speciation proceeded more slowly when mutations were neutral than when they were adaptive and in the presence of gene flow. We also uncovered several unexpected patterns, including a rever-

sal of the relationship between mutation rates and speciation duration expected under the evolutionary speed hypothesis at very high mutation rates, and the insensitivity of the genomic grey-zone of speciation to gene flow. Our theoretical and empirical results linking speciation duration to population size revive Mayr's verbal theory of genetic revolution. Specifically, speciation can occur rapidly in small populations in the absence of selection, due to the increased possibility of fixing incompatibilities, and empirical data from 196 plant species pairs are consistent with this non-ecological speciation scenario.

However, smaller populations are also more vulnerable to extinction and may collapse before speciation is completed. This trade-off underscores the need, in the future, to jointly consider speciation duration and lineage persistence to understand how microevolutionary processes translate into macroevolutionary speciation rates. The framework developed here provides a foundation for the development of such integrative analyses.

4 Materials and methods

4.1 Predictions under the holey adaptive landscape model

Following Gavrilets (1999), we use a system of ordinary differential equations (ODE) to approximate the mean polymorphism D_w (genetic distance between pairs of individuals in a population), the mean divergence D_b (genetic distance between pairs of individuals from two separate populations), and the number of distinct substitutions (fixed alleles) between the two populations k. Under the assumptions of linkage equilibrium and rare alleles, the within- and between populations mean compatibilities are given by (Gavrilets, 1999, eq. 12 and 15):

$$w_w = \frac{\Gamma(K+1, D_w)}{\Gamma(K+1)} \quad \text{and} \quad w_b = \frac{\Gamma(K+1-k, D_b-k)}{\Gamma(K+1-k)}$$

with $\Gamma(\cdot,\cdot)$ and $\Gamma(\cdot)$ the upper incomplete gamma function and complete gamma function, respectively. The effect of purifying selection against incompatibilities is captured by a selection coefficient that depends on the level of within-population polymorphism (Gavrilets, 1999, eq. 11):

$$s = \frac{e^{-D_w} D_w^K}{\Gamma(K+1, D_w)}.$$

In the absence of local adaptation, the dynamics follow the system of equations:

$$\begin{cases} \dfrac{\mathrm{d}D_w}{\mathrm{d}t} = -sD_w + 2\nu - \dfrac{D_w}{N} + 2m_e(D_b - D_w), \\ \dfrac{\mathrm{d}D_b}{\mathrm{d}t} = -s(D_b - k) + 2\nu + 2m_e(D_w - D_b), \\ \dfrac{\mathrm{d}k}{\mathrm{d}t} = 2\nu R 2^{-2Nm_e} - 2km_e R\left(\dfrac{e}{2}\right)^{2Nm_e}. \end{cases} \tag{2}$$

where N is the size of each population, ν is the number of mutations per individual per generation, denoted thereafter as mutation rate, $m_e := m \times w_b/w_w$ is the effective migration rate, m the migration rate (probability for an individual to have migrated from the other population at the previous generation), s is the selection coefficient defined above, and R a fixation coefficient capturing the relative speed of allele fixation compared to a hypothetical case without incompatibility. The expression of R is given by (Gavrilets, 1999, eq. 13b):

$$R = \frac{2e^{-S}\sqrt{S}}{\sqrt{\pi}\ \mathrm{erf}\left(\sqrt{S}\right)}, \tag{3}$$

with erf the error function and $S := Ns/2$.

In the particular case without migration ($m = m_e = 0$), and under the assumption that D_w reached an equilibrium (so R is constant), we can obtain an analytical solution for the duration of speciation by solving the equation for k, which reduces to $\frac{\mathrm{d}k}{\mathrm{d}t} = 2\nu R$. When the ancestral population splits, individuals are distributed randomly in the two

populations, and it is unlikely that an allele is absent in one population and fixed in the other. Therefore, if $t = 0$ represents the splitting time, the initial number of distinct substitutions is $k(0) = 0$, and thus $k(t) = 2\nu Rt$. Considering that speciation is achieved when $k(t) = K$, the duration of speciation is given by $\tau = \frac{K}{2\nu R}$. If the two populations have different sizes N_1 and N_2, and therefore different coefficients of fixation R_1 and R_2, the equation becomes equation 1.

To analyse the dynamics in the neutral scenarios, we numerically solved the system given in equation 2. In the presence of migration, speciation is not systematic. In this case, predicting whether speciation will occur or not based on the ODEs can be challenging and time consuming. We designed an alternative, fast approach, based on the bifurcation diagram for the equilibriums of this system. The detailed method is provided in supplementary text S1.

To analyse the dynamics in the allopatric scenario with local adaptation, we used the system of ODEs:

$$\begin{cases} \dfrac{\mathrm{d}D_w}{\mathrm{d}t} = -sD_w + 2\nu - \dfrac{D_w}{N} \\ \dfrac{\mathrm{d}D_b}{\mathrm{d}t} = 2\nu R_{\mathrm{LA}}, \\ \dfrac{\mathrm{d}k}{\mathrm{d}t} = 2\nu R_{\mathrm{LA}}. \end{cases} \tag{4}$$

with R_{LA} the coefficient of fixation in the case of local adaptation, given by (Gavrilets, 1999, eq 19):

$$R_{\mathrm{LA}} = \frac{4e^{-S(1-\alpha)^2}\sqrt{S}}{\sqrt{\pi}\left[\mathrm{erf}\left(\sqrt{S}(1+\alpha)\right) + \mathrm{erf}\left(\sqrt{S}(1-\alpha)\right)\right]}$$

with $\alpha = s_{\mathrm{LA}}/s$ denoting the importance of selection due to adaptation compared to the selection due to incompatibility. The equation for D_w is an approximation that ignores the effect of local adaptation (the equation is the same as in the neutral scenario with $m_e = 0$) and is justified by the fact that increases in polymorphism due to positive selection when alleles are in low frequency are counter-balanced by decreases when they are in high frequency. The equation for D_b is an asymptotic approximation (Gavrilets, 1999, eq. 13a); this equation shows that R_{LA} captures the acceleration of between-population genetic divergence induced by the divergent selection in the two populations.

We used Python 3.13 and the module `scipy` (Virtanen et al., 2020) to numerically solve the ODEs. We provide a tool to resolve the equations directly in command line, available on github.com/pierre-veron/HoleyAdaptSpeciation. In order to speed up convergence, we used some numerical approximations for some of the calculations, such as the fixation coefficient. Our implementation provides the option to use either the exact form or the approximation.

4.2 Simulations

The predictions above rely on several simplifications: rare allele assumption, linkage equilibrium, approximation of stochastic processes by deterministic equations, and additional approximations in the scenario with local adaptation. To verify the validity of these simplifications, we compared the output of the equations with stochastic simulations (supplementary text S2). We provide a tool to run the simulations in command line, available on github.com/pierre-veron/HoleyAdaptSpeciation. Without migration, we find that the equations predict well the dynamics observed under the simulations, in particular when we added recombination to the simulations to approximate the linkage equilibrium assumption (figures S7 to S10, configurations A2 and B2). In the neutral allopatric scenario, even simulations without recombination matched the predictions well (condition A1). In the presence of migration to the contrary (conditions C to H), outcomes are more stochastic, and some simulations lead to speciation while it is not expected based on our equations, especially when the migration rate is high and populations are small ($N = 100$ for each population) (conditions G and H). This is consistent with the deterministic approximation, which is not valid when populations are small.

4.3 Analyses of the grey zone of speciation

The grey zone (GZ) of speciation is the period before speciation during which the two populations are not fully compatible nor fully incompatible. We characterized the GZ by analyzing the between-population compatibility w_b as a function of either *(i)* time (the temporal GZ) or *(ii)* net divergence $D_a := D_b - D_w$ (the genomic GZ). The latter corresponds to the GZ obtained from empirical genomic data (e.g. Roux et al., 2016; Monnet et al., 2025). We fitted a decreasing sigmoid to these curves. The temporal GZ is thus characterized by $w_b(t) = 1/(1 + e^{\alpha(t-t_{GZ})})$, with α the slope, and t_{GZ} the timing, or position, of the temporal GZ. The genomic GZ is similarly characterized by $w_b(D_a) = 1/(1 + e^{\beta(D_a - D_{GZ})})$ with β the slope, and D_{GZ} the position, of the genomic GZ. We performed the fit using the function `curve_fit` from the package `scipy` and we checked that the sigmoid was a good approximation (coefficient of determination $R^2 > 0.971$).

4.4 Empirical link between population size and the duration of speciation

We used genomic estimates of various population genetic parameters from Monnet et al. (2025). These were obtained by applying the Demographic Inference with Linked Selection (DILS) model (Fraïsse et al., 2021) to 280 pairs of plant species (118 species or populations from 25 genera). DILS outputs a best supported model among ancient migration (AM), secondary contact (SC) and isolation with migration (IM); given our goal to retrieve estimates of speciation duration, we selected the 196 species pairs for which speciation was inferred to be complete, i.e., the AM model (see figure S3). We retrieved posterior distributions of the time of the split (T_{split}) and the time of cessation of gene flow (T_{AM}), and used their difference as our estimate of speciation duration (T_{spec}). We also retrieved the posterior distributions of the sizes of the ancestral (N_a)

and two descendant populations (N_{large} for the largest, and N_{small} for the smallest). We tested the relationship between the duration of speciation and the population sizes by fitting an ordinary least square (OLS) regression: $\log T_{\text{spec}} \sim \log N_{\text{large}} + \log N_{\text{small}} + \log N_a$ and repeated this over 400 samples.

The species pairs we used are possible pairs within each of 25 plant genera (Monnet et al., 2025). Individual species therefore appear in multiple pairs, generating a complex pattern of non-independence in the data, structured per genus. To account for this non-independence, we repeated the analysis after randomly permuting species identities within each genus. A potential correlation obtained for these permuted data cannot be explained by a true correlation with population sizes, but is rather due to repeated patterns in the data structure (see figure S3). Comparisons between empirical correlations and null model correlations (i.e., those obtained after random permutations) allow for the detection of "true correlations" not due to the way the data are organised. For each empirical sampled, we fitted an OLS on 100 randomised data and compared the obtained linear coefficients using a paired *t*-test (see table S1).

5 Aknowledgements

We thank Sergey Gavrilets, Camille Roux, Quentin Rougemont, and Emmanuel Schertzer for their help with the model or data. PV acknowledges funding from the *École polytechnique* and the *Institut des mathématiques pour la Planète Terre* (https://impt.math.cnrs.fr/).

6 Data availability

All scripts and data used in this analysis are available on Zenodo: https://doi.org/10.5281/zenodo.18683265. A GitHub repository containing the functions to run HAL predictions and simulations in command line is also available: https://github.com/pierre-veron/HoleyAdaptSpeciation/.

References

Afonso Silva, A. C., Maliet, O., Aristide, L., Nogués-Bravo, D., Upham, N., Jetz, W. and Morlon, H. (2025). Negative global-scale association between genetic diversity and speciation rates in mammals. *Nature Communications*, *16*(1), 1796.

Benton, M. J. and Pearson, P. N. (2001). Speciation in the fossil record. *Trends in Ecology & Evolution*, *16*(7), 405–411.

Bromham, L., Hua, X., Lanfear, R. and Cowman, P. F. (2015). Exploring the relationships between mutation rates, life history, genome size, environment, and species richness in flowering plants. *The American Naturalist*, *185*(4), 507–524.

Burbrink, F. T., Ruane, S., Rabibisoa, N., Raselimanana, A. P., Raxworthy, C. J. and Kuhn, A. (2023). Speciation rates are unrelated to the formation of population structure in Malagasy gemsnakes. *Ecology and Evolution*, *13*(8), e10344.

Christie, K., Fraser, L. S. and Lowry, D. B. (2022). The strength of reproductive isolating barriers in seed plants: Insights from studies quantifying premating and postmating reproductive barriers over the past 15 years. *Evolution*, *76*(10), 2228–2243.

Christie, K. and Strauss, S. Y. (2018). Along the speciation continuum: Quantifying intrinsic and extrinsic isolating barriers across five million years of evolutionary divergence in California jewelflowers. *Evolution*, *72*(5), 1063–1079.

Coyne, J. A. and Orr, H. A. (2004). *Speciation*. Sinauer Associates, Inc. Publ.

Etienne, R. S., Morlon, H. and Lambert, A. (2014). Estimating the duration of speciation from phylogenies. *Evolution*, *68*(8), 2430–2440.

Fraïsse, C., Popovic, I., Mazoyer, C., Spataro, B., Delmotte, S., Romiguier, J., Loire, É., Simon, A., Galtier, N., Duret, L., Bierne, N., Vekemans, X. and Roux, C. (2021). DILS: Demographic inferences with linked selection by using ABC. *Molecular Ecology Resources*, *21*(8), 2629–2644.

Gavrilets, S. (1997). Evolution and speciation on holey adaptive landscapes. *Trends in Ecology & Evolution*, *12*(8), 307–312.

Gavrilets, S. (1999). A dynamical theory of speciation on holey adaptive landscapes. *The American Naturalist*, *154*(1), 1–22.

Gavrilets, S. (2003). Perspective: Models of speciation: What have we learned in 40 years? *Evolution*, *57*(10), 2197–2215.

Gavrilets, S., Hai, L. and Vose, M. D. (1998). Rapid parapatric speciation on holey adaptive landscapes. *Proceedings of the Royal Society B*, *265*(1405), 1483–1489.

Giraud, T., Gladieux, P. and Gavrilets, S. (2010). Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution*, *25*(7), 387–395.

Goldie, X., Lanfear, R. and Bromham, L. (2011). Diversification and the rate of molecular evolution: No evidence of a link in mammals. *BMC Evolutionary Biology*, *11*(1), 286.

Harvey, M. G., Seeholzer, G. F., Smith, B. T., Rabosky, D. L., Cuervo, A. M. and Brumfield, R. T. (2017). Positive association between population genetic differentiation and speciation rates in New World birds. *Proceedings of the National Academy of Sciences*, *114*(24), 6328–6333.

Harvey, M. G., Singhal, S. and Rabosky, D. L. (2019). Beyond reproductive isolation: Demographic controls on the speciation process. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 75–95.

Hua, X. and Bromham, L. (2017). Darwinism for the genomic age: Connecting mutation to diversification. *Frontiers in Genetics*, *8*(12).

Huang, J.-P., Leavitt, S. D. and Lumbsch, H. T. (2018). Testing the impact of effective population size on speciation rates – a negative correlation or lack thereof in lichenized fungi. *Scientific Reports*, *8*(1), 1–6.

Kimura, M. and Ohta, T. (1971). On the rate of molecular evolution. *Journal of Molecular Evolution*, *1*(1), 1–17.

Lagomarsino, L. P., Condamine, F. L., Antonelli, A., Mulch, A. and Davis, C. C. (2016). The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist*, *210*(4), 1430–1442.

Landis, M. J., Quintero, I., Muñoz, M. M., Zapata, F. and Donoghue, M. J. (2022). Phylogenetic inference of where species spread or split across barriers. *Proceedings of the National Academy of Sciences*, *119*(13), e2116948119.

Lanfear, R., Ho, S. Y. W., Love, D. and Bromham, L. (2010). Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences*, *107*(47), 20423–20428.

Maliet, O., Hartig, F. and Morlon, H. (2019). A model with many small shifts for estimating species-specific diversification rates. *Nature Ecology & Evolution*, *3*(7), 1086–1092.

Maya-Lastra, C. A. and Eaton, D. A. R. (2021). *Genetic incompatibilities do not snowball in a demographic model of speciation*.

Mayr, E. (1942). *Systematics and the origin of species from the viewpoint of a zoologist*. Columbia University Press.

Mayr, E. (1963). *Animal Species and Evolution* (Harvard University Press).

Monnet, F., Postel, Z., Touzet, P., Fraïsse, C., Van de Peer, Y., Vekemans, X. and Roux, C. (2025). Rapid establishment of species barriers in plants compared with that in animals. *Science*, *389*(6765), 1147–1150.

Morlon, H., Andréoletti, J., Barido-Sottani, J., Lambert, S., Perez-Lamarque, B., Quintero, I., Senderov, V. and Veron, P. (2024). Phylogenetic insights into diversification. *Annual Review of Ecology, Evolution, and Systematics*, *55*, 1–21.

Nee, S. (2006). Birth-Death models in macroevolution. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 1–17.

Orr, H. A. and Orr, L. H. (1996). Waiting for speciation: The effect of population subdivision on the time to speciation. *Evolution*, *50*(5), 1742–1749.

Perez-Lamarque, B., Öpik, M., Maliet, O., Afonso Silva, A. C., Selosse, M.-A., Martos, F. and Morlon, H. (2022). Analysing diversification dynamics using barcoding data: The case of an obligate mycorrhizal symbiont. *Molecular Ecology*, *31*(12), 3496–3512.

Price, T. D. and Bouvier, M. M. (2002). The evolution of F_1 postzygotic incompatibilities in birds. *Evolution*, *56*(10), 2083–2089.

Quintero, I., Lartillot, N. and Morlon, H. (2024). Imbalanced speciation pulses sustain the radiation of mammals. *Science*, *384*(6699), 1007–1012.

Rabosky, D. L. (2016). Reproductive isolation and the causes of speciation rate variation in nature. *Biological Journal of the Linnean Society*, *118*(1), 13–25.

Rabosky, D. L. and Matute, D. R. (2013). Macroevolutionary speciation rates are decoupled from the evolution of intrinsic reproductive isolation in *Drosophila* and birds. *Proceedings of the National Academy of Sciences*, *110*(38), 15354–15359.

Rieseberg, L. H. and Willis, J. H. (2007). Plant Speciation. *Science*, *317*(5840), 910–914.

Rohde, K. (1992). Latitudinal gradients in species diversity: The search for the primary cause. *Oikos*, *65*(3), 514–527.

Rolland, J., Henao-Diaz, L. F., Doebeli, M., Germain, R., Harmon, L. J., Knowles, L. L., Liow, L. H., Mank, J. E., Machac, A., Otto, S. P., Pennell, M., Salamin, N., Silvestro, D., Sugawara, M., Uyeda, J., Wagner, C. E. and Schluter, D. (2023). Conceptual and empirical bridges between micro- and macroevolution. *Nature Ecology & Evolution*, *7*(8), 1181–1193.

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N. and Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, *14*(12), e2000234.

Schluter, D. and Pennell, M. W. (2017). Speciation gradients and the distribution of biodiversity. *Nature*, *546*(7656), 48–55.

Singhal, S., Huang, H., Grundler, M. R., Marchán-Rivadeneira, M. R., Holmes, I., Title, P. O., Donnellan, S. C. and Rabosky, D. L. (2018). Does population structure predict the rate of speciation? A comparative test across Australia's most diverse vertebrate radiation. *The American Naturalist*, *192*(4), 432–447.

Smyčka, J., Toszogyova, A. and Storch, D. (2023). The relationship between geographic range size and rates of species diversification. *Nature Communications*, *14*(1), 5559.

Veron, P., Andréoletti, J., Giraud, T. and Morlon, H. (2025). Speciation completion rates have limited impact on macroevolutionary diversification. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *380*(1919), 20230317.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … And SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272.

Wiens, J. J. (2024). Review: Speciation across life and the origins of biodiversity patterns. *Evolutionary Journal of the Linnean Society*, *3*(1), kzae025.

Wiens, J. J. and Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, *19*(12), 639–644.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress of Genetics*, *8*, 209–222.

Supplementary materials

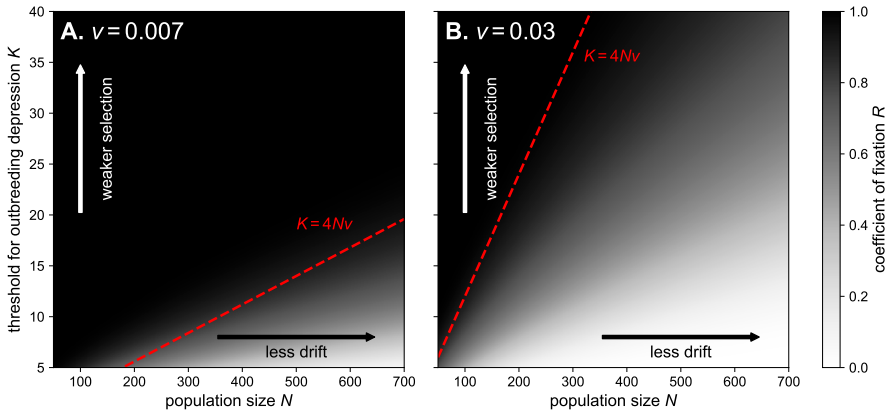
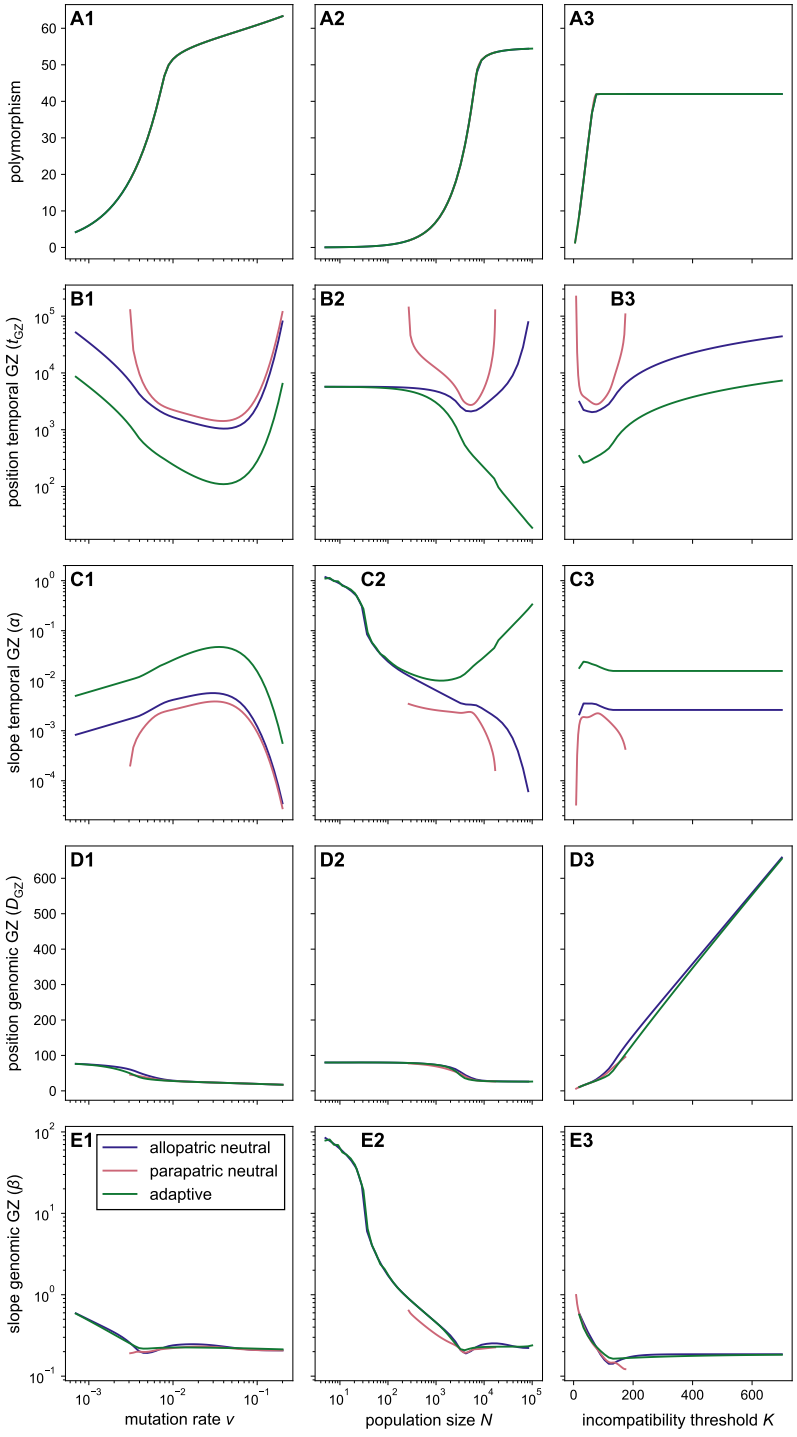


Figure S1. Values of the coefficient of fixation R (representing the relative speed of fixation of mutations compared to a case without purifying selection against incompatibilities, see equation 3) as a function of the population size N and the incompatibility threshold K, for two values of mutation rate ν (**A** and **B**). The dotted red line shows the rule of thumb proposed by Gavrilets (1999) for the limiting conditions where purifying selection against incompatibilities can be neglected ($R \approx 1$).



599

Figure S2. Details of the characteristics of the temporal and genomic grey zones (GZ) of speciation as a function of the mutation rate (**1**), the population size (**2**) and the incompatibility threshold (**3**). **A** shows the mean polymorphism, **B** and **C** show the position and slope of the temporal GZ and **D** and **E** show the position and the slope of the genomic GZ. Each line corresponds to a scenario of speciation.

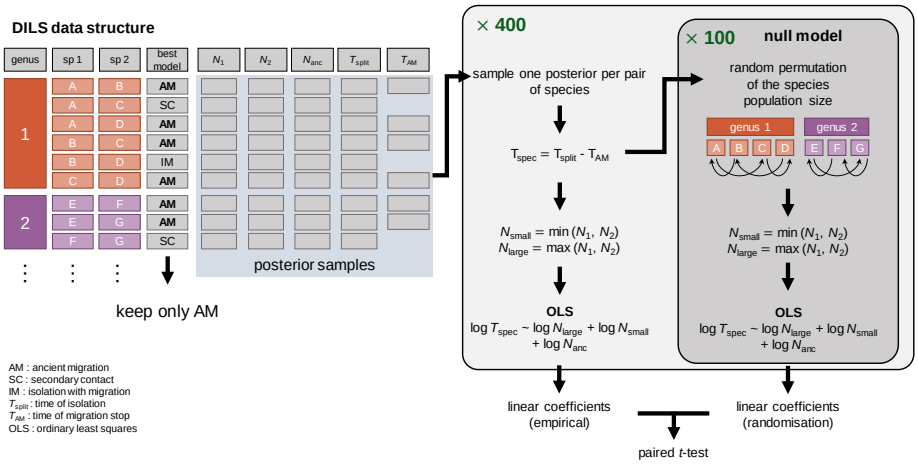


Figure S3. Illustration of the paired structure of the data obtained from DILS, and of the statistical approach used to test the significance of the regression between speciation duration and population sizes . We permuted species within each plant genus to construct a null model.

Table S1. Details of the paired t-test between the linear coefficient of the OLS regression on the empirical versus randomized data (n = 400 samples).

variable	**linear coefficient**		t	$\log_{10} p$-value
	empirical	**randomised**		
large pop. size	0.214 ± 0.053	0.102 ± 0.102	46.2	-161
small pop. size	0.250 ± 0.038	0.117 ± 0.083	78.6	-244
ancestral pop. size	-0.011 ± 0.054	0.113 ± 0.053	-85.1	-257

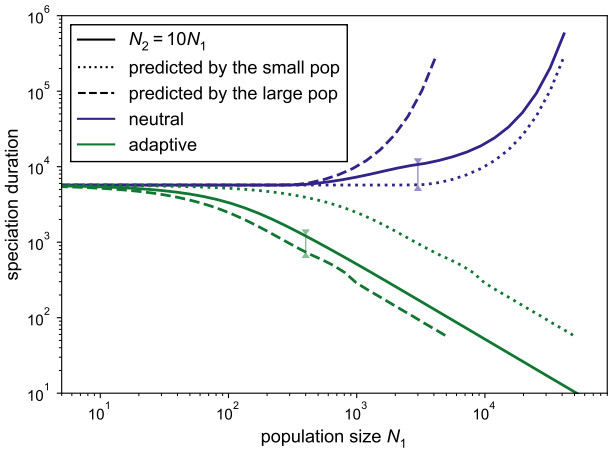


Figure S4. Duration of speciation under the holey adaptive landscape (HAL) allopatric models, in the case of asymmetric populations (solid lines), as a function of N_1 with $N_2 = 10N_1$, compared to the prediction with two populations with size N_1 (dotted line) and two populations with size N_2 (dashed line). In the absence of local adaptation, speciation duration is primarily driven and best predicted by the size of the smallest population (see the small difference between the solid and dotted blue lines, blue arrow); in the presence of local adaptation, it is instead best predicted by the size of the largest population (green arrow).

Supplementary text S1. Determination of the cases without speciation under the parapatric model

In the presence of migration, there are combinations of parameters for which speciation does not occur. Determining these cases by solving the ODEs in equation 2 is not straightforward and time consuming (in particular, speciation not occurring within the set time frame does not imply it will never occur). We therefore designed an alternative approach based on the analyses of the equilibria of the differential equations. For a given combination of parameters, we computed the equilibria by numerically finding the values D_b, D_w and k such that the time derivatives provided in equation 2 are zero. Depending on the parameter values, there are either two equilibrium conditions (one with large D_b, and one with small D_b) or one equilibrium (the two values coincide, see the bifurcation diagrams on figure S5). These equilibrium conditions are not necessarily met, as they require a specific combination of polymorphism, divergence and number of substitutions. The dynamics of $D_b(t)$ around these equilibria determine whether speciation occurs or not, with $D_b(t)$ diverging to ∞ when speciation occurs. We tested numerically the stability of the equilibrium conditions by resolving $D_b(t)$ with an initial condition close to the equilibrium value and D_w, k taken equal to the equilibrium values. In the cases with one equilibrium, we found that this equilibrium is unstable ($D_b \to \infty$ as $t \to \infty$; figure S6 cases 4 and 5) ; in the cases with two equilibria (figure S6 cases 1, 2 and 3), the lower equilibrium is stable (D_b converges to this equilibrium) and the higher equilibrium is unstable ($D_b \to \infty$). In practice, the initial divergence $D_b(0)$ is small compared to the equilibrium values, such that condition 3 is not met. Hence, in the parameter space with two equilibria, D_b converges to the lower, stable equilibrium and speciation does not occur. In the parameter space with one equilibrium, to the contrary, speciation occurs and D_b increases without limit. The bifurcation points shown on figure S5 (dotted lines) can thus be used to determine the parameter values at which there is a transition between scenarios with or without speciation (as in figure 2).

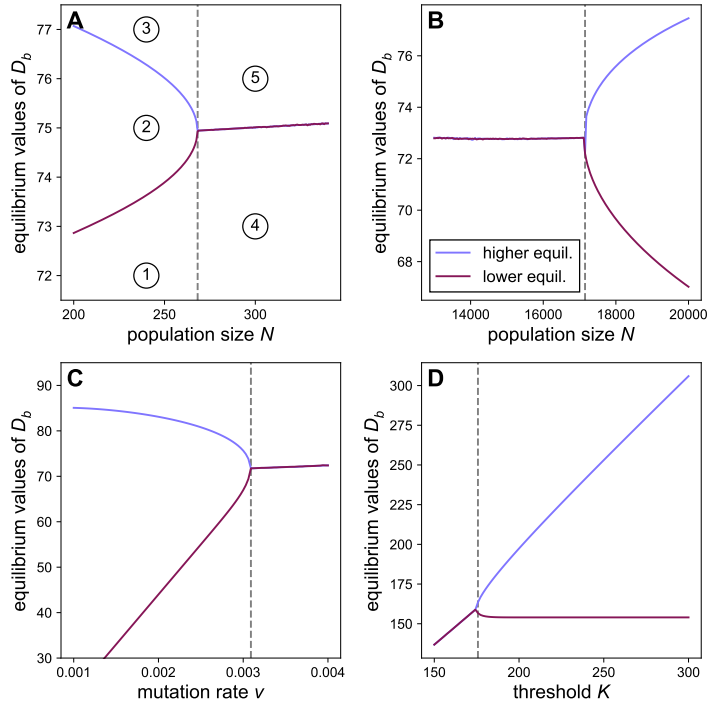


Figure S5. Bifurcation diagrams showing the equilibrium value(s) for D_b as a function of the population size (**A**: low values, **B**: high values), the mutation rate ν (**C**) and the threshold of incompatibility K (**D**). When one parameter varies, the others are kept constant: $\nu = 0.0007$, $N = 6000$, $m = 10^{-4}$ and $K = 80$. The numbers on panel **A** indicate initial conditions for which the dynamics of D_b are represented on figure S6.

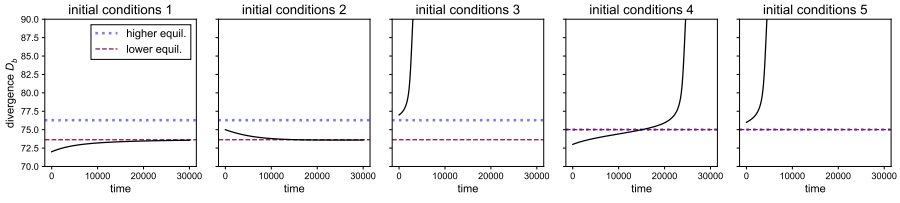


Figure S6. Dynamics of D_b when its initial value is close to the equilibrium and smaller than (1, 4), higher than (3, 5), or between (2) the equilibrium values, corresponding to 5 different regions of the bifurcation diagrams, indicated on figure S5A. The dotted lines indicate the equilibrium value(s). In each case the initial values of D_w and k are the equilibrium values.

Supplementary text S2. Comparison of the HAL deterministic model with simulations

The equations we used to investigate speciation under the HAL model rely on several simplifications, in particular a deterministic approximation (stochastic processes are modelled by ordinary differential equations), an assumption of linkage equilibrium (the allele frequency at a locus is assumed to be independent of the allele frequency at the other loci), a rare allele approximation (the frequency of alleles at polymorphic sites is assumed to be close to 0 or 1 most of the time), and additional approximations detailed in Materials and methods in the case with local adaptation. In reality, the stochasticity around the equilibrium can play a role in the duration of speciation, especially in small populations where the random reproduction of a small number of individuals can have a strong effect on the dynamics of the whole population. This is exacerbated in the case with migration, as the few individuals who migrate may or may not be compatible with individuals from the receiving population, with important consequences for the speciation process. Furthermore, the assumption of linkage equilibrium is not necessarily met, for example if the recombination rate is low. Gavrilets (1999) performed simulations without recombination and interpreted observed discrepancies between predictions and simulations as the effect of linked loci. We implemented here stochastic individual-based simulations that can account for recombination. To assess the conditions under which the equations are valid and the effect of the linkage equilibrium assumption, we compared the outcomes of the equations with simulations including (or not) recombination. Our simulation process is detailed in figure S7. figure S8 and figure S9 report the results of the comparison with "intermediate" ($N = 600$) and small ($N = 200$) population sizes, which are then summarized in figure S10. With larger population sizes, we expect predictions to be closer to simulations as the deterministic approximation is more valid.

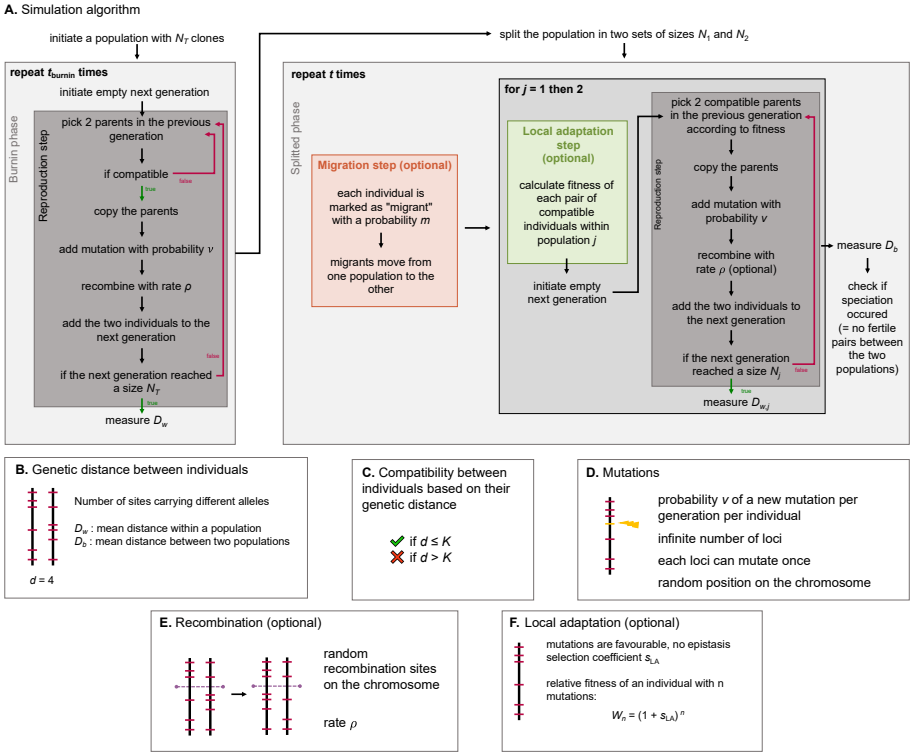
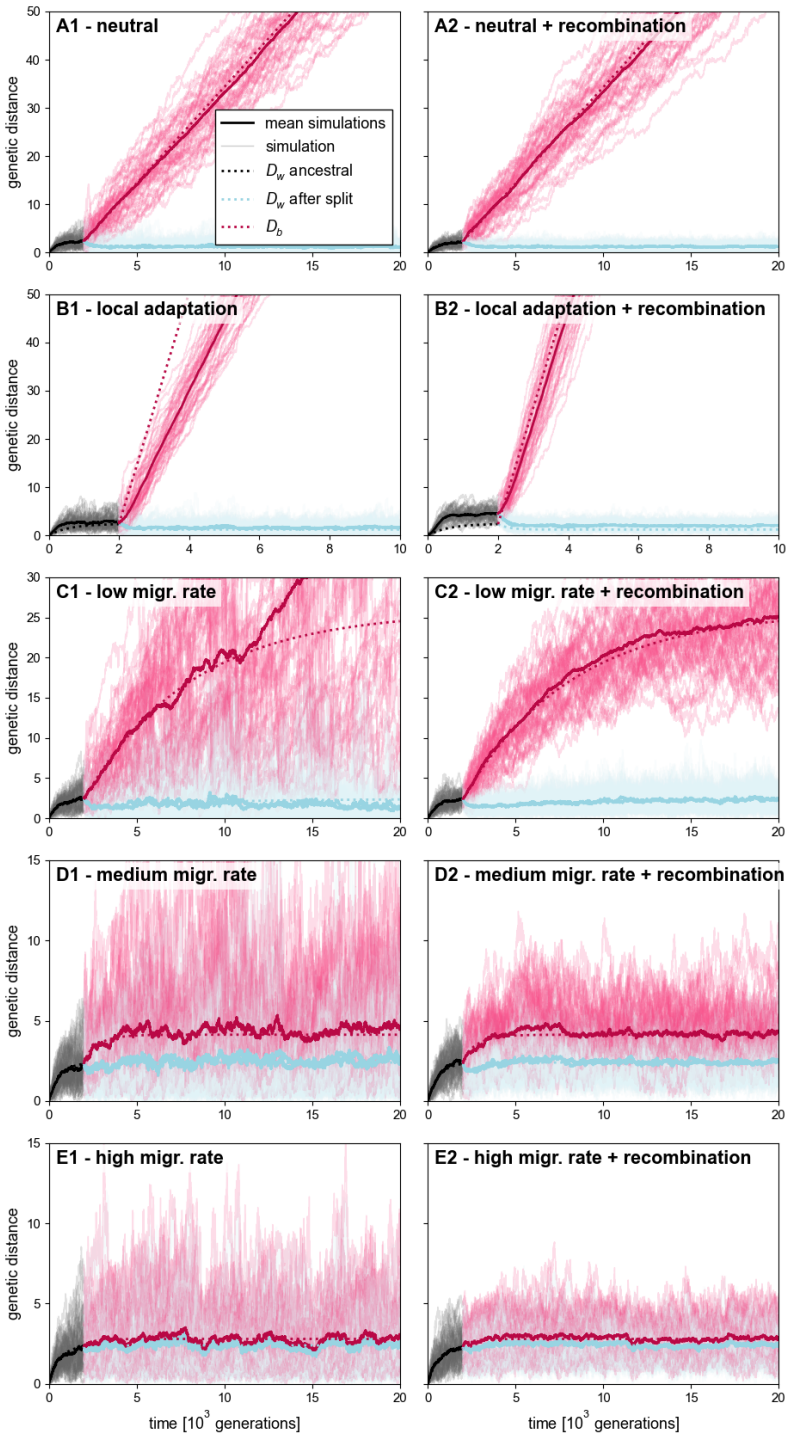


Figure S7. Illustration of the simulation algorithm (**A**) and the several components of the process (**B** to **F**). **A**: A burnin phase is used to allow the ancestral population to reach an equilibrium polymorphism. After this phase, the populations are splitted randomly . **B**: The distance between individuals is calculated as the number of sites carrying different alleles. **C**: The HAL model assumes that individuals can reproduce if their genetic distance is smaller than the incompatibility threshold. **D**: At each generation, each individual can mutate with probability μ. In this case, a site that never carried a mutation is altered. A gene cannot undergo the same mutation twice independently. **E**: For simulations with recombination, a part of the chromosome is randomly exchanged between the two offsprings of a pair of parents. **F**: For simulations with local adaptation, the fitness of the each compatible pair of individuals within each populations is calculated and is used to sample the parents during the reproduction step.



600

Figure S8. Dynamics of polymorphism and divergence under the holey adaptive landscape model obtained with the stochastic simulations (50 simulations for each configuration, solid thin lines ; the solid thick lines show the mean of the simulations) and with the deterministic approximations (dotted lines). The first column shows simulations without recombination and the second column shows simulations with recombination (recombination rate $\rho = 2$). The first row (**A**) corresponds to the allopatric neutral model, the second row (**B**) corresponds to the allopatric model with local adaptation ($s_{\mathrm{LA}} = 0.005$) and the last third rows (**C**, **D**, **E**) correspond to the model with low, medium and high migration (migration rate $m = 8.5 \times 10^{-5}, 0.001155$ and 0.00498 respectively). Parameters of the model: $\nu = 0.002, N_1 = N_2 = 300, K = 40$, burnin time 2000 generations.

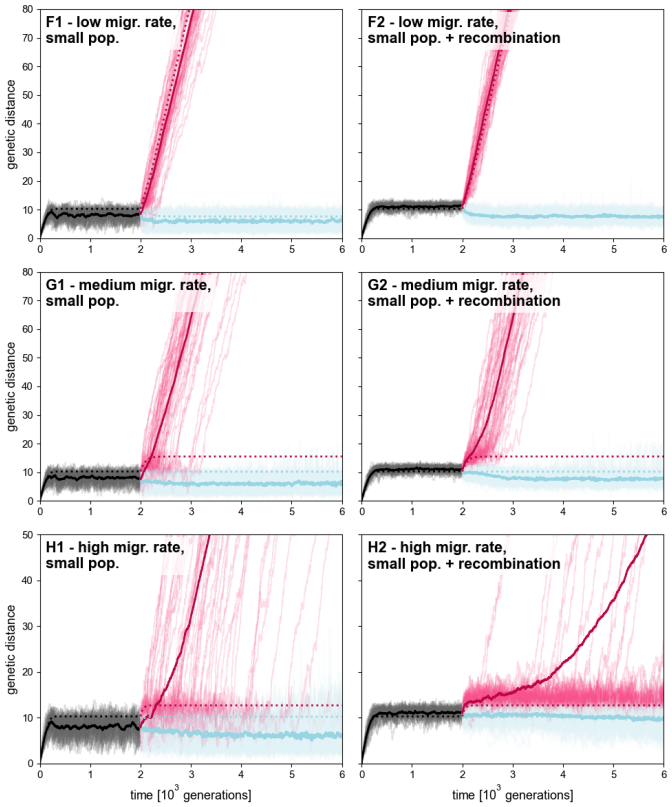


Figure S9. Same as figure S8 but with smaller populations : $\nu = 0.0384$, $N_1 = N_2 = 100$, $K = 20$ and migrations rates 0.001, 0.005 and 0.01 for rows **F**, **G** and **H** respectively.

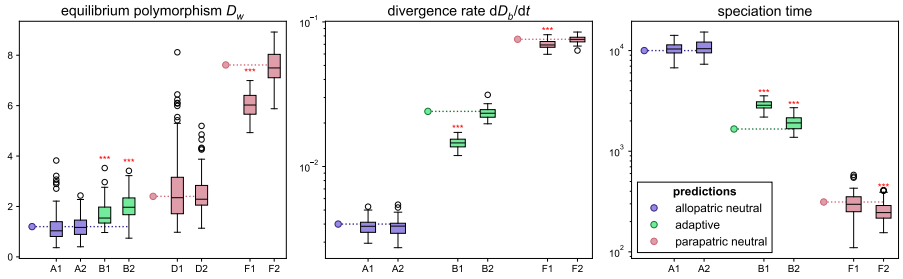


Figure S10. Comparison of the dynamics of the simulations and the predictions, based on (left) the equilibrium polymorphism D_w, (middle) the slope of $D_b(t)$ and (right) the duration of speciation. The boxplots indicate the distribution of the estimates from 50 simulations for each configuration, and the dot the expected value based on the deterministic prediction. The configurations (A1, A2, B1, B2, D1, D2, F1 and F2) are those of figures S8 and S9. The red asterisks indicate significant differences, assessed with *t*-test and corrected for multiple testing.