INSTITUT DE BIOLOGIE DE L'ÉCOLE NORMALE SUPÉRIEURE – PSL
ÉCOLOGIE SOCIÉTÉ ÉVOLUTION – UNIVERSITÉ PARIS-SACLAY
ÉCOLE DOCTORALE DIVERSITÉS ORIGINES NATURES – MNHN

Numéro attribué par la bibliothèque

_____

THÈSE

pour obtenir le grade de

DOCTEUR DU MUSÉUM NATIONAL D'HISTOIRE NATURELLE

Spécialité : Sciences de l'évolution

soutenue publiquement à Paris, le 25 juin 2026, par

Pierre Veron

Fossé micro–macro dans la recherche sur la spéciation : identifier les processus micro-évolutifs qui expliquent les différences de diversification entre lignées

Thèse dirigée par Hélène Morlon et Tatiana Giraud

JURY

Violaine Llaurens	Directrice de recherche *Collège de France, Paris*	Rapporteuse
Rampal Etienne	Professor *University of Groningen*	Rapporteur
Nicolas Bierne	Directeur de recherche *Université de Montpellier*	Examinateur
Guillaume Achaz	Professeur des universités *Université Paris-Cité*	Examinateur
Hélène Morlon	Directrice de recherche *École normale supérieure PSL, Paris*	Directrice de thèse
Tatiana Giraud	Directrice de recherche *Université Paris-Saclay, Gif-sur-Yvette*	Directrice de thèse

Title. Bridging the micro–macro gap in speciation research: identifying microevolutionary processes that explain differences in diversification between lineages

Keywords. macroevolution, microevolution, speciation, models

Abstract. Speciation is the process by which one species gives birth to two distinct species. This process is the source of biodiversity on Earth. Phylogenetic estimates of speciation rates have revealed a high heterogeneity across lineages. Given that speciation is the large-scale consequence of microevolutionary processes, we expect to explain the variation in speciation rates by differences in microevolutionary mechanisms.

This thesis aims to better understand the link between between microevolutionary processes and macroevolutionary speciation rates, using mathematical models and empirical data. In the first chapter, I use a speciation model based on the prediction of the genetic divergence and polymorphism between populations under different scenarios. This model predicts the duration of speciation and and the shape of the grey zone, i.e., when at which speed between-populations compatibility drops. By varying parameters that describe species traits, evolutionary processes, and speciation scenarios, I assess how these factors shape speciation dynamics. I show that the link between population size and speciation duration depends on the speciation scenario, in particular that faster speciation between smaller populations is typical of non-ecological speciation. A positive correlation between population size and speciation duration in plants, which I demonstrate using previously inferred demographic parameters on genomic datasets, suggests that speciation occurs primarily through a non-ecological scenario, i.e., that adaptation to different environments is not the main source of reproductive isolation.

In the second chapter, I focus on the influence of the different steps of speciation on the macroevolutionary process. I derive, mathematically, "equivalent" speciation and extinction rates arising from the protracted birth–death model. I show that the speed at which two distinct lineages become new species has in general relatively little influence on the equivalent speciation rate. For parameters estimated on a family of Australian lizards, my analyses suggest that the most limiting steps in speciation are the formation of populations and their survival, rather than the speed of speciation completion.

In this thesis, I emphasise that the micro–macro link depends on the scenario and context of speciation. Using theoretical models, I highlight the biological processes and the steps of speciation that seem the most relevant to explain the observed differences in speciation dynamics and speciation rates. Applying these models to empirical data makes it possible to quantify the relative influence of these processes. My thesis contributes to our understanding of how speciation occurs and what factors influence its rate at the macroevolutionary scale.

Titre. Fossé micro–macro dans la recherche sur la spéciation : identifier les processus micro-évolutifs qui expliquent les différences de diversification entre lignées

Mots-clés. macroévolution, microévolution, spéciation, modèles

Résumé. La spéciation est le processus par lequel une espèce donne naissance à deux espèces distinctes. Ce processus est à l'origine de la biodiversité sur Terre. Les estimations phylogénétiques des taux de spéciation ont montré que la fréquence des évènements de spéciation est très variable entre lignées. La spéciation étant la conséquence à large échelle de processus micro-évolutifs, on peut espérer expliquer la variabilité des taux de spéciation par des différences de mécanismes micro-évolutifs.

Cette thèse vise à mieux comprendre le lien entre entre processus micro-évolutifs et taux de spéciation à l'échelle macro-évolutive, en utilisant des modèles mathématiques et des données empiriques. Dans un premier chapitre, j'utilise un modèle de spéciation fondé sur la prédiction de la divergence et du polymorphisme génétiques entre population selon différents scénarios. Ce modèle prédit la durée de la spéciation et la forme de sa zone grise, c'est-à-dire quand et à quelle vitesse la compatibilité entre les populations diminue. En faisant varier des paramètres caractérisant les espèces et leur évolution ainsi que les scénarios de spéciation, j'évalue comment ces facteurs influencent la dynamique de la spéciation. Je montre que le lien entre taille de population et durée de spéciation dépend du scénario de spéciation, notamment que la spéciation rapide entre petites populations est caractéristique de la spéciation non-écologique. Une association positive entre taille de populations et durée de spéciation chez des plantes, que je mets en évidence grâce à des paramètres démographiques précédemment inférés, suggère que la spéciation a principalement lieu selon un scénario non-écologique, c'est-à-dire que l'adaptation à des environnements différents n'est pas la source principale de l'isolement reproducteur.

Dans un deuxième chapitre, je m'intéresse à l'influence des différentes étapes de la spéciation sur le processus macro-évolutif. Je calcule, mathématiquement, les taux de spéciation et d'extinction « équivalents » qui résultent d'un processus de naissance–mort retardé (c'est-à-dire, qui prend en compte le caractère non instantané de la spéciation). Je montre que la vitesse à laquelle deux lignées séparées deviennent de nouvelles espèces a en général une influence relativement faible sur le taux équivalent de spéciation. Pour les valeurs de paramètres estimées sur une famille de lézards d'Australie, mes analyses suggèrent que les étapes les plus limitantes dans le processus de spéciation sont la formation de populations et leur survie et non la vitesse d'achèvement de la spéciation.

Dans cette thèse, je souligne que le lien micro–macro dépend du scénario et du contexte de spéciation. En combinant modèles théoriques et données empiriques, je mets en évidence les processus biologiques et les étapes qui semblent les plus pertinents pour expliquer les différences observées dans les dynamiques et les taux de spéciation. Ma thèse contribue à mieux comprendre comment la spéciation se produit, et ce qui module la fréquence des évènements de spéciation à l'échelle macro-évolutive.

Contents

Acknowledgements — Remerciements

I would like to thank the members of the jury, who kindly accepted to evaluate my work. Violaine Llaurens, Rampal Etienne, Nicolas Bierne and Guillaume Achaz, it is a great honour to have you at my defense.

Je remercie mes directrices de thèse, Hélène Morlon et Tatiana Giraud. Hélène, lorsque je t'ai contactée pour construire un projet ensemble, tu m'as manifesté une grande confiance et tu m'as accordé beaucoup de liberté. Je te remercie de m'avoir partagé ta vision du métier de chercheur. Tu as fait preuve de beaucoup de patience pour expliquer la macroévolution à un étudiant pour qui tout était nouveau dans ce domaine, je t'en remercie chaleureusement. Tatiana, je suis reconnaissant pour tous tes conseils toujours très pertinents en génétique et évolution, pour ta relecture méticuleuse de chacun de mes travaux, pour ta gentillesse et ta disponibilité. Nos discussions vont me manquer.

Je souhaite remercier les membres de mon comité de suivi. Amandine Véber, je te remercie d'avoir pris le temps de m'aiguiller pour certaines questions de probabilités. Bert van Bocxlaer, je te remercie pour tes remarques sur mon travail, et pour l'invitation à Lille. Merci enfin à Nathalie Machon, responsable du doctorat au MNHN pour ta gentillesse et tes conseils.

Cette thèse a été possible grâce au soutien financier de l'École polytechnique et de l'Institut des mathématiques pour la Planète Terre. Je remercie les personnes de ces institutions qui ont accepté de financer mon projet.

J'adresse des remerciements chaleureux à toutes les personnes de l'IBENS sans qui ces quatre années n'auraient pas été aussi agréables. Benoît, ça a été un grand plaisir de discuter avec toi, de parler de recherche et d'enseignements et même d'enseigner ensemble. Olivia, I already miss your joyful presence at the lab and our discussions. Rosario, sunshine, I loved the moments we spent together, thank you for your support and your kind attentions. Joëlle, j'ai beaucoup aimé animer la fête de la science avec toi, et je te souhaite tous mes vœux de bonheur. Agathe, je te remercie pour ton aide

précieuse pour décrypter les articles de mathématiques parfois obscurs. J'ai pris du plaisir à travailler sur ce modèle de spéciation avec toi. J'adresse une pensée chaleureuse à Anaïs : travailler avec toi dans le cadre de ton stage a été une grande joie ! Jérémy, enfin, j'ai beaucoup apprécié t'avoir comme compagnon de thèse, que ce soit pour tes recommandations de cinéma, pour nos discussions passionnantes et pour le projet que nous avons eu ensemble. Merci pour tout cela ! I am grateful to Sophia, Viktor, Nils, Lucas, Mahwash, César, Kenneth, Paul, Amélie, Minghao, Mathilde, Octavio, Walker and all the (past) colleagues of the *Biodiv* team. Nina, j'ai adoré préparer le cours de mathématiques pour les biologistes avec toi ! Juan, organiser le séminaire de section avec toi était également un plaisir. Amaury, merci de faire vivre le département de biologie. Ton enthousiasme pour mon travail et ton invitation à présenter dans ton équipe m'ont touché.

J'aimerais également remercier les collègues de l'IDEEV, que j'avais le plaisir de retrouver presque tous les jeudis, après un périple pour atteindre le plateau de Saclay ! Élise, merci pour ta bonne humeur, tes encouragements et tes conseils. Quentin, je te remercie de m'avoir aidé sur des questions de génétique, et pour m'avoir fait part de tes réflexions sur la spéciation. Ton expertise sur les modèles démographiques m'a également beaucoup aidé. Elsa, quelle joie d'avoir avancé ensemble sur nos thèses, d'avoir partagé nos expériences de la recherche et de la vie. J'espère que nous continuerons de nous voir souvent ! Je garde un excellent souvenir de notre semaine à Barcelone, avec Loreleï et Alice également. Merci à Lou, Jeanne et Gabriela pour l'accueil dans le nouvel open-space. Je remercie Guislaine, ma marraine de thèse : tu as fait preuve d'une gentillesse et d'un soin très appréciés lors des moments plus difficiles. Jacqui : merci pour tous les délices, les conseils à la fois scientifiques et culinaires et la sortie champignons. Je pense également aux personnes qui font et ont fait vivre le labo : Adrien, Jeanne, Ricardo, Fanny, Agathe, Alodie, Jean-Philippe. Merci Mathis pour ta gentillesse et ton écoute. Arthur, merci d'avoir été un compagnon adorable à Roscoff et, avec Thomas, Maxence et d'autres, de faire de l'IDEEV un institut si plaisant.

J'adresse un remerciement aux personnes qui font fonctionner les universités. En particulier, je remercie Eleni, Béatrice, Patricia, Virginie (à l'IBENS), Zhor, Wisa (à l'IDEEV) et Mahjouba (au Muséum).

Je souhaite remercier toutes les personnes croisées au détour de conférences, visites et séminaires avec qui j'ai pu avoir des discussions très enrichissantes. Achille Lenglin, Camille Roux, Christelle Fraïsse et Xavier Vekemans, merci à vous pour votre accueil chaleureux à Lille, ça a été un plaisir de vous revoir à Roscoff. Votre expertise dans chacun de vos domaines m'a été d'une grande aide. Merci aux personnes en charge de la chaire « modélisation mathématique et biodiversité » qui organisent l'école de recherche à Aussois qui est chaque année l'occasion de rencontrer des personnes passionnantes.

Je remercie Séverine sans qui je n'en serais pas là aujourd'hui.

J'ai de la chance de pouvoir compter sur mes ami·e·s, qui m'ont apporté joie, respirations et conseils pendant cette thèse. Merci aux potes du lycée Faidherbe à Lille que j'ai eu le plaisir de revoir à mon arrivée à Paris : Chloé, Adel, Camille, Louis, Antoine. Dear Jakob, thank you for warming up by your presence our first freezing flat in Paris! Juan, merci pour ta tendresse et ton humour ! Je souhaite remercier aussi les merveilleux colocataires qui m'ont soutenu, nourri et chouchouté : Nico, Alice, Tristan, Marie, Luana, Tiffanie, Jaime et Catalina, vous êtes formidables. J'ai une pensée chaleureuse pour Chloé : t'avoir dans ma vie depuis le bar d'étage et la coloc en Suisse est une chance. Je remercie également Sébastien qui m'a accueilli chez lui dans le Sud et sur qui j'ai pu compter à tout moment. Maureen, tu es présente dans ma vie depuis si longtemps, et c'est toujours chouette de te retrouver.

Je souhaite remercier mes parents, Lisa, Loïc, Delphine et Gilles, pour votre soutien pendant toutes mes études jusqu'à maintenant.

À mon cher Thibault, qui depuis le début croit en moi et en notre relation. Merci d'avoir été à mes côtés tout le long du chemin, en dépit des difficultés. Je souhaite que notre relation continue de nous apporter joie, écoute, enthousiasme et découvertes.

Résumé détaillé en français

La spéciation est le processus par lequel une espèce donne naissance à deux espèces distinctes. Ce processus est à l'origine des millions d'espèces qui peuplent la planète Terre, mais il est encore mal compris. L'utilisation de modèles macro-évolutifs permet de découvrir des caractéristiques de la diversification des espèces, mettant notamment en évidence le fait que les évènements de spéciation ne sont pas répartis uniformément sur toutes les lignées. On observe en effet une hétérogénéité des taux de spéciation, avec des groupes d'espèces qui se diversifient beaucoup plus vite que d'autres. Dans la mesure où les évènements de spéciation sont la conséquence à large échelle de processus micro-évolutifs, c'est-à-dire de processus d'évolution au sein des espèces, on s'attend naturellement à expliquer cette variabilité des taux de spéciation par des différences de mécanismes micro-évolutifs entre les espèces. Par exemple, la vitesse d'accumulation d'isolement reproducteur entre deux populations, ou la capacité de dispersion des individus, devraient être des prédicteurs des taux de spéciation.

Il existe cependant différents contextes de spéciation, ce qui rend difficile la prédiction des taux de spéciation en se fondant sur les processus micro-évolutifs. On distingue trois grands contextes géographiques de spéciation selon le niveau de connexion entre les populations qui deviendront des espèces : allopatrique, où les populations sont séparées par une barrière géographique, sympatrique, où les populations vivent au même endroit, et parapatrique, où les populations sont séparées par une barrière qui permet quelques échanges. Ce contexte géographique influence le lien potentiel entre processus micro-évolutifs et taux de spéciation. Un autre facteur d'influence est l'adaptation locale à l'environnement, qui peut aboutir à l'isolement reproducteur par « spéciation écologique ». De plus, la plupart des prédictions sur les liens possibles entre caractéristiques des espèces et taux de spéciation restent spéculatifs. Certaines de ces prédictions sont parfois contradictoires, par exemple la capacité de dispersion des individus devrait diminuer l'isolement des populations, et donc la formation d'espèces, mais peut en même temps favoriser la survie de ces populations (et donc leur chance de se différencier jusqu'à devenir une nouvelle espèce), en leur permettant de coloniser des environnements.

Des approches cherchant des corrélations entre caractéristiques micro-évolutives et spéciation existent, mais leurs résultats sont souvent mitigés et ne permettent pas de généraliser un effet. Ainsi, une étude a trouvé un lien positif entre la capacité d'espèces d'oiseaux des Amériques à former des populations génétiquement différenciées et le taux de spéciation estimé de ces espèces, cependant ce résultat positif n'a pas été retrouvé dans d'autres groupes comme des lézards ou des orchidées. Enfin, l'augmentation de l'isolement reproducteur semble décorrélée des taux de spéciation macro-évolutifs, comme cela a été montré chez des oiseaux et des drosophiles, quand bien même cette étape est essentielle dans le processus de spéciation. Le recours à des modèles théoriques est dès lors utile pour clarifier sous quelles conditions on peut s'attendre à trouver un lien entre paramètres micro-évolutifs et taux de spéciation.

Cette thèse vise à mieux comprendre le lien entre micro- et macroévolution en utilisant des modèles mathématiques de spéciation. Dans un premier chapitre, j'utilise un modèle de spéciation qui prédit l'évolution de deux populations séparées, éventuellement connectées par de la migration, et dans lequel on fait l'hypothèse que deux individus peuvent se croiser et avoir une descendance fertile si leurs génotypes sont assez proches. Ce modèle, appelé « paysage adaptatif troué », permet de prédire d'une part la durée de spéciation, c'est-à-dire le temps nécessaire pour atteindre un isolement reproducteur total entre les populations, et d'autre part la zone grise de spéciation, c'est-à-dire l'évolution de leur compatibilité pendant cette durée. En faisant varier des paramètres caractérisant les espèces et leur évolution (taille de population, architecture génomique des barrières au flux de gènes, taux de mutation) ainsi que les scénarios de spéciation (allopatrie, parapatrie, spéciation écologique), je prédis l'effet de ces paramètres sur la dynamique de spéciation. Je montre que le lien entre taille de population et durée de spéciation dépend du scénario de spéciation que l'on considère, notamment que la spéciation rapide entre petites populations est caractéristique de la spéciation non-écologique. En étudiant l'histoire démographique précédemment inférée sur des données génomiques de plantes, je mets en évidence une association positive entre les tailles de population et la durée de spéciation entre des paires d'espèces proches. Ce résultat suggère que la spéciation chez les plantes dans ce jeu de données se produit principalement selon un scénario non-écologique, c'est-à-dire que l'adaptation à des environnements différents n'est en général pas la source de l'isolement reproducteur.

La zone grise de la spéciation, c'est-à-dire la dynamique de compatibilité entre des populations qui ne sont pas encore des espèces, mais qui ne sont pas parfaitement compatibles, connaît un intérêt scientifique croissant, notamment en raison de la présence d'hybrides entre lignées et de la capacité computationnelle d'inférer l'histoire évolutive de ces lignées. Dans ce même chapitre, j'étudie la forme de cette zone grise en fonction des paramètres biologiques et des scénarios de spéciation. En particulier, je mets en évidence l'importance de faire une distinction entre zone grise temporelle et zone grise

génomique. La première étudie l'évolution de la compatibilité entre populations au cours du temps tandis que la deuxième étudie la compatibilité en fonction de la divergence génétique entre ces populations. Dans les deux cas, la forme de la zone grise dépend des paramètres biologiques, mais le scénario de spéciation n'influence pas l'évolution de la compatibilité entre populations en fonction de la divergence, alors qu'il influence sa dynamique temporelle. Ce résultat indique qu'utiliser la divergence génomique comme *proxy* du temps peut conduire à des interprétations fausses, et il explicite quels processus influencent ou non la forme de cette zone grise.

Dans un deuxième chapitre, je m'intéresse plus spécifiquement à l'influence des différentes étapes de spéciation sur le processus final macro-évolutif. Pour cela, j'utilise le modèle de spéciation « naissance–mort retardé » (*protracted birth–death*) qui tient compte du fait que la spéciation n'est pas instantanée. Dans ce modèle, on se focalise sur les lignées, et on considère que la spéciation a lieu si une lignée donne naissance à une nouvelle lignée (on appelle cela l'initiation de la spéciation) et si cette lignée devient une nouvelle espèce (on appelle cela l'achèvement de la spéciation). Le processus prévoit aussi que la lignée puisse s'éteindre. Dans un temps fini donné, une espèce peut donc s'éteindre, donner naissance à une deuxième espèce ou bien simplement survivre sans évènement de spéciation. En étudiant les probabilités des trois différentes issues ainsi que l'espérance conditionnelle du temps d'attente de ces évènements, je calcule des taux effectifs de spéciation et d'extinction. Ces taux sont définis comme les taux d'un processus de naissance–mort classique qui aurait les mêmes propriétés mathématiques que le processus de naissance–mort retardé.

En regardant l'influence relative de chacun des paramètres utilisés dans le processus de naissance–mort retardé sur les taux effectifs de spéciation et d'extinction, je développe une méthode qui permet de déterminer quelle est l'étape la plus limitante dans le processus final de spéciation. Je montre ainsi que le taux d'achèvement de la spéciation, c'est-à-dire la vitesse avec laquelle deux lignées déjà séparées deviennent des nouvelles espèces, a une influence relativement faible sur le taux de spéciation équivalent. En appliquant cette méthode aux valeurs des paramètres estimées par une étude précédente sur une famille de lézards d'Australie, je montre que l'étape la plus limitante dans le processus de spéciation est l'initiation, suivie de la survie. L'achèvement de la spéciation, bien qu'essentielle, a une influence très faible sur le taux de spéciation macro-évolutif. Ce chapitre met en évidence l'importance de considérer toutes les étapes de la spéciation si l'on veut expliquer pourquoi les taux de spéciation sont si variables.

L'hétérogénéité des taux de spéciation interroge, dans la mesure où l'on n'explique pas encore pourquoi certains groupes se diversifient plus rapidement que d'autres. Lier la micro- et la macroévolution devrait permettre d'apporter une réponse à cette question étant donné que la spéciation émerge de processus micro-évolutifs. Dans cette thèse, je souligne que ce lien dépend du scénario et du contexte de spéciation. En utilisant

des modèles théoriques, je mets en évidence les processus biologiques et les étapes qui semblent les plus pertinents pour expliquer les différences observées dans les dynamiques de spéciation. L'application de ces modèles à des données empiriques permet de quantifier l'importance relative de ces processus et de répondre à des questions importantes sur comment la spéciation se produit, et ce qui module sa vitesse à l'échelle macro-évolutive.

Introduction

Speciation is the process by which a single species gives rise to two distinct species. This process is at the origin of the biodiversity on Earth, with millions of existing species and is therefore fundamental in biology. Speciation can be thought as by a split of one evolutionary lineage into two different lineages, a common representation in the field of systematics where phylogenetic trees describe the relationships between species. This perspective, called the macroevolutionary view of speciation, is useful to study the generation of biodiversity over long time scales, resulting in its current distribution in different parts of the tree of life and regions of the globe. However, speciation is not an instantaneous process and describing it as dichotomous splits is not adapted to the study of the population-level (microevolutionary) processes that lead to speciation.

The combination of macroevolutionary models and phylogenetic data allowed us to uncover patterns of diversification, in particular that speciation events are not equally frequent depending on the evolutionary groups or regions. This "speciation rate heterogeneity" should be explained by differences in the microevolutionary mechanisms of speciation, because microevolution underlies macroevolution events. The variability of contexts of speciation makes it challenging to generalise how speciation occurs and may prevent finding an empirical link between microevolutionary processes and macroevolutionary speciation rates. Many predictions of what should explain the variability of speciation rates exist, such as the speed of accumulation of reproductive isolation between populations, or the population sizes, but they often remain purely verbal. Theoretical models describing this scaling-up can be useful to clarify which links are expected (or not) between microevolutionary parameters and speciation rates.

1 MICROEVOLUTIONARY VIEW OF SPECIATION

In the microevolutionary perspective, speciation is the consequence of all the mechanisms that would prevent individuals from different populations to reproduce.

1.1 Reproductive isolation

For sexual organisms, the probability of successful reproduction between individuals decreases with their genetic distance (e.g., Castillo 2017; Turissini et al. 2017; Christie and Strauss 2018; Christie et al. 2022). This phenomenon is described as reproductive isolation (RI*). RI can take different forms depending on the organisms, and these forms can sometimes coexist: pre-mating RI that prevents genetically distant individuals from mating (e.g., mate preference, assortative mating; Freeman et al. 2022), postmating prezygotic RI involving a form of gamete selection favouring closely-related individuals (Garlovsky et al. 2024), and postzygotic RI involving a reduced fitness despite successful crossing (e.g., inviable zygote, developmental defects or fertile offspring; Reifová et al. 2023). When this last mechanism is a result of a maladaptation to a given environment, it is called extrinsic postzygotic RI (e.g., the wing patterns of hybrid *Heliconius* butterflies increase their risk of being caught by predators; Merrill et al. 2012); when this fitness is intrinsic to the organisms, it is called intrinsic postzygotic RI (Ravigné et al. 2010; e.g., pollen killer mechanisms in some plants, which reduce the fertility of male hybrids, Simon et al. 2016).

1.2 Modelling the dynamics of speciation

1.2.1 Dobzhansky-Muller incompatibilities

The achievement of RI between distinct populations takes time and the characterization of its accumulation over time is expected to help understanding how and at which pace speciation occurs. The classical view explaining the apparition of intrinsic postzygotic RI, is through genetic incompatibilities (also called genetic barriers). The model of Dobzhansky-Muller incompatibilities (DMIs), named after Dobzhansky (1936) and Muller (1942), also called Bateson-Dobzhansky-Muller incompatibilities (Bateson 1909), explain how genetic barriers appear between distinct populations as a result of epistatic genetic interactions.

In the simplest version of the model, two genes are involved and different alleles fix at each of those genes (either neutrally or due to adaptation; Unckless and Orr 2009) in the two populations, in an order such that certain combinations of alleles have never occurred in any population (see figure 1) and have therefore not been tested by natural selection. In the case of a secondary contact, the combination of alleles in a hybrid genotype can be disadvantageous or even lethal, i.e., causing an incompatibility. These incompatibilities can involve conflicts between mitochondrial and nuclear DNA, but incompatibilities within the nuclear DNA seem more common (Burton 2022).

* All acronyms used are listed on page 142.

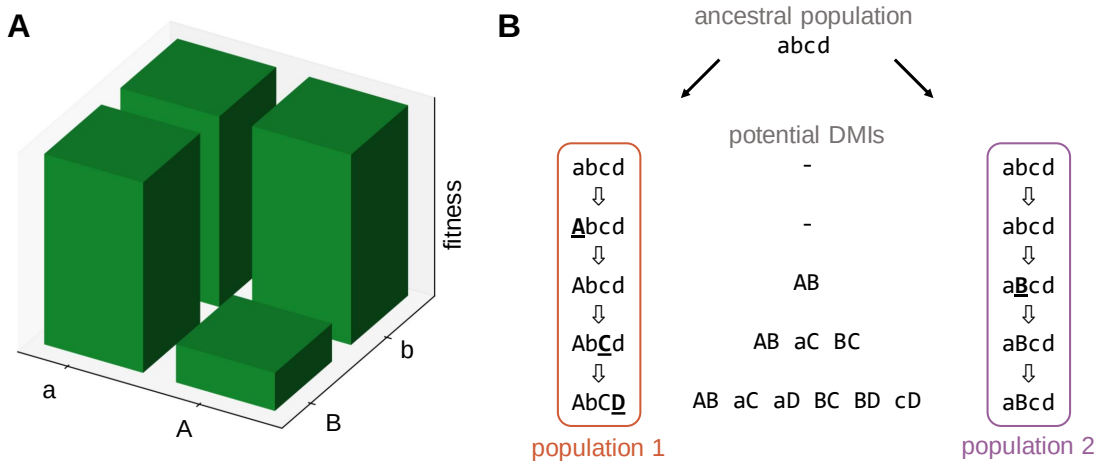


Figure 1 – Dobzhansky-Muller incompatibilities (DMI). A. Principle of a two-loci DMI on the combination of allele `AB`. The allelic combinations `ab`, `aB` and `Ab` have the same fitness, such that `A` and `B` can fix in distinct populations, but hybrids of these populations with allelic combination `AB` are unfit. **B.** Accumulation of potential DMIs between two separate populations. Here, 4 loci are represented. At each step, one of the two populations has a new substitution on a new locus (the allele is underlined), and the combinations of two substitutions that were never together in any population are considered as potential DMIs. In this model described in Orr (1995), the number of potential DMIs accumulates quadratically through time.

1.2.2 Speed of accumulation of DMIs

Using the DMI model between two separate populations, Orr (1995) predicted the dynamics of incompatibilities. In this model, populations accumulate substitutions either neutrally or by natural selection. Given that there is a large number of possible substitutions, it is assumed that each new substitution alters a site that was never altered before. Hence, each gene has two possible alleles, the ancestral allele (indicated by a lowercase letter in figure 1) and the altered allele (indicated by an uppercase letter). An example of an ancestral population with a genotype `abcd` that evolve in two populations with genotypes `AbCD` and `aBcd` is illustrated in figure 1B. Given that each substitution appears after the previous one is fixed, it is potentially incompatible with all the substitutions that previously occurred in the other population. Consequently, the k^{th} substitution can potentially cause $k-1$ DMIs. If the number of substitutions occur regularly with time and each potential DMI has the same probability of being one, the number of substitutions is thus proportional to the square of the time after split (Orr 1995; Orr and Turelli 2001). This expectation, called the "snowball" theory because of the acceleration of DMIs with time, has been largely debated, both empirically and theoretically.

Direct measures of DMIs suggest that they accumulate faster than linearly, for instance in *Drosophila* (Matute et al. 2010) or *Solanum* (Sherman et al. 2014). But this kind of test is not frequent, because it requires identifying the loci responsible for the DMIs, which is done for few organisms and requires a good knowledge of the genetic background of those incompatibilities (Städler et al. 2011). This is done, for instance, by crossing heterozygous individuals from different populations or species, and mapping the genotypes that survive or do not survive (Presgraves et al. 2003). Using stochastic simulations, Maya-Lastra and Eaton (2021) noted that the snowball effect is a consequence of two major assumptions of Orr's model: the infinite number of sites and the absence of demography. The first assumption states that each new mutation alters a new locus and that no locus mutates independently in the two populations. The second assumption is that the population is finite, i.e., the sites causing DMIs are never polymorphic within populations and are thus never subject to purifying selection. Indeed, as some combination of alleles are likely to cause DMIs, the genotype carrying these combinations may be selected against within a population. Maya-Lastra and Eaton (2021) showed that relaxing one of these assumptions actually cancels the acceleration of accumulation of DMIs with time, but introduces more variability in the speed of accumulation of incompatibilities compared to Orr's model. Their simulations also show that the beginning of the accumulation of DMIs can look like a quadratic function, which still raises the question of the shape of the accumulation of genetic incompatibilities, possibly explaining why the empirical studies focusing on few DMIs find a probable snowball.

1.2.3 Speed of accumulation of RI

Other studies do not focus directly on the DMIs but look at the empirical link between the genetic differentiation of populations and the sterility of hybrids or another measure of RI. In such studies, the effect — rather than the number — of incompatibilities is quantified. In a review, Gourbière and Mallet (2010) conclude that the "snowball" effect of RI in empirical data is not general. In *Drosophila*, the level of RI does not accumulate faster than linearly, but anurans (frogs) exhibit evidence of a snowball effect (Mendelson et al. 2004). In addition to the arguments of Maya-Lastra and Eaton (2021), there are mechanisms that could explain the absence of a snowball effect, such as reinforcement (see next paragraph) or the absence of epistatic effect (i.e., incompatibility only due to independent single genes as in two allele diploid model with underdominance; Gavrilets 2003).

Reinforcement is the process by which prezygotic barriers appear as a consequence of natural selection, in cases when hybrids have lower fitness (Coyne and Orr 2004). In this case, individuals avoiding mating with members of a different population have a better reproductive success than those who mate indistinctly (Marshall et al. 2002).

Gourbière and Mallet (2010) predict that reinforcement reduces the accumulation of substitutions responsible for RI after some time, because the hybrids are less abundant in the population and the selective pressure against them are weaker. In the latter case, we expect that the number of substitutions causing incompatibility would slow down. Mendelson et al. (2004) highlight that the absence of acceleration in the measured level of RI could be due to the effect of multiple incompatibilities: if DMIs accumulate faster than linearly but their effect is not additive, then we would not necessarily observe an acceleration of RI with respect to the genetic divergence.

The biological species concept

Working on speciation requires to use a practical definition of what we call a species. There are several definitions of a "species", and these definitions result in different boundaries (de Queiroz 2007), which in turn affect the estimation of macroevolutionary rates. In this thesis, we focus on speciation in sexually reproducing organisms and do not treat the question of speciation in viruses, bacteria or archea for instance. We will use the definition of a species by Ernst Mayr (1942), a "group of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups" (biological species concept). Although many other definitions of a species have been proposed (Mayden 1997), some being more adapted to non-sexual organisms or in palaeontology, we decided to use this definition because species emerging from RI is the most broadly used concept in speciation models, which is motivated by evidence of RI between diverging groups. A frequent criticism of the biological species concept is that RI is not systematically tested when defining a species. However, recent development of population genetics inference tools allowed to test whether gene flow occurs between putative species (Rougemont et al. 2017; Fraïsse et al. 2021; Burban et al. 2024). The detection of gene flow indicates that RI is not achieved, but its absence does not necessarily implies full RI because of the possibility of geographic barriers.

1.3 Modes of speciation

The geographical context of speciation is important to understand the forms of RI that accumulate between incipient species. In the DMI model by Orr (1995) for instance, the populations accumulating incompatibility must not be in contact, as they fix neutrally substitutions that would be disadvantageous if any hybrid would form. In the models with reinforcement however, the incipient species must be in contact to allow for natural selection to favour forms of assortative mating. Hence, biologists make distinctions

between the degrees of geographic isolation between the populations during speciation (Coyne and Orr 2004), that are called “modes of speciation”.

1.3.1 Allopatric speciation

In the allopatric mode of speciation (figure 2), speciation occurs after the formation of a geographical barrier isolating two populations. RI evolves without contact or gene flow between them. This barrier can be for instance the consequence of climatic events (glaciation, change of sea level) or geological ones (orogenesis, continental drift). For example, the formation of the Isthmus of Panama led to the allopatric speciation of a large quantity of marine organisms 3 millions years ago (Lessios 1998). Allopatric speciation is frequent in islands or archipelagos (Thorpe et al. 2010; Wang et al. 2016), but also on continents, for instance in fragmented habitats like mountains (Boucher et al. 2016) or lakes (April et al. 2013).

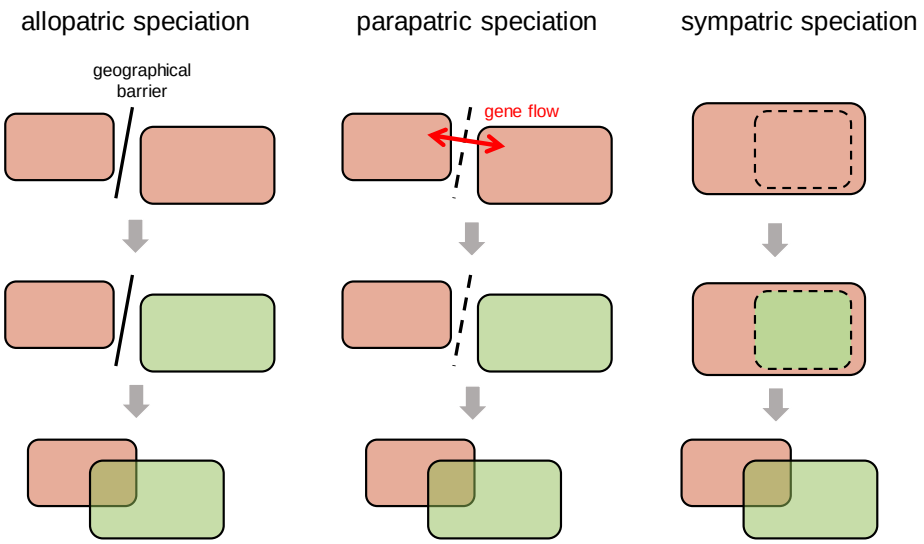


Figure 2 – Different modes of speciation.

1.3.2 Sympatric speciation

In sympatric speciation, RI emerges between groups that live in the same geographical place (figure 2), and the individuals from those groups could have the same opportunity of mating between or within their group (Gavrilets 2003). Kondrashov and Mina (1986) define sympatric speciation as a situation in which “the probability of mating between two individuals depends on their genotypes only”. This mode of speciation is not as obvious as the allopatric mode because we generally observe that closely related species live in different geographical zones, and it has long been considered as less likely from a

theoretical perspective. Indeed, models of sympatric speciation require another explanation for the non-random mating of individuals, such as a form of specialisation for resources driven by competition (Dieckmann and Doebeli 1999), environmental gradients (Doebeli and Dieckmann 2003), sexual selection or sexual conflict (Gavrilets and Waxman 2002; Bouinier et al. 2026) or a form of assortative mating (Higgs and Derrida 1992). As there is no external factor of isolation in sympatric models of speciation, an assumption of a form of isolating mechanism must be made to explain the formation of species (Papadopulos et al. 2011), as suggested by the evidence of faster apparition of intrinsic RI in taxa with overlapping ranges (Coyne and Orr 2004).

Assuming a form of isolating mechanism of populations in contact is less straightforward than the allopatric model, where divergence naturally emerges between the populations, but it does not mean that sympatric speciation is not possible in nature. In a review, Bolnick and Fitzpatrick (2007) compiled empirical evidence of sympatric speciation, for instance between cichlid fishes in small lakes or Lord Howe island palms. Sympatric speciation is also frequent for parasites or symbionts, because a form of specialisation in host directly results in reproductive isolation if the organisms mate in their host. Forbes et al. (2017) put together examples of phytophagous insects that tend to be highly specialized in species of plant, and forms that live or lay eggs on a different plant would be reproductively isolated. This mode of speciation is also frequent in fungi, for instance in Ascomycetes fungi responsible of plant parasitism, where mating occurs after the plant infection (Giraud et al. 2010). In those cases, specialisation allows to form new isolated lineages and cause speciation.

1.3.3 Parapatric speciation

Allopatric speciation requires that the two populations' ranges do not overlap and that there is no exchange between them. This latter assumption, sometimes called "strict allopatry", does not seem realistic for organisms with a large capacity of dispersion, or adjacent populations. The parapatric mode of speciation is a context in which two (or more) populations are geographically separated, but some gene flow exist between the two during the speciation process. The parapatric model is distinct from the sympatric model in the fact populations live in distinct geographic areas and between-populations mating, although possible, is less probable than within-population mating (Coyne and Orr 2004). This non-random mating is controlled by the propensity of the organisms to disperse, that is generally formalized with a migration rate (Gavrilets et al. 1998; Gavrilets 2000; Yamaguchi and Iwasa 2016), or by a mating neighbourhood in cases of spatially explicit models (de Aguiar 2017). Migration makes speciation more difficult compared to strict allopatry, because gene flow can homogenize the two pools and prevent the accumulation of reproductive barriers; on the other hand, migration allows the colonisation of new geographic regions, potentially triggering speciation. Modelling

the full speciation process (colonisation of new environment and evolution in parapatry), Yamaguchi and Iwasa (2016) found that there is an optimal migration rate maximizing the rate of formation of new species.

Coyne and Orr (2004) mention that the evidence of parapatric speciation is hard to find, notably because looking at species actual range does not allow to make a difference between allopatry and parapatry. The development of inference tools based on population genetic models now allows to distinguish between those cases. For instance, Niemiller et al. (2008) reconstructed the history of migration between species of subterranean and surface salamanders of the genus *Gyrinophilus* in Texas. They found that the model where the populations were isolated, but with migration, is best supported, providing evidence of parapatric speciation in this species complex. Distinction between parapatric and sympatric speciation may be difficult to make in practice, as there is a continuum between allopatric (no gene flow) and sympatric speciation (unlimited gene flow) with parapatry in the middle (Gavrilets 2003). Recent coalescent models have allowed to show that gene flow occurred frequently during speciation, for instance in plants (Papadopulos et al. 2011), *Drosophila* (Yusuf et al. 2026), fungi (Hartmann et al. 2020), or in marine groups like whales and rorquals (Árnason et al. 2018).

1.3.4 Likelihood of each mode

In a meta-analysis of range overlap among more than 3000 pairs of sister species, Hernández-Hernández et al. (2021) assessed the likelihood of the different modes of speciation among different groups (fungi, plants, animals). The assumption that is commonly made is that if sister pairs ranges do not overlap, speciation between the two was probably allopatric, otherwise it probably occurred in sympatry or parapatry. With this methodology, the authors found that allopatry was dominant in animals, but less in fungi (24%) and plants (30%). However, a limitation of such studies is that the current range overlap is only a partial information on the range overlap at the speciation phase. It is possible, for instance, that speciation occurred in allopatry and was followed by a dispersion to common areas, as on the illustration figure 2, or the opposite, as proven by Yusuf et al. (2026) in *Drosophila* where a majority of pairs of species that live today in allopatry show evidence of gene flow in the past. In fact, the analysis of geography alone only provides an idea of the past distribution of species, and the use of demographic models such as the one in Fraïsse et al. (2021) allows to recover the most likely scenario of speciation, see subsection 1.5.

1.4 Ecological versus non-ecological speciation

In all modes of speciation, the populations or groups must accumulate divergence (either genetically or in traits) so that speciation is complete. The processes explaining

the accumulation of RI is however debated and varies among lineages. In the scenario of ecological speciation, RI evolves between populations by "divergent natural selection arising from differences between ecological environments" (Schluter 2009). This hypothesis is ancient and was even intuited by Darwin (1859) who thought that natural selection has a role in the formation of species.

Alternatively, RI between populations can accumulate without divergent selection, a process called neutral, non-ecological or mutation-order (Mani and Clarke 1990) speciation. In this process, RI is caused by genetic divergence or genetic incompatibilities between two populations that evolve separated but in similar environments, or their accumulation is decoupled from the ecological selection of traits. As stated by Rundell and Price (2009), this process is possible, with some examples of close species living in similar ecological environments, or ecological divergence occurring after speciation.

Using trait difference between sister pairs in vertebrates, Anderson and Weir (2022) tested several models expected to reflect whether speciation is ecological or not. If speciation is ecological, traits should diverge and show adaptation around two optima, otherwise traits should converge around one optimum (if the traits are under selection) or evolve as a Brownian motion (if the traits are not under selection). They found that, over the tested traits for which a form of selection was detected, only a minority of species pairs (median: 9.9%) supported divergent selection. This result shows that the divergence observed between species is not due to divergent adaptation for the traits and the species considered here. Although this study does not specify the mechanism of RI in itself, the difficulty of finding traits evolving under divergent selective forces suggests that RI may not generally be the consequence of evolution under divergent selection. This study however does not prove the absence of divergent selection in all species, as in a few pairs of species some traits show a signal of divergent selection.

Ecological speciation is probably an existing phenomenon and was until recently considered as the default speciation mode, but its dominance may not be so obvious, at least in vertebrates. The question of this mode of speciation is important, as the evolutionary mechanisms behind the accumulation of RI may strongly change the predictions of the link between microevolutionary parameters and the speed of speciation, as we will see in section 3 of the introduction and in the rest of the thesis.

1.5 Population genetic inference methods can inform speciation dynamics

The demographic history of a species (e.g., changes in population sizes, events of population splits, migration) influence the frequency of alleles in populations. Inference methods such as ∂a∂i (Gutenkunst et al. 2010), fastsimcoal (Excoffier et al. 2021), DILS (Fraïsse et al. 2021), RIDGE (Burban et al. 2024) take advantage of this and allow to

reconstruct the past demography of species based on molecular data. These softwares use inference techniques such as approximate Bayesian computation (ABC) to estimate the parameters of the demographic models: simulated outputs of the models are compared to empirical data using summary statistics (e.g., the site-frequency spectrum, number of single-nucleotide polymorphisms, population genetic differentiation, genetic divergence).

In particular the software DILS (for "demographic inference with linked selection") allows to compare the data to four main scenarios of gene flow between two populations, see figure 3.

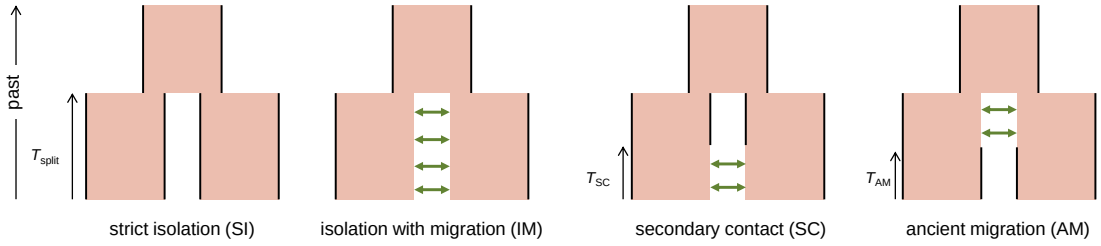


Figure 3 – Scenarios under the DILS model. Green arrows indicate gene flow. Simplified from Fraïsse et al. (2021).

In addition to finding the scenario that best corresponds to the data, DILS provides a posterior distribution of the parameters of the chosen model (for instance in the AM model, the time when populations split, the time when gene flow stops, the size of ancestral and descendant populations, etc.).

These models are very helpful to understand speciation dynamics. First, they allow to test the existence of gene flow between the populations, which is informative on the geographic context of speciation (allopatric, parapatric; Yusuf et al. 2026) and on the nature of the compared taxa (population or species; Roux et al. 2016). Second, the inferred parameters of the model can be exploited to calculate the timing of speciation, with the assumption that in the AM model, cessation of gene flow corresponds to a full RI between populations, and thus, the completion of speciation.

2 MACROEVOLUTIONARY VIEW OF SPECIATION

In macroevolution, i.e., in the field of evolution that focuses on large time scale processes operating above the species level, speciation is mostly considered as a succession of instantaneous "cladogenic" events corresponding to the splitting of lineages (Monnet 2023). This succession of events is then modelled as the realization of a stochastic process characterized by its rate

2.1 What are speciation rates?

The speciation rate is the number of speciation events occurring in given lineage per unit of time (typically 1 million years). Dated species phylogenies can be useful to estimate these rates. A dated phylogeny is a branching tree representing the ancestor-descendant relationships between lineages, with nodes corresponding to speciation events, and the length of branches indicating time. The simplest model than can be used to estimate speciation rates from a phylogeny is the Yule (1925) process, characterized by a single speciation (or birth) rate λ, constant across the tree and through time. This rate is expressed in number of species (spp) per million years (Myr).

By looking at the number of lineages through time (LTT) in a salamander phylogeny, Nee et al. (1994) found that the extinction of lineages induced an apparent increase of the birth rate close to the present, a phenomenon called "pull of the present". This effect is explained by the fact that on a reconstructed phylogeny (a phylogeny with only extant species), lineages that appeared close to the present have less chance to have gone extinct and are therefore more sampled. If we only considered a pure-birth model without extinction, this effect could be misinterpreted as an increase of the birth rate close to the present. The authors showed that taking accounting for extinction is fundamental to precisely estimate the speciation rate, opening the way to the use of birth–death (BD) processes for the phylogenetic analysis of diversification.

Terminology

speciation rate (λ) number of speciation events per lineage per unit of time
extinction rate (μ) number of extinction events per lineage per unit of time
net diversification rate speciation rate $-$ extinction rate
turnover rate (ε) extinction rate $/$ speciation rate
homogeneous rates rates do not vary between lineages
constant rates rates do not vary through time

2.2 Unbalanced phylogenetic trees

It was rapidly identified that simple BD models with homogeneous rates were not sufficient to accurately represent phylogenetic relationships between lineages. Notably, empirical phylogenies indicate that the distribution of the number of species on each side of a node most often diverges from the expectation under a diversification model with homogeneous rates, a phenomenon called "phylogenetic imbalance". For example, Heard (1992) compared tree imbalance measured on 195 trees across multiple groups to the values expected under a random pure birth process and found that the trees are generally more unbalanced than expected under a homogeneous pure birth process. The

author concluded that this discrepancy could be explained by variable speciation rates across the lineages.

This model incongruence has been particularly well illustrated by Alfaro et al. (2009) on a phylogeny of jawed vertebrates. In this paper, the authors showed that the large variation in the number of species within groups of vertebrates is unlikely to be explained by a homogeneous BD process, and that the occurrence of punctual shifts in the values of the diversification rates for some groups is more likely. Using a method called "modelling evolutionary diversification using stepwise Akaike information criterion" (MEDUSA), the authors identified some groups with high net diversification rates (Neoaves, Percomorpha fishes including the cichlids, lizard and snakes) and some other groups with low net diversification rates (coelacanths, crocodylians and tuataras). Although high net diversification rates do not translate necessarily into high speciation rate (it can also be due to a low extinction rate), some of the groups for which the turnover rate could be accurately estimated have distinct speciation rates.

2.3 Phylogenetic diversification models with heterogeneous rates uncover the variability in speciation rates

Given the inability of homogeneous BD models to accurately represent empirical phylogenies, a variety of models accounting for changes in diversification rates have been developed. Rabosky (2014) proposed the "Bayesian analysis of macroevolutionary mixtures" model (BAMM) in which lineages can experience transitions between different diversification regimes, called "rate shifts". Under this assumption, all the descendants of a given lineage have the same speciation and extinction rates, until a rate shift happens. Using a Bayesian framework, this model allows an estimation of the position of the rate shifts on the tree and the value of the speciation rates, and its implementation in a R package made this inference tool very popular. However BAMM relies on the assumption that the variability of speciation rates can be explained by a few major rate shifts, linked for instance to key innovations, while the speciation rates are expected to be influenced by many factors, and thus may vary more frequently by small amounts.

In order to account for these gradual changes in speciation rates, Maliet et al. (2019) proposed a model called "cladogenetic diversification rate shift" (ClaDS) in which speciation and extinction rates are specific to each lineage and are sampled randomly at each speciation event around the rate of their parental lineage. This assumption is motivated by the view that speciation speciation is associated with modifications of species-characteristics (adaptation to a different niche, change of range size, change of phenotype) that themselves influence the propensity of species to generate new lineages or survive. In this process, the speciation rate λ of a lineage is log-normally distributed around a value determined by the speciation rate of the parental lineage and a trend

capturing the deterministic time-variation of speciation rates (see figure 4A). In the end, an inference of the parameters of that process, using a Markov chain Monte Carlo (MCMC), allows to accurately estimate a lineage-specific speciation rate on a phylogeny.

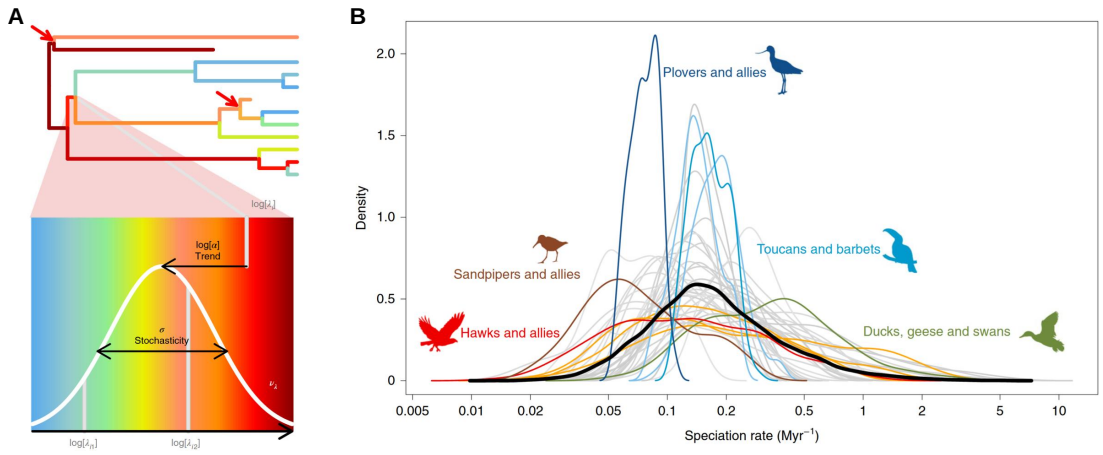


Figure 4 – The Cladogenetic diversification rate shift (ClaDS) model. A. Principle of ClaDS: speciation rates vary at each branching event, inherited from the parent lineage with a temporal deterministic trend given by α and a stochasticity given by σ. The color on each branch indicates the value of the speciation rate. **B.** Distributions of branch-specific speciation rates inferred by ClaDS on a bird phylogeny, for different clades of birds (colored lines) and for all birds (black line). Figures from Maliet et al. (2019).

More recent models allowing for a continuous variation of diversification rates — instead of discrete changes at the branching events — were developed. This class of models called "birth–death diffusion model" (BDD; Quintero et al. 2024) was motivated by the assumption that diversification rates may be influenced by a myriad of factors, both in the environment or specific to the lineages and may change through time continuously, as defended by Title et al. (2025). In those models, the diversification rates vary as a geometric Brownian process along the branches, constrained by inheritance of the rates from the parental lineage. The inference of the rates under this model on phylogenies is possible with a Bayesian data augmentation technique, and was applied to a mammal phylogeny by Quintero et al. (2024).

In a review to which I contributed, Morlon et al. (2024) compiled the distribution of speciation rates among many organisms in the tree of life, showing that these rates typically vary between 0.01 and 1.0 spp Myr^{-1}. Although the comparison between the clades in this meta-analysis must be made with caution because the inference methods are different, the variation within some clades is striking: some groups exhibit a variation of more than two orders of magnitude (orchids, rosids, ray-finned fishes, birds), see figure 5.

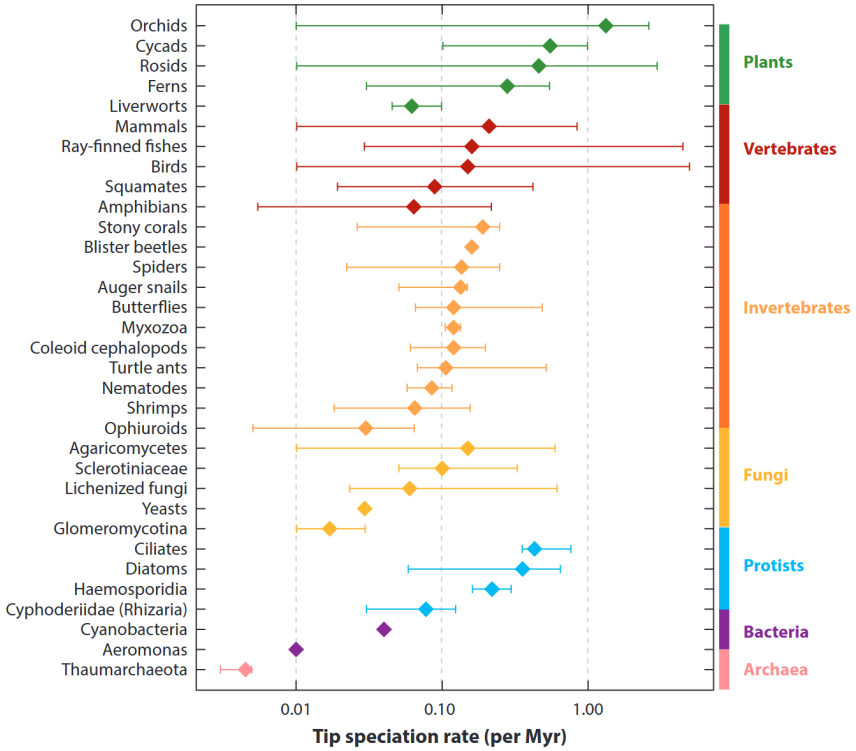


Figure 5 – Speciation rates estimated from phylogenetic approach compiled from different studies. For each clade, the mean (or estimated median), minimum and maximum speciation rate are represented. Figure by Benoît Perez-Lamarque, published in Morlon et al. (2024).

In particular, Maliet et al. (2019) applied ClaDS to 42 bird phylogenies, which allowed to identify some clades having a large variability of speciation rates such as Accipitridae (a group of large prey birds) for which speciation rates vary between 0.013 and 1.2 spp Myr^{-1}(see figure 4B, red line). In fact, most of the variation of bird speciation rates can be explained by variation within clades. A similar analysis run on squamates found a speciation rate at least twice larger in snakes (0.18 spp Myr^{-1}) than lizards (0.08 spp Myr^{-1}; Title et al. 2024). Stiller et al. (2022) applied ClaDS on Syngnathidae (a family of 300 species of fishes that include seahorses) and found speciation rates estimated at the tips are distributed between 0.05 and 0.25 spp Myr^{-1}. In mammals, Afonso Silva et al. (2025) found with ClaDS a comparable speciation rate variability than the one in birds. Using BDD in mammals too, Quintero et al. (2024) found smaller within-clade variation, but large between-clade variation: primates and rodents, for instance, exhibit an elevated speciation rate (ca. 1 spp Myr^{-1} and 0.5 spp Myr^{-1}, respectively) compared to other mammals (ca. 0.1 spp Myr^{-1} in Scandentia, < 0.1 spp Myr^{-1} in Protheria).

3 ATTEMPTS TO EXPLAIN MACROEVOLUTIONARY SPECIATION RATES WITH MICROEVOLUTIONARY PROCESSES

3.1 Predictions

Intuitively, all the biological processes involved in the speciation stages (formation of new populations, genomic or phenotypic divergence, mechanisms of RI) could have an effect on the large-scale diversification patterns. Indeed, we can expect that microevolutionary processes acting on smaller temporal scales (natural selection, mutations, genetic drift) scale up to macroevolutionary, time scales.

For example, given that population size influences various microevolutionary forces (e.g., drift and selection), we can expect that it also influences speciation rate. Mayr (1963) predicted that speciation would be triggered by population bottlenecks, through "genetic revolutions", and thus would be more frequent in smaller populations. However, Orr and Orr (1996) predicted the opposite: larger populations experience stronger natural selection so speciation should be more frequent.

Other predictions exist on which factors may favour or hinder speciation: Coyne and Orr (2004) mention for instance that sexual selection or reinforcement should accelerate speciation, and since they are the often expected in sympatric speciation, this last context should lead to faster speciation, compared to allopatric speciation. Biotic traits, like the ability to disperse, are expected to have an influence on speciation rates, however, the direction of that influence is debated. As pointed out by Coyne and Orr, dispersal ability reduces gene flow but also reduces the chance to colonise a new environment. Similarly, characteristic relative to species, like their range or the degree of range fragmentation should impact speciation.

In a perspective article, Dynesius and Jansson (2014) proposed multiple biological traits (e.g., dispersal ability, sexual behaviour, population size, etc.) and environmental factors (environmental gradients, spatial barriers, climatic variability, etc.) that may have an influence on speciation rates. The authors considered that speciation rate depends on three phases of the speciation process, called "controls of speciation": (*i*) the splitting of population, (*ii*) the speed of speciation, denoting the duration of the phase where RI accumulates and (*iii*) the persistence of lineages in the early stages. They also assessed the predicted effect of each proposed factor on each phase of the speciation process. For instance, the dispersal ability may reduce the propensity of species to split and accumulate RI but increase their persistence. Half of the proposed factors of speciation have opposite effects: they may favour one stage of speciation and hinder another and yet it is difficult to conclude what would be their effect on the macroevolutionary speciation rate. This is

the case for dispersal ability, or specialization for instance. According to the predictions of Dynesius and Jansson (2014), the other half of the factors have a concurring effect on speciation stages and may therefore be good candidates for predictors of speciation rates. For instance the authors expect larger population size or geographic range size to favour all stages of the speciation process, by promoting easier splitting into several populations, increasing selection efficiency and reducing the extinction risk during the speciation process.

However the predictions of Dynesius and Jansson (2014) remained mostly verbal and not necessarily backed-up by empirical evidence. The idea that efficient selection in larger populations speeds-up speciation relies on the assumption that speciation occurs mostly by the adaptation of lineages to distinct environments (ecological speciation), which is debated. Although ecological speciation has occurred in some lineages, whether most speciation events are driven by divergent adaptation is not clear (Anderson et al. 2025).

Rabosky (2016) predicted that species able to accumulate faster RI should have higher speciation rates, under the assumption that everything else is equal. But he also emphasised the fact that speciation is a multi-step process, and the effect of microevolutionary processes affecting the speed of RI may in fact be mitigated by other phases of speciation. Similarly to Dynesius and Jansson (2014) and other authors (Mayr 1963; Etienne and Rosindell 2012; Harvey et al. 2019), he described speciation as a result of three steps (initiation, RI, survival) on which microevolutionary processes may have different effects.

3.2 Studies attempting to connect diversification rates to speciation processes give mixed results

In order to test the verbal predictions mentioned in the previous paragraphs, some correlative approaches have been proposed to find a relationship between macroevolutionary speciation rates and characteristics of species possibly impacting speciation (see figure 6).

3.2.1 Effect of speciation initiation

The first step of speciation (population formation or speciation initiation) is characterized by a rate of population subdivision, that can be measured through proxies (see figure 6B). Harvey et al. (2017) proposed to measure the rate of population formation as the number of new genetically distinguishable populations per unit of time. More precisely, they applied a coalescent model to trees from population genetic and phylogeographic data from 173 species of birds from the Americas. They found that each species is split in 1 to 24 populations, which gives a rate of population differentiation between 0

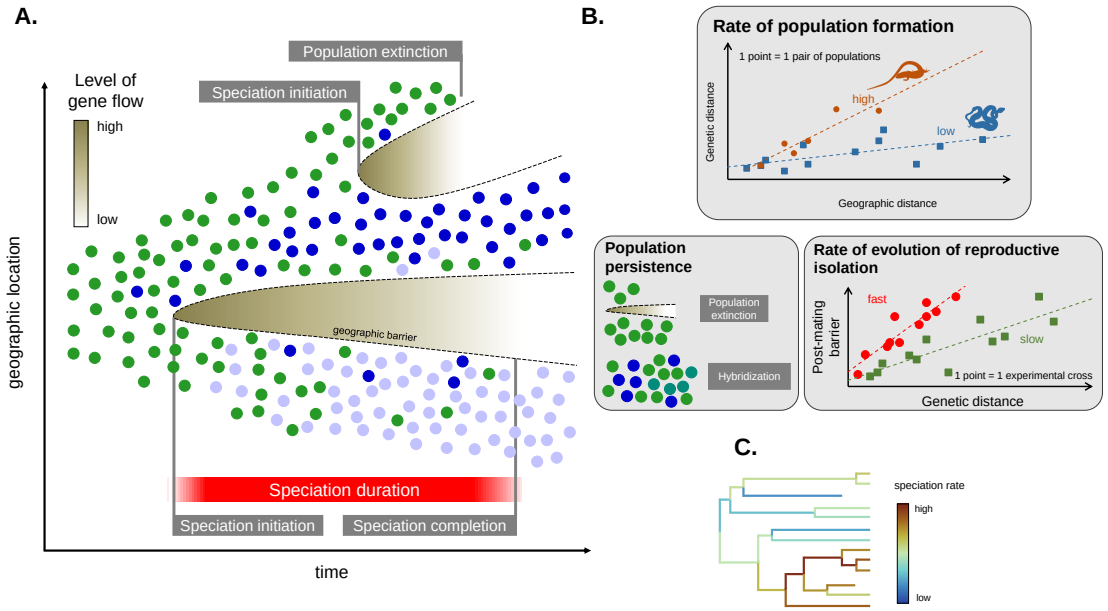


Figure 6 – The different steps of the speciation process can modulate macroevolutionary speciation rates. A. Different steps of the speciation process, from speciation initiation to speciation completion. Each dot represents one individual, colored by its genotype. The shaded regions correspond to geographic barriers, with color indicating the degree of gene flow within the grey zone of speciation. **B.** Schematic representation of how rates of key steps of the speciation process can be estimated: The slope of the relationship between geographic distance and genetic distance is indicative of the rate of population formation. Here, better dispersers (snakes) have a lower rate of population formation than poor dispersers (lizards); population persistence is estimated from the probability of extinction or hybridization; and the slope of the relationship between genetic distance and postzygotic reproductive isolation (RI) (e.g., measured as the percentage of inviable or sterile offspring in experimental crosses) is indicative of the rate of evolution of RI. **C.** Together, rates of population formation, population persistence, and evolution of RI explain macroevolutionary speciation rates, as estimated on a phylogeny. Silhouettes from `phylopic.com`. Figure by Pierre Veron, as a contribution to the review by Morlon et al. (2024).

and 6.64 divergences per Myr. In parallel, the authors estimated a clade-dependent speciation rate using BAMM. They found that the speciation rate was positively correlated with the population differentiation rate, accounting phylogenetic relatedness between lineages. This correlation suggests that bird species that are able to differentiate faster into distinct populations are also prone to speciate faster, although the variability in speciation rates remain largely not explained and could be due to other factors.

Similarly, Weeks and Claramunt (2014) estimated a dispersal ability on birds endemic to an Australian archipelago using their wing aspect ratio. They compared this measure to a phylogenetic estimate of diversification rate and found that better dispersers tend to have a lower diversification rate. Although the authors did not directly estimate the rate of population formation, but a proxy for dispersion, this result is in line with the conclusion from Harvey et al. (2017) because better dispersers are expected to form less isolated populations.

However, these analyses on birds do not seem to be generalisable to all lineages. Singhal et al. (2018) measured the rate of population differentiation in Australian squamates in a different way from Harvey et al. (2017), accounting for the geographical distance between populations. This proxy, illustrated on figure 6B, consists in measuring the correlation between the geographic distance and the genetic divergence between sampled individuals within each species, with the assumption that higher slopes characterize species that tend to form more isolated populations, because the homogenizing effect of gene flow is weaker. This method is called "isolation by distance" (IBD). Contrary to the previous study with birds, this study found no link between speciation rate inferred with BAMM and the rate of population formation. This negative result was replicated on other groups of squamates in the Neotropical Cerrado by the same authors, with speciation rate inferred with ClaDS (Singhal et al. 2022), and by other authors in a family of snakes in Madagascar (Burbrink et al. 2023). A similar result but with slightly different methods was found by Kisel et al. (2012) in orchids.

Although the absence of correlation in the previous studies could be due to a lack of statistical power, their consistency across taxonomical and geographical realms suggests that the rate of formation of population isolates, as measured above, does not influence speciation as a "rate limiting factor", either due to the mode of speciation in squamates or to the fact that the measure of IBD does not capture other population isolation mechanisms than isolation by distance. For instance, populations can be structured by physical barriers or habitat fragmentation instead of geographical distance only.

3.2.2 Effect of accumulation of reproductive isolation

The other steps of speciation have also been considered as rate limiting factors. The rate at which populations evolve RI is expected to have a direct influence on

speciation rates, because (everything else being equal) two populations where individuals become reproductively incompatible faster should also evolve into distinct species faster. An empirical measure of the rate of accumulation of RI is the relationship between reproductive barriers — estimated as the percentage of inviable of sterile offspring in experimental crosses — and the genetic distance between the crossed individuals (see figure 6B). Rabosky and Matute (2013) studied the relationship between the estimated rate of accumulation of RI and the speciation rate inferred by BAMM, both in *Drosophila* and birds. Using a statistical method accounting for evolutionary relatedness between clades (phylogenetic generalized least squares, PGLS), they found no correlation between the rate of accumulation of RI estimated on clades and the lineage-specific speciation rate: lineages that evolve RI faster do not seem to speciate faster.

3.2.3 Effect of population persistence

This general absence of strong link between proxy for speciation process and speciation rate has been debated. Harvey et al. (2019) wrote that the effect of population persistence (see figure 6B) may be strong on the speciation process. "Population persistence" here encompasses the non-extinction (survival) of the populations that can generate new species and the maintenance of their isolation (a secondary contact may prevent the differences between the populations to accumulate ; although hybridization can sometimes favour speciation, see Peñalba et al. 2024). Because the accumulation of RI is not instantaneous, i.e., it can take thousands or millions of years (Harvey et al. 2019), the persistence of populations is a good candidate to be a factor influencing speciation rates, as suggested by models (e.g., Gavrilets et al. 2000). But population persistence is difficult to estimate and I am not aware of studies documenting its correlation with speciation rate. This gap in the micro–macro research on speciation is largely due to the difficulty of documenting the history of populations, and particularly their extinction.

3.3 Protracted speciation: speciation takes time

A possible limitation of the classical BD processes used to estimate speciation rates from phylogenetic trees, is the fact that they consider speciation as an instantaneous process. Speciation is indeed a multi-step process and factors that have an influence on one of these steps may have a different influence on another step. Rosindell et al. (2010) proposed to decouple these steps when we model speciation. Their model, "protracted speciation" (*protracted* means prolonged) assumes that speciation is the result of the following scenario: first, a population is formed (at this time, it is called an "incipient species"), then, after a transition time, this population is considered to be a new species (this process is called "speciation completion").

Etienne and Rosindell (2012) proposed a formalisation of this model, the protracted

birth–death model (PBD), described as a stochastic process where each lineage can go extinct with a given extinction rate, generate a new incipient species with a rate of initiation and this incipient lineage can become a good species with a rate of completion. The PBD model is described on figure 7, with rates of initiation, completion and extinction named b, λ and μ, following the notations by Etienne et al. (2014). It has to be noted here that λ and μ do not correspond to the birth and death rate of a BD process: λ is the rate at which one incipient species become a good one and μ is the rate of extinction of a lineage (and not of a species). Some variants of this model exist in which the rates of initiation are different depending on the status of the lineage (good or incipient species), which then involve five rates denoted λ_1, λ_2, λ_3, μ_1 and μ_2 (Etienne and Rosindell 2012).

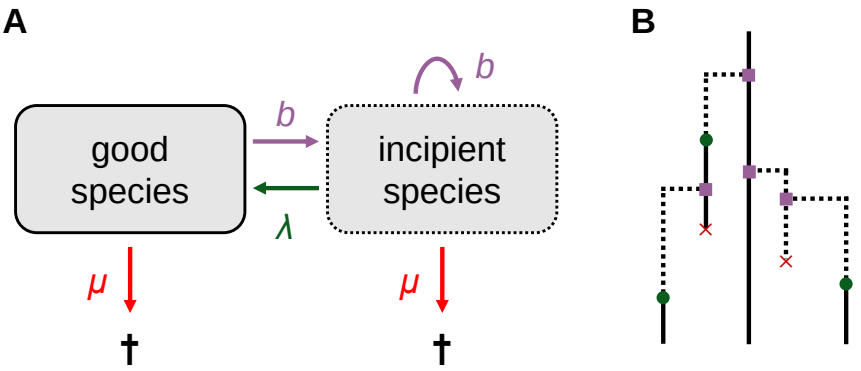


Figure 7 – Protracted birth–death (PBD) model. A. Transition between good species and incipient species, with b the initiation rate, λ the completion rate and μ the extinction rate. Adapted from Etienne et al. (2014). **B.** Example of a population phylogeny under the PBD model. Dotted lineages: incipient species; solid lineages: good species; purple square: initiation; green circle: completion; red cross: extinction.

The protracted speciation model brings an explanation to the tree shape observed in empirical phylogenies, with longer branches close to the present (stemmy trees), as found for example in multiple bird phylogenies (Phillimore and Price 2008). Indeed, since the lineages close to the present are likely to have generated incipient lineages that did not have time to complete speciation, looking at species phylogenies make the impression of less branching events close to the tips. Using a likelihood expression developped by Lambert et al. (2014), Etienne et al. (2014) proposed an inference machinery to estimate the parameters of the PBD process from a species phylogeny. However, they show that the three parameters of the model are not all identifiable. Later, based on the state-dependent speciation and extinction (SSE) framework, Hua et al. (2022) developed ProSSE, a model that allows for the estimation of the rates of initiation, extinction and speciation completion from a population-level (instead of a species-level) phylogeny. To obtain such a phylogeny for an empirical system, we need to sample and identify

all phylogeographic lineages of a group of species, using for instance multispecies coalescent trees. Applied on a phylogeny of Australian rainbow skinks (a group of lizards), this model found an initiation rate of 0.27 spp Myr^{-1}, a completion rate of 0.31 spp $Myr^{-1}$and an extinction rate close to zero.

4 THESIS OUTLINE

The missing link between micro and macroevolution in speciation research raises several questions. First, the effect of microevolutionary processes in the accumulation of RI is unclear. There are multiple verbal predictions on the role of gene flow, population size, adaptation on the tempo at which RI accumulates, but they have not been formerly modelled. The effect, for instance, of population size on speciation has been predicted, but only under the assumption of ecological speciation and needs to be compared to the effect under other modes of speciation. The validation of these predictions with empirical data is also lacking.

Second, the view of speciation as a multi-step process is generally admitted, but understanding the effect of microevolutionary processes on each on these steps is not enough to understand their effect on speciation rate at the macroevolutionary level. Indeed, if a characteristic of a species has a positive effect on one of the steps of speciation (for instance on the rate of accumulation of RI) but hinders another aspect (for instance reduces population survival), it is more challenging to actually estimate their "net effect" (Dynesius and Jansson 2014) on speciation rate. One needs to understand the relative influence of each of the speciation steps on the overall speciation rate. In summary: do some species speciate faster because they are able to form new isolated populations, because these populations accumulate RI faster, or because these populations persist longer?

In the first chapter of my PhD manuscript "**Rapid speciation in small populations challenges the dominance of ecological speciation**", I use an integrative model of accumulation of RI to understand how microevolutionary parameters (population size, mutation rate, local adaptation and migration) influence the duration of speciation and the shape of the grey zone. I show that making a distinction between the different modes of speciation (ecological, allopatric, parapatric) is crucial to predict the effect of these parameters on the tempo of speciation. I compare the expectation of the link between population size and speciation duration with empirical data estimated from plant genomes and find a positive association between population size and speciation duration, which I interpret as a signature of non-ecological speciation.

In the second chapter "**Speciation completion rates have limited impact on macroevolutionary diversification**", I estimate the relative influence of each speciation step

using the protracted birth–death (PBD) framework. The PBD process assumes that speciation occurs after two lineages split (initiation) and become reproductively isolated (completion). This random process is controlled by several rates (rates of initiation, completion, extinction), but I calculate equivalent birth and death rates that would result in a comparable process. Using these equivalent birth and death rates, I show which of the parameters of the PBD model influence mostly the birth (i.e., the macroevolutionary speciation) rate. An application of this method to the PBD rates estimated on Australian rainbow skinks in Hua et al. (2022), shows that speciation in this group is mostly limited by speciation initiation, and not by speciation completion.

In the appendix "**Linking diversification rates to the speciation process**", I review the attempts to link speciation rates and proxies for different speciation steps. This part is based on a section I contributed to in a published review.

CHAPTER I

Rapid speciation in small populations challenges the dominance of ecological speciation

This chapter consist of the manuscript of an article by Pierre Veron[*,#], Anaïs Spire[*], Agathe Chave-Lucas[*], Tatiana Giraud[#] and Hélène Morlon[*]. The manuscript is currently under evaluation and is available on bioRxiv (Veron et al. 2026).

In the end of the 1990's, Sergey Gavrilets proposed a model of speciation inspired from the rugged fitness landscape by Wright (1932), but showed that, considering multiple dimensions in the genotype space (typically, DMIs with a large number of loci), the fitness could be considered almost flat with valleys of null fitness. With this view, separate populations evolve along ridges of high fitness and can be, after some time, separated by valleys or "holes" of null fitness that cannot be crossed. This view was new compared to the classical rugged landscape point of view, because it did not require populations to evolve in valleys of low fitness for speciation to occur. This model called "holey adaptive landscape" (HAL) allows a mathematically tractable modelling of speciation dynamics, which very few models allow.

The formalisation of the HAL was proposed in Gavrilets et al. (1998), inspired from a speciation model by Higgs and Derrida (1992). In this model, two individuals can have a fertile offspring if their genetic distance is smaller than a threshold. The "fitness"

∗ Institut de Biologie de l'École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France.
Écologie Société et Évolution, CNRS, Université Paris-Saclay, AgroParisTech, Gif-sur-Yvette, France.

in this model is therefore the reproductive success of a given pair, or in an equivalent way, the individual fitness of a hybrid depending on its heterozygosity. In a first article, Gavrilets et al. (1998) described the possibility to reach speciation in a parametric setting of the model under certain conditions, using stochastic simulations. But the founder work on the HAL model is certainly the result of 1999 in which Gavrilets derived a set of ordinary differential equations (ODE) predicting the evolution of populations under this model. The subtlety to consider for the mathematical formulation of this model is the counter-selection induced by genetic incompatibilities. Mutations do not carry *a priori* a selective advantage or disadvantage, but a rare genotypes may have a reduced effective fitness due to the difficulty to find compatible partners. There is therefore a selection within populations which can alter the substitution rate, depending on the genetic diversity within this population. Gavrilets approximated this effect with a coefficient R and derived an analytical expression for this coefficient as a function of the mean polymorphism within a population.

Gavrilets (1999) does not really focus on the time for speciation itself or the influence of the biological parameters, such as population size or mutation rate on the dynamics. In this chapter we use Gavrilets' equations to predict speciation duration, either analytically in the case without migration, or by numerical resolution in the case with migration. We go further in the analysis because we use the numerical resolution of the set of ODE to characterize when and how fast reproductive isolation (RI) accumulates between two separate populations, during the phase called "grey zone" of speciation by Roux et al. (2016). With my co-authors, we were particularly interested in the effect of biological parameters on speciation duration and the grey zone. We therefore explored the parameters space and characterized how population size, mutation rate and the threshold for genetic incompatibility influence the duration of speciation, for three different scenarios. We find that the relationship between speciation duration and population size depends on the scenario of speciation: if the mutations responsible for the genetic incompatibility are adaptive (ecological speciation), larger populations speciate faster; however, when these mutations are not adaptive, speciation can be faster in small populations.

With this work, we show that some classical verbal predictions can actually be misleading, as for instance larger mutation rates do not always lead to faster speciation, and the shape of the grey zone — if considered as a function of the genetic distance — does not depend on the scenario of speciation. Finally, we use demographic data from plants from the study of Monnet et al. (2025) to show that smaller populations tend to speciate faster, indicating that ecological speciation is likely not the valid scenario in most of the species considered in this data set.

This chapter emerged from a collaborative project and I had great pleasure to work with Anaïs Spire, whom I supervised as a M1 intern, and Agathe Chave-Lucas, a PhD student in my institute. In this work, Anaïs helped a lot to implement stochastic

simulations for the scenario with migration, and implemented the equations by Gavrilets for this scenario in a numerical solver. With this work, she could explore a large parameter space that paved the way to our analyses for the "parapatric" speciation scenario. Agathe Chave-Lucas provided her expertise to adapt the equations of Gavrilets in more complex cases, like asymmetric population sizes or migration rates. She also developed the approach for predicting the occurrence of speciation without solving the whole system, based on the bifurcation diagram. This method, described in supplementary text S I.1, allowed for a more precise and efficient prediction of the cases without speciation. I personally created the code for the numerical predictions and the stochastic simulations based on Gavrilets' ODE, and used the results of these simulations for this manuscript, both for the speciation duration and the grey zone. For the empirical part, I used the results of the demographic software "demographic inference with linked selection" (DILS) applied on plant genomes, published by Monnet et al. (2025). I developed the statistical method to account for the particular structure of the data, in order to test the empirical relationship between population size and speciation duration.

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 duration 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.

Keywords speciation, micro–macro, holey adaptive landscape, plants

I.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 spp Myr^{-1} in birds (Maliet et al. 2019), and from 0.005 to 1.5 spp 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 (RI) (Rabosky and Matute 2013), population divergence rates (Harvey et al. 2017; Harvey et al. 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 (whether the mutations that confer a local selective advantage are neutral or deleterious in the other population), 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).

I.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 do not interbreed, as a result of either post- or prezygotic incompatibilities (e.g., hybrid inviability, assortative mating). Hence K is the maximum number of loci on which two individuals can be different and still have a chance to mate successfully. This formulation is motivated by the decrease in the reproductive success of two individuals with the genetic distance that separates them (Price and Bouvier 2002; Rabosky and Matute 2013; Christie and Strauss 2018; Christie et al. 2022); it provides a mathematically tractable way to represent high dimensional, correlated adaptive landscapes (Gavrilets 1997, figure I.1A), and to account for the multigenic origin of incompatibilities. It focuses on the consequence of incompatibilities — reproductive isolation — rather than on their nature (e.g. trait-based mate choice, Dobzhansky-Muller incompatibilities; Dobzhansky 1936; Muller 1942). 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 I.1B). In the "parapatric neutral" scenario, the two populations can exchange alleles through migration, with a migration rate m. In the "allopatric adaptive" scenario, mutations confer a selective advantage in the local population in which they arise, with coefficient s_{LA}.

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 I.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, would successfully interbreed if they were in contact. 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 would successfully interbreed with any individual from

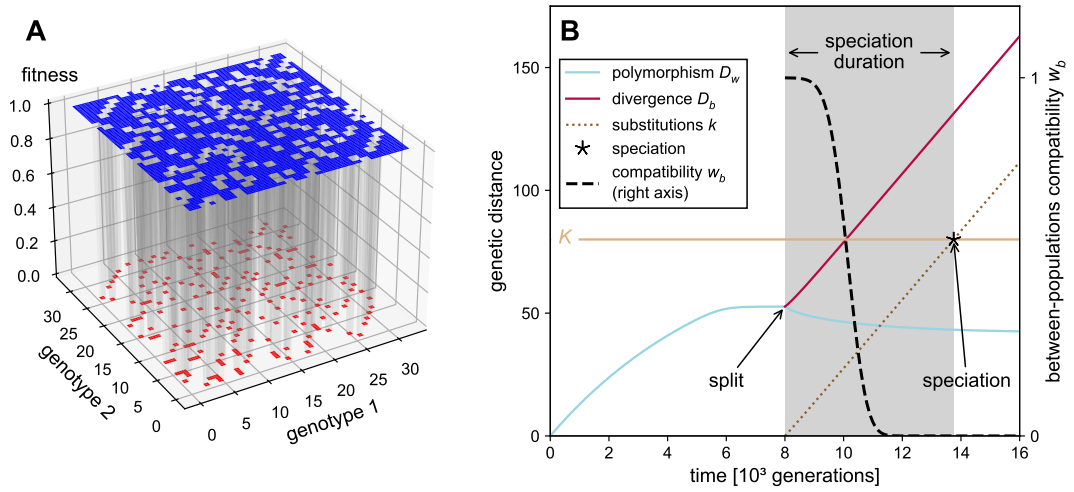


Figure I.1 – The holey adaptive landscape. **A.** Illustration of the holey adaptive landscape (HAL) with 5 polymorphic sites (32 different genotypes) and a compatibility threshold $K = 3$. Populations can evolve along nearly neutral ridges (in blue) and end up being separated by fitness holes (in red). **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 $N_T = 6000$, $N_1 = N_2 = 3000$, $\nu = 0.007$, $K = 80$.

the other population, should the two populations come into contact (i.e., $w_b = 0$; Mayr 1942). This reproductive isolation can result from either pre- or postzygotic barriers, and occurs if (and when) k reaches K (figure I.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 S I.2 and figures S I.7 to S I.10).

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 :

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

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 I.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 S I.1), reducing within-population polymorphism (figure I.2, lower row) and affecting the duration of speciation (figure I.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 I.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 I.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 S I.1; figure I.2 upper row, pink curves).

High mutation rates are generally thought to increase the rate at which populations acquire substitutions and RI, 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 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 I.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 I.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

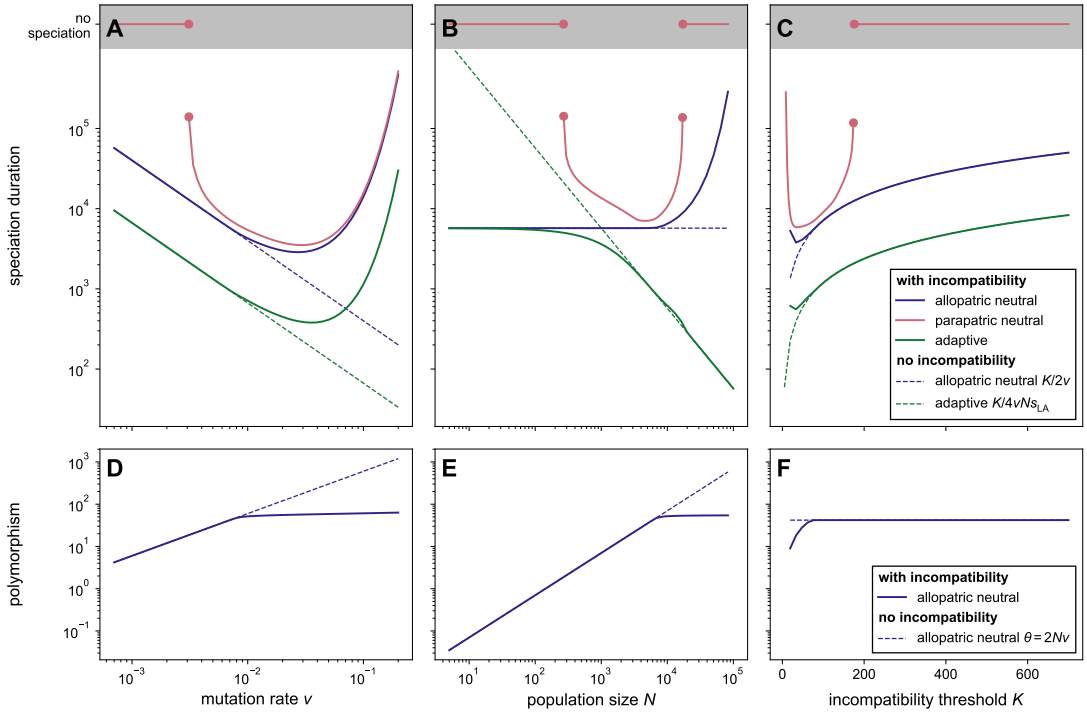


Figure I.2 – Expected speciation duration and polymorphism. Expected speciation duration (**upper row**) and polymorphism (**lower row**) under the holey adaptive landscape model (HAL; solid lines), as a function of mutation rate (**A**, **D**), population size (**B**, **E**) and incompatibility threshold (**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 $s_{\mathrm{LA}} = 0.001$.

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 I.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 I.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 I.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 I.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 S I.4). 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 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. These genomic changes can result in pre- and/or postzygotic barriers, which have the same consequence of leading to the reproductive isolation of genetically distant individuals. 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 I.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 I.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 S I.1 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 I.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 (GZ), 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 I.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 I.3A and figure I.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 I.3A2, in green), and that migration enforces a slow compatibility drop (figure I.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 I.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 S I.2), which makes it more likely to find pairs of individuals that are not compatible between the two

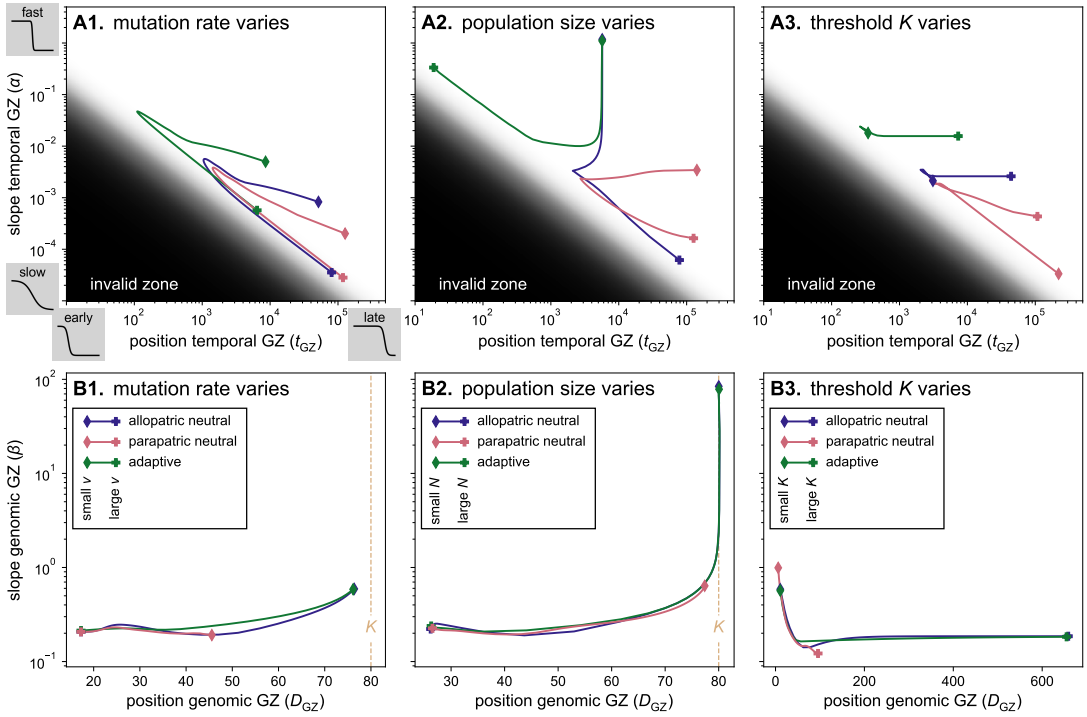


Figure I.3 – The grey zone of speciation. Characteristics of the temporal (**A**) and the genomic (**B**) grey zone (GZ) of speciation, with varying mutation rate (**1**), population size (**2**) and incompatibility threshold (**3**). The temporal GZ (resp. genomic GZ) are characterized by a slope α (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 $w_b = 0.75$).

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 I.3B versus figure I.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 RI is achieved at a smaller level of divergence in plants compared to animals. They hypothesized that this difference could be explained by RI occurring with fewer reproductive barriers in plants. We show here that the number of loci required to achieve RI (K) indeed has a large effect on the timing of the grey zone (figure I.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.

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 I.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 I.4). 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 I.2B, right-most part) where polymorphism is expected to be independent of population size (figure I.2E, right-most part). The empirical correlations are therefore consistent with expectations under non-ecological speciation and large population sizes.

The consistency of the empirical correlation between population size and speciation duration with expectations under a non-ecological speciation scenario does not contradict years of empirical evidence for local adaptation in plants (Fournier-Level et al. 2011; Lowry 2012). Instead, it suggests a decoupling between the traits under local adaptation and the genetic differences causing reproductive isolation (Frayer et al. 2025), such that non-ecological rather than ecological processes control the duration of speciation. Under this scenario, as incompatibilities accumulate between populations, less gene flow will counteract local adaptation, and ecological differences between sister species (as exemplified in van der Niet and Johnson 2009) will arise more easily. The ecological differences would thus be facilitated by reproductive isolation rather than serving as its

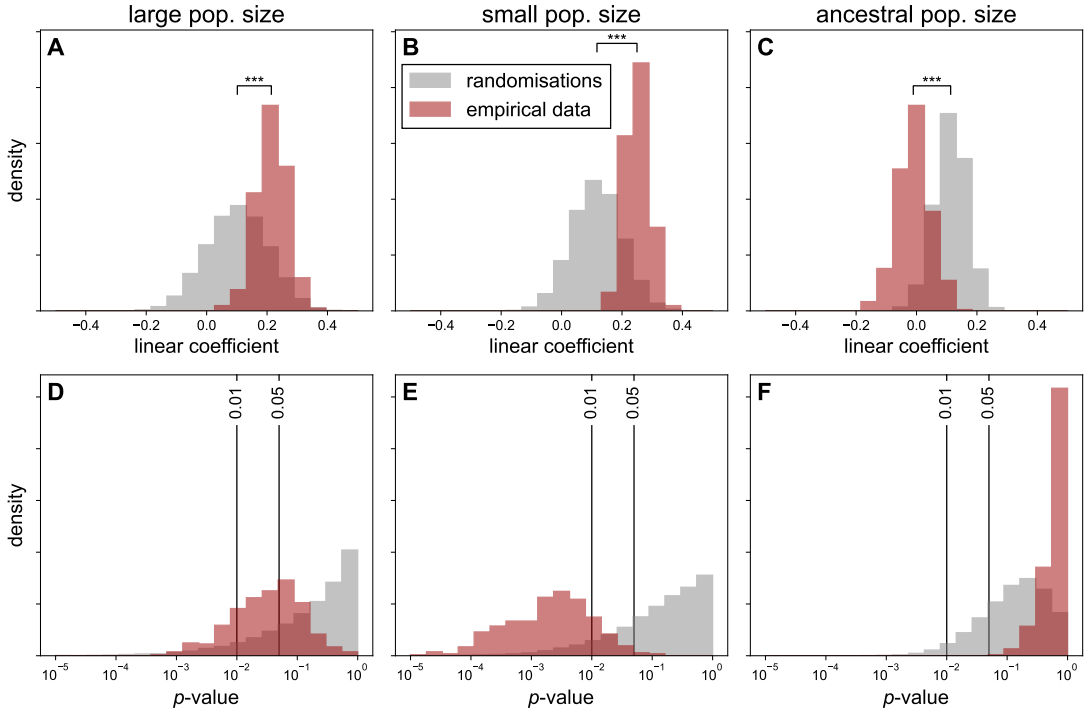


Figure I.4 – Correlation between speciation duration and population size in plant data. Distribution of linear coefficients (**A**, **B**, **C**) and p-values (**D**, **E**, **F**) for the correlation between speciation duration and population size across posterior samples (red) and randomisations (grey). 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 S I.1).

initial cause. An alternative explanation for the observed positive correlation between speciation duration and population size is that factors promoting the accumulation of incompatibilities are stronger in small populations, such as selection and geographical isolation. For example, small populations may frequently originate from long-distance dispersal into novel habitats, leading to increased geographic isolation. Upon arrival in markedly different environments, they would also experience stronger, more divergent selection. If small populations generally arose as a result of such long-distance dispersal, we would expect less gene flow in smaller populations. We did not find any significant relationship between ancestral population size and ancestral migration rates estimated from the plant genomic data (figure S I.12); however, this could simply be due to the difficulty of inferring past migration rates (Fraïsse et al. 2021). Regardless of the mechanisms at play, our results suggest that one should not necessarily expect speciation to be faster in larger populations.

I.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 reversal 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 complete. 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.

I.4 MATERIALS AND METHODS

I.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:

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

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):

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

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 $\mathrm{d}k/\mathrm{d}t = 2\nu R$. 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 $\tau = K/2\nu R$. 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 I.1.

To analyse the dynamics in the neutral scenarios, we numerically solved the system given in equation I.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 S I.1.

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

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

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. In this model, the mutations responsible for incompatibilities have a selective advantage in the population in which they appear. In this allopatric model the two populations do not exchange any

gene, such that the selective status of mutations in the other population (i.e., neutral, deleterious or advantageous) is irrelevant.

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 a GitHub repository (Veron 2026). 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.

I.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 S I.2). We provide a tool to run the simulations in command line, available on a GitHub repository (Veron 2026). 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 S I.7 to S I.10, 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.

I.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

$R^2 > 0.971$).

I.4.4 Empirical link between population size and 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 S I.3). 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 S I.3). 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 S I.1).

ACKNOWLEDGEMENTS

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/).

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/.

CHAPTER II

Speciation completion rates have limited impact on macroevolutionary diversification

This chapter is an article by Pierre Veron[*,#,§], Jérémy Andréoletti[*], Tatiana Giraud[#] and Hélène Morlon[*], published in the issue "'A mathematical theory of evolution': phylogenetic models dating back 100 years" of *Philosophical Transactions of the Royal Society B* (Veron et al. 2025), one century after the publication of Yule (1925).

Several authors (Mayr 1963; Dynesius and Jansson 2014; Rabosky 2016) describe speciation as the outcome of a three-steps process (initiation, persistence, completion). From the macroevolutionary perspective, phylogenies are usually described with birth–death (BD) models, i.e., with instantaneous speciation and extinction. Linking micro- and macroevolution requires then to adapt macroevolutionary models to this biological reality. Protracted speciation models, proposed by Rosindell et al. (2010) transcribe these multiple steps in a phylogenetic model. In this framework, speciation occurs after an initiation and a completion, and it can fail because of the extinction of the populations before completion.

A formalisation of the protracted speciation model was proposed by Etienne and Rosindell (2012) as the protracted birth–death (PBD) model, which can be summarised

∗ Institut de Biologie de l'École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France.
Écologie Société et Évolution, CNRS, Université Paris-Saclay, AgroParisTech, Gif-sur-Yvette, France.
§ Muséum National d'Histoire Naturelle, Paris, France

with three rates, the rate of initiation, the rate of completion and the rate of extinction of populations. This stochastic birth–death process generates a phylogeny where each lineage corresponds either to a "good" species, or an "incipient" species. Then this tree can be transformed into a classical species phylogeny by considering only the good species.

A lot of micro-evolutionary studies focus mostly on the completion step, for instance by looking at the rate of accumulation of reproductive isolation. However this sole consideration could be limiting when we try to explain why speciation rates vary, notably because the other steps of speciation could also influence it. Rabosky (2016) uses the term of "rate-limiting factor" call a potential step of speciation that may explain variations in macroevolutionary speciation rates. The use of the PBD model is helpful to enlighten this question of the rate-limiting factor of speciation.

In this chapter, I use the mathematical properties of the PBD process to estimate the relative influence of each of the steps of speciation — through its corresponding rate — on the macroevolutionary speciation rate. To do so, I calculate equivalent birth and death rates from the parameters of the PBD process and estimate how a change of one of these parameters influences speciation rate. I then apply this method to parameters of protracted speciation estimated by Hua et al. (2022) on a clade of lizards, and conclude that the completion rate has in fact very little influence on the macroevolutionary speciation rate, although this step is essential in the speciation process.

This work is a collaboration with Jérémy Andréoletti, former PhD student supervised by Hélène Morlon. Jérémy used his expertise to implement the numerical simulations of the phylogenetic trees under the different processes (PBD and BD with the estimated equivalent rates) in order to verify the validity of the approximation I made in calculating equivalent BD rates. He also very usefully proofread my calculations and estimated the tip speciation rates on the generated phylogenies. This chapter is organized as follows: in the materials and methods I briefly explain how I calculate the equivalent BD rates, with two variants: time-varying BD rates (varBD) and a time-constant BD rates, and how I estimate the relative influence of the PBD parameters. I then present the results of this analysis and discuss them. In the appendix of this article (section II.5), I explain the details of the mathematical derivations used to calculate the equivalent BD rates. Supplementary figures that do not contain important result but additional analyses or examples of numerical resolutions are provided in the end of the manuscript, page 127.

ABSTRACT

Standard birth–death processes used in macroevolutionary studies assume instantaneous speciation, an unrealistic premise that limits the interpretation of speciation and extinction rates. The protracted birth–death (PBD) model instead assumes that speciation involves two steps: initiation and completion. In order to understand their respective influence on macroevolutionary speciation rates, we compute a standard time-varying birth–death scenario that is "equivalent" to the PBD model in terms of speciation and extinction probabilities. First, we find a sharp decline in the equivalent birth rate near the present, indicating that rates estimated at the tips of phylogenies may not accurately reflect the underlying speciation process. Second, the completion rate controls the timing of the decay rather than the asymptotic equivalent rates. The equivalent birth rate in the past scales with the speciation initiation rate, with a scaling factor depending mostly on the population extinction rate. Our results suggest that the rates of population formation and extinction may often play a larger role than the speed of accumulation of reproductive isolation in modulating speciation rates. Our study establishes a theoretical framework for understanding how microevolutionary processes combine to explain the diversification of species on macroevolutionary time scales.

Keywords speciation, macroevolution, microevolution, phylogeny

II.1 INTRODUCTION

birth–death models are widely used to understand the diversification of species groups; in this context, births represent speciation events, i.e., the emergence of two daughter species from an ancestral one, and deaths represent species extinction. This specific use of birth–death models is particularly widespread to interpret both fossil (Silvestro et al. 2014) and phylogenetic data (Stadler 2013; Morlon et al. 2024) in terms of diversification dynamics. In birth–death models, speciation is considered to be an instantaneous phenomenon, represented as a branching event following a Poisson point process.

Despite the widespread use of birth–death models to represent speciation and extinction events, speciation is not instantaneous. Speciation requires an initial isolation of populations (speciation initiation) followed by the accumulation of genetic barriers to gene flow until speciation is complete. Speciation can initiate and fail before completion, for example because of secondary contact, or because isolated populations go extinct before speciation completion (Coyne and Orr 2004; Rosenblum et al. 2012; Dynesius and Jansson 2014). The whole speciation process may take hundreds of thousands up to several millions of years (Benton and Pearson 2001; Etienne and Rosindell 2012; Etienne et al. 2014; Hua et al. 2022).

Ignoring the fact that speciation takes time by using standard birth–death models has non trivial consequences for our understanding of diversification dynamics. For example, when standard birth–death models are used in combination with phylogenetic trees of extant species to estimate speciation and extinction rates, the "protracted" nature of speciation may be misinterpreted as a speciation rate slowdown towards the present (Etienne and Rosindell 2012; Moen and Morlon 2014).

Etienne and Rosindell (2012) pioneered the development of the so-called protracted birth–death model (PBD). Instead of assuming that speciation is instantaneous as in the standard birth–death model, this model assumes that there are events of speciation *initiation* corresponding to the formation of *incipient* species that eventually become *good* species after a random *completion* time. An incipient lineage is a lineage that is not yet considered as a different species from the ancestral lineage. For sexually reproducing organisms, the completion time is the time it takes for lineages to achieve reproductive isolation. Each lineage is thus subject to initiation, extinction and completion events. Although the PBD model is phenomenological, the protracted nature of speciation represents the effect of mechanisms such as the accumulation of genetic incompatibilities between lineages (for instance, Bateson-Dobzhanski-Muller incompatibilities), gene flow between incipient species and adaptation that are not explicit in this model but contribute to modulate the completion time (Etienne et al. 2014).

The protracted birth–death model has several advantages over the the standard birth–death model. First, it is biologically more realistic, and it thus unsurprisingly produces phylogenies that are closer to empirical phylogenies than those produced by the standard birth–death model (Etienne and Rosindell 2012). Specifically, it produces phylogenies that are less tippy (fewer recent speciation events) than those arising from the standard birth–death model. Second, it allows the integration of intraspecific processes that lead to speciation. For example, the matching competition birth–death model (MCBD, Aristide and Morlon 2019) integrates the effect of intraspecific competition on character displacement leading to speciation by modeling character displacement in incipient lineages.

Despite the advantages of the protracted birth–death model, the overwhelming majority of phylogenetic analyses of diversification use the standard birth–death model. Most available models for phylogenetic analyses of diversification are versions of the standard birth–death model, with birth and death rates that can vary in time and/or across lineages (Morlon et al. 2024). The protracted birth–death model can be fitted to empirical phylogenies; however not all of its parameters can be reliably estimated from a phylogeny (Etienne et al. 2014), which limits its usefulness. Recently, Hua et al. (2022) showed that the parameters of a protracted speciation model (slightly different from PBD) can be accurately estimated from population-level (rather than species-level) phylogenies; however such phylogenies remain rare. Fitting standard birth–death models thus remains the norm in phylogenetic analyses of diversification.

If speciation takes time but is estimated by fitting standard birth–death models to phylogenies, which assume instantaneous speciation, what do resulting speciation and extinction rate estimates actually represent? We can expect that speciation rate estimates will be higher when rates of speciation initiation and completion are higher, and rates of extinction of incipient species are lower, but precisely answering this question requires to establish an analytical relationship between the parameters of the protracted and standard birth–death models. To our knowledge, such a relationship has not yet been established.

Elucidating the relationship between the parameters of the protracted and standard birth–death models is important not only to clarify the meaning of speciation rates estimated from phylogenies, but also to understand the microevolutionary processes that modulate these rates (Morlon et al. 2024). Indeed, macroevolutionary speciation rates (estimated from phylogenies) vary by orders of magnitude (Maliet et al. 2019; Quintero et al. 2024), but the processes underlying this variation remain unclear. Efforts to find empirical correlations between macroevolutionary speciation rates and rates of population formation or evolution of reproductive isolation have not been conclusive; a proposed explanation is that this expected correlation is erased by the frequent extinction of incipient species (Rabosky and Matute 2013; Singhal et al. 2018; Singhal et al. 2022). The idea is that, population survival rather than population formation and the

accumulation of reproductive barriers may be the factor "limiting" speciation.

More generally, each of speciation initiation, speciation completion and population survival may be the process limiting macroevolutionary speciation rates in some situations but not others (Rabosky 2016). For instance, a lineage that has a propensity to accumulate fast reproductive isolation but does not experience frequent population splits might not have a high speciation rate: here, the rate of speciation completion is not limiting. Hence, factors acting on speciation initiation, the accumulation of reproductive isolation and the extinction of incipient species can have non trivial outcomes in terms of speciation rate.

Here, we obtain a mathematical link between the parameters of the protracted and standard birth–death diversification models by computing "equivalent" speciation and extinction rates, meant to represent the macroevolutionary outcomes (macroevolutionary speciation and extinction rates) of the protracted birth–death process. More precisely, our equivalent rates generate the same speciation and extinction probabilities as the protracted process. As we discuss in the paper, this is distinct from computing "congruent" (sensu Louca and Pennell 2020) speciation and extinction rates that would provide the same likelihood of extant species trees. Our main interest here is on the speciation and extinction rate outcomes of the protracted birth–death process, beyond what can be inferred from phylogenies.

II.2 MATERIALS AND METHODS

II.2.1 Birth–death and protracted birth–death models

The standard birth–death (BD) model involves two rates (figure II.1**A**): the birth (speciation) rate λ and the death (extinction) rate μ. The protracted birth–death (PBD) model as defined in Etienne et al. (2014) involves 5 rates (figure II.1**B**) : the rate of speciation initiation from a good species λ_1, the rate of completion λ_2, the rate of speciation initiation from an incipient species λ_3, the rate of extinction of a good species μ_1 and the rate of extinction of an incipient species μ_2.

By convention, time elapses in the direction of increasing values. The process begins at time 0 with one good lineage and runs until the present at time T. We now consider an intermediate time $T-t$, and introduce $p_E^G(t)$ and $p_S^G(t)$ the probabilities that, for any given good lineage, the first event is an extinction or a speciation event, respectively, and $p_N^G(t)$ the probability that none of these events occur during a time interval of length t (figure II.1**C**). Similarly, $p_E^I(t)$ and $p_C^I(t)$ are the probabilities that, during a time interval of length t, the first event is respectively an extinction, or a completion event, when starting with an incipient lineage; $p_N^I(t)$ is the probability that none of these events occur during the time interval (figure II.1**D**). By "extinction" we mean the direct extinction of

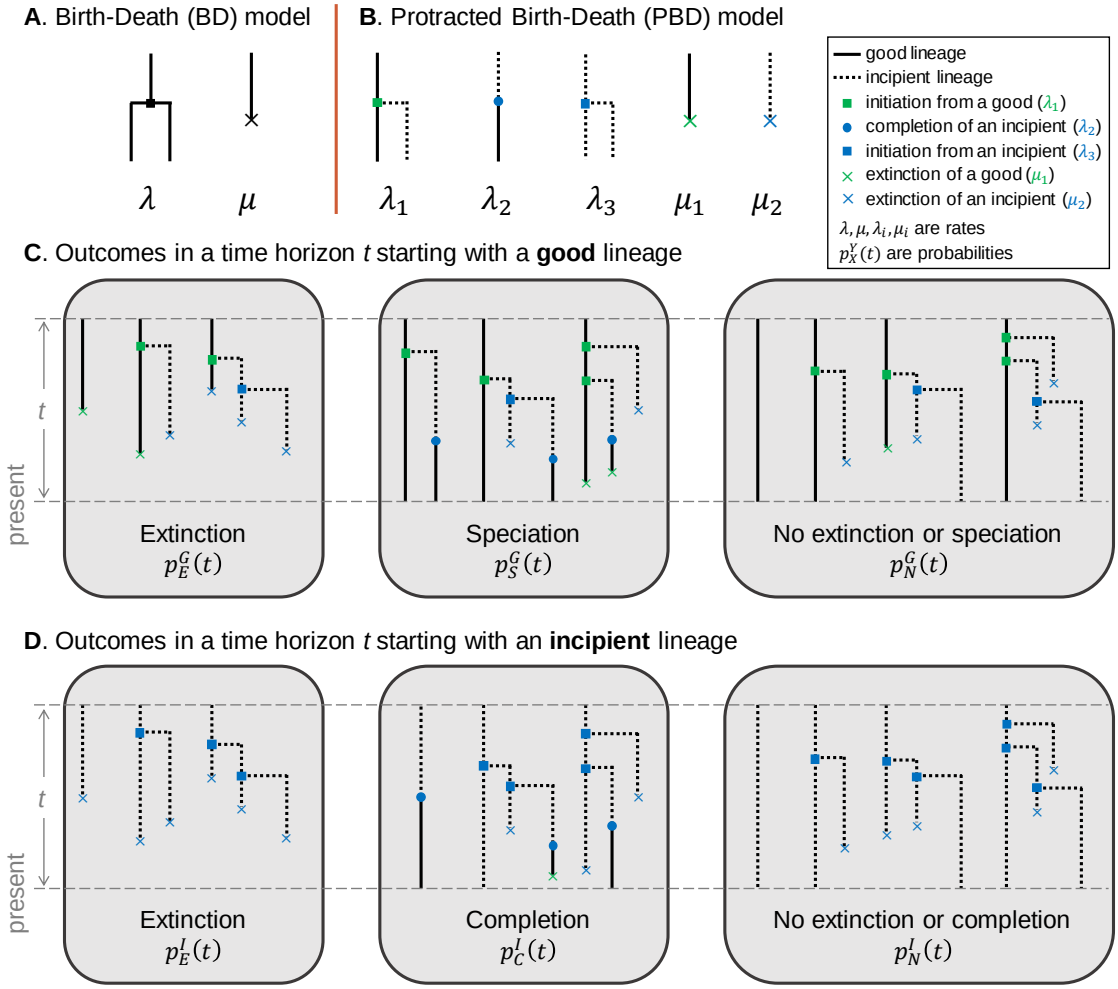


Figure II.1 – The birth–death (BD) and protracted birth–death (PBD) models. Illustrations of the rates involved in the BD model (**A**), and the PBD model (**B**). Possible outcomes of the PBD process in a fixed time horizon, starting from a good lineage (**C**) or an incipient lineage (**D**); non exhaustive examples of events leading to these outcomes are shown.

the lineage in question, or the extinction of all possible descendants of this lineage in the time interval considered. Similarly, by "completion" we mean the direct completion of the lineage in question, or the completion of an incipient daughter lineage in the time interval under consideration.

In case of multiple events occurring in the time interval (e.g., speciation followed by extinction), we consider only the first event. For instance the third case of speciation in figure II.1**C** (where the lineages all go extinct after speciation) is considered as a speciation event and is recorded in $p_S^G(t)$. We only consider the events of speciation and extinction and not the intermediate steps of speciation initiation, because we are interested in comparing these outcomes with those of the standard birth–death process. In particular, an initiation event followed by an extinction of the two lineages will be considered as an extinction. We will however keep track of the initiation events when computing the speciation and extinction probabilities.

II.2.2 Equivalent time-dependent BD rates

Definitions of equivalent birth and death rates, and relation with probabilities of speciation and extinction under the PBD process

Given a PBD process with fixed parameter values running from 0 to T, we assume that the probabilities of speciation and extinction within an infinitesimal time interval $[t-\mathrm{d}t,t]$ can be written as $\hat{\lambda}(t)\,\mathrm{d}t$ and $\hat{\mu}(t)\,\mathrm{d}t$, for any time t. We call the quantities $\hat{\lambda}(t)$ and $\hat{\mu}(t)$ the time-dependent equivalent birth and death rates. Given a good lineage alive at time $T-t-\mathrm{d}t$, the probability that the first event occurring within the time interval $[T-t-\mathrm{d}t,T]$ is a speciation event is given by:

$$p_S^G(t+\mathrm{d}t) = \underbrace{\hat{\lambda}(T-t)\,\mathrm{d}t}_{(i)} + \underbrace{(1-\hat{\lambda}(T-t)\,\mathrm{d}t-\hat{\mu}(T-t)\,\mathrm{d}t)}_{(ii)}\underbrace{p_S^G(t)}_{(iii)}$$

with (i) the probability of speciation within the small time interval $[T-t-\mathrm{d}t,T-t]$, (ii) the probability of no speciation nor extinction within the small time interval $[T-t-\mathrm{d}t,T-t]$ and (iii) the probability that the first event occurring within the time interval $[T-t,T]$ is a speciation event, conditioned on the existence of the lineage at the time $T-t$. Similarly we have:

$$p_E^G(t+\mathrm{d}t) = \hat{\mu}(T-t)\,\mathrm{d}t + (1-\hat{\lambda}(T-t)\,\mathrm{d}t-\hat{\mu}(T-t)\,\mathrm{d}t)p_E^G(t).$$

Hence, we have the dynamical system

$$\begin{cases} \frac{\mathrm{d}p_S^G(t)}{\mathrm{d}t} = \hat{\lambda}(T-t) - (\hat{\lambda}(T-t)+\hat{\mu}(T-t))p_S^G(t) \\ \frac{\mathrm{d}p_E^G(t)}{\mathrm{d}t} = \hat{\mu}(T-t) - (\hat{\lambda}(T-t)+\hat{\mu}(T-t))p_E^G(t) \end{cases}$$

which allows us to express the time-dependent birth rate:

$$\hat{\lambda}(t) = \frac{\left(1 - p_E^G(T-t)\right) \left.\frac{\mathrm{d}p_S^G}{\mathrm{d}t}\right|_{T-t} + p_S^G(T-t) \left.\frac{\mathrm{d}p_E^G}{\mathrm{d}t}\right|_{T-t}}{p_N^G(T-t)} \tag{II.1}$$

and death rate:

$$\hat{\mu}(t) = \frac{\left(1 - p_S^G(T-t)\right) \left.\frac{\mathrm{d}p_E^G}{\mathrm{d}t}\right|_{T-t} + p_E^G(T-t) \left.\frac{\mathrm{d}p_E^G}{\mathrm{d}t}\right|_{T-t}}{p_N^G(T-t)}. \tag{II.2}$$

In what follows, we compute the probabilities $p_S^G(t)$ and $p_E^G(t)$ for any time $t \in [0,T]$, which provides us with the equivalent rates.

By assuming that the probabilities of speciation and extinction within an infinitesimal time interval $[t - \mathrm{d}t, t]$ can be written as $\hat{\lambda}(t)\,\mathrm{d}t$ and $\hat{\mu}(t)\,\mathrm{d}t$, we have assumed that these probabilities depend only on time, and not on the history of the lineage considered. This is an approximation, because (under the PBD process) good lineages carry the history of their incipient species. An old lineage is indeed more likely to have a pool of incipient lineages (for instance, when λ_3 is high, their number increases exponentially with time) and therefore less likely to go extinct. The equations we used for the probabilities of speciation $p_S^G(t)$ and extinction $p_E^G(t)$ correspond to the case of a good lineage with no incipient species, and are approximations otherwise. We expect that these approximations will affect mainly the extinction rate, as we ignore the buffering effect on extinction of incipient lineages that exist at the time when the rate is computed. The equivalent extinction rate could therefore overestimate extinction probabilities.

Speciation and extinction probabilities under the PBD process

In order to compute $p_S^G(t)$ and $p_E^G(t)$ for any time $t \in [0,T]$, we first need to compute probabilities associated with incipient lineages.

Starting from an incipient lineage, the two exclusive possible ways leading to extinction within a horizon t are (figure II.2, upper left panel): (a) the first event to occur within the time horizon is an extinction or (b) the lineage forms two incipient lineages after a time $u \leq t$ and both incipient lineages go extinct within the remaining time $t - u$. The probability for an incipient lineage and its potential descendants to go extinct within a time t thus satisfies the following equation:

$$\begin{aligned} p_E^I(t) = &\int_0^t \mu_2 e^{-\Lambda u}\,\mathrm{d}u && \text{figure II.2(a)} \\ &+ \int_0^t \lambda_3 e^{-\Lambda u} (p_E^I(t-u))^2\,\mathrm{d}u && \text{figure II.2(b)} \end{aligned} \tag{II.3}$$

where $\Lambda = \lambda_2 + \lambda_3 + \mu_2$.

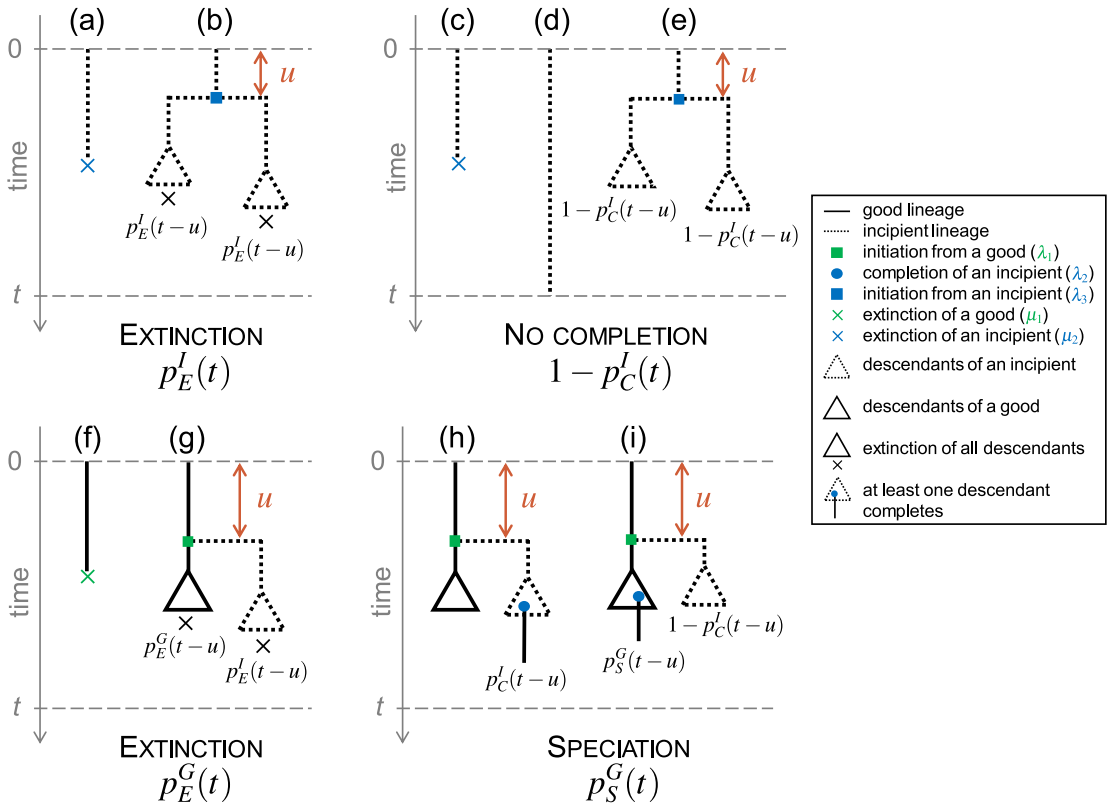


Figure II.2 – Alternative scenarios by which main outcomes occur in the protracted birth–death model. Upper panels: starting from an incipient lineage at time 0, decomposition of the possible exclusive ways of extinction (left) and no completion (right) before a given time t. Bottom panels: starting from a good lineage at time 0, decomposition of the possible exclusive ways of extinction (left) and speciation (right) before a given time t. Dotted lines represent incipient lineages, solid lines represent good lineages. Triangles summarize the subtrees containing the potential descendants of an ancestor lineage, with the condition that all of them go extinct within the remaining time (indicated by a cross), or that one of them completes speciation (indicated by a blue dot), or that none of these events occur. The probabilities written under the triangles correspond to the probability of the described event.

The solution to this integral equation is (subsection II.5.1):

$$p_E^I(t) = \frac{1}{\lambda_3}\sqrt{\frac{c(\Lambda-k)e^{kt}+\Lambda/k}{\left(ce^{kt}+1/k\right)^2}} - \frac{\Lambda(k-\Lambda)+2\mu_2\lambda_3}{2}$$

with $k = \sqrt{\Lambda^2 - 4\mu_2\lambda_3}$ and $c = \frac{2}{k-\Lambda} - \frac{1}{k}$.

Instead of $p_C^I(t)$, it is easier to calculate $1 - p_C^I(t)$, the probability of non-completion. Starting from an incipient lineage, the three exclusive ways not to complete speciation are (figure II.2, upper right panel): (c) the first event to occur is the extinction of the lineage within a time t, (d) survival of the lineage during a time t without completion, extinction or initiation, or (e) initiation of speciation after a time $u \leq t$ and non-completion of any of the daughter lineages within the remaining time $t-u$. The probability of speciation completion of an incipient lineage (or any of its potential descendant) within a time t thus satisfies the following equation:

$$\begin{aligned} 1 - p_C^I(t) = & \int_0^t \mu_2 e^{-\Lambda u}\,\mathrm{d}u && \text{figure II.2(c)} \\ & + e^{-\Lambda t} && \text{figure II.2(d)} \\ & + \int_0^t \lambda_3 e^{-\Lambda u}(1 - p_C^I(t-u))^2\,\mathrm{d}u && \text{figure II.2(e)} \end{aligned} \qquad \text{(II.4)}$$

This equation can be solved numerically by solving an ordinary differential equation (subsection II.5.2).

Starting now from a good lineage, the two exclusive possible ways leading to extinction as first event within a horizon t are (figure II.2, bottom left panel): (f) the lineage directly goes extinct within a time t or (g) the lineage forms an incipient lineage and both daughter lineages (one good and one incipient) and their potential descendants die within a time $t-u$ without speciation. The probability for a good lineage and its potential descendants to go extinct within a time t thus satisfies the following equation:

$$\begin{aligned} p_E^G(t) = & \frac{\mu_1}{\Theta}(1 - e^{-\Theta t}) && \text{figure II.2(f)} \\ & + \int_0^t \lambda_1 e^{-\Theta u} p_E^I(t-u) p_E^G(t-u)\,\mathrm{d}u. && \text{figure II.2(g)} \end{aligned} \qquad \text{(II.5)}$$

where $\Theta = \lambda_1 + \mu_1$.

This equation can be solved numerically, provided that we already have a solution for $p_E^I(t)$ (subsection II.5.3).

Starting from a good lineage, the two exclusive ways in which speciation occurs within a time t are (figure II.2, bottom right panel): (h) the lineage initiates speciation after a time $u \leq t$, and the incipient lineage completes speciation within the remaining time $t-u$, or (i) the lineage initiates speciation at a time $u \leq t$, the completion of this incipient lineage fails, and the good lineage speciates within the remaining time $t-u$.

The probability that a good lineage fulfils speciation within a time t thus satisfies the following equation:

$$p_S^G(t) = \int_0^t \lambda_1 e^{-\Theta u} \times \quad \text{initiation at time } u$$
$$\left[p_C^I(t-u) \right. \quad \text{figure II.2(h)}$$
$$\left. +(1-p_C^I(t-u))p_S^G(t-u)\right] \mathrm{d}u \quad \text{figure II.2(i)} \qquad \text{(II.6)}$$

This equation can be solved numerically (subsection II.5.4).

For all numerical integrations, we used the module `SciPy` in Python (Virtanen et al. 2020).

After solving these equations numerically and using equation II.2, we obtain the time-dependent equivalent birth rate $\hat{\lambda}(t)$ and death rate $\hat{\mu}(t)$.

II.2.3 Equivalent constant BD rates

Defined as above, the equivalent rates $\hat{\lambda}(t)$ and $\hat{\mu}(t)$ depend on time, even though the parameters of the PBD are constant. However, as we will show later and in agreement with previous results (Etienne and Rosindell 2012), the time-dependency is particularly manifest towards the present. Intuitively, given enough time, constant rates of initiation, completion and extinction result in constant equivalent speciation and extinction rates. However, towards the present (i.e., towards the end of the PBD process), incipient lineages did not have enough time to complete speciation, resulting in a decline in speciation rates. As one of our main goals here is to understand how speciation initiation, completion and extinction translate into macroevolutionary speciation and extinction rates, we now introduce equivalent constant BD rates, meant to represent the relationship between the parameters of the PBD and BD process far from the present.

We define equivalent constant BD rates, $\tilde{\lambda}$ (birth rate) and $\tilde{\mu}$ (death rate), as the parameters of the standard BD process such that the probability of speciation and the expected time to speciation match those under the PBD process with rates λ_1, λ_2, λ_3, μ_1, and μ_2, starting from a good lineage. The probability of extinction does not bring any additional information since extinction and speciation events are complementary over infinite time. Intuitively, we expect that the equivalent time-dependent birth and death rates tend towards $\tilde{\lambda}$ and $\tilde{\mu}$ in the past; the equivalent time-dependent birth rate then declines toward the present.

We shown that equivalent constant-time rates are given by the following expressions (see details subsection II.5.5):

$$\tilde{\lambda} = (1-\pi)\lambda_1 \quad \text{and} \quad \tilde{\mu} = \mu_1 \qquad \text{(II.7)}$$

where

$$\pi = \frac{\lambda_2 + \lambda_3 + \mu_2}{2\lambda_3} \left(1 - \sqrt{1 - 4\frac{\lambda_3 \mu_2}{(\lambda_2 + \lambda_3 + \mu_2)^2}} \right)$$

is the probability of non-completion of an incipient lineage.

Equation II.7 shows that, far from the present, the equivalent death rate is exactly the rate of extinction of good species, and the equivalent birth rate is directly proportional to the rate of speciation initiation from a good species, with a coefficient of proportionality that represents the probability of completion without time horizon and depends on the rates specific to incipient lineages (initiation, completion and extinction). As in the case of the time-dependent equivalent rates, these rates correspond to the case of a process starting with a good lineage without incipient lineages, and the equivalent extinction rate may thus overestimate actual extinction probabilities under the PBD.

In order to better understand the influence of each parameter of the PBD process on the equivalent constant birth rate $\tilde{\lambda}$, we calculate partial derivatives of this function with respect to the different parameters. A high partial derivative with respect to a given parameter reflects a strong influence of this parameter on the equivalent birth rate, and therefore that the corresponding step of the speciation process may be limiting. We compute the relative influence of a given parameter as the ratio between the absolute partial derivative with respect to this parameter and the sum of the absolute partial derivatives with respect to all other parameters. We perform these analyses both for the simplified PBD model where good and incipient lineages have the same rates of initiation ($\lambda_1 = \lambda_3$) and extinction ($\mu_1 = \mu_2$) and for the full PBD model. Detailed calculations are provided in subsection II.5.6.

II.2.4 Simulations under the PBD process and equivalent BD processes

Although the equivalent rates were designed to construct BD processes that approach the PBD process, these processes are not identical. We used simulations to compare reconstructed trees (i.e., trees of extant species) generated by the PBD process and their equivalent BD processes. In each of these simulations, we consider the trees of extant good species, disregarding extinct and incipient lineages. We compared trees simulated under the constant-rate PBD model with trees simulated under the corresponding time constant (BD) and time-varying (varBD) equivalent BD models. We independently varied each of the 5 parameters of the PBD model, each taking 5 values (the default value and 2 above, 2 below, with amplitude chosen to guarantee computational tractability), resulting in 25 parameter combinations. For each combination, we computed the trajectories of equivalent time-dependent birth and death rates over 15 million years — approximated by piecewise-constant birth and death rates over 200 intervals — and simulated 500

tree replicates using the R library `TreeSim` (Stadler 2011). The values of these rates are given in figure S II.4. Each tree simulation was conditioned on the survival of two extant lineages (up to the failure of 100 simulation attempts for each replicate), starting from a single stem branch. We also generated the same number of trees under the PBD model for each combination of parameters using the `PBD` package (Etienne and Rosindell 2012). Finally, we simulated trees under the BD model with equivalent constant rates, using the package `TreeSim`. We expect these simulations to deviate the most from those obtained under the PBD process. We compared the outputs of the simulations under the three models in terms of species richness at present (SR), tree shape and tree topology.

To analyze tree shape, we used the γ statistic (Pybus and Harvey 2000), computed with the package `ape` (Paradis and Schliep 2019). The γ statistics quantifies the relative position of the internal nodes of a tree and compares it to the expectations under a pure-birth (Yule) model. $\gamma > 0$ corresponds to trees where internal nodes are closer to the tips than expected under Yule's model, while $\gamma < 0$ corresponds to trees where internal nodes are closer to the root.

To analyze tree topology, we used the stairs2 balance index (Norström et al. 2012), computed with the package `treestats` (Janzen and Etienne 2024). This statistic measures the mean size ratio between the smaller and larger pending subtree for all vertices. Stairs2 is higher for trees with more balanced subtrees and lower for more imbalanced trees. The stairs2 statistic has been shown to perform well (Khurana et al. 2023; Kersting et al. 2025) and is less sensitive to tree size than other statistics such as Aldous' β (Aldous 2001).

II.2.5 Tip speciation rate estimates

As we will show later and already mentioned, the equivalent birth rate declines close to the present, suggesting that speciation rates estimated at the tips of phylogenies may poorly reflect the underlying speciation process. To evaluate this effect, we computed the widely used DR (diversification rate) statistic (Jetz et al. 2012; Title and Rabosky 2019) at the tip of all the trees we simulated under the PBD process, using the package `epm` (Title et al. 2022). Next, for each tree, we compared the median DR over all extant leaves to the constant-rate equivalent birth rate. In order to evaluate deviations due solely to the use of DR (a biased estimator of speciation rate), we also computed DR at the tips of trees simulated under the constant-rate equivalent BD process.

II.2.6 Ability to recover equivalent constant BD rates by fitting the BD model to truncated trees

If we acknowledge that speciation in nature usually takes time, birth and death rates estimates obtained by fitting a constant-rate BD model to empirical reconstructed trees are hard to interpret. As noted above, under a constant rate PBD model, we expect equivalent birth rates to approach the equivalent constant birth rate $\tilde{\lambda}$ in the past, and to decline closer to the present. Hence, we can expect that speciation rate estimates obtained by fitting a BD model to the entire tree will have intermediate values below $\tilde{\lambda}$. However, fitting a BD model to to older parts of the tree should provide good estimates of the equivalent constant BD rates.

To test this expectation, we truncated the phylogenies simulated under the PBD process in subsection II.2.4 at different time points in the past, and fitted a constant BD model to these truncated phylogenies, using the dedicated function `fit_bd_in_past` (Lewitus et al. 2018; Perez-Lamarque et al. 2022) from the R package `RPANDA` (Morlon et al. 2016). We fitted a constant-rate BD model to both entire trees and truncated trees "sliced" at 17 regularly spaced time points between the present and 4 million years before the present. Finally, we compared the speciation and extinction rate estimates obtained with various truncation times to the analytically-derived equivalent constant BD rates.

II.3 RESULTS

II.3.1 Equivalent time-dependent BD rates

We used equations II.1 and II.2 and numerical solutions of the equations describing the probabilities of speciation, completion and extinction with time to derive equivalent time-dependent birth and death rates (thereafter simply referred to as birth and death rates) for a large range of parameter values.

We find that birth rates decrease close to the present, reaching 0 at present (time $t \to T$, see figure II.3). Death rates depend less on time and can be considered almost constant with time if we neglect a small decrease followed by a short increase when $t \to T$. In the past ($t \to 0$), birth and death rates converge to the values predicted by our analytical expression of the equivalent constant BD rates (equation II.7). This provides indirect (graphical) evidence that the constant equivalent birth–death rates can be considered as asymptotic rates of the time-dependent equivalent birth–death rates.

Initiation rates have a strong effect on the birth rate (figure II.3A). In the past, birth rates converge to values scaling with the initiation rate, all other rates being equal, as expected from equation II.7. As expected, lower completion rates result in lower birth rates, and an effect of the protracted nature of speciation that extends further into the past

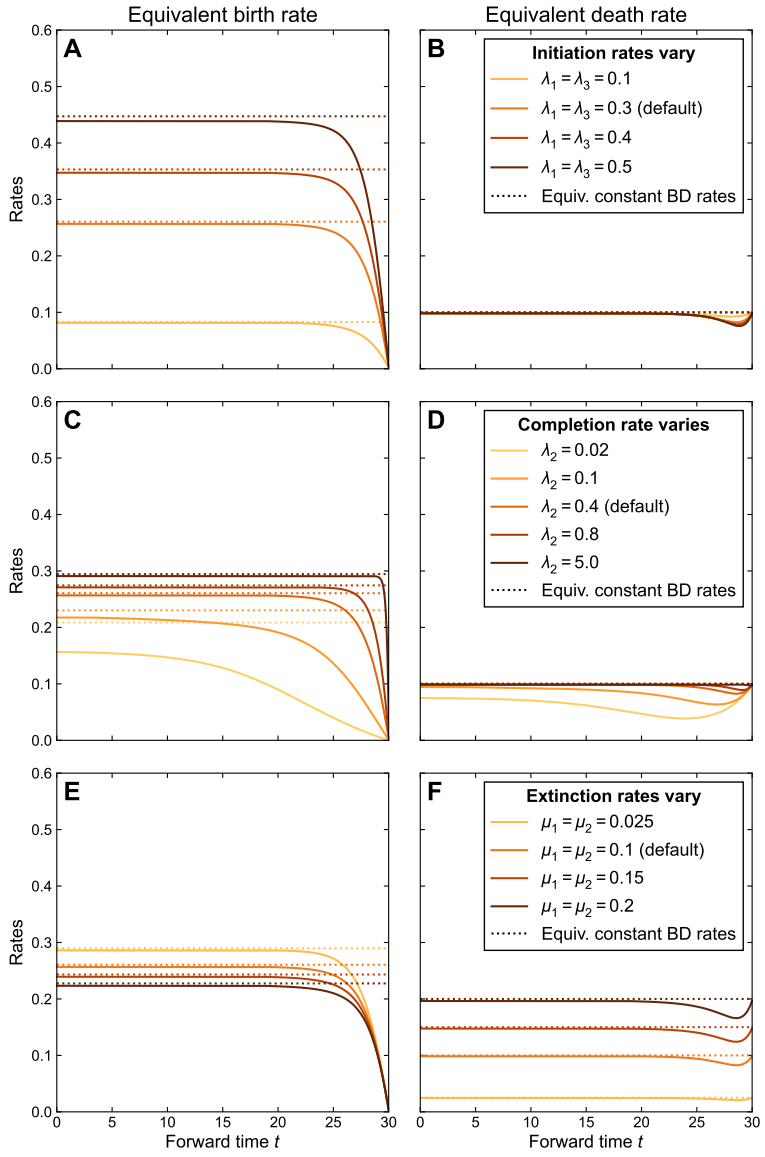


Figure II.3 – Influence of the parameters of the protracted birth–death (PBD) process on equivalent birth–death (BD) rates. Solid lines represent equivalent BD rates derived from equations II.1 and II.2 as a function of time for different values of PBD rates. Dotted lines represent constant equivalent BD rates derived analytically from equation II.7 for the same PBD parameters. In the top/middle/bottom row the initiation/completion/extinction rates vary, with the other rates constant (default values are $\lambda_1 = \lambda_3 = 0.3, \lambda_2 = 0.4, \mu_1 = \mu_2 = 0.1$). In figures S II.2 and S II.3, we calculated the same rates with the 5 parameters varying independently. $t = 30$ corresponds to the present, $t = 0$ to the past.

(panel C). In the limit $\lambda_2 \to \infty$, the model converges to a pure BD model with constant rates, except very close to the present. Indeed, with high completion rates, incipient lineages complete speciation very fast and with high probability, so speciation occurs as soon as an initiation event occurs. Finally, birth rates are lower when extinction rates are higher, except closer to the present where the effect of extinction diminishes (panel E). This effect is entirely due to the extinction of incipient lineages (μ_2), as the extinction rate of good lineages (μ_1) has virtually no effect on the birth rate (figure S II.2). The higher extinction rate of incipient lineages renders speciation less likely, as incipient species more often go extinct before completing speciation.

Death rates closely match extinction rates (figure II.3F), entirely due to the effect of the extinction rate of good lineages, as expected from equation II.7; indeed, the extinction rate of incipient lineages has no effect on the death rate (figure S II.2).

II.3.2 Equivalent constant BD rates

As shown by equation II.7 and already described above, the equivalent constant death rate equals the extinction rate of good species; the equivalent constant birth rate scales with speciation initiation, it increases with the completion rate and decreases with the extinction rate of incipient lineages (figure II.3, dashed lines). We better characterized the influence of each step of the speciation process (i.e., initiation, survival of incipient species and completion) on $\tilde{\lambda}$, interpreted as the macroevolutionary speciation rate, by computing relative partial derivatives. This allows us to identify the steps of the PBD process that may limit macroevolutionary speciation rates figure II.4. We identify that speciation initiation is limiting when its rate is low and the completion rate is high (regions 1 and 2). Speciation completion is limiting when its rate is low compared to the other parameters (regions 3 and 5). An increase in the population extinction rate has most effect when this rate is low (region 6) and when initiation rate is high compared to the completion rate (regions 4 and 5).

II.3.3 Trees generated by the PBD process and equivalent BD processes

We compared the size, shape and topology of trees generated under the PBD model to those of trees generated under their equivalent time-constant and time-varying BD models (figure II.5). As expected, tree size (SR) increases with higher rates of speciation initiation and completion, and decreases with higher rates of extinction of good and incipient lineages (figure II.5, top row). These trends are well captured by both the equivalent time-varying and time-constant models, with the exception of the increase in species richness with the rate of initiation from incipient lineages λ_3. Compared to the PBD model, the equivalent BD models produce larger trees when λ_3 is small, and

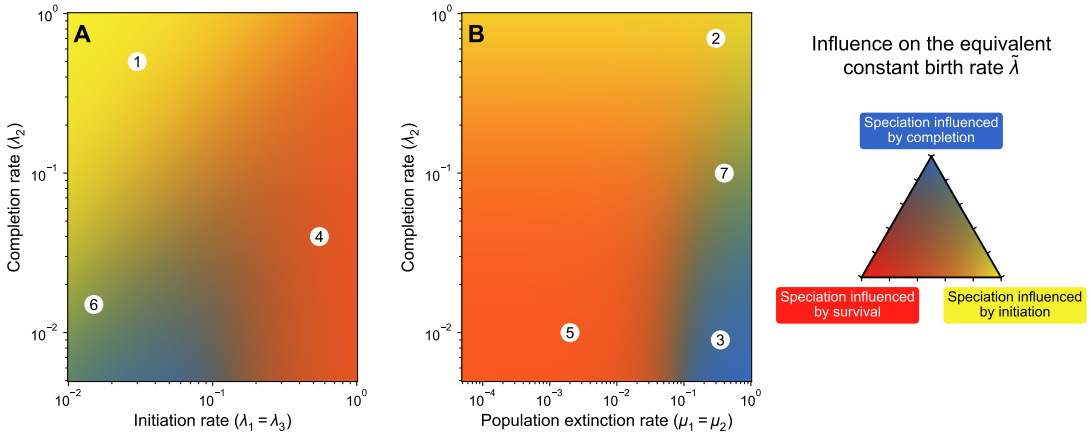


Figure II.4 – Relative influence of the parameters of the PBD model on the equivalent constant birth rate. Colors indicate which of the PBD process among initiation, completion and population extinction limits the equivalent constant birth rate $\tilde{\lambda}$ most, as a function of the initiation and completion rates (**A**) and the population extinction and completion rates (**B**), with the color code explained on the triangle in the right. A yellow region (1 or 2) indicates a combination of parameters where the most influential parameter on the birth rate is the rate of initiation. A blue region (3) indicates a combination of parameters where the most influential parameter is the rate of completion λ_2. An orange region (4 or 5) indicates a combination of parameters where both population extinction and speciation initiation are influential. Green regions (6 or 7) indicate situations where both speciation initiation and completion have a similar influence on the birth rate. In all cases, the rates of initiation and completion have a positive influence on the birth rate and the rate of population extinction has a negative influence. The detailed methods are explained in subsection II.5.6 and the values of the relative influence are provided in figure S II.5. When they do not vary, default values of the parameters are 0.1.

smaller trees when λ_3 is large. λ_3 has little influence on the size of trees generated under equivalent BD models, consistently with the weak influence of λ_3 on the equivalent rates (see figure S II.3C and D).

As expected given unachieved speciation close to the present, the PBD process results in trees with negative γ values (reflecting long terminal branches), unless the completion or extinction rates are high (figure II.5, middle row). γ decreases with increasing ratios of speciation initiation to speciation completion rates, i.e., when the protracted nature of the speciation process is more pronounced. The equivalent variable-rate BD model captures these trends relatively well, while the constant-rate BD model produces trees with positive γ values, indicating nodes closer to the tips. The drop towards 0 in the equivalent time-variable birth rates captures the shortage of speciation events close to

the present induced by the protracted process, except when the rate of initiation from an incipient lineage (λ_3) gets large. In this case the PBD process produces trees with increasingly long branch lengths, while the distribution of nodes generated under the equivalent time-dependent BD model remains stable.

Under most PBD parameters, stairs2 values of tree topology are stable, close to 0.65, and well reproduced by both the time-constant and the time-variable equivalent BD models. Higher values of stairs2 (reflecting more balanced trees) are observed in parameter ranges leading to small trees, and likely reflect tree size rather than a true difference in tree balance.

Deviations observed between trees simulated under the PBD process and those simulated under the equivalent time-varying BD model under some λ_3 values likely come from the approximation we made when establishing the link between equivalent rates and the probabilities of speciation and extinction under the PBD (see section II.2.2). If, as expected, the equivalent extinction rate overestimates extinction probabilities, this explains why we obtain smaller trees with less negative γ values, as extinction pushes nodes towards the present (the "pull of the present"; Nee et al. 1994). This is especially pronounced if λ_3 is high; in this case, there are many incipient lineages that are not accounted for (the ones that originated before the time at which the rates are calculated).

II.3.4 Tip speciation rate estimates

DR tip speciation rates estimated on trees simulated under the PBD process are generally below equivalent constant birth rates (figure S II.8), as expected given the decrease in the equivalent birth rate near the present. This is not due to biases linked to the use of the DR statistics, as DR computed on trees generated under the equivalent BD process closely match equivalent birth rates.

II.3.5 Recovery of equivalent BD rates by fit to truncated PBD trees

Our expectation that fits of a constant rate BD model to the "old" part of trees generated under a PBD process would provide good estimates of the equivalent BD rates was verified (figure S II.7). When fitting a constant rate BD model to the entire tree (i.e., when the truncation time is zero), the estimated speciation and extinction rates are well below the expected equivalent constant rates. However when truncation time increases, these estimates converge to the expected equivalent rates. In the case of the extinction rate, estimates remain slightly below the expected equivalent rates, supporting our intuition that equivalent extinction rates should overestimate actual extinction events. The convergence occurs with a truncation time relatively close to the present, consistent with the observed time of decline of the equivalent time-dependent BD rates. This recent

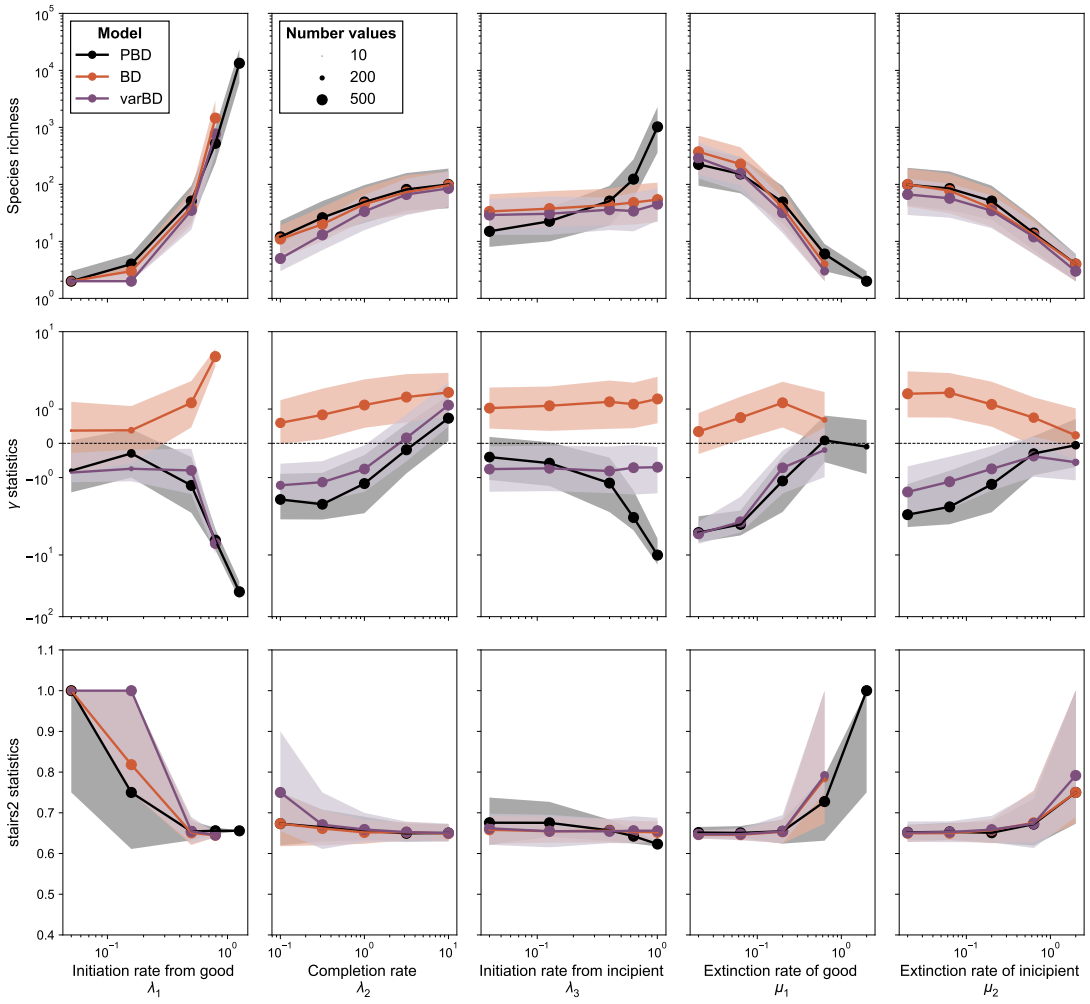


Figure II.5 – Statistics of trees generated under the protracted birth–death (PBD) process and its equivalent birth–death (BD) processes. By row: species richness SR, γ shape statistic and stairs2 balance index for trees generated under the three models (PBD, equivalent constant rate BD and equivalent time varying rate BD) for different values of the parameters of the PBD model. In each column, only one of the PBD parameters varies, with the others held constant (default $\lambda_1 = 0.5, \lambda_2 = 1.0, \lambda_3 = 0.4, \mu_1 = 0.2, \mu_2 = 0.2$). For each set of parameters of the PBD model, BD and varBD trees were generated under equivalent birth and death rates computed using equations II.1, II.2 and II.7. The line corresponds to the median of statistics across all 200 replicates, the shaded area indicates the first and last quartile of the statistics. The size of the dots indicates the number of valid data for which the statistics could be calculated.

truncation time appears optimal, as estimates are not as good when more of the tree is truncated, probably due to a loss of statistical power with decreasing data size.

II.4 DISCUSSION

In this study we derived predictions, under the protracted birth–death (PBD) model of diversification, for "equivalent" birth and death rates, meant to represent macroevolutionary speciation and extinction rates. We showed that the equivalent rates — in particular the birth rate — vary when the process approaches the present, but can be considered constant when far from the present. Our analytical predictions of the rates in the past allowed us to explore the importance of each step of the speciation process (i.e., initiation, survival, completion) in modulating the macroevolutionary speciation rate. We found that the initiation and survival rates in general play a much larger role than the completion rate. In addition, we showed that the constant equivalent birth rate can be estimated by fitting a standard birth–death process on truncated reconstructed trees. This opens the possibility to relate estimates of macroevolutionary speciation rates, which have major consequences for the build up of diversity over geological time scales, to the microevolutionary processes that modulate speciation initiation, population survival and speciation completion.

Our equivalent rates are distinct from "congruent" rates, defined by Louca and Pennell (2020) as a combination of birth and death rates that result in the same likelihood on reconstructed trees. First, the equivalent BD process with variable rates is only an approximation of the PBD process (see the following paragraph), and therefore the equivalent BD model is not strictly speaking congruent to the PBD model. Second, we found a unique solution to our equations, demonstrating the unicity of equivalent rates: they are the only one that yield the same speciation and extinction probabilities at all time. Any other scenario in the same congruence class would have the same likelihood on reconstructed trees, but not the same speciation and extinction probabilities. An implication of this observation is that a time-varying birth–death scenario fitted on a reconstructed tree is not guaranteed to provide the equivalent rates. The criterion we chose to define equivalent rates is also distinct from the criterion used by Pannetier et al. (2021) to find a time-dependent BD model that approximates the diversity-dependent BD model. In the latter paper, the authors computed rates yielding the same expected diversity-through-time (DTT). We expect our equivalent rates to produce the same DTT as that of the underlying PBD process, but we have not demonstrated that there are no other combinations of rates that also yield the same DTT.

As expected given previous results on the protracted birth–death process (Etienne and Rosindell 2012), we find that equivalent birth rates decline to zero as time approaches the present. Close to the present, speciation is less likely to occur because it requires

a delay. The decay in equivalent birth rate starts earlier when the completion rate is lower. Trees simulated under the equivalent time-dependent BD model have similar characteristics to trees simulated under the PBD process, in terms of tree size and shape (figure II.5), meaning that the equivalent BD process captures the dynamics of speciation and extinction induced by the protracted model relatively well. It is however important to remember that the PBD and equivalent BD processes are not entirely interchangeable. PBD is a process with memory, which induces an age dependency of speciation and extinction rates that is not captured by the equivalent BD process. For example "old" species are more likely to have accumulated incipient lineages, which buffer their extinction risk, but this cannot be captured by an age-independent BD process. More generally, the approximations made in our equivalent BD process lead to an overestimation of the frequency of extinction events occurring under the PBD process, explaining some of the deviations observed when comparing simulated trees, in particular for high values of the rate of initiation from incipient lineages (λ_3).

Before equivalent speciation rates start to drop, they can be considered virtually constant. In this regime, we derived analytical relationships between the equivalent birth and death rates and the parameters of the PBD process. These relationships show that we can expect macroevolutionary extinction rates to provide a relatively good approximation of the rate at which species go extinct, except if the rate of initiation from incipient lineages is large. They also show that macroevolutionary speciation rates are directly proportional to speciation initiation rates, with a coefficient of proportionality that depends on the rates of completion, initiation and extinction of incipient lineages. Hence, all aspects of the speciation process play a role in modulating macroevolutionary speciation rates, but some play a larger role than others, and which ones will likely depend on the ecology, genetics and biogeography of the species group considered. Indeed, by identifying regimes within which each aspect of the speciation process is expected to be the most influential (figure II.4), we found that the rates of speciation initiation and population extinction often are the limiting factors. Completion rates are only limiting when they are very small and at intermediate values of speciation initiation rates, or high population extinction rates. Hence, a species that accumulates fast reproductive isolation will be characterized by a higher completion rate but this will not necessarily have a strong effect on the speciation rate. In their analysis of Australian rainbow skinks, Hua et al. (2022) estimated rates of speciation initiation and completion of 0.27 and 0.31, respectively, with low extinction rates. In this area of parameter space, speciation rates are clearly not limited by the speed at which reproductive isolation is acquired (relative influence $\sim 10^{-7}$), but more by the initiation of speciation (68%) and the persistence of lineages (32%) (figure S II.9).

Our results have implications for the phylogenetic analysis of diversification. Indeed, ignoring the fact that speciation takes time by fitting standard BD models to empirical

phylogenies necessarily leads to model misspecification (e.g., when fitting a constant rate BD model) or misinterpretation (e.g., when interpreting the effect of protraction as a speciation rate decline). In the recent years, there has been a surge in the use of tip-rate speciation estimates, i.e., estimates of speciation rates at present across the tips of a phylogeny, obtained with statistics such as the diversification rate (DR; Jetz et al. 2012) statistic, or models such as the Bayesian analysis of macroevolutionary mixtures (BAMM; Rabosky 2014), the cladogenetic diversification shift (ClaDS; Maliet et al. 2019; Maliet and Morlon 2021), or the birth–death diffusion (BDD; Quintero et al. 2024). Future work investigating how speciation rates estimated with these methods on trees generated under the PBD process compare with equivalent rates will be useful. We conducted this exercise here with the DR statistic, and found that the obtained rates generally underestimate the equivalent constant birth rate, reflecting the convergence to zero of the equivalent birth rate close to the present. Hence, tip rate estimates are particularly challenging to interpret in terms of the underlying speciation process. Speciation rates estimated on deeper parts of the phylogeny approximate equivalent rates quite well, suggesting that these estimates indeed reflect the macroevolutionary outcome of the combined processes of initiation, survival and completion. Ideally, we would need to account for the protracted nature of speciation in every model used to infer diversification rates from phylogenies. By incorporating a protracted process in the State-dependent Speciation-Extinction (SSE) framework, Hua et al. (2022) made progress in this direction. A workaround is to truncate phylogenies, a method that has previously been used in another context (Phillimore and Price 2008). However, when truncating actual species splits to avoid the effect of protraction close to the present, this requires employing inference methods that account for this truncation. These have been implemented for only a limited set of models, such as the constant-rate, time-dependent and environment-dependent models developed in `RPANDA` (Morlon et al. 2016). More systematically implementing this truncation option in diversification models would be useful. Our tests indicate that truncating phylogenetic trees at intermediate time points yields the most accurate estimates of equivalent rates. As the optimal truncation time will depend on the extent to which speciation is protracted (i.e., the completion rate), we recommend to apply truncation at various time points and choose the time at which the estimated rates reach a plateau.

Our results also have implications for the ongoing effort to relate macroevolutionary speciation rates to microevolutionary processes (Rabosky 2016; Harvey et al. 2019; Rolland et al. 2023; Morlon et al. 2024). Indeed, microevolutionary processes act individually on each step of the speciation process (initiation, survival, completion), and our expressions of equivalent rates quantify how these combine to modulate macroevolutionary speciation rates. For example, the apparent decoupling between tip rate estimates of speciation and the speed at which reproductive isolation is acquired (Rabosky and Matute 2013; Freeman et al. 2022) is not surprising given our expectation that completion rates have a limited effect on the macroevolutionary speciation rate (and the limitations

of tip rate estimates mentioned before). Etienne et al. (2014) observed a weak correlation between speciation completion rates estimated with the PBD model and speciation rates estimated with a BD model, on both simulated and bird phylogenies, especially when the completion rate was low. The authors interpreted this result as the limits of estimating speciation rates with a BD process when speciation is protracted. Our results suggest that the decoupling is in fact real and expected. We would expect the correlation to be especially weak when the completion rate is high rather than low, and the discrepency between these expectations and Etienne et al.'s observations could be linked to the limits of estimating speciation rates with a BD process. We, however, expect to find a correlation between macroevolutionary speciation rates and the rate of population formation, which reflects speciation initiation. This is indeed what was found by Harvey et al. (2017); the lack of correlation in other studies (Singhal et al. 2018; Singhal et al. 2022; Burbrink et al. 2023) may reflect cases when speciation is modulated by the rate of population extinction rather than by the rate of speciation initiation, as expected if the initiation rate is high and/or the population extinction rate is low.

By taking into account the time it takes to complete speciation, protracted birth–death models provide more biologically realistic models than standard birth–death models, and a theoretical framework for understanding how each step of the speciation process influences macroevolutionary speciation rates. A more systematic account of the protracted nature of speciation in phylogenetic analyses of diversification would both improve our estimates of diversification rates and our understanding of how microevolutionary processes combine to modulate macroevolutionary speciation rates.

II.5 APPENDIX

II.5.1 Resolution of the probability of extinction of an incipient lineage

In equation II.3, let us do the variable change $v = t - u$ to get rid of the variable t in the integral:

$$\begin{aligned} p_E^I(t) &= \frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + \int_0^t \lambda_3 e^{-\Lambda(t-v)}(p_E^I(v))^2 \,\mathrm{d}v \\ &= \frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + \lambda_3 e^{-\Lambda t}\int_0^t e^{\Lambda v}(p_E^I(v))^2 \,\mathrm{d}v \\ &= \frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + \lambda_3 e^{-\Lambda t} f(t) \end{aligned}$$

with f defined with, for all $t \geq 0$:

$$f(t) := \int_0^t e^{\Lambda v}(p_E^I(v))^2 \,\mathrm{d}v. \tag{II.8}$$

We note that

$$\begin{aligned} f'(t) &= e^{\Lambda t}(p_E^I(t))^2 \\ &= e^{\Lambda t}\left(\frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + \lambda_3 e^{-\Lambda t} f(t)\right)^2. \end{aligned}$$

This is a non-linear ordinary differential equation (ODE) of first order, with the initial condition $f(0) = 0$.

$$\begin{aligned} \Lambda\varphi(t) + \varphi'(t) &= \left[\frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + \lambda_3\varphi(t)\right]^2 \\ \Leftrightarrow \quad \varphi'(t) &= \left[\frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + \lambda_3\varphi(t)\right]^2 - \Lambda\varphi(t) \end{aligned} \tag{II.9}$$

Let us check that the function

$$\varphi(t) := \frac{1}{\lambda_3^2\left(ce^{kt} + 1/k\right)} - \frac{\Lambda(k-\Lambda) + 2\mu_2\lambda_3}{2\lambda_3^2\Lambda} + \frac{\mu_2}{\lambda_3\Lambda}e^{-\Lambda t}.$$

with $k = \sqrt{\Lambda^2 - 4\lambda_3\mu_2}$ and $c > 0$ satisfies this last ODE. Inside the bracket of the right-hand side (RHS) of equation II.9 we notice the simplification:

$$\begin{aligned}
&\frac{\mu_2}{\Lambda}(1-e^{-\Lambda t})+\lambda_3\varphi(t)\\
&\quad=\frac{\mu_2}{\Lambda}(1-e^{-\Lambda t})+\frac{1}{\lambda_3(ce^{kt}+1/k)}-\frac{\Lambda(k-\Lambda)+2\mu_2\lambda_3}{2\lambda_3\Lambda}+\frac{\mu_2}{\Lambda}e^{-\Lambda t}\\
&\quad=\frac{1}{\lambda_3(ce^{kt}+1/k)}+\frac{\Lambda-k}{2\lambda_3}.
\end{aligned}$$

So the RHS of equation II.9 is:

$$\begin{aligned}
&\left[\frac{\mu_2}{\Lambda}(1-e^{-\Lambda t})+\lambda_3\varphi(t)\right]^2-\Lambda\varphi(t)\\
&\quad=\frac{1}{\lambda_3^2}\left[\frac{1}{(ce^{kt}+1/k)^2}+\frac{\Lambda-k}{ce^{kt}+1/k}+\frac{(\Lambda-k)^2}{4}\right]-\frac{\Lambda}{\lambda_3^2(ce^{kt}+1/k)}\\
&\qquad+\frac{\Lambda(k-\Lambda)+2\mu_2\lambda_3}{2\lambda_3^2}-\frac{\mu_2}{\lambda_3}e^{-\Lambda t}\\
&\quad=\frac{1}{\lambda_3^2(ce^{kt}+1/k)^2}+\frac{\Lambda-k}{\lambda_3^2(ce^{kt}+1/k)}+\frac{(\Lambda-k)^2}{4\lambda_3^2}-\frac{\Lambda}{\lambda_3^2(ce^{kt}+1/k)}\\
&\qquad+\frac{2\Lambda(k-\Lambda)+4\mu_2\lambda_3}{4\lambda_3^2}-\frac{\mu_2}{\lambda_3}e^{-\Lambda t}
\end{aligned}$$

and since $(\Lambda-k)^2+2\Lambda(k-\Lambda)+4\mu_2\lambda_3=k^2-\Lambda^2+4\mu_2\lambda_3=0$ given the definition of k, this simplifies into:

$$\begin{aligned}
&=-\frac{k}{\lambda_3^2(ce^{kt}+1/k)}+\frac{1}{\lambda_3^2(ce^{kt}+1/k)^2}-\frac{\mu_2}{\lambda_3}e^{-\Lambda t}\\
&=\frac{-kce^{kt}}{\lambda_3^2(ce^{kt}+1/k)^2}-\frac{\mu_2}{\lambda_3}e^{-\Lambda t}\\
&=\varphi'(t).
\end{aligned}$$

so the suggested φ is a solution of the ODE equation II.9. The condition $\varphi(0)=0$ imposes $c=\frac{2}{k-\Lambda}-\frac{1}{\Lambda}$. Finally, we have the solution of the original ODE:

$$f(t)=e^{\Lambda t}\left[\frac{1}{\lambda_3^2\left(ce^{kt}+\frac{1}{k}\right)}-\frac{\Lambda(k-\Lambda)+2\mu_2\lambda_3}{2\lambda_3^2\Lambda}\right]+\frac{\mu_2}{\lambda_3\Lambda}.$$

From equation II.8 we have

$$f'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_0^t e^{\Lambda v}(p_E^I(v))^2\,\mathrm{d}v\right) = e^{\Lambda t}(p_E^I(t))^2.$$

So the probability of extinction $p_E^I(t)$ can be retrieved

$$\begin{aligned} p_E^I(t) &= \sqrt{e^{-\Lambda t} f'(t)} \\ &= \frac{1}{\lambda_3}\sqrt{\frac{c(\Lambda-k)e^{kt}+\Lambda/k}{\left(ce^{kt}+1/k\right)^2} - \frac{\Lambda(k-\Lambda)+2\mu_2\lambda_3}{2}}. \end{aligned}$$

II.5.2 Resolution of the probability of completion of an incipient lineage

To get rid of the variable t in the integral of equation II.4, we do the change of variable $v = t-u$. This gives us:

$$\begin{aligned} 1-p_C^I(t) &= \frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + e^{-\Lambda t} + \int_0^t \lambda_3 e^{-\Lambda(t-v)}\left(1-p_C^I(v)\right)^2\mathrm{d}v \\ &= \frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + e^{-\Lambda t} + e^{-\Lambda t}\int_0^t \lambda_3 e^{\Lambda v}\left(1-p_C^I(v)\right)^2\mathrm{d}v \end{aligned}$$

we now do a logarithmic change of variable in the integral, $z = e^{\Lambda v}$:

$$\begin{aligned} &= \frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + e^{-\Lambda t} + e^{-\Lambda t}\int_1^{e^{\Lambda t}} \lambda_3 z\left(1-p_C^I\left(\frac{\log z}{\Lambda}\right)\right)^2\frac{1}{\Lambda z}\,\mathrm{d}z \\ &= \frac{\mu_2}{\Lambda}(1-e^{-\Lambda t}) + e^{-\Lambda t}\left(1+\frac{\lambda_3}{\Lambda}g\left(e^{\Lambda t}\right)\right) \end{aligned}$$

with, for $x \geq 1$:

$$g(x) := \int_1^x \left(1-p_C^I\left(\frac{\log z}{\Lambda}\right)\right)^2\mathrm{d}z.$$

Noting that $g'(x) = [1-p_C^I(\log x/\Lambda)]^2$, we have:

$$g'(x) = \left[\frac{\mu_2}{\Lambda}\left(1-\frac{1}{x}\right) + \frac{1}{x}\left(1+\frac{\lambda_3}{\Lambda}g(x)\right)\right]^2.$$

This is a non-linear ODE. With the initial condition $g(1) = 0$, it can be solved numerically and we retrieve the probability of completion of an incipient lineage with

$$p_C^I(t) = 1-\sqrt{g'\left(e^{\Lambda t}\right)}.$$

II.5.3 Resolution of the probability of extinction of a good lineage

From equation II.5 we do the same change of variable as in the previous part $v = t - u$:

$$p_E^G(t) = \frac{\mu_1}{\Theta}(1 - e^{-\Theta t}) + \lambda_1 e^{-\Theta t} \int_0^t e^{\Theta v} p_E^I(v) p_E^G(v)\, \mathrm{d}v$$

then with $z = e^{\Theta v}$ we have:

$$\begin{aligned}
&= \frac{\mu_1}{\Theta}(1 - e^{-\Theta t}) + \lambda_1 e^{-\Theta t} \int_1^{e^{\Theta t}} z p_E^I\left(\frac{\log z}{\Theta}\right) p_E^G\left(\frac{\log z}{\Theta}\right) \frac{1}{\Theta z}\, \mathrm{d}z \\
&= \frac{\mu_1}{\Theta}(1 - e^{-\Theta t}) + \frac{\lambda_1}{\Theta} e^{-\Theta t} \int_1^{e^{\Theta t}} p_E^I\left(\frac{\log z}{\Theta}\right) p_E^G\left(\frac{\log z}{\Theta}\right) \mathrm{d}z \\
&= \frac{\mu_1}{\Theta}(1 - e^{-\Theta t}) + \frac{\lambda_1}{\Theta} e^{-\Theta t} h(e^{\Theta t})
\end{aligned}$$

with, for $x \geq 1$

$$h(x) = \int_1^x p_E^I\left(\frac{\log z}{\Theta}\right) p_E^G\left(\frac{\log z}{\Theta}\right) \mathrm{d}z.$$

We note that

$$\begin{aligned}
h'(x) &= p_E^I\left(\frac{\log x}{\Theta}\right) p_E^G\left(\frac{\log x}{\Theta}\right) \\
&= p_E^I\left(\frac{\log x}{\Theta}\right) \times \left[\frac{\mu_1}{\Theta}\left(1 - \frac{1}{x}\right) + \frac{\lambda_1}{\Theta}\frac{1}{x} h(x)\right]. \qquad \text{(II.10)}
\end{aligned}$$

With the condition $h(1) = 0$, equation II.10 gives the ODE satisfied by h. If we solve this equation numerically we can retrieve $p_E^G(t)$:

$$p_E^G(t) = \frac{h'(e^{\Theta t})}{p_E^I(t)}.$$

II.5.4 Resolution of the probability of speciation of a good lineage

From equation II.6, we do the same change of variable $v = t - u$, then $z = e^{\Theta v}$:

$$\begin{aligned}
p_S^G(t) &= \lambda_1 e^{-\Theta t} \int_0^t e^{\Theta v} \left[p_C^I(v) + (1 - p_C^I(v)) p_S^G(v)\right] \mathrm{d}v \\
&= \lambda_1 e^{-\Theta t} \int_1^{e^{\Theta t}} z \left[p_C^I\left(\frac{\log z}{\Theta}\right) + \left(1 - p_C^I\left(\frac{\log z}{\Theta}\right)\right) p_S^G\left(\frac{\log z}{\Theta}\right)\right] \frac{1}{\Theta z}\, \mathrm{d}z \\
&= \frac{\lambda_1}{\Theta} e^{-\Theta t} \int_1^{e^{\Theta t}} \left[p_C^I\left(\frac{\log z}{\Theta}\right) + \left(1 - p_C^I\left(\frac{\log z}{\Theta}\right)\right) p_S^G\left(\frac{\log z}{\Theta}\right)\right] \mathrm{d}z \\
&= \frac{\lambda_1}{\Theta} e^{-\Theta t} m(e^{\Theta t})
\end{aligned}$$

with, for $x \geq 0$

$$m(x) := \int_1^x \left[p_C^I\left(\frac{\log z}{\Theta}\right) + \left(1 - p_C^I\left(\frac{\log z}{\Theta}\right)\right) p_S^G\left(\frac{\log z}{\Theta}\right)\right] \mathrm{d}z.$$

We note that

$$m'(x) = p_C^I\left(\frac{\log x}{\Theta}\right) + \left(1 - p_C^I\left(\frac{\log x}{\Theta}\right)\right) p_S^G\left(\frac{\log x}{\Theta}\right) \tag{II.11}$$

$$= p_C^I\left(\frac{\log x}{\Theta}\right) + \left(1 - p_C^I\left(\frac{\log x}{\Theta}\right)\right) \frac{\lambda_1}{\Theta}\frac{1}{x} m(x). \tag{II.12}$$

With the initial condition $m(1) = 0$, the equation II.12 gives the ODE satisfied by m. If we solve this equation numerically we can retrieve $p_S^G(t)$ with the derivative of the solution m, from equation II.11:

$$p_S^G(t) = \frac{m'(e^{\Theta t}) - p_C^I(t)}{1 - p_C^I(t)}.$$

We show an example of the obtained numerical solutions of p_E^I, p_C^I, p_E^G and p_S^G and empirical probabilities from simulations in figure S II.1.

II.5.5 Calculation of the time-constant BD rates under the PBD model

Speciation probability and expected time for speciation under the BD model

Under the BD model with parameters λ and μ, the probability of speciation of a given lineage is the probability that a birth event occurs before a death event, i.e.,

$$\mathbf{P}(\text{speciation}) = \frac{\lambda}{\lambda + \mu}. \tag{II.13}$$

Conditionally on speciation, the expected time T it takes for a given lineage to speciate is

$$\mathbb{E}[T|\text{speciation}] = \frac{1}{\lambda + \mu} \tag{II.14}$$

since T has the same distribution as $\min(X,Y)$ with $X \hookrightarrow \mathscr{E}(\lambda)$ and $Y \hookrightarrow \mathscr{E}(\mu)$ two independent clock variables representing speciation and extinction events. The minimum of two exponential independent processes is distributed exponentially with a parameter equal to the sum of the rates.

Speciation probability and expected time for speciation under the PBD model

Under the PBD model, a good lineage speciates if it generates at least one incipient lineage and one of these incipient lineages or one of the descendants of these lineages completes speciation.

Considering an incipient lineage L that has still no descendant, there are two outcomes: (*i*) the lineage or one of its descendants at least complete speciation (we will denote this event as "complete"), or (*ii*) neither the lineage nor any of its descendants complete speciation before dying (we will denote this event as "does not complete").

Let us denote $\pi := \mathbb{P}(L \text{ does not not complete})$ and N the number of direct descendants of the lineage before it dies or completes speciation, and L and $L_1, ..., L_N$ those lineages. All the possible events that can happen to L (completion at rate λ_2, new incipient lineage at rate λ_3 or extinction at rate μ_2) are independent point process. The first event to happen is therefore a point process with rate $\lambda_2+\lambda_3+\mu_2$ and the probability that it is the formation of an incipient lineage is $\lambda_3/(\lambda_2+\lambda_3+\mu_2)$. We can thus decompose:

$$\begin{aligned}\pi &= \sum_{n=0}^{+\infty} \mathbb{P}(L \text{ has } n \text{ descendants})\mathbb{P}(L \text{ dies}) \prod_{i=1}^{n} \mathbb{P}(L_i \text{ does not complete}) \\ &= \sum_{n=0}^{+\infty} \left(\frac{\lambda_3}{\lambda_2+\lambda_3+\mu_2}\right)^n \frac{\mu_2}{\lambda_2+\lambda_3+\mu_2}\pi^n.\end{aligned}$$

Therefore

$$\pi = \frac{\mu_2}{\lambda_2+\lambda_3+\mu_2} \times \frac{1}{1-\frac{\pi\lambda_3}{\lambda_2+\lambda_3+\mu_2}} \Leftrightarrow \pi^2 - \pi\frac{\lambda_2+\lambda_3+\mu_2}{\lambda_3} + \frac{\mu_2}{\lambda_3} = 0.$$

The solution of this equation in $[0,1]$ is

$$\pi = \frac{1}{2}\frac{\lambda_2+\lambda_3+\mu_2}{\lambda_3}\left(1-\sqrt{1-4\frac{\lambda_3\mu_2}{(\lambda_2+\lambda_3+\mu_2)^2}}\right). \tag{II.15}$$

Let us now consider a good lineage. Ignoring the proportion π of incipient daughter lineages that will not complete speciation, we can consider the filtered point process with rates of successful initiation $(1-\pi)\lambda_1$ and extinction μ_1. Therefore the probability of speciation is the probability that a successful initiation occurs before the extinction, given by:

$$\mathbf{P}(\text{speciation}) = \frac{(1-\pi)\lambda_1}{\mu_1+(1-\pi)\lambda_1}. \tag{II.16}$$

The speciation time is the branching time of the first successful incipient lineage to speciate. Therefore the speciation time is distributed as the result of the filtered point process and has the expected value:

$$\mathbb{E}[T|\text{speciation}] = \frac{1}{(1-\pi)\lambda_1 + \mu_1}. \tag{II.17}$$

Expression of the equivalent rates of speciation and extinction

The equivalent rates of BD are the rates such that $\mathbf{P}(\text{speciation})$ and $\mathbb{E}[T|\text{speciation}]$ are equal in both models. To do so, we set $\tilde{\lambda}$ and $\tilde{\mu}$ such that the pairs of expressions in equation II.13/II.16 and II.14/II.17 are equal. This gives us:

$$\begin{aligned}\tilde{\lambda} &= \frac{\mathbf{P}(\text{speciation})}{\mathbb{E}[T|\text{speciation}]} \\ &= (1-\pi)\lambda_1\end{aligned} \tag{II.18}$$

and

$$\begin{aligned}\tilde{\mu} &= \frac{1-\mathbf{P}(\text{speciation})}{\mathbb{E}[T|\text{speciation}]} \\ &= \mu_1.\end{aligned}$$

II.5.6 Analysis of parameters limiting the equivalent birth rate

We measure the influence of the variation in each of the parameters of the PBD model on the equivalent birth rate $\tilde{\lambda}$. Parameters that influence $\tilde{\lambda}$ the most can be considered potentially limiting, as they can drive $\tilde{\lambda}$ down depending on their values. We calculated the partial derivatives of $\tilde{\lambda}$, given by the expression equation II.18, with respect to the PBD parameters $\lambda_1, \lambda_2, \lambda_3$ and μ_2 (not with respect to μ_1, denoting the rate of extinction of a good lineage, because it has no influence on the equivalent birth rate).

$$\frac{\partial\tilde{\lambda}}{\partial\lambda_1} = (1-\pi), \quad \frac{\partial\tilde{\lambda}}{\partial\lambda_2} = -\lambda_1\frac{\partial\pi}{\partial\lambda_2}, \quad \frac{\partial\tilde{\lambda}}{\partial\lambda_3} = -\lambda_1\frac{\partial\pi}{\partial\lambda_3},$$
$$\text{and} \quad \frac{\partial\tilde{\lambda}}{\partial\mu_2} = -\lambda_1\frac{\partial\pi}{\partial\mu_2}.$$

In what follows, we explicit the expressions (*i*, *ii* and *iii*) of the three partial derivatives of π with respect to the three parameters of the incipient lineages. From equation II.15 we remind $\pi = \Lambda\left(1-\sqrt{1-4\lambda_3\mu_2/\Lambda^2}\right)/2\lambda_3$ with $\Lambda = \lambda_2+\lambda_3+\mu_2$.

$$(i)\quad \frac{\partial\pi}{\partial\lambda_2} = \frac{1}{2\lambda_3}\left(1-\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}\right) + \frac{\Lambda}{2\lambda_3}\left(\frac{-\frac{8\lambda_3\mu_2}{\Lambda^3}}{2\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}}\right)$$
$$= \frac{\pi}{\Lambda} - \frac{2\mu_2}{\Lambda\sqrt{\Lambda^2-4\lambda_3\mu_2}},$$
$$(ii)\quad \frac{\partial\pi}{\partial\lambda_3} = \frac{1}{2}\frac{\lambda_3-\Lambda}{\lambda_3^2}\left(1-\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}\right) + \frac{\Lambda}{2\lambda_3}\times\left(-\frac{\partial u}{\partial\lambda_3}\frac{1}{2\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}}\right)$$

with u defined as $1-4\frac{\lambda_3\mu_2}{\Lambda^2}$, so $\frac{\partial u}{\partial\lambda_3} = 4\mu_2\frac{\lambda_3-\lambda_2-\mu_2}{\Lambda^3}$, hence

$$\frac{\partial\pi}{\partial\lambda_3} = -\frac{\lambda_2+\mu_2}{2\lambda_3^2}\left(1-\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}\right) + \frac{\Lambda}{4\lambda_3}\frac{4\mu_2\Lambda}{\Lambda^3\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}}$$
$$= -\frac{\lambda_2+\mu_2}{\Lambda\lambda_3}\pi + \frac{\mu_2}{\lambda_3\Lambda^2}\frac{\lambda_2+\mu_2-\lambda_3}{\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}},\quad \text{and}$$
$$(iii)\quad \frac{\partial\pi}{\partial\mu_2} = \frac{1}{2\lambda_3}\left(1-\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}\right) - \frac{\Lambda}{2\lambda_3}\frac{\partial u}{\partial\mu_2}\frac{1}{2\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}}$$

with $\frac{\partial u}{\partial\mu_2} = -4\lambda_3\frac{\lambda_2+\lambda_3-\mu_2}{\Lambda^3}$, so

$$\frac{\partial\pi}{\partial\mu_2} = \frac{\pi}{\Lambda} + \frac{\lambda_2+\lambda_3-\mu_2}{\Lambda^2\sqrt{1-4\frac{\lambda_3\mu_2}{\Lambda^2}}}.$$

We then use the simplified framework of the PBD model where all rates of initiation are equal ($b := \lambda_1 = \lambda_3$) and all rates of population extinction are equal ($e := \mu_1 = \mu_2$). By the chain rule:

$$\frac{\partial\tilde{\lambda}}{\partial b} = \frac{\partial\tilde{\lambda}}{\partial\lambda_1} + \frac{\partial\tilde{\lambda}}{\partial\lambda_3} \quad \text{and} \quad \frac{\partial\tilde{\lambda}}{\partial e} = \frac{\partial\tilde{\lambda}}{\partial\mu_1} + \frac{\partial\tilde{\lambda}}{\partial\mu_2}.$$

We then evaluate the relative influence of a parameter as the ratio between the absolute partial derivatives. For instance the relative influence of the rate of initiation b is:

$$\text{relative influence}(b) = \frac{\left|\frac{\partial\tilde{\lambda}}{\partial b}\right|}{\left|\frac{\partial\tilde{\lambda}}{\partial b}\right| + \left|\frac{\partial\tilde{\lambda}}{\partial\lambda_2}\right| + \left|\frac{\partial\tilde{\lambda}}{\partial e}\right|}.$$

We provided a detailed plot of the values of this relative influence of the three simplified parameters in figure S II.5. The summary of these relative influences is provided on figure II.4. A similar analysis was also done with the detailed model (with $\lambda_1 \neq \lambda_3$ and $\mu_1 \neq \mu 2$ a priori, figure S II.6).

ACKNOWLEDGEMENTS

We thank Thibault Juillard, Amandine Véber and Amaury Lambert for mathematical advice, and Benoît Perez-Lamarque who kindly shared his code to estimate the BD rates on truncated phylogenies.

We also thank Tanja Stadler, Matt Pennell and Rampal Etienne for their comments of a previous version of the manuscript.

CODE AVAILABILITY

The scripts used to make these analyses are available on `https://doi.org/10.5281/zenodo.14001155` and on a GitHub repository (Veron and Andréoletti 2025).

CHAPTER III

Discussion

III.1 SYNTHESIS OF THE RESULTS

In the first chapter, I investigated how mutation rates, population sizes and the genomic architecture of speciation change the dynamics of speciation between two distinct populations, potentially connected by migration. The holey adaptive landscape (HAL) as formalized by Gavrilets (1999), i.e., through a set of deterministic ODEs, has the advantage of being rapidly solvable numerically, and in the same time highly modular, to include for example local adaptation and asymmetric population sizes.

I found that the influence of the microevolutionary parameters on speciation duration depends on the scenario of speciation and is far from trivial. Population size is the most striking example: faster speciation in large populations, as predicted by Dynesius and Jansson (2014), is actually only true in the cases of ecological speciation. In the other cases (allopatric neutral or parapatric neutral speciation), speciation duration is not a decreasing function of population size. Genomic parameters of plant demographic models indicate a positive correlation between population size and duration of speciation, suggesting that the main speciation scenario is not the ecological one.

In the second chapter, I showed, using the protracted birth–death (PBD) model, that the rate of speciation completion has a minor influence on the macroevolutionary rate of speciation. By looking at the mathematical properties of the stochastic process under the PBD model, I estimated time-dependent and constant equivalent birth and death rates, that reflect the fact that speciation is not instantaneous and the several steps of speciation (initiation, survival of populations, completion of speciation) can change the probability of its success. I assessed the effect of the parameters of the PBD models on

these equivalent rates.

The rate of completion has an influence on the rate of speciation in the recent past, by modulating the speed at which it decreases through time. I estimated the relative influence of each speciation step (initiation, completion, survival) on the macroevolutionary speciation rate. Applying this method to parameters of the PBD process inferred by Hua et al. (2022) on a phylogeny of Australian skinks allowed to conclude indeed that the completion rate has a very small influence on the macroevolutionary speciation rate. Speciation initiation and, to a lesser extant, the survival of populations are the steps that modulate the speciation rate over long time-scales.

III.2 NET EFFECT OF POPULATION SIZE ON SPECIATION

In my first chapter, I showed that population size has a negative effect on the duration of speciation, at least in the data produced by Monnet et al. (2025). To do so, I relied on the output of the software "Demographic inference with linked selection" (DILS; Fraïsse et al. 2021). This software uses approximate Bayesian computation (ABC) to (*i*) compare several scenarios of demography between (here) two populations and (*ii*) numerically estimate the parameters of these scenarios, including the population sizes, the time when populations split and the time when gene flow started and ended in the past. Here, I used the output of the software where the best supported model is the one where gene flow stopped (ancient migration, AM), and I took the duration of the period over which gene flow occurred as a measure of speciation duration, with the assumption that the cessation of gene flow corresponds to the completion of speciation.

The interpretation of the parameters inferred using ABC on this kind of models is subject to debate. Although the publication introducing DILS shows that the ABC retrieves quite accurately simulated population sizes and time (e.g., of split and cessation of gene flow; Fraïsse et al. 2021), one could argue that these parameters only correspond to theoretical models. In natural systems indeed, neither population sizes nor migration rates are piecewise constant, and the populations can have an internal structure. However, the 39 summary statistics of the DILS model measured on the plant genomic data, and used in the ABC, are well explained, especially with the posterior distribution refined using neural networks (figure S I.11). This shows that, although the reality may be naturally more complex and the model may not be able to distinguish between several scenarios producing the same summary statistics, the software captures well what explains the variability of the genetic samples of the populations pairs that are considered.

Notably, using a population genetic model to compare the speciation dynamics

between several pairs of species/populations has the advantage of providing comparable results, which allows to derive general trends. Even though the speciation geographical and ecological contexts vary among the studied genera, taking the cessation of gene flow as the moment of speciation allows to have a standard definition of speciation. Naturally, this is an approximation of the reality. We could indeed imagine that some populations do not experience gene flow because of a geographic barrier between them and not because of intrinsic incompatibilities. The assumption made here allows to obtain general tendency on the link between speciation duration and population size, and should be taken with caution, in particular if used individually to obtain the "true" speciation duration for a given pair of taxa.

In this chapter, I found that larger populations tend to speciate faster. However, I only considered one step of the speciation process, and did do not test or predict the effect of population size on the other steps. In the HAL model, the isolation of populations simply occurs *ex nihilo*, independently of species characteristics such as population size, geographic distribution and dispersal capacity. The second chapter of this thesis showed that this step of "speciation initiation" has a big influence on macroevolutionary speciation rates. Harvey et al. (2019) predicted that demography influences the "metapopulation structure", which has a tight link with speciation initiation. We could think of different effects of population size: on one hand smaller populations would be associated to more patchy distributions, on the other hand larger populations would tend to have more opportunities to colonise new environments. It would be interesting to to test a potential relationship between the capacity of species to form isolated populations (as measured by Harvey et al. 2017; Singhal et al. 2022 or Burbrink et al. 2023) and population size estimates to enlighten this question.

An expected positive effect of population size on population survival seems more obvious, and I showed in the second chapter that population extinction is one of the strongest predictors of speciation rate. A future direction of research would be to integrate a risk of extinction — for instance using stochastic population models as reviewed by Ovaskainen and Meerson (2010) — into models of speciation such as the HAL model. Small populations may indeed speciate faster, but since they are more prone to extinction, the overall association between population size and speciation rate remains an open question.

III.3 Assessing the importance of ecological speciation

Ecological speciation denotes the scenario where RI evolves between populations by "divergent natural selection arising from differences between ecological environments"

(Schluter 2009). The question of whether speciation occurs mostly by this scenario is debated (Rundle and Nosil 2005; Nosil and Flaxman 2010; Anderson and Weir 2022). The existence of sister species living in different environments is indicative of ecological speciation, but it is not an evidence, as adaptation can take place while, or after, speciation happened, but not be its primary driver.

In my first chapter, I used the model by Gavrilets (1999) to predict the dynamics of speciation under the scenario of ecological speciation and under a neutral scenario. In this model, RI emerges because genetically distant individuals do not interbreed. The difference between the neutral scenario and the ecological scenario is the way alleles fix on the locus responsible for the incompatibility. In the neutral model, alleles are under no selective pressures — except the pressures induced by the incompatibilities themselves. In the scenario of ecological speciation, the alleles are under positive selection, a process called "local adaptation". Although a premise for ecological speciation is a divergent selection between two environments, a positive selection is equivalent in the HAL model because of the infinite allele assumption and the absence of migration in this scenario. Indeed, mutations that appear in the first population will not be tested by natural selection in the second population.

In the HAL model, there is no assumption on the mechanisms that would cause the incompatibilities. The only assumption is that genetically distant individuals do not mate or do not have fertile offspring. This could be induced by pre-mating RI (mate choice), postzygotic RI (a mechanism like DMI) or anything in-between (or a combination of factors, see Christie and Strauss 2018). In the case with local adaptation, we still assume that the differences explain the incompatibilities, but their accumulation is favoured by a positive selection within each population. This process is close to the definition of a "magic trait", i.e., a phenotypic trait that is under divergent selection and associated to a non-random mating (Servedio et al. 2011). This is the case, for instance, when the adaptation of a pathogen to a host controls the partners with whom it can reproduce (Giraud et al. 2010). The HAL model with local adaptation actually covers a broader scenario of ecological speciation, because the incompatibilities between individuals include multiple loci. This model may be closer to the reality of *Heliconius* butterflies for example, where wing color pattern probably acts as a magic trait, but the assortative mating is actually based on a multitude of traits, not only the wing color pattern (Mérot et al. 2015; Mérot et al. 2017).

The selective regime of alleles responsible for genetic incompatibilities has a consequence, notably on the effect of population size on speciation duration. In their work on the expected controls of speciation rate, Dynesius and Jansson (2014) predicted that larger population should speciate faster, but they implicitly assumed that speciation occurs after mutations under positive selection fix in the populations, which is not always the case. Selection tends to be less efficient in small populations, due to genetic drift

(Hartl and Clark 2007), regardless of the direction of that selection (positive or negative). As a consequence, ecological speciation is faster in larger populations compared to smaller populations ; and neutral speciation is faster in smaller populations compared to larger populations.

I used this dependence of speciation duration on population size to test the dominance of ecological speciation, using plant demographic data, inferred from plant genomes by Monnet et al. (2025), and concluded that ecological speciation may not be the dominant scenario in the tested plants. This result is in line with the analysis by Anderson and Weir (2022), where signals of divergent selection between sister species of vertebrates was found weak. However, there are alternative explanations for the observed patterns in the association between population size and the duration of speciation. First, what I observed as "duration of speciation" may actually reflect an alternative scenario, with gene flow being suppressed because of external factors instead of intrinsic incompatibilities, as discussed in the previous section. Second, a confounding factor cannot be excluded. For instance, one could imagine that smaller populations are associated with the colonisation of more distant areas, which themselves are associated with less gene flow because of geographic isolation. This last intuition could be tested by looking at a correlation between population size and migration rate, however the inference of migration rate under demographic models remains difficult.

III.4 SPECIATION LIMITED BY POPULATION SEPARATION

In the second chapter I showed that speciation rates are mostly influenced by speciation initiation in the framework of the PBD model. But what is speciation initiation? In their founder publication on the protracted speciation model, Rosindell et al. (2010) defined speciation initiation as the formation of an "incipient species", that is yet not considered as a distinct species from its sister lineage. This incipient species may become distinct in the future, which requires the acquisition of a form of independent evolution from its sister lineage.

A natural interpretation of the initiation phase is the separation of one population into two distinct populations. A population is a group of individuals that have a certain degree of disjunction from the rest of the species, because they live in a same area and tend to reproduce within this group (Wells and Richmond 1995). The formation of population is considered as the starting step of speciation, as considered by Dynesius and Jansson (2014) under the term of "splitting".

The empirical relationship between the rate at which populations split and the macroevolutionary speciation rates is not clear. Harvey et al. (2017) found a positive association

between the rate of population formation and the speciation rate. The rate of population formation is estimated by fitting a coalescent model to a phylogeny of populations. With this method, the obtained value corresponds to the number of newly formed population per lineage and per million year. But, in another study on lizards, Singhal et al. (2018) proposed another form of population differentiation rate based on geographic isolation, called isolation by distance (IBD). IBD relies on the relationship between geographical isolation between individuals and their genetic divergence. The slope of that relationship, called rate of IBD, should characterize the propensity of a species to form new populations — and is indeed correlated with mobility-related traits — but is uncorrelated to the macroevolutionary speciation rate (see also Singhal et al. 2022; Burbrink et al. 2023).

A possible explanation for the absence of correlation between population differentiation and speciation rate is the fact that these studies use the tip-speciation rate as a variable. In models like ClaDS, speciation rates vary across time and lineages, either by a general trend, or due to a random variation at each branching event. The tip-speciation rate is a measure of this variable at the leaves of the phylogenetic tree, i.e., at the present for each extant species. In my second chapter, I show that the "equivalent" macroevolutionary speciation rate under a PBD process converges to zero at present, because the probability that a lineage initiates and completes speciation in a small amount of time is negligible. The effect of speciation initiation is therefore less noticeable in the equivalent speciation rates measured close to the present. Tip-speciation rate estimates do probably not reflect an existing dependency between the rates of population differentiation and speciation.

We note, however, that the correlation between population differentiation and speciation rate measured by Harvey et al. (2017) is significant and positive. The contrast with the non-significant results of Singhal et al. (2018), Singhal et al. (2022), and Burbrink et al. (2023) could come from several differences between the studies: (*i*) the use of BAMM, in which the speciation rates vary by shifts instead of changing at each branching event like ClaDS, (*ii*) the geographic scale, Harvey et al. studied a population across a large continent and (*iii*) the difference in the measure of population differentiation. Harvey et al. indeed used a coalescent method to actually estimate the number of populations, whereas in the IBD method, the measure reflects the capacity of isolation of the species. I suggest that, although speciation rate is limited by the initiation phase, the formation of populations is a large-scale process that is not well reflected by the rate of IBD. It should be interesting to assess the degree of gene flow between the patches where the individuals have been sampled in those studies, to see if they correspond to actual populations. As predicted by Dynesius and Jansson (2014), the persistence of the populations through time also needs to be taken into account if we want to predict the speciation rate.

In summary, although the speciation initiation phase has a large effect on the speciation rate, the speciation initiation rate is not easy to estimate on natural populations

in a way it corresponds to its definition in the PBD model. The various measures of population differentiation may not correspond to a proxy for this rate. A solution is to directly fit a protracted speciation model, for instance using the ProSSE framework, but this requires an extensive sampling of the populations of a group of species (Hua et al. 2022).

III.5 THE GREY ZONE OF SPECIATION

Roux et al. (2016) popularized the term "grey zone of speciation", a concept that is also encountered in the form of "speciation continuum" (de Queiroz 2007) to denote a region of genetic divergence where pairs of taxa are semi-isolated, meaning that the reproductive compatibility between them is reduced, but not completely absent. Semi-isolated taxa are characterized by regions of the genome with gene flow coexisting with regions without gene flow. The term has then been generalized to situations where finding species boundaries is challenging (Fernández-Álvarez et al. 2023; Chan et al. 2026).

Using theoretical models to predict how population compatibility decays with genetic divergence, as I did in the first chapter, allows to better understand the factors shaping this grey zone. I indeed showed that, under the HAL model, the speciation scenario (allopatric without adaptation, allopatric with adaptation or parapatric) has little influence on the evolution of between-populations compatibility, as measured by Roux et al. (2016). I also showed that the decay in compatibility starts at lower levels of genetic divergence, and is slower, for larger mutation rates or population sizes. Naturally, the parameter representing the genetic architecture of incompatibilities influences the shape of the grey zone too.

Although Roux et al. (2016) found that the grey zone was fairly similar across animals, Monnet et al. (2025) found a difference between animals and plants. Indeed, they found that the genetic divergence where the grey zone occurred is smaller in plants compared to animals. They put this result in parallel with the higher speciation rates observed in plants compared to vertebrates and invertebrates (see figure 5). Our results provide some insight on this interpretation. The shape of the grey zone reflects when and how RI accumulates, which is tightly linked to the "completion rate" that is used in the PBD framework and in the second chapter of this thesis. In this work, I showed, based on the PBD rates estimated by Hua et al. (2022) on a group of lizards, that the completion rate is not expected to have a big influence on speciation rate. That is also supported by a former correlative study on birds and *Drosophila*, in which Rabosky and Matute (2013) found no link between macroevolutionary speciation rate and the speed of accumulation of RI. If we now consider plants and assume, as suggested by Monnet et al. (2025), that they have a larger completion rate (because RI accumulates faster), the influence of completion rate on macroevolutionary speciation is expected to be even

smaller (see figure II.4). Hence, the difference in accumulation of RI observed between plants and animal is likely not sufficient to explain their difference in macroevolutionary speciation rates.

The weak influence of the completion rate on macroevolutionary speciation does not however make the study of the grey zone uninteresting. The question of how fast RI accumulates is a fundamental question itself, and it can still help to explain the micro–macro gap. We highlight in the previous section that the rate of formation of incipient species is a crucial parameter to explain a variability in speciation rate, but understanding how populations become species is necessary to clarify what we actually call an incipient species on the field. If plants indeed accumulate RI faster, due to phenomena like polyploidization or reinforcement that can occur in a few generations for instance (Long et al. 2026), less isolation between populations is required to initiate speciation. On the contrary, for animals, we would expect incipient species to be more isolated populations.

III.6 Concluding remarks

The link between micro- and macroevolution in speciation has long been studied and debated. In the reference book on speciation by Coyne and Orr (2004), a full chapter is dedicated to this topic. One possible reason for this interest is that the micro–macro link defies our initial intuition: we think we understand how species form, but we do not understand why some groups speciate very frequently and others much more rarely. We are tempted to look for correlations between different factors, but we generally find surprising results. Debates have taken place — and continue to take place — between proponents of different modes of speciation. Above all, establishing the micro–macro link requires bridging the gap between different scientific fields, each with their own methods, perspectives and vocabulary.

I believe that mathematical models can provide great contribution to address important questions behind the micro–macro gap in speciation. They help us refine and test our intuitions. I tried to contribute to this debate by using existing, and I hope, not too complex mathematical models, and comparing their output to real data. This allowed me to advance our understanding of how speciation occurs and what should influence macroevolutionary speciation rates.

This work confirms that focusing on the accumulation of reproductive barriers only is not sufficient. I hope it will encourage future research to account for the complexity of the geographical structure of populations, as well as their potential extinction. The protracted nature of speciation should also been taken into account in a more systematic way. Although I do not think a general model explaining speciation rates across the tree

of life exists, I am convinced that simple models including key processes can improve our understanding of patterns of speciation observed empirically.

APPENDIX

Linking diversification rates to the speciation process

This part is the section 8 of the review by Morlon et al., to which I contributed as a part of my PhD project. Some elements written here can be redundant with the introduction.
The full review is published in: Hélène Morlon, Jérémy Andréoletti, Joëlle Barido-Sottani, Sophia Lambert, Benoît Perez-Lamarque, Ignacio Quintero, Viktor Senderov and Pierre Veron (2024). Phylogenetic insights into diversification. *Annual Review of Ecology, Evolution, and Systematics* 55, 1–21.

Most phylogenetic diversification approaches consider speciation as an instantaneous event and estimate a macroevolutionary speciation rate, which quantifies the average number of successful speciation events per species per time unit. These approaches have revealed that speciation rates can vary by several orders of magnitude even within taxonomically restricted species groups, and they have been heavily used to identify intrinsic or extrinsic factors that correlate with speciation rates, as discussed above. A more mechanistic understanding of how these factors influence species formation to modulate speciation rates requires a full consideration of the speciation process (figure 6A; Rabosky 2016; Hua and Bromham 2017; Harvey et al. 2019; Rolland et al. 2023). For macroorganisms with sexual reproduction, the most common process of species formation begins when population subsets become isolated (speciation initiation or population formation); their survival in isolation must then be maintained over a period significant enough that the accumulation of differences prevents interbreeding between individuals

from different populations or makes their hybrids less fit (reproductive isolation, RI, or speciation completion) (figure 6A; Coyne and Orr 2004). The macroevolutionary speciation rate thus intuitively depends on the frequencies at which populations form, come back into contact again as one species, and go extinct, as well as the time it takes to achieve RI.

In the last decade, several studies have investigated the empirical relationship between macroevolutionary speciation rates, as estimated at the tips of phylogenies, and characteristics of species expected to reflect aspects of the speciation process. The initial and final stages of the speciation process, namely population formation and RI, have received the most attention. The rate of population formation has been measured using two proxies: the number of genetically distinguishable populations within species divided by species age (Harvey et al. 2017) and the slope of the relationship between geographic distance and the genetic divergence between populations (figure 6B; Singhal et al. 2018; Singhal et al. 2022; Burbrink et al. 2023). While a significant positive correlation between this rate of population formation and speciation rate was found in birds (Harvey et al. 2017), it does not seem to hold in squamates (Singhal et al. 2018; Singhal et al. 2022; Burbrink et al. 2023). The rate at which populations evolve RI has been measured as the slope of the relationship between reproductive barriers — estimated as the percentage of inviable or sterile offspring in experimental crosses — and the genetic distance between the crossed individuals (figure 6B; Rabosky and Matute 2013). In birds and *Drosophila*, some lineages evolve RI much faster than others, yet this does not seem to result in faster speciation rates (Rabosky and Matute 2013). These somewhat unexpected, and sometimes contradictory results, suggest that the persistence of populations and their spatial isolation over time, which are not captured by proxies of population formation and the rate at which populations evolve RI, may play a larger role than the two latter processes in modulating speciation rates (Rabosky 2016; Harvey et al. 2019).

A better understanding of these results, and ultimately of the processes that modulate macroevolutionary speciation rates as estimated from phylogenies, requires first a better theoretical and empirical understanding of how macroevolutionary speciation rates are related to the rates of speciation initiation; population extinction; and speciation duration, i.e., the time it takes to evolve RI. Theoretically, this amounts to establishing a mathematical link between the parameters of the instantaneous birth–death process used to estimate macroevolutionary speciation rates and those of the protracted birth–death process, which includes rates of speciation initiation and population extinction, as well as speciation duration (Etienne and Rosindell 2012). To our knowledge, such a link has yet to be formally established. Empirically, one challenge arises from the fact that it is not possible to estimate all the parameters of the protracted birth–death process from a species-level molecular phylogeny of extant species (Etienne et al. 2014). Nevertheless, speciation duration can be estimated from extant phylogenies (Etienne et al. 2014) or

using fossil data (Benton and Pearson 2001), and it could be insightful, for example, to investigate whether this duration is inversely correlated to the macroevolutionary speciation rate and/or to the rate at which species evolve RI, as could be expected. In addition, a recent study suggests that the parameters of the protracted birth–death process can be accurately estimated from a population-level (rather than species-level) phylogeny (Hua et al. 2022). While such phylogenies remain rare, this paves the way for empirical tests of the relationship between macroevolutionary speciation rates and rates of speciation initiation, population extinction, and speciation duration.

Second, we need to better understand how each of the main stages of the speciation process is influenced by key characteristics of a species such as its population size, dispersal capacity, mutation rate, and the selection pressures it experiences in its biotic and abiotic environment. The goal here is not to improve our estimates of speciation rates but rather to better understand how species attributes modulate these rates. Particularly relevant factors are interactions with other species, such as competition, parasitism, or mutualism, or with abiotic factors, such as climatic changes, orogenesis, or habitat fragmentation. A wealth of theoretical and empirical studies from population dynamics, population genetics, and speciation theory exist to build upon (Rabosky 2016; Harvey et al. 2019). For example, how biotic and abiotic factors influence the extinction risks of populations or species has been studied at length (Wiens and Slaton 2012; Chichorro et al. 2019), and numerous studies have investigated the mechanisms by which genetic incompatibilities lead to speciation (for a review, see Satokangas et al. 2020), the effect of processes such as migration and niche partitioning on the accumulation of RI (Westram et al. 2022), and the dynamics of accumulation of RI in nature (for a review, see Kulmuni et al. 2020). However, these studies have often focused on only one stage of the speciation process, while it is becoming clear that all stages need to be considered simultaneously if we are to understand how microevolutionary processes modulate macroevolutionary speciation rates. This is key, because some characteristics of species can accelerate one step of the speciation process while decelerating another, which complexifies expectations. For example, species that disperse readily can colonize new environments, which may increase rates of speciation initiation (see for instance the "great speciator" Zosteropidae; Moyle et al. 2009), but at the same time lead to more gene flow between populations, which may increase the time it takes to complete speciation (Claramunt et al. 2011). Similarly, species with large population sizes may be less likely to fix nearly neutral mutations that end up causing genetic incompatibilities between individuals from diverging populations, which may increase speciation duration (Gavrilets 2000; Maya-Lastra and Eaton 2021) but at the same time decrease the probability that populations go extinct before completing speciation. Quantitative predictions of the direction and magnitude of the effect of such species characteristics on speciation initiation, population survival, and RI, backed up with empirical evidence, can then be used to predict their effect on the macroevolutionary speciation rate, provided the relationship between the latter

and the former has been clarified (c.f. previous paragraph). If these macroevolutionary predictions are confirmed by phylogenetic diversification models, we will have made significant progress in elucidating the main factors that modulate speciation rates.

Supplementary materials for Chapter I

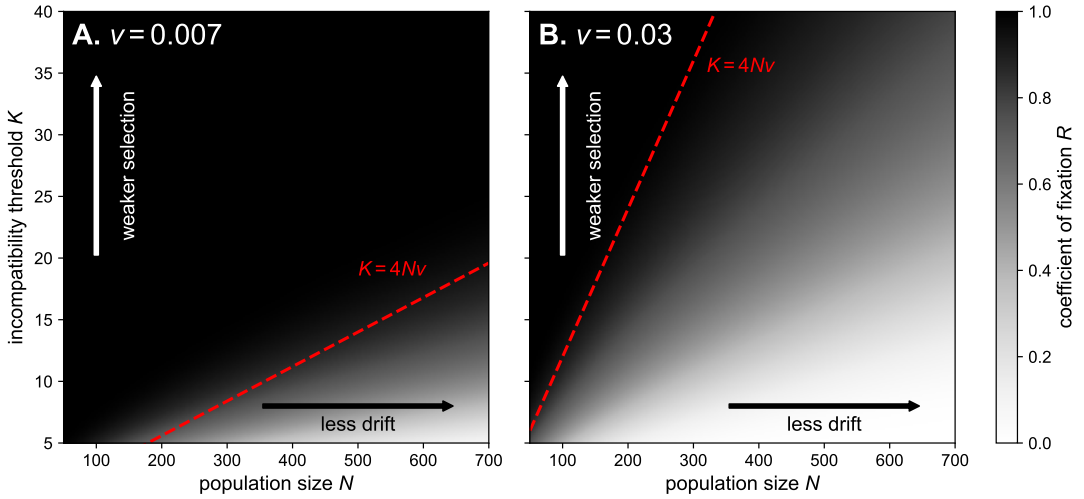


Figure S I.1 – Values of the coefficient of fixation R (representing the relative speed of fixation of mutations compared to a case without purifying selection against incompatibilities, see equation I.3) as a function of the population size N and the incompatibility threshold K, for two values of mutation 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 ($R \approx 1$).

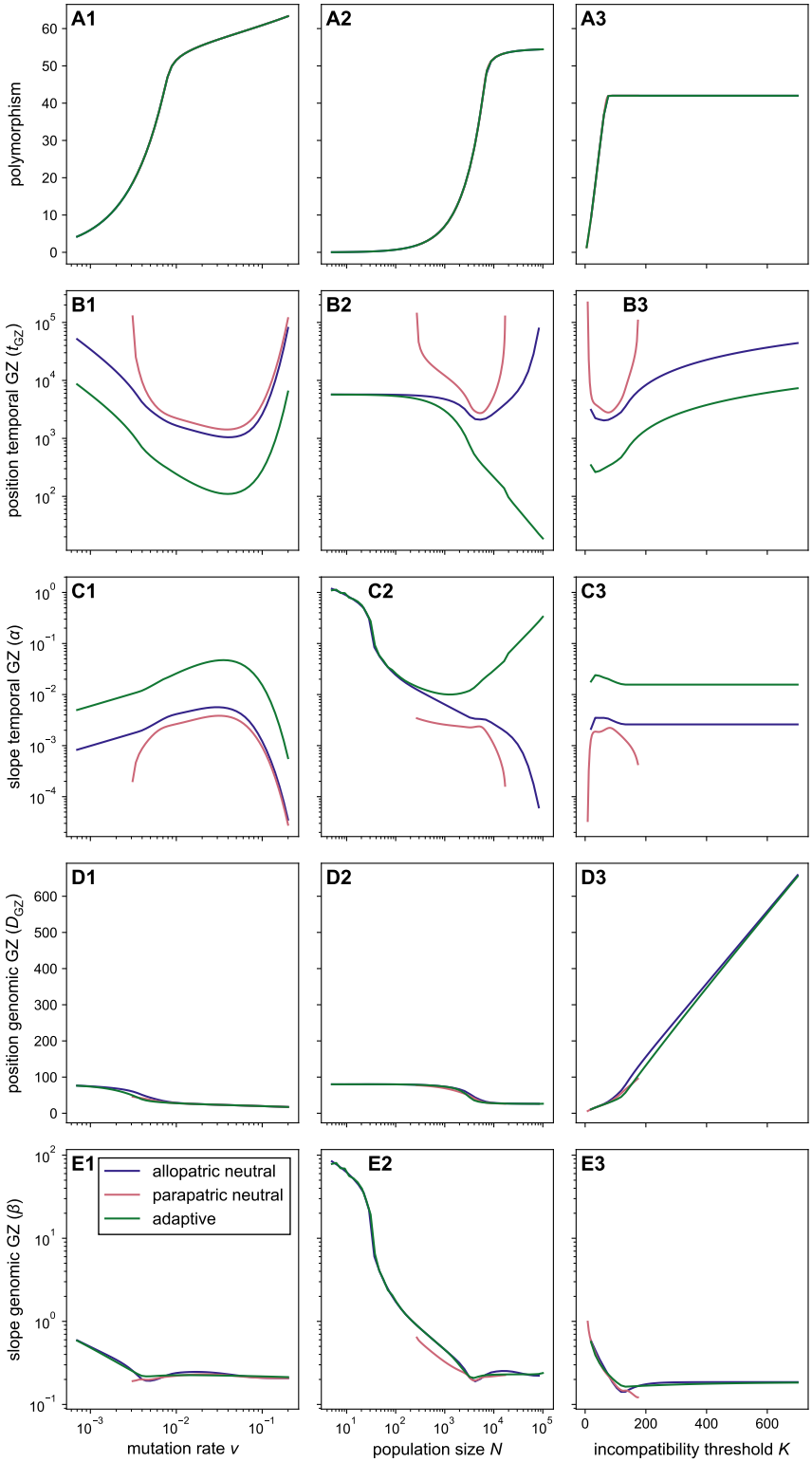

A1
A2
A3
B1
B2
B3
C1
C2
C3
D1
D2
D3
E1
E2
E3
polymorphism
position temporal GZ (t_{GZ})
slope temporal GZ (α)
position genomic GZ (D_{GZ})
slope genomic GZ (β)
allopatric neutral
parapatric neutral
adaptive
mutation rate v
population size N
incompatibility threshold K

Figure S I.2 – 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.** We show the mean polymorphism (**A**), the position (**B**) and slope (**C**) of the temporal GZ, the position (**D**) and slope (**E**) of the genomic GZ. Each line corresponds to a scenario of speciation.

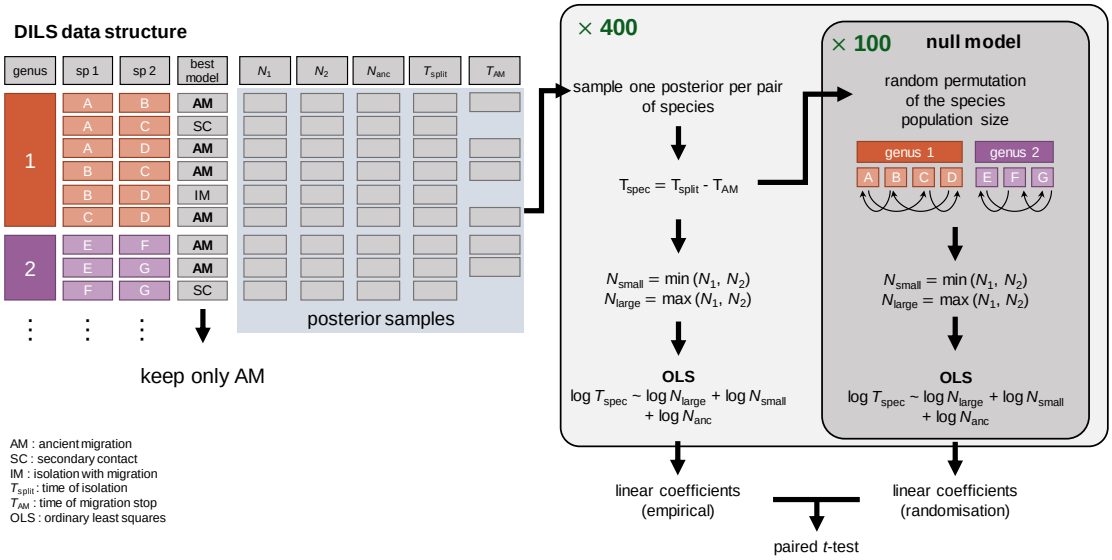


Figure S I.3 – Illustration of the paired structure of the data obtained from DILS, and of the statistical approach used to test the significance of the regression between speciation duration and population sizes. We permuted species within each plant genus to construct a null model.

Table S I.1 – 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

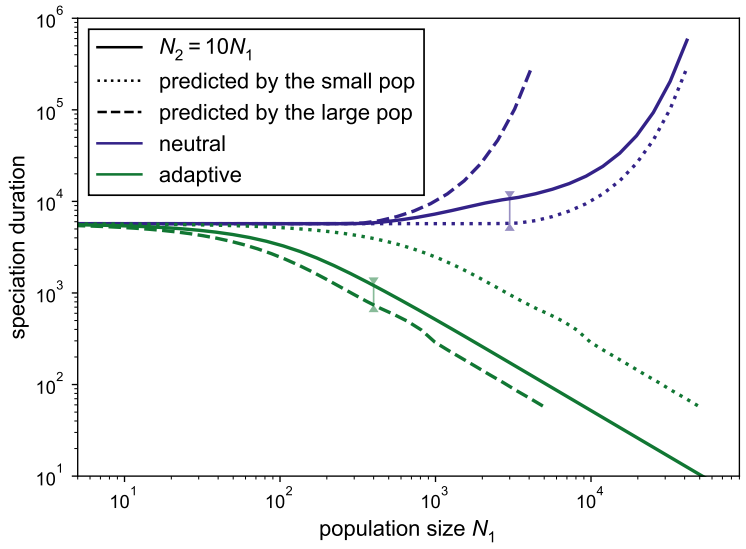


Figure S I.4 – Duration of speciation under the holey adaptive landscape (HAL) allopatric models, in the case of asymmetric populations (solid lines), as a function of N_1 with $N_2 = 10N_1$, compared to the prediction with two populations with size N_1 (dotted line) and two populations with size N_2 (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 S I.1 – 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 I.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 I.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 S I.5). 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 S I.6 cases 4 and 5); in the cases with two equilibria (figure S I.6 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 S I.5 (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 I.2).

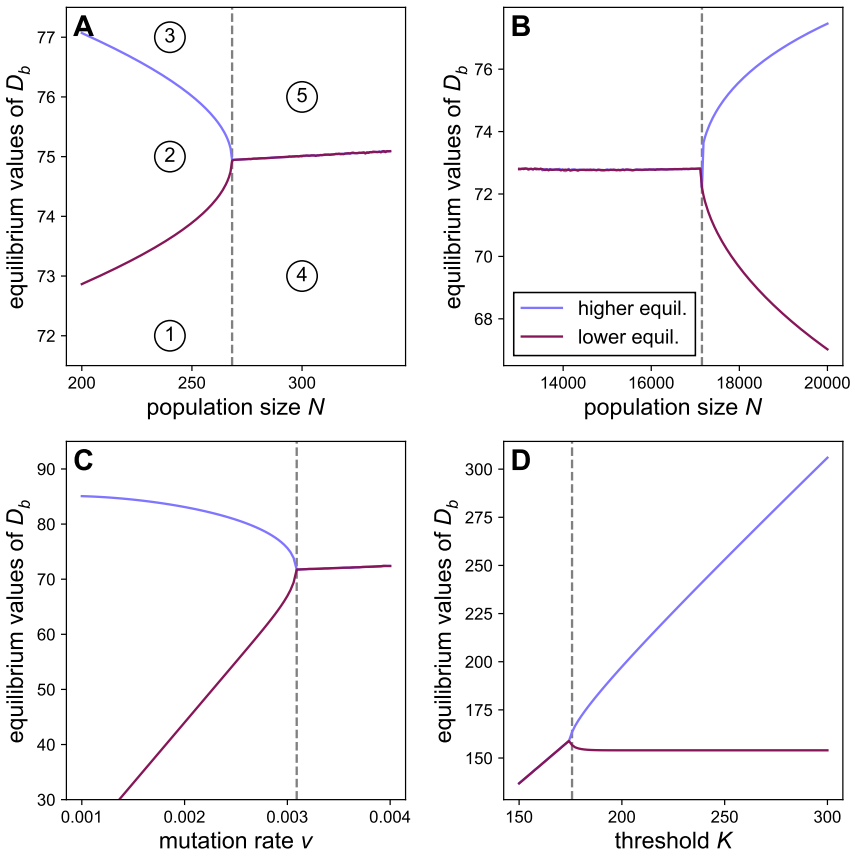


Figure S I.5 – Bifurcation diagrams showing the equilibrium value(s) for D_b as a function of the population size (**A**: low values, **B**: high values), the mutation rate ν (**C**) and the threshold of incompatibility K (**D**). When one parameter varies, the others are kept constant: $\nu = 0.0007$, $N = 6000$, $m = 10^{-4}$ and $K = 80$. The numbers on panel **A** indicate initial conditions for which the dynamics of D_b are represented on figure S I.6.

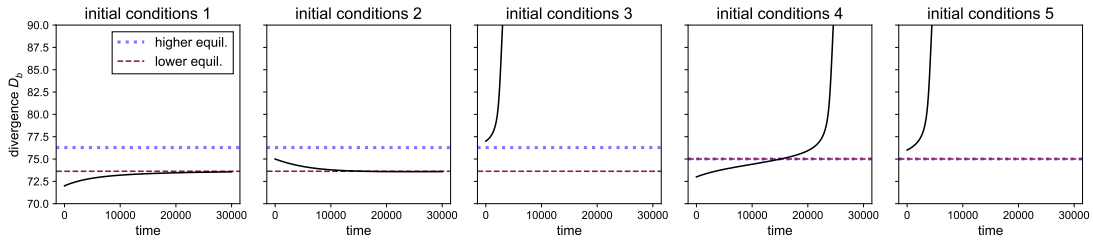


Figure S I.6 – Dynamics of D_b when its initial value is close to the equilibrium and smaller than (**1** and **4**), higher than (**3** and **4**), or between (**2**) the equilibrium values, corresponding to 5 different regions of the bifurcation diagrams, indicated on figure S I.5A. The dotted lines indicate the equilibrium value(s). In each case the initial values of D_w and k are the equilibrium values.

Supplementary text S I.2 – 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 S I.7. figure S I.8 and figure S I.9 report the results of the comparison with "intermediate" ($N = 600$) and small ($N = 200$) population sizes, which are then summarized in figure S I.10. With larger population sizes, we expect predictions to be closer to simulations as the deterministic approximation is more valid.

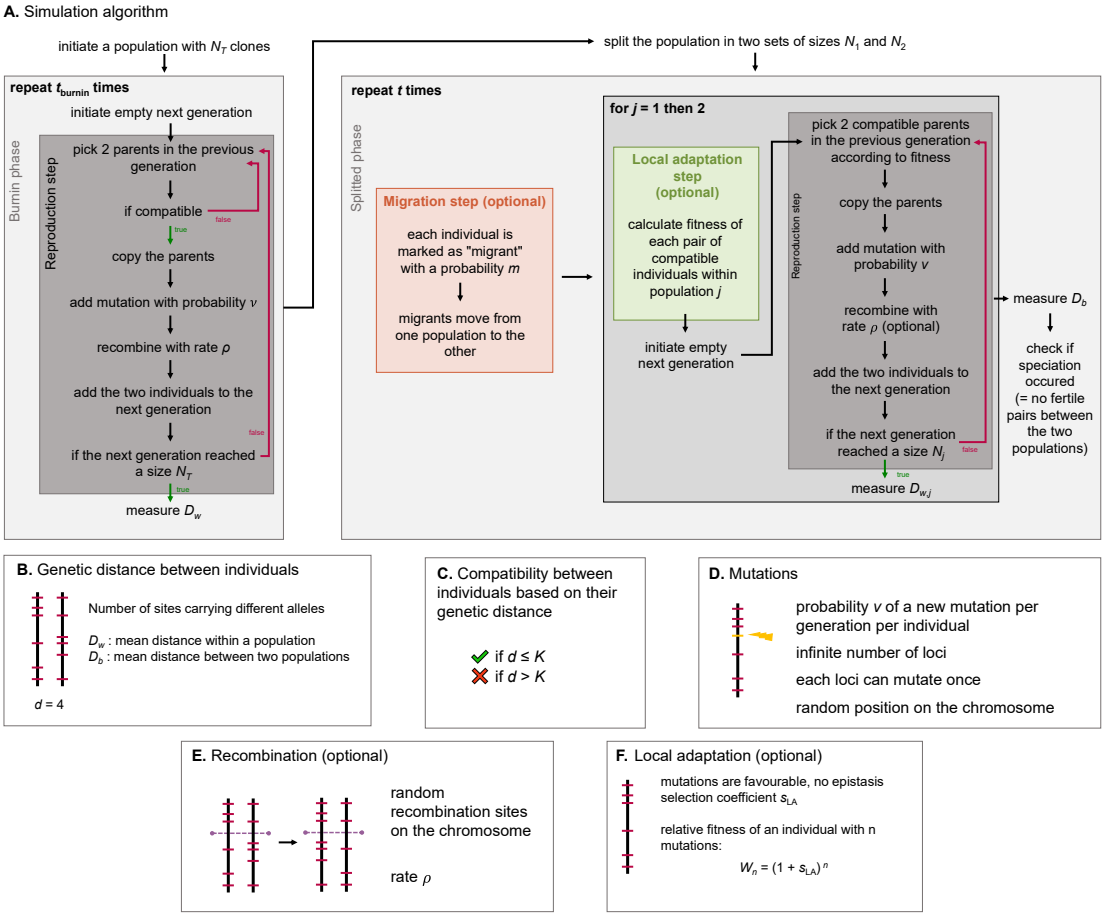


Figure S I.7 – Illustration of the simulation algorithm (**A**) and the several components of the process (**B** to **F**). **A.** A burnin phase is used to allow the ancestral population to reach an equilibrium polymorphism. After this phase, the populations are splitted randomly . **B.** The distance 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.

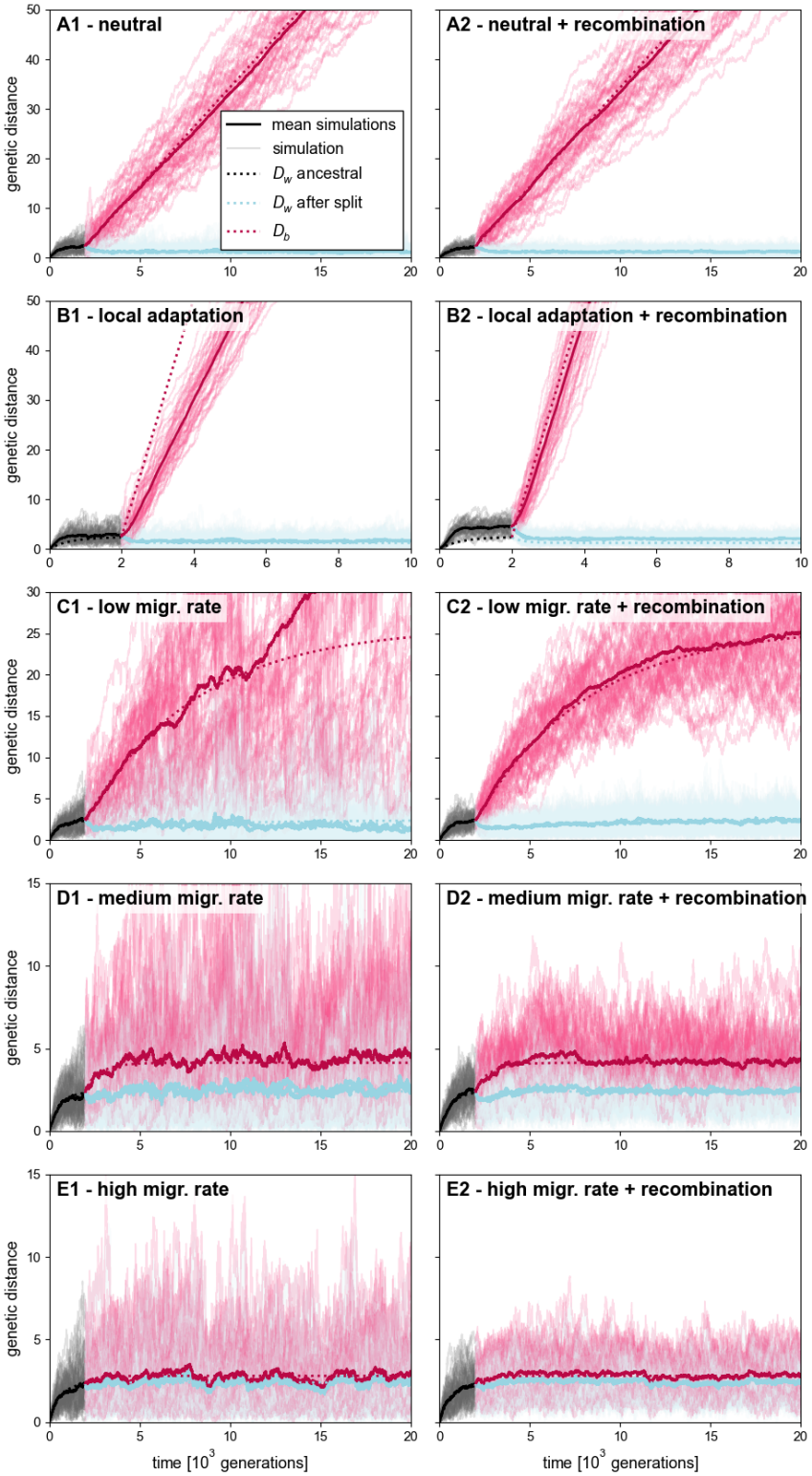

A1 - neutral
A2 - neutral + recombination
B1 - local adaptation
B2 - local adaptation + recombination
C1 - low migr. rate
C2 - low migr. rate + recombination
D1 - medium migr. rate
D2 - medium migr. rate + recombination
E1 - high migr. rate
E2 - high migr. rate + recombination
mean simulations
simulation
D_w ancestral
D_w after split
D_b
genetic distance
time [10^3 generations]

Figure S I.8 – Dynamics of polymorphism and divergence under the HAL 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.

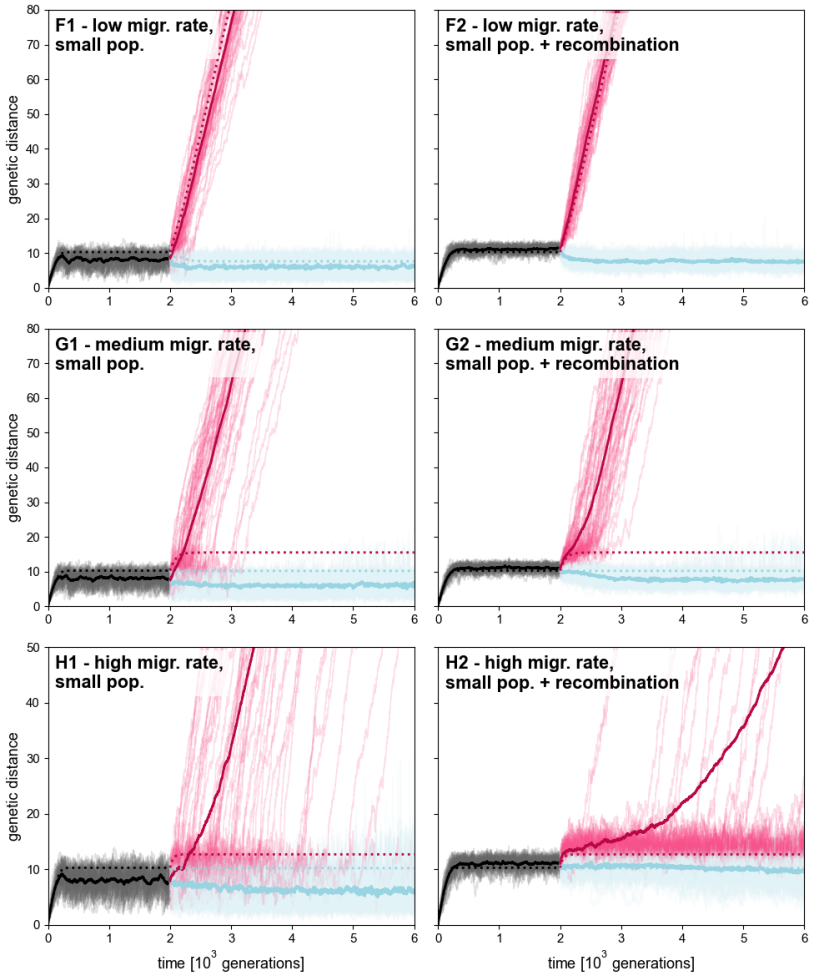


Figure S I.9 – Same as figure S I.8 but with smaller populations : $\nu = 0.0384$, $N_1 = N_2 = 100$, $K = 20$ and migrations rates 0.001 (**F**), 0.005 (**G**) and 0.01 (**H**).

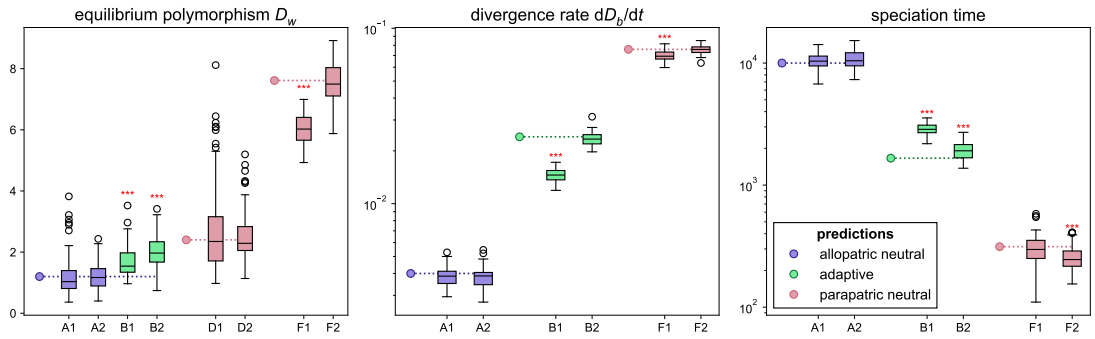
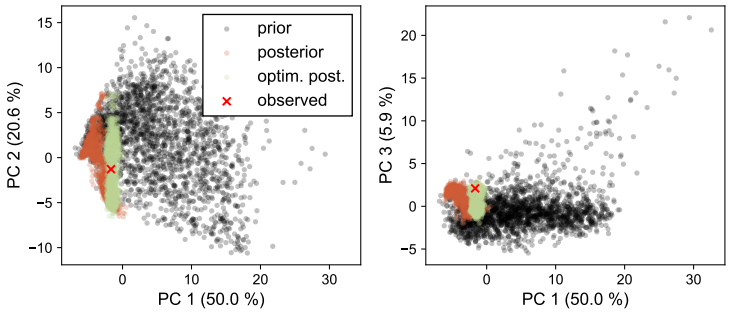
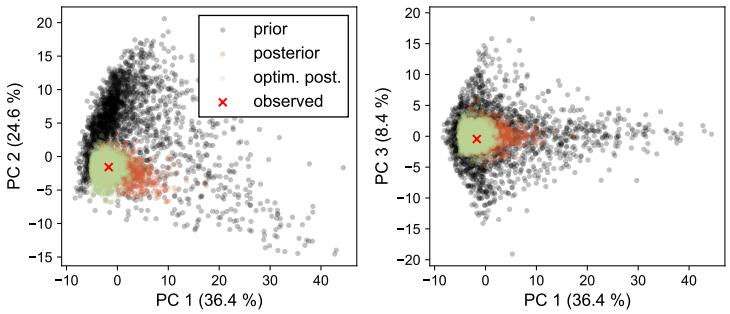


Figure S I.10 – Comparison of the dynamics of the simulations and the predictions, based on the equilibrium polymorphism D_w (left), the slope of $D_b(t)$ (middle) and the duration of speciation (right). The boxplots indicate the distribution of the estimates from 50 simulations for each configuration, 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 S I.8 and S I.9. The red asterisks indicate significant differences, assessed with t-test and corrected for multiple testing.

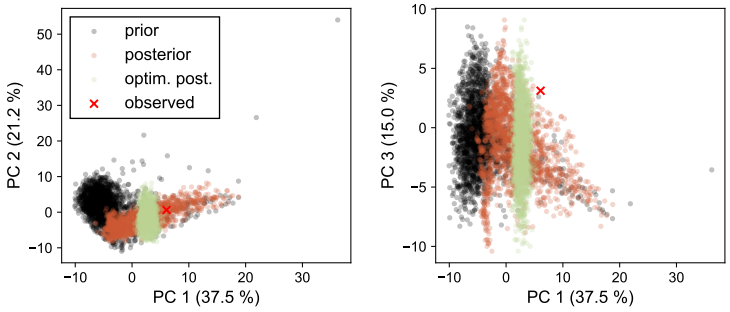
Actinidia arguta — A. chinensis (AM)



Ficus itoana — F. microdictya (AM)



Lupinus arboreus — L. misticola (AM)



Nepenthes ephippiata — N. hemsleyana (AM)

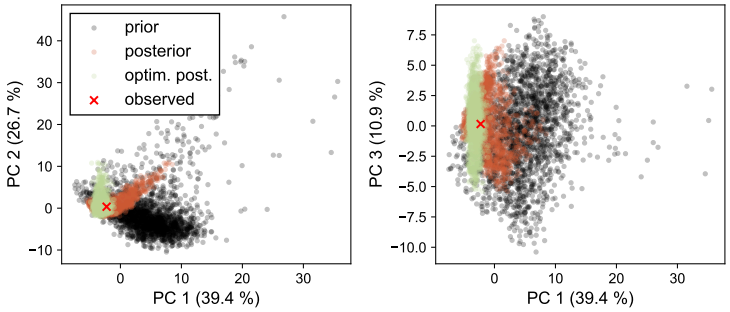


Figure S I.11 – Goodness-of-fit for the DILS inference. Principal components analysis of the summary statistics used for the inference of the demographic parameters by Monnet et al. (2025) for 4 pairs of plants. We show the distribution of the summary statistics obtained with (*i*) the prior parameters, (*ii*) the posterior parameters, (*iii*) the refined "optimized" parameters as well as the observed summary statistics in the data. We randomly chose 4 pairs belonging to different genera for which the supported model is ancient migration (AM).

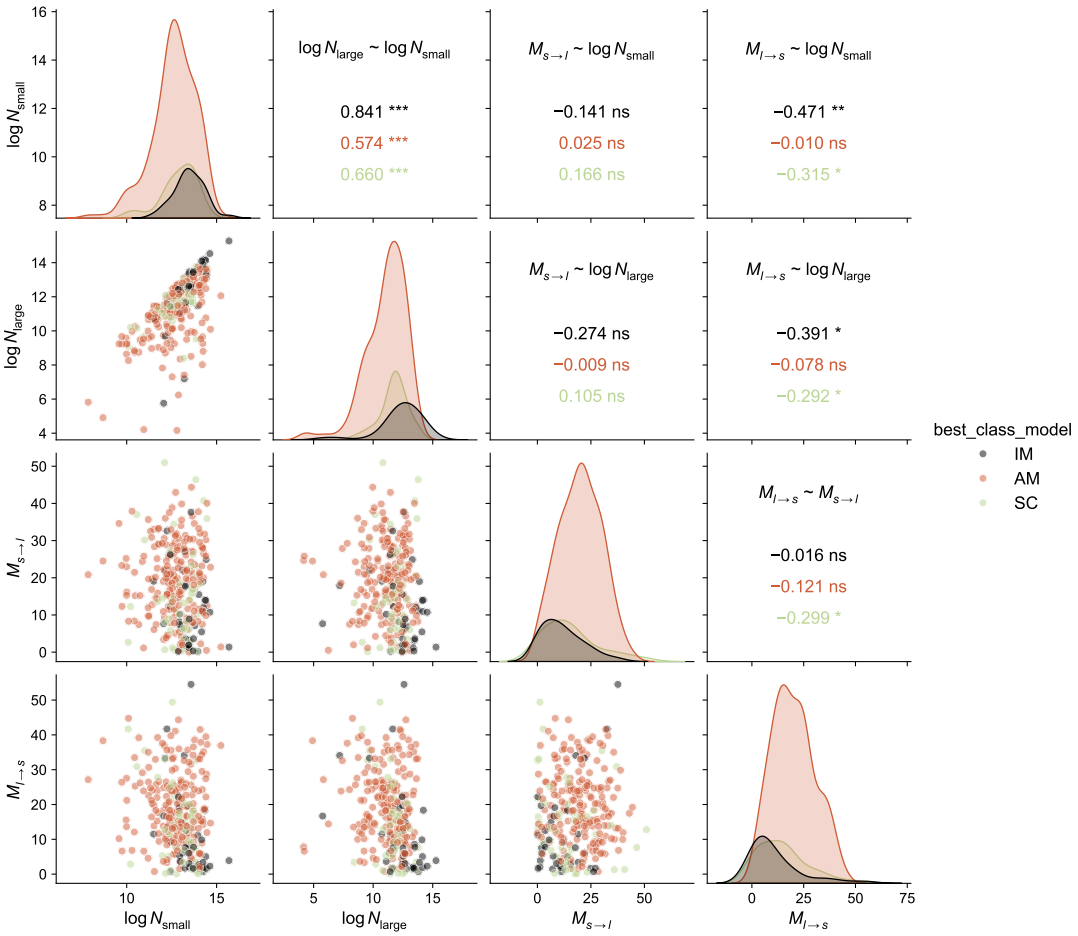


Figure S I.12 – Correlation between average log-population size and migration rates inferred by DILS in the data by Monnet et al. (2025). $M_{s\to l}$ correspond to the average number of migrants going from the small population to the large population at each generation. The upper corner indicates Spearman's correlation coefficient.

Supplementary materials for Chapter II

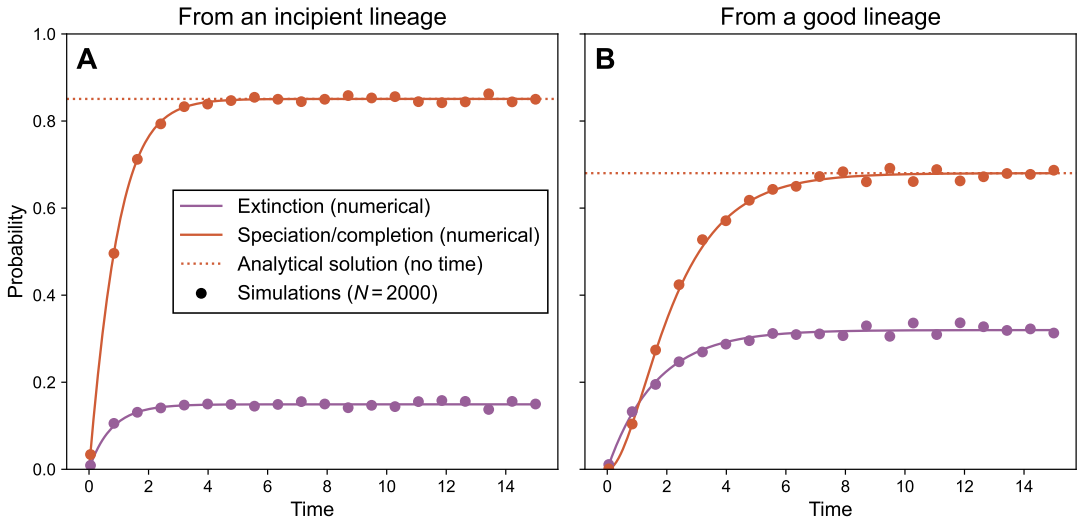


Figure S II.1 – Completion or speciation and extinction probabilities within a time horizon t calculated numerically (solid lines) **or empirically estimated on simulations** (dots). **A.** Probabilities starting with an incipient lineage (p_E^I, p_C^I). **B.** Probabilities starting with a good lineage (p_E^G, p_S^G). Each dot corresponds to $N = 2000$ replicates (each one starts with one lineage and runs for a finite time t) and shows the frequency of replicates ending by a speciation or extinction event within a time horizon t. The dotted line corresponds to the analytical expression of the speciation probability (equation II.15 for an incipient lineage and equation II.16 for a good lineage). Parameters used: $\lambda_1 = 0.5, \lambda_2 = 0.8, \lambda_3 = 0.4, \mu_1 = \mu_2 = 0.2$, same notations as in Etienne and Rosindell (2012) and in the chapter II.

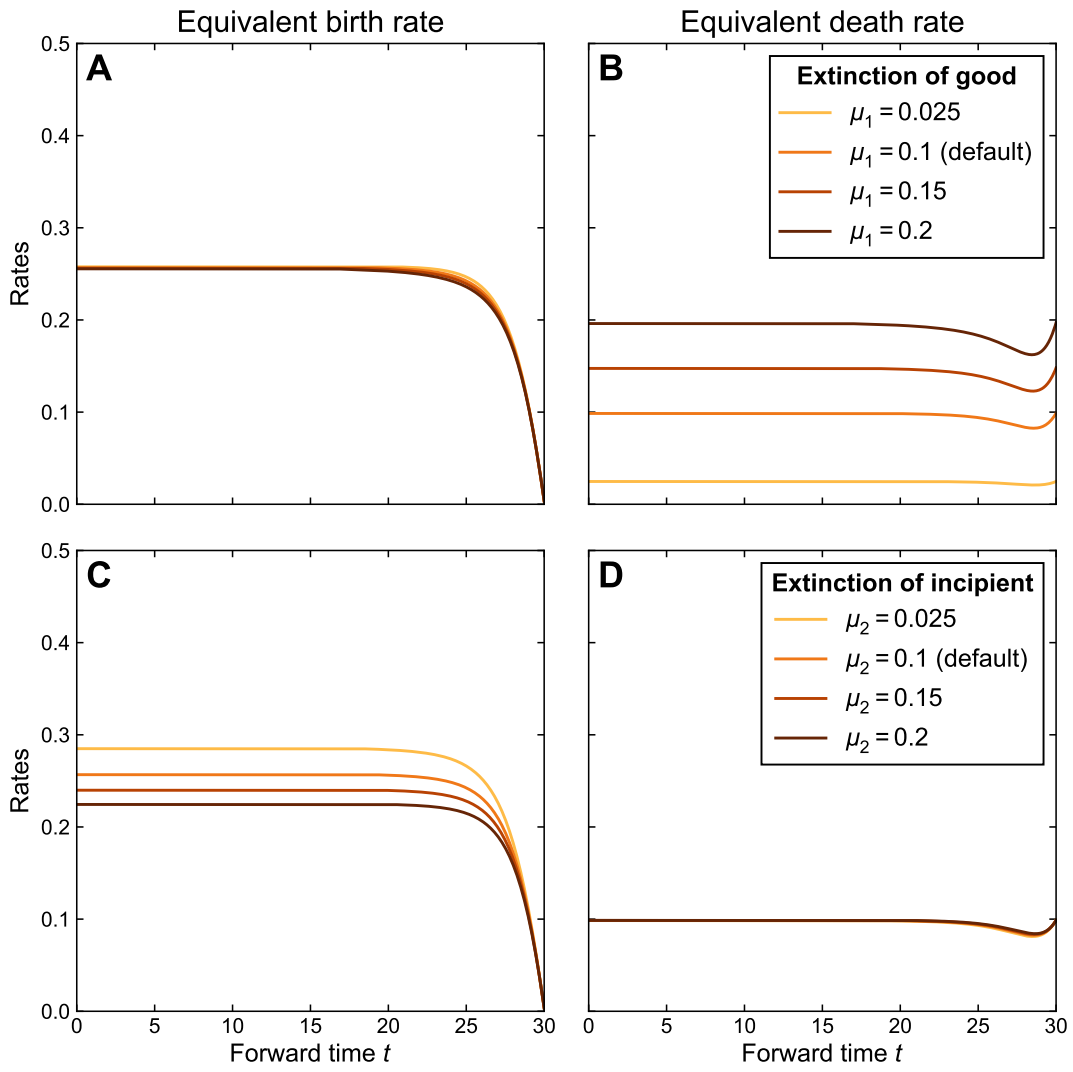


Figure S II.2 – Values of the time-dependent birth–death rates derived from equations II.1 and II.2 as a function of time **as both extinction rates of the PBD model μ_1 and μ_2 vary independently.** In the top row (**A** and **B**) the extinction rate of good lineages varies while in the bottom row (**C** and **D**) the extinction rate of incipient lineages varies. When one rate varies, the other four rates are kept constant (default values are $\lambda_1 = \lambda_3 = 0.3, \lambda_2 = 0.4, \mu_1 = 0.1, \mu_2 = 0.1$).

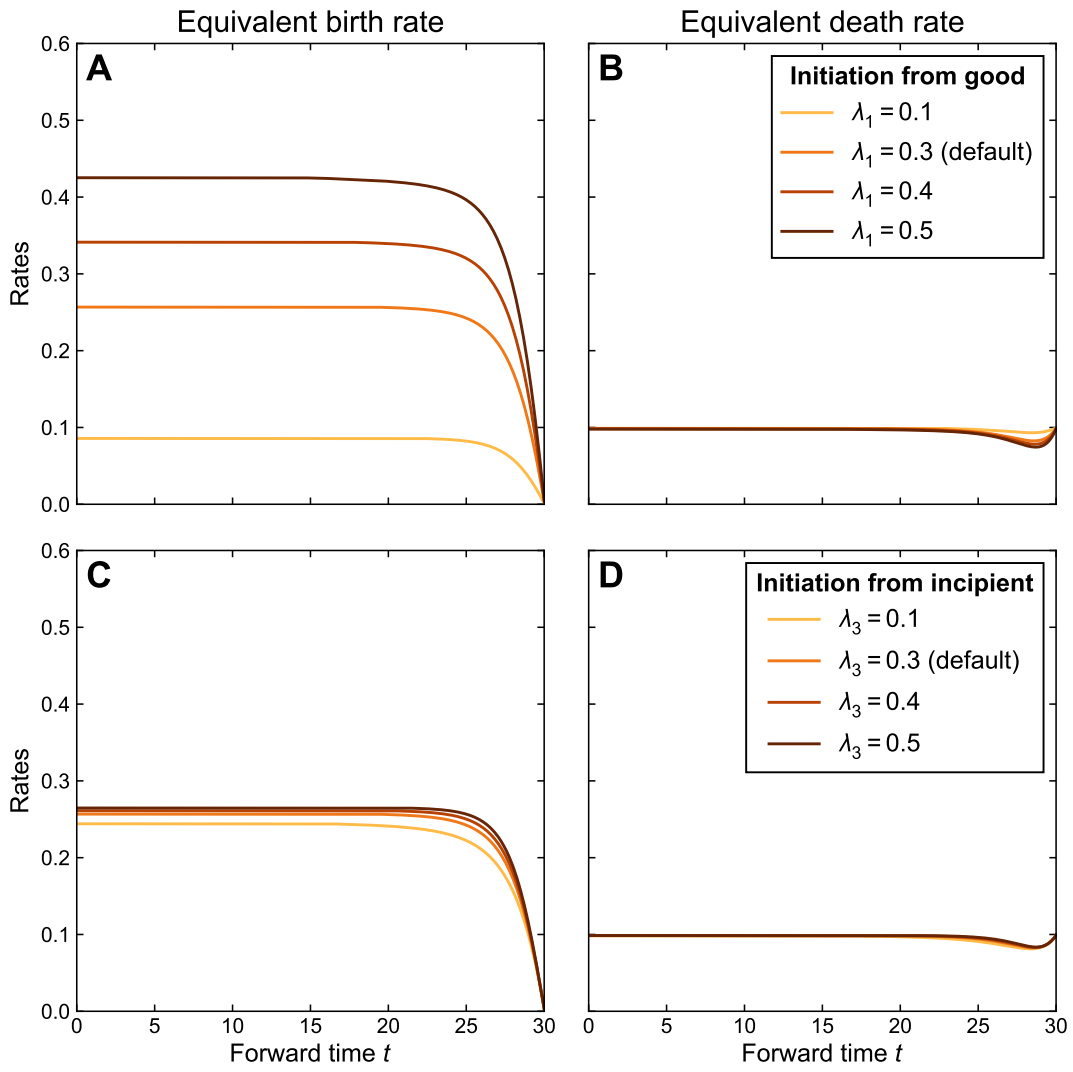


Figure S II.3 – Values of the time-dependent BD rates derived from equations II.1 and II.2 as a function of time **as both initiation rates of the PBD model (λ_1 and λ_3) vary independently**. In the top row (**A** and **B**) the initiation rate from good lineages varies while in the bottom row (**C** and **D**) the initiation rate from incipient lineages varies. When one rate varies, the other four rates are kept constant (default values are $\lambda_1 = 0.3, \lambda_3 = 0.3, \lambda_2 = 0.4, \mu_1 = \mu_2 = 0.1$).

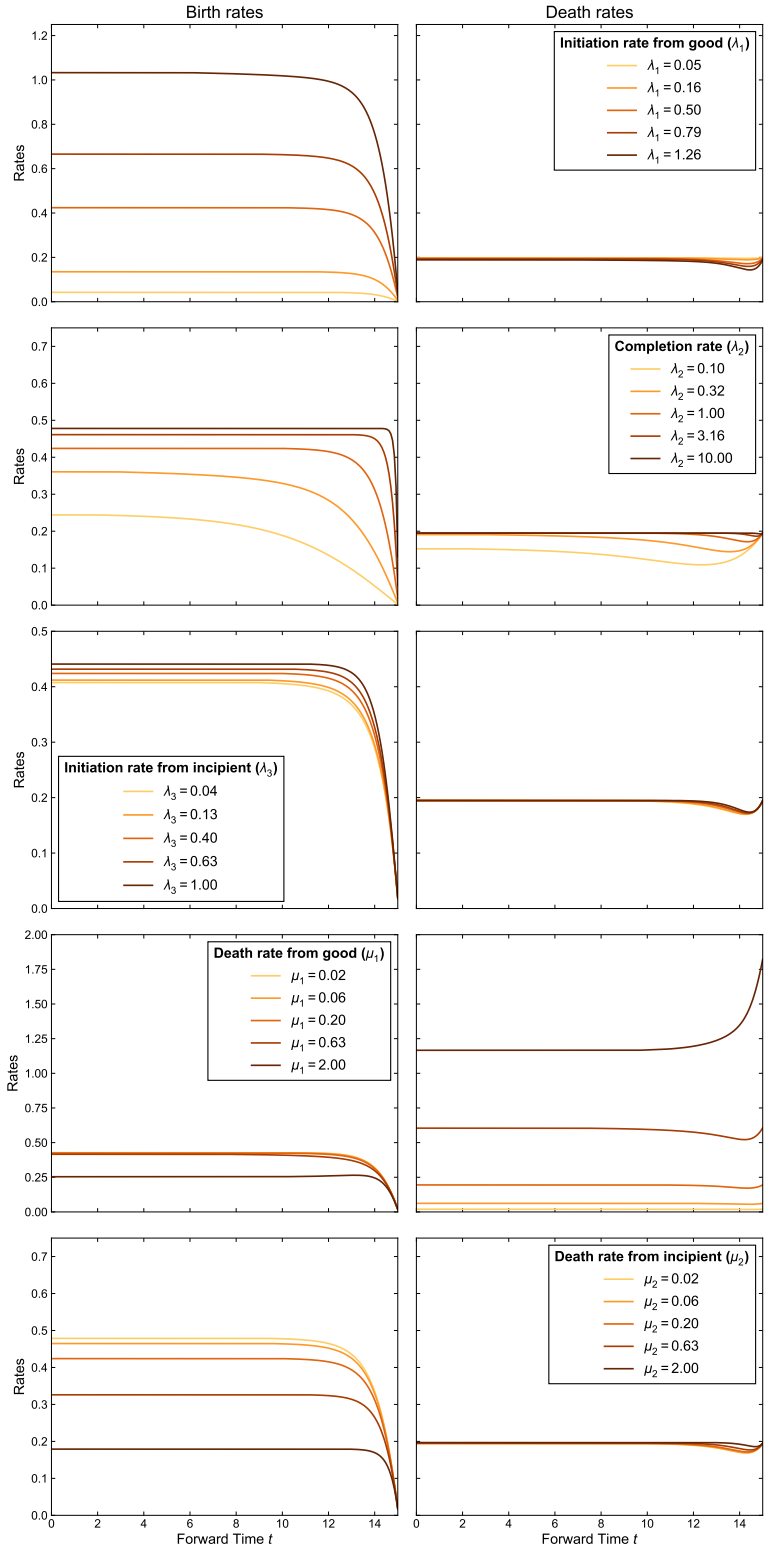

Birth rates
Death rates
Rates
Initiation rate from good (λ_1)
$\lambda_1 = 0.05$
$\lambda_1 = 0.16$
$\lambda_1 = 0.50$
$\lambda_1 = 0.79$
$\lambda_1 = 1.26$
Completion rate (λ_2)
$\lambda_2 = 0.10$
$\lambda_2 = 0.32$
$\lambda_2 = 1.00$
$\lambda_2 = 3.16$
$\lambda_2 = 10.00$
Initiation rate from incipient (λ_3)
$\lambda_3 = 0.04$
$\lambda_3 = 0.13$
$\lambda_3 = 0.40$
$\lambda_3 = 0.63$
$\lambda_3 = 1.00$
Death rate from good (μ_1)
$\mu_1 = 0.02$
$\mu_1 = 0.06$
$\mu_1 = 0.20$
$\mu_1 = 0.63$
$\mu_1 = 2.00$
Death rate from incipient (μ_2)
$\mu_2 = 0.02$
$\mu_2 = 0.06$
$\mu_2 = 0.20$
$\mu_2 = 0.63$
$\mu_2 = 2.00$
Forward Time t
Forward Time t

Figure S II.4 – Time-dependent equivalent BD rates used in the simulations. On each row, only one parameter of the PBD model varies. The left column shows the birth rate and the right column shows the death rate. Each curve corresponds to a set of parameters of the PBD model. The resulting equivalent BD rates are used to generate trees under a time-dependant BD model that are summarized in figure II.5.

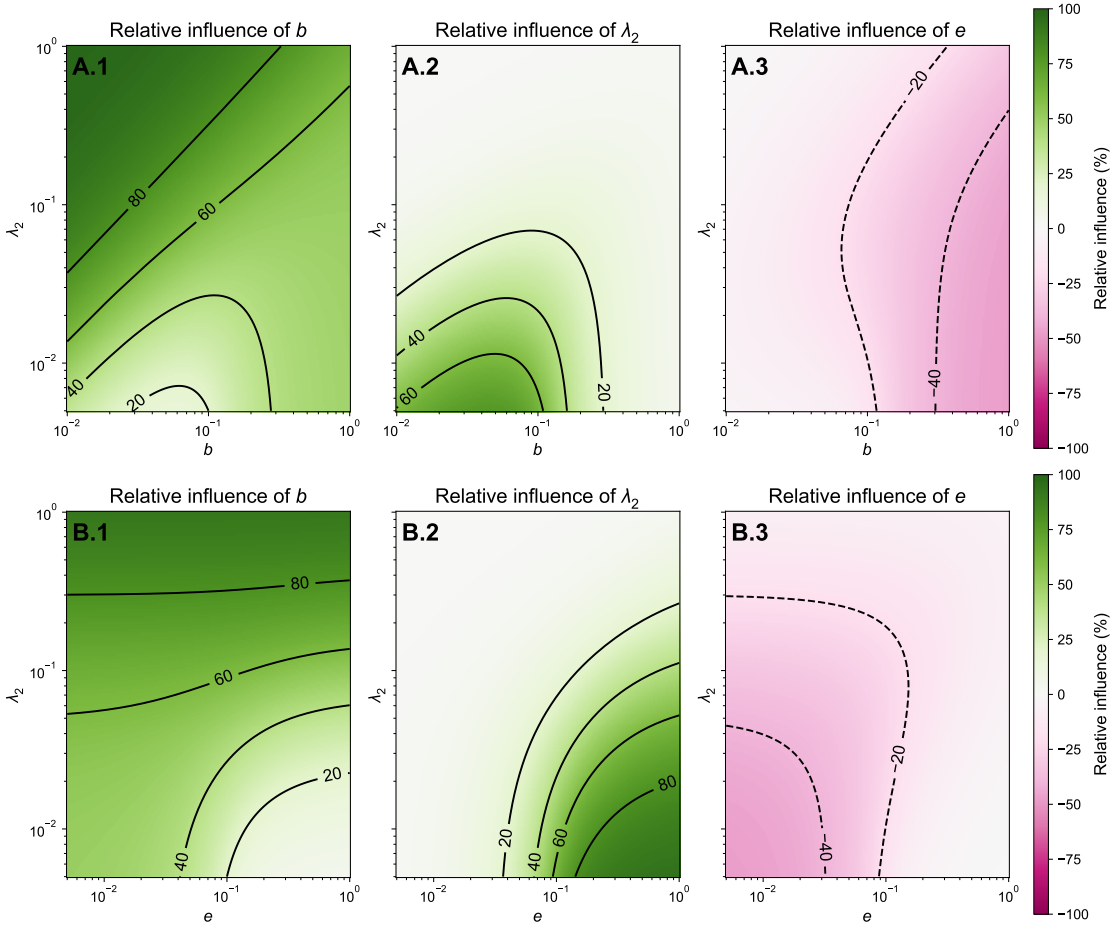


Figure S II.5 – Relative influence of the parameters of the *simplified* PBD model ($b := \lambda_1 = \lambda_3$ and $e := \mu_1 = \mu_2$) **on the constant equivalent birth rate** $\tilde{\lambda}$, as defined in subsection II.5.6, as a function of (b, λ_2) (**A**) and (e, λ_2) (**B**). When unspecified, the values of the rates of the PBD model are taken at 0.1. The values shown here are summarized in figure II.4.

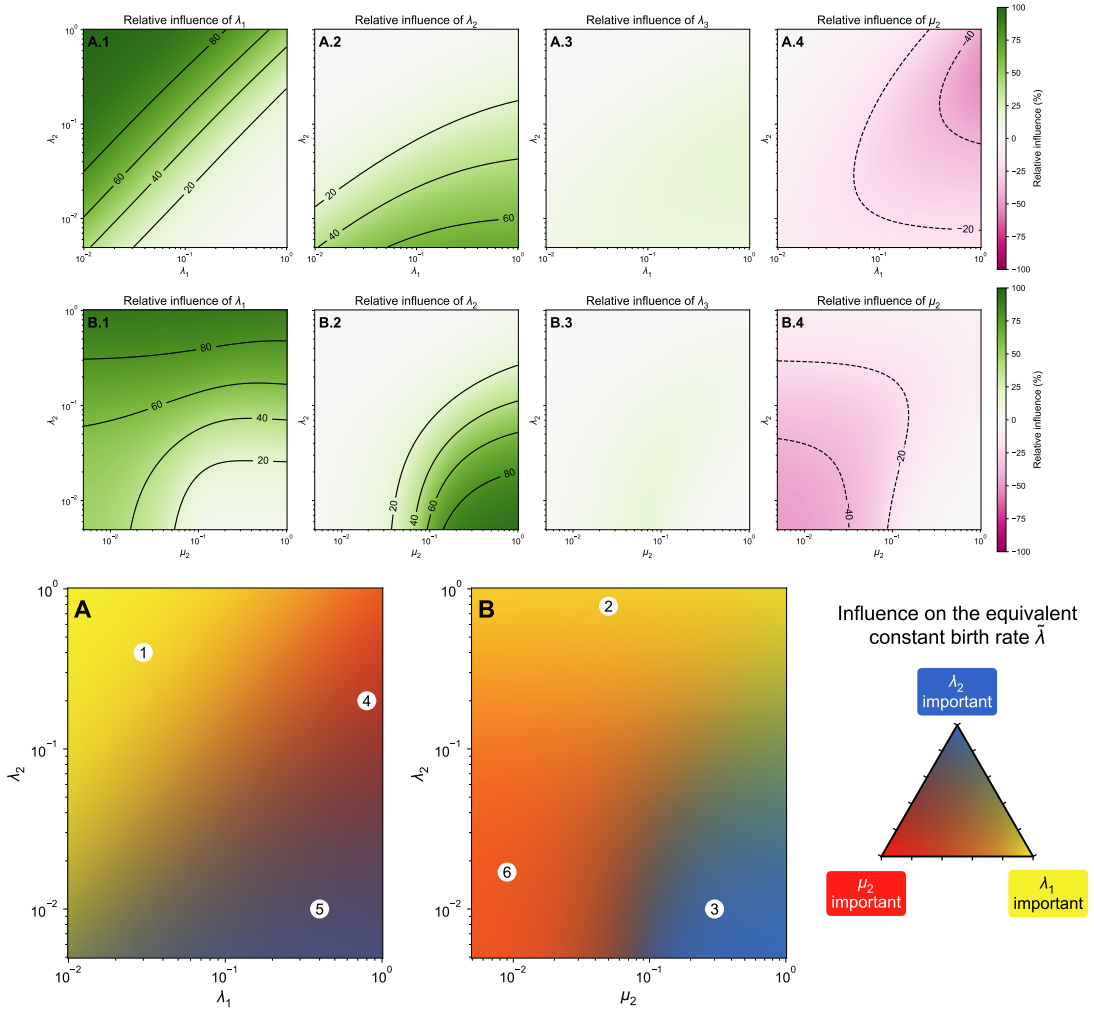
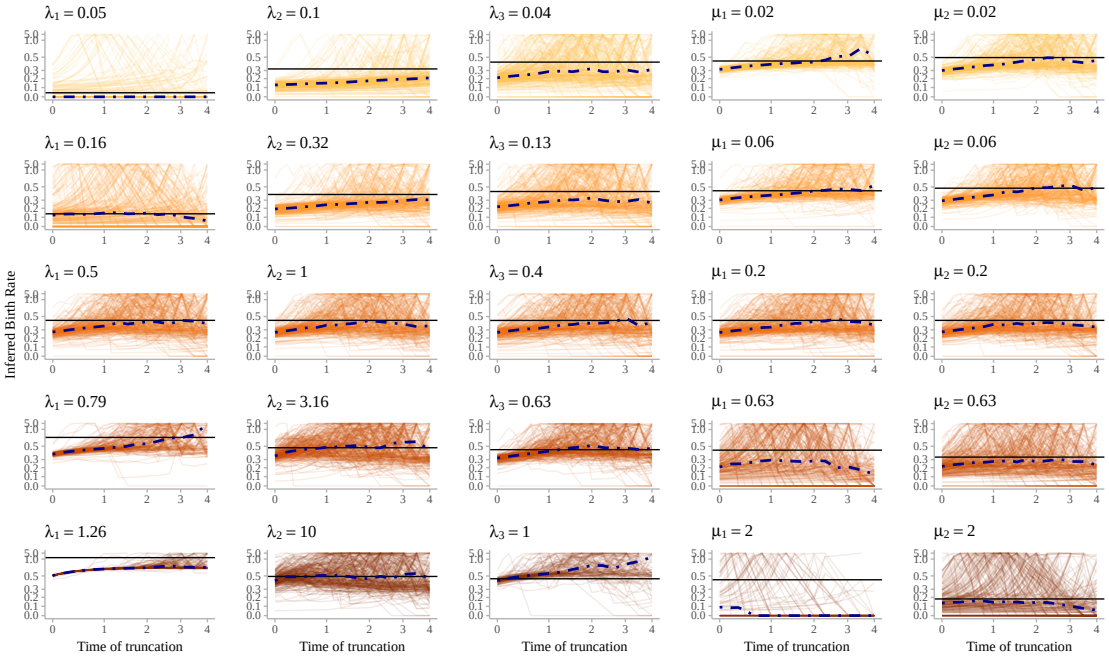
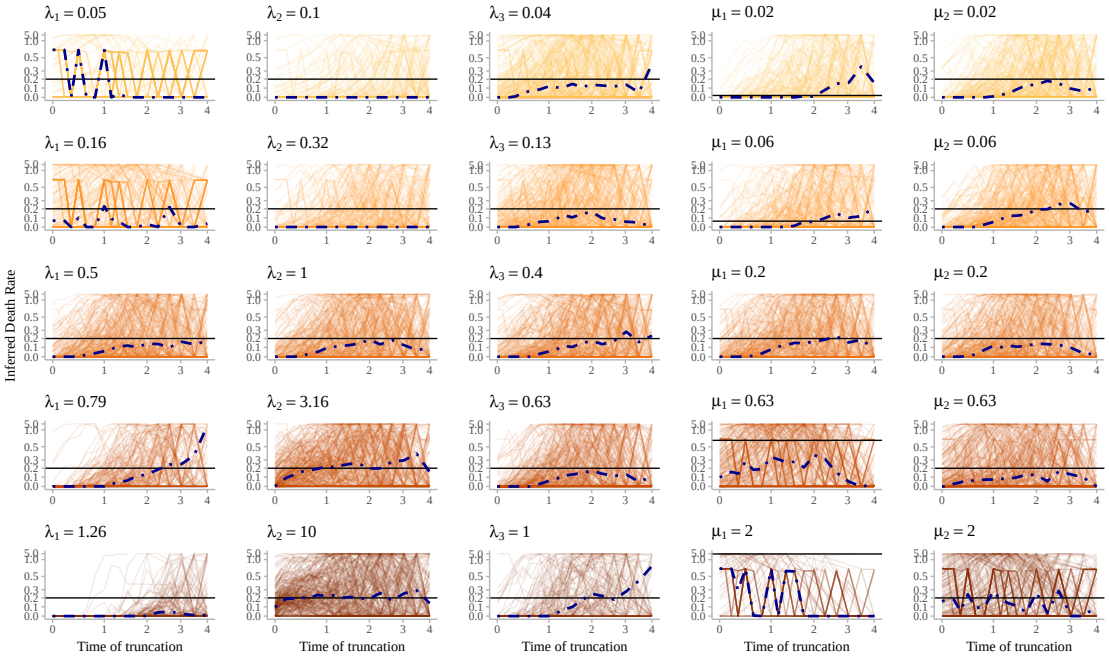


Figure S II.6 – Relative influence of the parameters of the PBD model on the constant equivalent birth rate $\tilde{\lambda}$, as defined in subsection II.5.6, as a function of (λ_1, λ_2) (**A**) and (μ_2, λ_2) (**B**). When unspecified, the values of the rates of the PBD model are set at 0.1. In the lower plot, colors indicate which of the parameters among λ_1, λ_2 and μ_2 are most limiting in the variation of the equivalent constant birth rate $\tilde{\lambda}$, with the color code explained in the triangle on the right. A yellow region (1 or 2) indicates a combination of parameters in which the most influential parameter on the birth rate is the rate of initiation λ_1. A blue region (3) indicates a combination of parameters for which the most influential parameter is the rate of completion λ_2. A red region (4) indicates a combination of parameters where the most influential parameter is the rate of extinction of incipient lineages μ_2. Purple regions (5), orange regions (6) or green regions indicate situations in which several parameters have a comparable influences on the birth rate. In all cases, the influences of λ_1 and λ_2 are positive and the influence of μ_2 is negative.



a – Inferred birth rates (λ) across truncation lengths and replicates.



b – Inferred death rates (μ) across truncation lengths and replicates.

Figure S II.7 – Correspondence between the equivalent time-constant BD rates and inferred BD rates on truncated phylogenies generated under the PBD model. Each column shows the effect of independently modifying a given PBD parameter, with the default value on the third row. For each replicate phylogeny, a classical BD model was fitted on the reconstructed phylogeny and a modified BD model was fitted on the same phylogeny with truncated terminal branches, for increasing truncation lengths. The dot-dashed line corresponds to the median rate across replicates. The horizontal line corresponds to the equivalent time-constant birth or death rate as defined in equation II.7. Note that the default BD model almost always underestimates the equivalent (asymptotic) rates, except for the highest completion rate ($\lambda_2 = 10$), while they are usually recovered with intermediate truncation times.

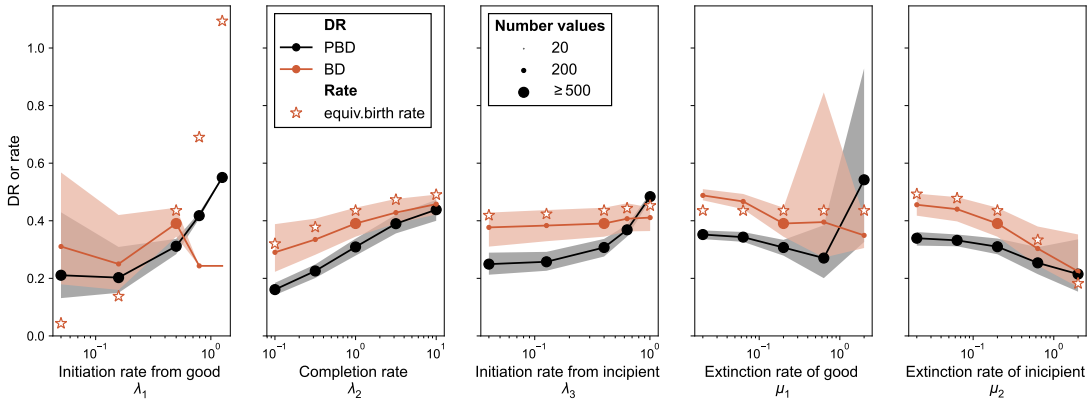


Figure S II.8 – Diversification rate (DR) statistics on simulated trees under the (PBD) and birth–death (BD) process with the equivalent rates. The trees are the same as in figure II.5. In each column, only one of the PBD parameters varies, with the others held constant (default $\lambda_1 = 0.5, \lambda_2 = 1.0, \lambda_3 = 0.4, \mu_1 = 0.2, \mu_2 = 0.2$). For each set of parameters of the PBD model, trees were also generated under equivalent birth and death rates computed using equation II.7. On each tree, we computed the median DR across all tips of the tree. The dots and shaded area indicate the median and interquartile ranges of the median DR. The stars indicate the equivalent constant birth rate $\tilde{\lambda}$ as defined in equation II.7, which is used to generate the BD trees. The DR statistics is defined in Jetz et al. (2012).

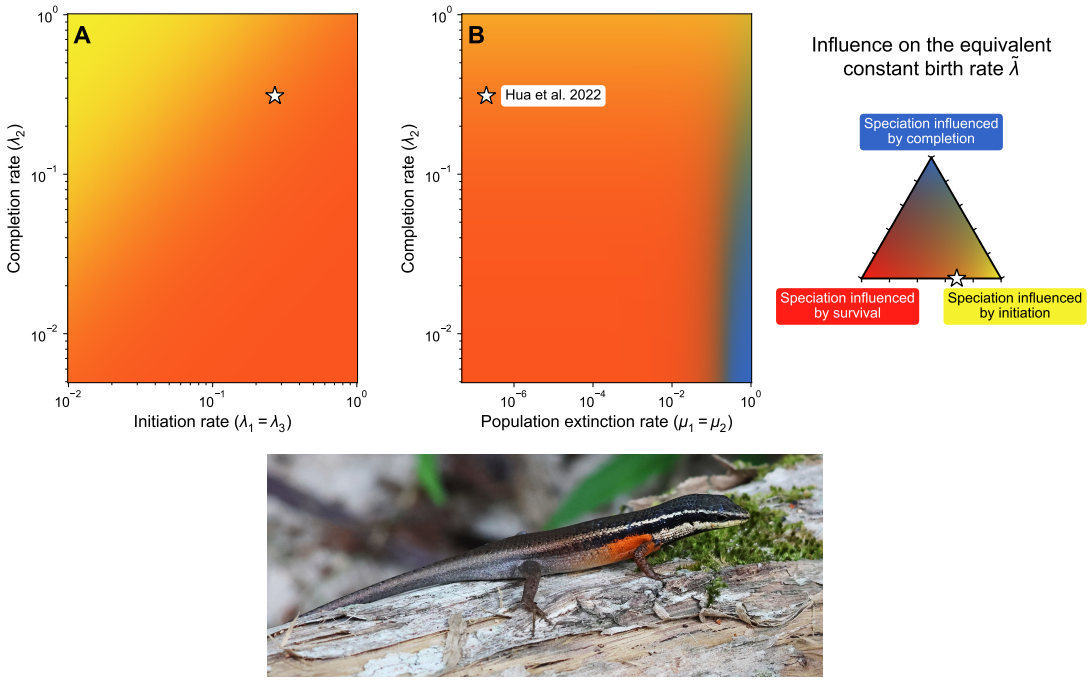


Figure S II.9 – Relative influence of the parameters of the PBD model on the equivalent constant birth rate, with the values fitted on the phylogeny of Australian Rainbow Skinks in Hua et al. (2022) using ProSSE. Colors indicate which of the PBD process among initiation, completion and population extinction limits the equivalent constant birth rate $\tilde{\lambda}$ most, as a function of the initiation and completion rates (**A**) and the population extinction and completion rates (**B**), with the color code explained on the triangle in the right. The stars indicate the parameters used as default, taken from the study by Hua et al. (2022) ($b = 0.27, \lambda_2 = 0.31, e = 2 \times 10^{-7}$) where the author fitted a ProSSE model (a model slightly different from PBD) to a Rainbow Skinks phylogeny of 41 species. In all cases, the rates of initiation and completion have a positive influence on the birth rate and the rate of population extinction has a negative influence. For the combination of parameters indicated by the stars, we find that the relative influences are 0.682 for b, 1.10×10^{-7} for λ_2 and 0.318 for e.
Carlia rostralis © Graham Winterflood 2024, *iNaturalist* CC BY-SA 4.0.

List of figures

List of acronyms

ABC approximate Bayesian computation.
AM ancient migration.

BAMM Bayesian analysis of macroevolutionary mixtures.
BD birth–death.
BDD birth–death diffusion.

ClaDS cladogenetic diversification rate shift.

DILS demographic inference with linked selection.
DMI Dobzhanski–Muller incompatibility.
DR diversification rate.
DTT diversity through time.

GZ grey zone.

HAL holey adaptive landscape.

IBD isolation by distance.
IM isolation with migration.

LTT lineage through time.

MCMC Markov chain Monte Carlo.
Myr million years.

ODE ordinary differential equation.
OLS ordinary least squares.

PBD protracted birth–death.
PGLS phylogenetic generalized least squares.
ProSSE protracted state-dependent speciation and extinction.

RI reproductive isolation.

SC secondary contact.
spp species.
SR species richness.
SSE state-dependent speciation and extinction.

varBD time-varying birth–death.

Bibliography

Afonso Silva, Ana C., Odile Maliet, Leandro Aristide, David Nogués-Bravo, Nathan Upham, Walter Jetz and Hélène Morlon (2025). Negative global-scale association between genetic diversity and speciation rates in mammals. *Nature Communications* 16, 1796.

de Aguiar, Marcus A. M. (2017). Speciation in the Derrida-Higgs model with finite genomes and spatial populations. *Journal of Physics A: Mathematical and Theoretical* 50.8, 085602.

Aldous, David J. (2001). Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statistical Science* 16.1, 23–34.

Alfaro, Michael E., Francesco Santini, Chad Brock, Hugo Alamillo, Alex Dornburg, Daniel L. Rabosky, Giorgio Carnevale and Luke J. Harmon (2009). Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences* 106.32, 13410–13414.

Anderson, Sean A. S., Sachin Kaushik and Daniel R. Matute (2025). The comparative analysis of lineage-pair traits. *Systematic Biology* 75.2, 328–343.

Anderson, Sean A. S. and Jason T. Weir (2022). The role of divergent ecological adaptation during allopatric speciation in vertebrates. *Science* 378.6625, 1214–1218.

April, Julien, Robert H. Hanner, Anne-Marie Dion-Côté and Louis Bernatchez (2013). Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Molecular Ecology* 22.2, 409–422.

Aristide, Leandro and Hélène Morlon (2019). Understanding the effect of competition during evolutionary radiations: an integrated model of phenotypic and species diversification. *Ecology Letters* 22.12, 2006–2017.

Árnason, Úlfur, Fritjof Lammers, Vikas Kumar, Maria A. Nilsson and Axel Janke (2018). Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Science Advances* 4.4, eaap9873.

Bateson, William (1909). "Heredity and variation in modern lights". In: *Darwin and Modern Science*. Cambridge University Press, pp. 85–101. ISBN: 978-0-511-69395-3.

Benton, Michael J. and Paul N. Pearson (2001). Speciation in the fossil record. *Trends in Ecology & Evolution* 16.7, 405–411.

Bolnick, Daniel I. and Benjamin M. Fitzpatrick (2007). Sympatric speciation: Models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics* 38, 459–487.

Boucher, Florian C., Niklaus E. Zimmermann and Elena Conti (2016). Allopatric speciation with little niche divergence is common among alpine Primulaceae. *Journal of Biogeography* 43.3, 591–602.

Bouinier, Titouan, Arthur Brunaud, Charline Smadi and Violaine Llaurens (2026). *Evolution of divergent daily temporal niches shaped by male-male competition can generate sympatric speciation*. bioRxiv: 2024.07.31.601896. Pre-published.

Bromham, Lindell, Xia Hua, Robert Lanfear and Peter F. Cowman (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.

Burban, Ewen, Maud I. Tenaillon and Sylvain Glémin (2024). RIDGE, a tool tailored to detect gene flow barriers across species pairs. *Molecular Ecology Resources* 24.4, e13944.

Burbrink, Frank T., Sara Ruane, Nirhy Rabibisoa, Achille P. Raselimanana, Christopher J. Raxworthy and Arianna Kuhn (2023). Speciation rates are unrelated to the formation of population structure in Malagasy gemsnakes. *Ecology and Evolution* 13.8, e10344.

Burton, Ronald S. (2022). The role of mitonuclear incompatibilities in allopatric speciation. *Cellular and Molecular Life Sciences* 79.2, 103.

Castillo, Dean M. (2017). Factors contributing to the accumulation of reproductive isolation: A mixed model approach. *Ecology and Evolution* 7.15, 5808–5820.

Chan, Kin Onn, Dario N. Neokleous, Shahrul Anuar, Rafe M. Brown, Carl R. Hutter, Indraneil Das and Stefan T. Hertwig (2026). A genomic perspective on cryptic species reveals complex evolutionary dynamics in the gray zone of the speciation continuum. *Systematic Biology*, syag001.

Chichorro, Filipe, Aino Juslén and Pedro Cardoso (2019). A review of the relation between species traits and extinction risk. *Biological Conservation* 237, 220–229.

Christie, Kyle and Sharon Y. Strauss (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.

Christie, Kyle, Linnea S. Fraser and David B. Lowry (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.

Claramunt, Santiago, Elizabeth P. Derryberry, J. V. Remsen Jr and Robb T. Brumfield (2011). High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proceedings of the Royal Society B: Biological Sciences* 279.1733, 1567–1574.

Coyne, Jerry A. and H. Allen Orr (2004). *Speciation*. Sunderland, Massachusetts: Sinauer Associates, Inc. Publ. ISBN: 0-87893-089-2.

Darwin, Charles (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

de Queiroz, Kevin (2007). Species concepts and species delimitation. *Systematic Biology* 56.6, 879–886.

Dieckmann, Ulf and Michael Doebeli (1999). On the origin of species by sympatric speciation. *Nature* 400.6742, 354–357.

Dobzhansky, Theodosius (1936). Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21.2, 113–135.

Doebeli, Michael and Ulf Dieckmann (2003). Speciation along environmental gradients. *Nature* 421.6920, 259–264.

Dynesius, Mats and Roland Jansson (2014). Persistence of within-species lineages: A neglected control of speciation rates. *Evolution* 68.4, 923–934.

Etienne, Rampal S. and James Rosindell (2012). Prolonging the past counteracts the pull of the present: Protracted speciation can explain observed slowdowns in diversification. *Systematic Biology* 61.2, 204–213.

Etienne, Rampal S., Hélène Morlon and Amaury Lambert (2014). Estimating the duration of speciation from phylogenies. *Evolution* 68.8, 2430–2440.

Excoffier, Laurent, Nina Marchi, David Alexander Marques, Remi Matthey-Doret, Alexandre Gouy and Vitor C Sousa (2021). fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* 37.24, 4882–4885.

Fernández-Álvarez, Fernando Á., Gustavo Sanchez, Diego Deville, Morag Taite, Roger Villanueva and A. Louise Allcock (2023). Atlantic oceanic squids in the "grey speciation zone". *Integrative and Comparative Biology* 63.6, 1214–1225.

Forbes, Andrew A., Sara N. Devine, Alaine C. Hippee, Eric S. Tvedte, Anna K. G. Ward, Heather A. Widmayer and Caleb J. Wilson (2017). Revisiting the particular role of host shifts in initiating insect speciation. *Evolution* 71.5, 1126–1137.

Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt and A. M. Wilczek (2011). A map of local adaptation in *Arabidopsis thaliana*. *Science* 334.6052, 86–89.

Fraïsse, Christelle et al. (2021). DILS: Demographic inferences with linked selection by using ABC. *Molecular Ecology Resources* 21.8, 2629–2644.

Frayer, Megan E., Nemo V. Robles, María José Rodríguez-Barrera, Jenn M. Coughlan and Molly Schumer (2025). The molecular evolutionary basis of species formation revisited. *Trends in Genetics* 41.12, 1068–1095.

Freeman, Benjamin G., Jonathan Rolland, Graham A. Montgomery and Dolph Schluter (2022). Faster evolution of a premating reproductive barrier is not associated with faster speciation rates in New World passerine birds. *Proceedings of the Royal Society B: Biological Sciences* 289.1966, 20211514.

Garlovsky, Martin D. et al. (2024). Synthesis and scope of the role of postmating prezygotic isolation in speciation. *Cold Spring Harbor Perspectives in Biology* 16.10, a041429.

Gavrilets, Sergey (1997). Evolution and speciation on holey adaptive landscapes. *Trends in Ecology & Evolution* 12.8, 307–312.

Gavrilets, Sergey, Li Hai and Michael D. Vose (1998). Rapid parapatric speciation on holey adaptive landscapes. *Proceedings of the Royal Society B: Biological Sciences* 265.1405, 1483–1489.

Gavrilets, Sergey (1999). A dynamical theory of speciation on holey adaptive landscapes. *The American Naturalist* 154.1, 1–22.

Gavrilets, Sergey, Randal Acton and Janko Gravner (2000). Dynamics of speciation and diversification in a metapopulation. *Evolution* 54.5, 1493–1501.

Gavrilets, Sergey (2000). Waiting time to parapatric speciation. *Proceedings of the Royal Society B: Biological Sciences* 267.1461, 2483–2492.

Gavrilets, Sergey and David Waxman (2002). Sympatric speciation by sexual conflict. *Proceedings of the National Academy of Sciences* 99.16, 10533–10538.

Gavrilets, Sergey (2003). Perspective: Models of speciation: What have we learned in 40 years? *Evolution* 57.10, 2197–2215.

Giraud, Tatiana, Pierre Gladieux and Sergey Gavrilets (2010). Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution* 25.7, 387–395.

Giraud, Tatiana (2021). "Génération de la biodiversité : que sait-on sur la nature et l'origine de la vie ? Pourquoi y a-t-il autant d'espèces sur terre ?" Lesson. *Dynamique de la biodiversité et évolution* (Collège de France, Paris).

Goldie, Xavier, Robert Lanfear and Lindell Bromham (2011). Diversification and the rate of molecular evolution: no evidence of a link in mammals. *BMC Evolutionary Biology* 11.1, 286.

Gourbière, Sébastien and James Mallet (2010). Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". *Evolution* 64.1, 1–24.

Gutenkunst, Ryan N., Ryan D. Hernandez, Scott H. Williamson and Carlos D. Bustamante (2010). Diffusion approximations for demographic inference: dadi. *Nature Precedings*.

Hartl, Daniel L. and Andrew G. Clark (2007). *Principles of Population Genetics*. Fourth edition. Sunderland, Mass: Sinauer Associates. ISBN: 978-0-87893-308-2.

Hartmann, Fanny E., Ricardo C. Rodríguez de la Vega, Pierre Gladieux, Wen-Juan Ma, Michael E. Hood and Tatiana Giraud (2020). Higher gene flow in sex-related chromosomes than in autosomes during fungal divergence. *Molecular Biology and Evolution* 37.3, 668–682.

Harvey, Michael G., Glenn F. Seeholzer, Brian Tilston Smith, Daniel L. Rabosky, Andrés M. Cuervo and Robb T. Brumfield (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, Michael G., Sonal Singhal and Daniel L. Rabosky (2019). Beyond reproductive isolation: Demographic controls on the speciation process. *Annual Review of Ecology, Evolution, and Systematics* 50.1, 75–95.

Heard, Stephen B. (1992). Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution* 46.6, 1818–1826.

Hernández-Hernández, Tania, Elizabeth C. Miller, Cristian Román-Palacios and John J. Wiens (2021). Speciation across the tree of life. *Biological Reviews* 96.4, 1205–1242.

Higgs, Paul G. and Bernard Derrida (1992). Genetic distance and species formation in evolving populations. *Journal of Molecular Evolution* 35.5, 454–465.

Hua, Xia and Lindell Bromham (2017). Darwinism for the genomic age: Connecting mutation to diversification. *Frontiers in Genetics* 8.12.

Hua, Xia, Tyara Herdha and Conrad J. Burden (2022). Protracted speciation under the state-dependent speciation and extinction approach. *Systematic Biology* 71.6, 1362–1377.

Huang, Jen-Pan, Steven D. Leavitt and H. Thorsten Lumbsch (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.

Janzen, Thijs and Rampal S. Etienne (2024). Phylogenetic tree statistics: A systematic overview using the new R package 'treestats'. *Molecular Phylogenetics and Evolution* 200, 108168.

Jetz, Walter, Gavin H. Thomas, Jeffrey B. Joy, Klaas Hartmann and Arne O. Mooers (2012). The global diversity of birds in space and time. *Nature* 491.7424, 444–448.

Kersting, Sophie J., Kristina Wicke and Mareike Fischer (2025). Tree balance in phylogenetic models. *Philosophical Transactions of the Royal Society B: Biological Sciences* 380.1919, 20230303.

Khurana, Mark P., Neil Scheidwasser-Clow, Matthew J. Penn, Samir Bhatt and David A. Duchêne (2023). The limits of the constant-rate Birth–Death prior for phylogenetic tree topology inference. *Systematic Biology* 73.1, 235–246.

Kimura, Motoo and Tomoko Ohta (1971). On the rate of molecular evolution. *Journal of Molecular Evolution* 1.1, 1–17.

Kisel, Yael, Alejandra C. Moreno-Letelier, Diego Bogarín, Martyn P. Powell, Mark W. Chase and Timothy G. Barraclough (2012). Testing the link between population genetic differentiation and clade diversification in Costa Rican Orchids. *Evolution* 66.10, 3035–3052.

Kondrashov, Alexey S. and Mikhail V. Mina (1986). Sympatric speciation: when is it possible? *Biological Journal of the Linnean Society* 27.3, 201–223.

Kulmuni, Jonna, Roger K. Butlin, Kay Lucek, Vincent Savolainen and Anja Marie Westram (2020). Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1806, 20190528.

Lagomarsino, Laura P., Fabien L. Condamine, Alexandre Antonelli, Andreas Mulch and Charles C. Davis (2016). The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist* 210.4, 1430–1442.

Lambert, Amaury, Hélène Morlon and Rampal S. Etienne (2014). The reconstructed tree in the lineage-based model of protracted speciation. *Journal of Mathematical Biology* 70, 367–397.

Landis, Michael J., Ignacio Quintero, Martha M. Muñoz, Felipe Zapata and Michael J. Donoghue (2022). Phylogenetic inference of where species spread or split across barriers. *Proceedings of the National Academy of Sciences* 119.13, e2116948119.

Lanfear, Robert, Simon Y. W. Ho, Dominic Love and Lindell Bromham (2010). Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences* 107.47, 20423–20428.

Lessios, Harilaos (1998). "The first stage of speciation as seen in organisms separated by the Isthmus of Panama". In: *Endless Forms: Species and Speciation*. Oxford University Press, pp. 186–201. ISBN: 978-0-19-770084-6.

Lewitus, Eric, Lucie Bittner, Shruti Malviya, Chris Bowler and Hélène Morlon (2018). Clade-specific diversification dynamics of marine diatoms since the Jurassic. *Nature Ecology & Evolution* 2.11, 1715–1723.

Long, Zhiqin, Xu Zhang, Yue Yu, Gregory L. Owens, Diana J. Rennison, Clarisse Palma-Silva, Daniel Ortiz-Barrientos and Loren H. Rieseberg (2026). Speciation with gene flow. *Plants, People, Planet*.

Louca, Stilianos and Matthew W. Pennell (2020). Extant timetrees are consistent with a myriad of diversification histories. *Nature* 580.7804, 502–505.

Lowry, David B. (2012). Ecotypes and the controversy over stages in the formation of new species. *Biological Journal of the Linnean Society* 106.2, 241–257.

Maliet, Odile, Florian Hartig and Hélène Morlon (2019). A model with many small shifts for estimating species-specific diversification rates. *Nature Ecology & Evolution* 3.7, 1086–1092.

Maliet, Odile and Hélène Morlon (2021). Fast and accurate estimation of species-specific diversification rates using data augmentation. *Systematic Biology* 71.2, 353–366.

Mani, G. S. and Bryan Campbell Clarke (1990). Mutational order: a major stochastic process in evolution. *Proceedings of the Royal Society B: Biological Sciences* 240.1297, 29–37.

Marshall, Jeremy L., Michael L. Arnold and Daniel J. Howard (2002). Reinforcement: the road not taken. *Trends in Ecology & Evolution* 17.12, 558–563.

Matute, Daniel R., Ian A. Butler, David A. Turissini and Jerry A. Coyne (2010). A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329.5998, 1518–1521.

Maya-Lastra, Carlos A. and Deren A. R. Eaton (2021). *Genetic incompatibilities do not snowball in a demographic model of speciation*. bioRxiv: 2021.02.23.432472. Pre-published.

Mayden, Richard L (1997). "A hierarchy of species concepts: the denouement in the saga of the species problem". In: *Species: The Units of Diversity*. Chapman & Hall, pp. 381–423. ISBN: 978-0-412-63120-7.

Mayr, Ernst (1942). *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York: Columbia University Press. ISBN: 978-0-674-86250-0.

— (1963). *Animal Species and Evolution*. Harvard University Press. ISBN: 978-0-674-86530-3.

Mendelson, Tamra C., Brian D. Inouye and Mark D. Rausher (2004). Quantifying patterns in the evolution of reproductive isolation. *Evolution* 58.7, 1424–1433.

Mérot, Claire, Brigitte Frérot, Ene Leppik and Mathieu Joron (2015). Beyond magic traits: Multimodal mating cues in *Heliconius* butterflies. *Evolution* 69.11, 2891–2904.

Mérot, Claire, Camilo Salazar, Ray M. Merrill, Chris D. Jiggins and Mathieu Joron (2017). What shapes the continuum of reproductive isolation? Lessons from *Heliconius* butterflies. *Proceedings of the Royal Society B: Biological Sciences* 284.1856, 20170335.

Merrill, Richard M., Richard W. R. Wallbank, Vanessa Bull, Patricio C. A. Salazar, James Mallet, Martin Stevens and Chris D. Jiggins (2012). Disruptive ecological selection on a mating cue. *Proceedings of the Royal Society B: Biological Sciences* 279.1749, 4907–4913.

Moen, Daniel and Hélène Morlon (2014). Why does diversification slow down? *Trends in Ecology & Evolution* 29.4, 190–197.

Monnet, François (2023). "Speciation dynamics: contrasts between plants and animals". PhD thesis. Université de Lille, Universiteit Gent.

Monnet, François, Zoé Postel, Pascal Touzet, Christelle Fraïsse, Yves Van de Peer, Xavier Vekemans and Camille Roux (2025). Rapid establishment of species barriers in plants compared with that in animals. *Science* 389.6765, 1147–1150.

Morlon, Hélène, Eric Lewitus, Fabien L. Condamine, Marc Manceau, Julien Clavel and Jonathan Drury (2016). RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods in Ecology and Evolution* 7.5, 589–597.

Morlon, Hélène, Jérémy Andréoletti, Joëlle Barido-Sottani, Sophia Lambert, Benoît Perez-Lamarque, Ignacio Quintero, Viktor Senderov and Pierre Veron (2024). Phylogenetic insights into diversification. *Annual Review of Ecology, Evolution, and Systematics* 55, 1–21.

Moyle, Robert G., Christopher E. Filardi, Catherine E. Smith and Jared Diamond (2009). Explosive Pleistocene diversification and hemispheric expansion of a "great speciator". *Proceedings of the National Academy of Sciences* 106.6, 1863–1868.

Muller, Hermann J. (1942). "Isolating mechanisms, evolution, and temperature". In: *Biology Symposium*. Vol. 6, pp. 71–125.

Nee, Sean, Edward C. Holmes, Robert M. May and Paul H. Harvey (1994). Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society B: Biological Sciences* 344.1307, 77–82.

Nee, Sean (2006). Birth-Death models in macroevolution. *Annual Review of Ecology, Evolution, and Systematics* 37.1, 1–17.

Niemiller, Matthew L., Benjamin M. Fitzpatrick and Brian T. Miller (2008). Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: Gyrinophilus) inferred from gene genealogies. *Molecular Ecology* 17.9, 2258–2275.

Norström, Melissa M., Mattia C. F. Prosperi, Rebecca R. Gray, Annika C. Karlsson and Marco Salemi (2012). PhyloTempo: A set of R scripts for assessing and visualizing temporal clustering in genealogies inferred from serially sampled viral sequences. *Evolutionary Bioinformatics* 8, 261–269.

Nosil, Patrik and Samuel M. Flaxman (2010). Conditions for mutation-order speciation. *Proceedings of the Royal Society B: Biological Sciences* 278.1704, 399–407.

Orr, H. Allen (1995). The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139.4, 1805–1813.

Orr, H. Allen and Lynne H. Orr (1996). Waiting for speciation: the effect of population subdivision on the time to speciation. *Evolution* 50.5, 1742–1749.

Orr, H. Allen and Michael Turelli (2001). The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55.6, 1085–1094.

Ovaskainen, Otso and Baruch Meerson (2010). Stochastic models of population extinction. *Trends in Ecology & Evolution* 25.11, 643–652.

Pannetier, Théo, César Martinez, Lynsey Bunnefeld and Rampal S. Etienne (2021). Branching patterns in phylogenies cannot distinguish diversity-dependent diversification from time-dependent diversification. *Evolution* 75.1, 25–38.

Papadopulos, Alexander S. T., William J. Baker, Darren Crayn, Roger K. Butlin, Ralf G. Kynast, Ian Hutton and Vincent Savolainen (2011). Speciation with gene flow on Lord Howe Island. *Proceedings of the National Academy of Sciences* 108.32, 13188–13193.

Paradis, Emmanuel and Klaus Schliep (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35.3, 526–528.

Peñalba, Joshua V. et al. (2024). The role of hybridization in species formation and persistence. *Cold Spring Harbor Perspectives in Biology* 16.12, a041445.

Perez-Lamarque, Benoît, Maarja Öpik, Odile Maliet, Ana C. Afonso Silva, Marc-André Selosse, Florent Martos and Hélène Morlon (2022). Analysing diversification dynamics using barcoding data: The case of an obligate mycorrhizal symbiont. *Molecular Ecology* 31.12, 3496–3512.

Phillimore, Albert B. and Trevor D. Price (2008). Density-dependent cladogenesis in birds. *PLoS Biology* 6.3, e71.

Presgraves, Daven C., Lakshmi Balagopalan, Susan M. Abmayr and H. Allen Orr (2003). Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* 423.6941, 715–719.

Price, Trevor D. and Michelle M. Bouvier (2002). The evolution of F_1 postzygotic incompatibilities in birds. *Evolution* 56.10, 2083–2089.

Pybus, Oliver G. and Paul H. Harvey (2000). Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society B: Biological Sciences* 267.1459, 2267–2272.

Quintero, Ignacio, Nicolas Lartillot and Hélène Morlon (2024). Imbalanced speciation pulses sustain the radiation of mammals. *Science* 384.6699, 1007–1012.

Rabosky, Daniel L. and Daniel R. Matute (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.

Rabosky, Daniel L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One* 9.2, e89543.

— (2016). Reproductive isolation and the causes of speciation rate variation in nature. *Biological Journal of the Linnean Society* 118.1, 13–25.

Ravigné, Virginie et al. (2010). "La spéciation". In: *Biologie Évolutive*. De Boeck, pp. 165–210. ISBN: 978-2-8073-0296-9.

Reifová, Radka et al. (2023). Mechanisms of intrinsic postzygotic isolation: from traditional genic and chromosomal views to genomic and epigenetic perspectives. *Cold Spring Harbor Perspectives in Biology* 15.10, a041607.

Rieseberg, Loren H. and John H. Willis (2007). Plant speciation. *Science* 317.5840, 910–914.

Rohde, Klaus (1992). Latitudinal gradients in species diversity: The search for the primary cause. *Oikos* 65.3, 514–527.

Rolland, Jonathan et al. (2023). Conceptual and empirical bridges between micro- and macroevolution. *Nature Ecology & Evolution* 7.8, 1181–1193.

Rosenblum, Erica Bree et al. (2012). Goldilocks meets Santa Rosalia: An ephemeral speciation model explains patterns of diversification across time scales. *Evolutionary Biology* 39.2, 255–261.

Rosindell, James, Stephen J. Cornell, Stephen P. Hubbell and Rampal S. Etienne (2010). Protracted speciation revitalizes the neutral theory of biodiversity. *Ecology Letters* 13.6, 716–727.

Rougemont, Quentin, Pierre-Alexandre Gagnaire, Charles Perrier, Clémence Genthon, Anne-Laure Besnard, Sophie Launey and Guillaume Evanno (2017). Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology* 26.1, 142–162.

Roux, Camille, Christelle Fraïsse, Jonathan Romiguier, Yoann Anciaux, Nicolas Galtier and Nicolas Bierne (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology* 14.12, e2000234.

Rundell, Rebecca J. and Trevor D. Price (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution* 24.7, 394–399.

Rundle, Howard D. and Patrik Nosil (2005). Ecological speciation. *Ecology Letters* 8.3, 336–352.

Satokangas, Ina, Simon H. Martin, Heikki Helanterä, Jari Saramäki and Jonna Kulmuni (2020). Multi-locus interactions and the build-up of reproductive isolation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1806, 20190543.

Schluter, Dolph (2009). Evidence for ecological speciation and its alternative. *Science* 323.5915, 737–741.

Schluter, Dolph and Matthew W. Pennell (2017). Speciation gradients and the distribution of biodiversity. *Nature* 546.7656, 48–55.

Servedio, Maria R., G. Sander Van Doorn, Michael Kopp, Alicia M. Frame and Patrik Nosil (2011). Magic traits in speciation: "magic" but not rare? *Trends in Ecology & Evolution* 26.8, 389–397.

Sherman, Natasha A., Anna Victorine, Richard J. Wang and Leonie C. Moyle (2014). Interspecific tests of allelism reveal the evolutionary timing and pattern of accumulation of reproductive isolation mutations. *PLoS Genetics* 10.9, e1004623.

Silvestro, Daniele, Jan Schnitzler, Lee Hsiang Liow, Alexandre Antonelli and Nicolas Salamin (2014). Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Systematic Biology* 63.3, 349–367.

Simon, Matthieu, Stéphanie Durand, Natacha Pluta, Nicolas Gobron, Lucy Botran, Anthony Ricou, Christine Camilleri and Françoise Budar (2016). Genomic conflicts that cause pollen mortality and raise reproductive barriers in *Arabidopsis thaliana*. *Genetics* 203.3, 1353–1367.

Singhal, Sonal, Huateng Huang, Maggie R. Grundler, María R. Marchán-Rivadeneira, Iris Holmes, Pascal O. Title, Stephen C. Donnellan and Daniel L. Rabosky (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.

Singhal, Sonal, Guarino R. Colli, Maggie R. Grundler, Gabriel C. Costa, Ivan Prates and Daniel L. Rabosky (2022). No link between population isolation and speciation rate in squamate reptiles. *Proceedings of the National Academy of Sciences* 119.4, e2113388119.

Smyčka, Jan, Anna Toszogyova and David Storch (2023). The relationship between geographic range size and rates of species diversification. *Nature Communications* 14, 1796.

Stadler, Tanja (2011). Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences* 108.15, 6187–6192.

— (2013). Recovering speciation and extinction dynamics based on phylogenies. *Journal of Evolutionary Biology* 26.6, 1203–1219.

Städler, Thomas, Ana Marcela Florez-Rueda and Margot Paris (2011). Testing for "snowballing" hybrid incompatibilities in *Solanum*: Impact of ancestral polymorphism and divergence estimates. *Molecular Biology and Evolution* 29.1, 31–34.

Stiller, Josefin, Graham Short, Healy Hamilton, Norah Saarman, Sarah Longo, Peter Wainwright, Greg W. Rouse and W. Brian Simison (2022). Phylogenomic analysis of Syngnathidae reveals novel relationships, origins of endemic diversity and variable diversification rates. *BMC Biology* 20.1, 75.

Thorpe, Roger S., Yann Surget-Groba and Helena Johansson (2010). Genetic tests for ecological and allopatric speciation in Anoles on an island archipelago. *PLoS Genetics* 6.4, e1000929.

Title, Pascal O. and Daniel L. Rabosky (2019). Tip rates, phylogenies and diversification: What are we estimating, and how good are the estimates? *Methods in Ecology and Evolution* 10.6, 821–834.

Title, Pascal O., Donald L. Swiderski and Miriam L. Zelditch (2022). EcoPhyloMapperR package for integrating geographical ranges, phylogeny and morphology. *Methods in Ecology and Evolution* 13.9, 1912–1922.

Title, Pascal O. et al. (2024). The macroevolutionary singularity of snakes. *Science* 383.6685, 918–923.

Title, Pascal O., L. Francisco Henao-Díaz, Rosana Zenil-Ferguson and Thais Vasconcelos (2025). An evolving view of lineage diversification. *Systematic Biology* 75.3, 391–405.

Turissini, David A., Joseph A. McGirr, Sonali S. Patel, Jean R. David and Daniel R. Matute (2017). The rate of evolution of postmating-prezygotic reproductive isolation in *Drosophila*. *Molecular Biology and Evolution* 35.2, 312–334.

Unckless, Robert L. and H. Allen Orr (2009). Dobzhansky–Muller incompatibilities and adaptation to a shared environment. *Heredity* 102.3, 214–217.

van der Niet, Timotheüs and Steven D. Johnson (2009). Patterns of plant speciation in the Cape floristic region. *Molecular Phylogenetics and Evolution* 51.1, 85–93.

Veron, Pierre and Jérémy Andréoletti (2025). *PBD_analog*. GitHub repository. URL: https://github.com/pierre-veron/PBD_analog.

Veron, Pierre, Jérémy Andréoletti, Tatiana Giraud and Hélène Morlon (2025). Speciation completion rates have limited impact on macroevolutionary diversification. *Philosophical Transactions of the Royal Society B: Biological Sciences* 380.1919, 20230317.

Veron, Pierre (2026). *HoleyAdaptSpeciation*. GitHub repository. URL: https://github.com/pierre-veron/HoleyAdaptSpeciation.

Veron, Pierre, Anaïs Spire, Agathe Chave-Lucas, Tatiana Giraud and Hélène Morlon (2026). *Rapid speciation in small populations challenges the dominance of ecological speciation*. bioRxiv: 2026.02.19.706750. Pre-published.

Virtanen, Pauli et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods* 17, 261–272.

Wang, Le, Zi Yi Wan, Huan Sein Lim and Gen Hua Yue (2016). Genetic variability, local selection and demographic history: genomic evidence of evolving towards allopatric speciation in Asian seabass. *Molecular Ecology* 25.15, 3605–3621.

Weeks, Brian C. and Santiago Claramunt (2014). Dispersal has inhibited avian diversification in Australasian archipelagoes. *Proceedings of the Royal Society B: Biological Sciences* 281.1791, 20141257.

Wells, Jeffrey V. and Milo E. Richmond (1995). Populations, metapopulations, and species populations: What are they and who should care? *Wildlife Society Bulletin* 23.3, 458–462.

Westram, Anja M., Sean Stankowski, Parvathy Surendranadh and Nicholas H. Barton (2022). What is reproductive isolation? *Journal of Evolutionary Biology* 35.9, 1143–1164.

Wiens, Delbert and Michèle R. Slaton (2012). The mechanism of background extinction. *Biological Journal of the Linnean Society* 105.2, 255–268.

Wiens, John J. and Michael J. Donoghue (2004). Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* 19.12, 639–644.

Wiens, John J. (2024). Review: Speciation across life and the origins of biodiversity patterns. *Evolutionary Journal of the Linnean Society* 3.1, kzae025.

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

Yamaguchi, Ryo and Yoh Iwasa (2016). Smallness of the number of incompatibility loci can facilitate parapatric speciation. *Journal of Theoretical Biology* 405, 36–45.

Yule, G. Udny (1925). A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society B: Biological Sciences* 213.3, 21–87.

Yusuf, Leeban H., Dominik R. Laetsch, Konrad Lohse and Michael G. Ritchie (2026). Genomic analyses in *Drosophila* do not support the classic allopatric model of speciation. *Evolution Letters* 10.2, 186–194.