



HAL
open science

**Dynamique éco-évolutive de la biodiversité :
Comprendre et prédire l'impact des fluctuations
environnementales à de multiples échelles
spatio-temporelles**

Elizabeth Barthelemy

► **To cite this version:**

Elizabeth Barthelemy. Dynamique éco-évolutive de la biodiversité : Comprendre et prédire l'impact des fluctuations environnementales à de multiples échelles spatio-temporelles. Biodiversité. Université Grenoble Alpes [2020-..], 2023. Français. NNT : 2023GRALV016 . tel-04144726

HAL Id: tel-04144726

<https://theses.hal.science/tel-04144726>

Submitted on 28 Jun 2023

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : CSV- Chimie et Sciences du Vivant

Spécialité : Biodiversité-Ecologie-Environnement

Unité de recherche : Laboratoire Interdisciplinaire de Physique

Dynamique éco-évolutive de la biodiversité: Comprendre et prédire l'impact des fluctuations environnementales à de multiples échelles spatio-temporelles

Eco-evolutionary biodiversity dynamics: understanding and predicting the influence of environmental heterogeneity at multiple temporal and spatial scales

Présentée par :

Elizabeth BARTHELEMY

Direction de thèse :

François MUNOZ

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Directeur de thèse

Rapporteurs :

Hélène MORLON

DIRECTRICE DE RECHERCHE, CNRS délégation Paris Centre

François MASSOL

DIRECTEUR DE RECHERCHE, CNRS délégation Hauts-de-France

Thèse soutenue publiquement le **7 mars 2023**, devant le jury composé de :

François MUNOZ

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Directeur de thèse

Philippe CHOLER

DIRECTEUR DE RECHERCHE, CNRS délégation Alpes

Examineur

Nicolas MOUQUET

DIRECTEUR DE RECHERCHE, CNRS délégation Occitanie Est

Examineur

Olivier FRANÇOIS

PROFESSEUR DES UNIVERSITES, Grenoble INP

Président

Hélène MORLON

DIRECTRICE DE RECHERCHE, CNRS délégation Paris Centre

Rapporteuse

François MASSOL

DIRECTEUR DE RECHERCHE, CNRS délégation Hauts-de-France

Rapporteur





Thèse effectuée au sein du **Laboratoire Interdisciplinaire de Physique**

Université Grenoble Alpes (UGA)
140 Rue de la Physique, Saint-Martin-d'Hères
38402 SAINT MARTIN d'HÈRES Cedex
France

*«The unpredictable and the predetermined unfold together to make everything the way it is.
It's how nature creates itself, on every scale, the snowflake and the snowstorm.»*

—Tom Stoppard, Arcadia, Act I - Scene IV

Remerciements

Pour l'honneur qu'ils me font en acceptant de juger ce travail, je tiens à remercier les membres du jury.

Je tiens à remercier mon directeur de thèse et mentor François Munoz, de m'avoir fait confiance dès le début. Merci d'avoir reconnu et façonné cette curiosité particulière propre à tout chercheur, de m'avoir guidé et m'avoir tant appris au long de ces cinq années à travailler ensemble.

Ce travail n'aurait pas été possible dans le soutien de l'Université Grenoble Alpes et de l'école doctorale CSV dont l'aide financière m'a permis de me consacrer sereinement à l'élaboration de cette thèse.

Je remercie également Claire Fortunel, qui m'a été d'un soutien inestimable, pour la gentillesse qu'elle m'a toujours témoigné autant comme collaboratrice que comme membre de mon comité de suivi.

Je souhaite également remercier Geoffroy Reboud et Ana Pichelin pour leurs contributions à ce travail.

Pour m'avoir accueilli à bras ouverts au sein de son laboratoire, pour sa bienveillance et sa disponibilité tout au long de ma thèse, je tiens à remercier Bahram Houchemandzadeh. Je remercie également toute l'équipe BIOP, à commencer par Hans Gielsman pour son accueil chaleureux, mais aussi Alexandre Dawid, Noël Scaramozzino et Stephan Lacour ainsi que tous les doctorants de l'équipe: Andréa Pavlou, Maike Sangster, Nassab Reda et Thibault Clavier.

Ce travail n'aurait pu être mené à bien sans la disponibilité et le soutien précieux que m'ont témoigné Michel Leboutin, Sandrine Nadau et Kristina Slavcheva. Je tiens également à remercier Mathilde Van Melle-Gateau, de m'avoir ouvert les portes de son bureau et pour tous les bons moments passés en sa compagnie.

Je remercie chaleureusement tous les doctorants et non-permanents du laboratoire, avec qui j'ai passé d'excellents moments, lors de soirées, de pauses déjeuner ou au détour d'un couloir. Merci à toutes les personnes qui ont fait de mon séjour au LIPhy une expérience riche et heureuse.

Au terme de ce parcours, je remercie enfin celles et ceux qui me sont chers et sans qui je n'aurais jamais eu la force de compléter ce travail.

Je souhaiterais remercier Mme Sylvie Brachet à qui je dois un dénouement heureux à cette thèse, merci de m'avoir guidé dans les moments les plus difficiles de cette expérience.

Je remercie également Marie Grange, d'avoir été ma grande sœur de thèse et pour son intarissable

Remerciements

bonne humeur.

Merci à Marie-Jo Diebolt et Aline Cissé, sans qui je ne me serais découvert une passion pour la peinture. L'atelier a été mon refuge lors de cette thèse, je sais que c'est grâce à votre amitié.

Je tiens à remercier mes amis Lily Baffrey et Olivier Lafont d'avoir été à mes côtés, et pour m'avoir appris qu'un trésor, comme leur amitié, se découvre par hasard.

Je souhaite également remercier Maxime Jaunatre d'avoir été le meilleur collaborateur avec lequel je puisse travailler, un ingénieur remarquable et un encore meilleur ami.

Merci à Marie Fontana, pour l'amitié précieuse qu'elle me porte et sans qui mon monde serait considérablement plus étroit.

Je tiens également à remercier ma meilleure amie, Manon Dzikowski, dont la constance dans ma vie est inestimable. Merci d'avoir toujours été là, même si c'était juste pour m'écouter râler.

Je remercie ma famille, et en particulier mes grands parents à qui je dois mes premiers souvenirs de la nature. Ma curiosité pour le vivant est née des moments passés avec eux. Et bien qu'il se peut que ma grand-mère ne comprenne pas grand chose à ce travail, il y a une pensée pour elle derrière chaque mot.

Un immense merci à mon frère et mes parents, dont l'amour et le soutien m'ont encouragé dans ma quête d'éducation, dans l'espoir qu'ils me pardonneront désormais de parler comme un chauffeur de poids lourd.

Enfin et surtout, tous mes remerciements à mon conjoint, Pierrick Morin, qui m'a cru capable de tout ce travail bien avant moi. Pour son soutien, sa patience et son amour quotidien qui n'a jamais failli malgré deux thèses (la sienne puis la mienne, non mais vraiment quelle idée!). C'est promis, il n'y en aura pas de troisième.

Liste des articles

Chapitre I. Développements algorithmiques et statistiques pour l'étude des dynamiques éco-évolutives

- Barthelemy, E., Jaunatre, M., and Munoz, F. “ecophylo: Simulating and assessing eco-evolutionary dynamics under past environmental changes in Python”. *submitted after revision to Plos Computational Biology*.
- Barthelemy, E., Denelle, P., Blanchard, G., Violle, C., and Munoz, F. “From biodiversity patterns to processes and back again: solving the inverse problem in ecology”.

Chapitre II. Influence des processus écologiques sur la diversité locale des communautés d'espèces

- Barthelemy, E., Gaba, S., Munoz, F. “Dispersal limitation and environmental filtering interplay in shaping the assembly of weed plant communities”. *in prep*
- Reboud, G., Barthelemy, E., Munoz, F. “The spatial occupancy and functional similarity of neighboring trees determine tree growth in a tropical rainforest”. *in prep*

Chapitre III. Influence des fluctuations environnementales passées sur la diversité régionale actuelle d'espèces

- Barthelemy, E., Fortunel, C., Jaunatre, M., Munoz, F. “Imprints of past habitat area reduction on extant taxonomic, functional and phylogenetic composition”. *Frontiers in Ecology and Evolution*. Vol. 9, (2021), DOI: 10.3389/fevo.2021.634413.
- Barthelemy, E., Pichelin, A., Munoz, F. “Shapes of regional trait distributions resulting from the interplay between phylogenetic imbalance and the pace of trait evolution”. *submitted to Journal of Theoretical Ecology*
- Barthelemy, E., Ranaivosoa, R. M. O., Zoelinirina, N. K., Munoz, F. “Diversification and divergence of rainforest woody plants in South India and Madagascar relate to geomorphological history”. (2022), DOI: 10.1111/jbi.14348.

Table des matières

Remerciements	v
Liste des articles	vii
Table des matières	ix
Liste des figures	xiii
Introduction	3
1 Le lien entre patrons et processus comme point de convergence des sciences de la biodiversité	3
1.1 Comprendre le vivant	3
1.2 Le problème inverse ou la question de l'inférence	11
2 De multiples processus impliquent de multiples échelles	12
2.1 Structuration de la biodiversité dans l'espace	12
2.2 Diversification de la biodiversité dans le temps	19
3 Vers l'intégration des processus à de multiples échelles spatio- temporelles dans un cadre unifié	22
3.1 La théorie neutre comme première intention	23
3.2 Limites de la théorie neutre	24
3.3 Développer de nouveaux modèles éco-évolutifs hérités de la théorie neutre	24
4 Comprendre et prédire l'impact de fluctuations environnementales à de multiples échelles spatiales et temporelles	27
Références	29
I Développements algorithmiques et statistiques pour l'étude des dynamiques éco-évolutives	35
Résumé du chapitre	37
1 ecophylo: Simulating and assessing eco-evolutionary dynamics under past environmental changes in Python.	39
1 Introduction	40
2 Design and Implementation	41
2.1 Coalescent-based simulation of assemblage genealogies	41
2.2 Phylogenetic reconstruction and species abundances	43

3	Results	43
3.1	Examining the effect of past demographic fluctuations	45
3.2	Examining the effect of the pace of evolution	45
3.3	Inferring eco-evolutionary mechanisms	46
4	Availability and future directions	48
4.1	Modeling out-of-equilibrium diversification dynamics in a fluctuating environment	48
4.2	Inference of eco-evolutionary dynamics from observed biodiversity patterns	49
4.3	Limitations and future perspectives	50
	References	51
1.A	Python scripts	55
2	From biodiversity patterns to processes and back again: solving the inverse problem in ecology	59
1	Introduction – solving the inverse problem in ecology	60
2	The curse of conditionality	62
3	Drawing more robust conclusions from virtual experiments	63
4	All models are wrong, all of which are useful	64
5	Resolving the conundrum of neutral and non-neutral processes	64
6	Conclusion	66
	References	67
II	Influence des processus écologiques sur la diversité locale des communautés d'espèces	71
	Résumé du chapitre	73
3	Dispersal limitation and environmental filtering interplay in shaping the assembly of weed plant communities	75
1	Introduction	76
2	Materials and Methods	77
2.1	Study site and sampling	77
2.2	Functional traits	77
2.3	Delimitating local weed communities and defining the regional species pool	78
2.4	Choice of summary statistics	78
2.5	Estimating parameters of neutral and niche-based dynamics in weed communities	78
2.6	Model misclassification and cross validation	79
3	Results	80
3.1	Selecting a model of weed community assembly	80
3.2	Parameters of neutral and niche-based assembly in weed communities	81

4	Discussion	81
	References	86
3.A	Supplementary Figures	89
4	The spatial occupancy and functional similarity of neighboring trees determine tree growth in a tropical rainforest	91
1	Introduction	92
2	Material and Methods	94
	2.1 Study site and sampling	94
	2.2 Neighborhood Crowding Index (NCI)	94
	2.3 Bayesian multilevel model implementation	94
3	Results	95
4	Discussion	97
	References	98
III	Influence des fluctuations environnementales passées sur la diversité régionale actuelle d'espèces	101
	Résumé du chapitre	103
5	Imprints of past habitat area reduction on extant taxonomic, functional and phylogenetic composition	105
1	Introduction	106
2	Material and methods	108
	2.1 Simulating phylogenies and trait values in extant assemblages	108
	2.2 Characterizing extant taxonomic, functional and phylogenetic diversity	109
3	Results	112
	3.1 Taxonomic diversity	112
	3.2 Phylogenetic diversity	113
	3.3 Functional composition	114
4	Discussion	114
	4.1 Relative species abundances keep track of (more) recent habitat changes	115
	4.2 Phylogenetic diversity keeps track of (more) ancient habitat changes	116
	4.3 Functional composition reflects the interplay between trait evolution and past diversification in phylogenies	116
	4.4 Process-based eco-evolutionary modelling	117
	4.5 Perspectives - From a regional pool to local communities: integrating the influence of assembly processes on local community composition	118
	References	119
5.A	Supplementary Figures	123
6	Shapes of regional trait distributions resulting from the interplay between phylogenetic imbalance and the pace of trait evolution	125
1	Introduction	126

2	Material and methods	127
2.1	Simulating phylogenies and characterizing their imbalance	127
2.2	Simulating functional trait distributions	128
3	Results	128
4	Discussion	130
4.1	Evolutionary legacies in extant functional composition	130
4.2	Lessons for community ecology	131
4.3	Limits and Perspectives	131
	References	132
7	Diversification and divergence of rainforest woody plants in South India and Madagascar relate to geomorphological history	135
1	Introduction	136
2	Material and Methods	137
2.1	Indian and Malagasy tropical rainforest data	137
2.2	Phylogenetic diversity per period	138
2.3	Null model-based testing	139
2.4	Contribution of taxa to phylogenetic diversity per period	139
3	Results	140
4	Discussion	141
	References	143
7.A	Supplementary Figures	146
7.B	Supplementary Tables	147
7.C	R Scripts	147
	Discussion	153
1	Intégration des processus éco-évolutifs à de multiples échelles	153
1.1	<i>ecophylo</i> : héritier de la théorie neutre?	153
1.2	Perspectives de modélisation futures	156
2	Le lien entre processus et patrons	159
2.1	Quelles approches en écologie et biogéographie ?	159
2.2	Simulation pour l'inférence: une nouvelle approche?	160
2.3	Le cas des dynamiques éco-évolutives: quels patrons de diversité pour l'inférence?	164
3	Conclusion	165
	Références	166
	Annexes	172

Liste des figures

INTRODUCTION	2
I.1 Hiérarchie des différents niveaux d'organisation auxquels correspondent différentes disciplines en écologie	4
I.2 Diversité α , β et γ	5
I.3 Distribution d'abondance (gauche) ou de rang-abondance (droite) des espèces d'arbres tropicaux (DBH>10cm) recensées à Barro Colorado Island (Panama)	6
I.4 La première esquisse de Darwin d'un arbre phylogénétique tirée de son First Notebook on Transmutation of Species (1837)	7
I.5 Un chronogramme des lépidoptères. Dans ce type d'arbre phylogénétique, la longueur des branches est proportionnelle au temps géologique.	8
I.6 Dynamique de patches de métacommunautés permettant une coexistence régionale entre une espèce compétitrice (A) et une espèces colonisatrice (B) par le biais d'un compromis entre compétition et colonisation permis par la migration	10
I.7 Illustration traditionnelle de la série de filtres hiérarchiques qui réduisent l'ensemble des membres du pool régional aux membres d'une communauté locale.	13
I.8 Répartition de la végétation du Mont Chimborazo en fonction de l'altitude. Cette carte constitue la pièce maîtresse de l'ouvrage fondamental de Humboldt: "Essai sur la géographie des plantes", publié en 1807, et représente également le début de la biogéographie (Bonpland 1807.	16
I.9 Dessin original de la ligne de Wallace montrant la division entre les flores et les faunes asiatiques et australasiennes.	17
I.10 <i>Carex maritima</i> . Zermatt, Riffelsee (Kanton Wallis, Suisse), 2700m	22
I.11 Sur les propriétés du coalescent	26
CHAPITRE I - Développements algorithmiques et statistiques pour l'étude des dynamiques éco-évolutives	36
Article 1. <i>ecophylo</i>: Simulating and assessing eco-evolutionary dynamics under past environmental changes in Python	38
1.1 Conceptual diagram illustrating the primary components and associated parameters to the eco-evolutionary model <i>ecophylo</i>	42
1.2 Variation in simulated taxonomic diversity	44
1.3 Variation in the simulated number of species with regard to the point mutation rate (μ), protractedness (τ) and the size of the assemblage (N_e). The dashed line represents the total species richness while color indicate the number of species with a given number of individuals (abundance class).	46

1.4	Cross-validation analysis with dimension reduction of simulated communities using the constant size model (A,C) and that with a change in size occurring 30,000 generations ago (B,D) based on either species richness, Shannon's and Simpson's diversity indexes and the number of species by octave of abundance class (A,B) or the same statistics transformed into the PCAIV components trained on the 100 closest simulations (C,D). Vertical dashed lines represent the misprediction errors for each of the 100 considered pseudo-observations and given a tolerance threshold of 1%.	47
-----	--	----

CHAPITRE II - Influence des processus écologiques sur la diversité locale des communautés d'espèces **72**

Article 3. Dispersal limitation and environmental filtering interplay in shaping the assembly of weed plant communities **75**

3.1	Model selection. Proportion of communities that were assigned to either a neutral model of community assembly (light gray) versus a model with additional stabilizing environmental filtering (dark gray) depending on the considered trait, on the field crop and on whether communities were located in the field center or at the margin.	80
3.2	Mean posterior migration rate m of the weed assembly model (with additional environmental filtering) depending on different traits (Height, Specific Leaf Area, and Ellenberg F, L and N) and according to the type to the type of crop and position relative to the field: either in the center or at the field margin (where there is no tillage)	82
3.3	Mean posterior optimal trait value (t_{opt}) of the environmental filtering component of the weed assembly model depending on different traits (Height, Specific Leaf Area, and Ellenberg F, L and N) and according to the type to the type of crop and position relative to the field: either in the center or at the field margin (where there is no tillage)	83
3.4	Mean posterior σ of the environmental filtering component of the weed assembly model depending on different traits (Height, Specific Leaf Area, and Ellenberg F, L and N) and according to the type to the type of crop and position relative to the field: either in the center or at the field margin (where there is no tillage)	84
3.S1	Misclassification analysis. For each trait, and given each crop type and relative position in the field, dark grey bars represent the proportion of weed communities for which the environmental filtering assembly model is deemed satisfactory (i.e. correctly identified in more than 70% of cases).	89

Article 4. The spatial occupancy and functional similarity of neighboring trees determine tree growth in a tropical rainforest **90**

4.1	Diagram of the different tropical tree strata and associated constraints in terms of light availability and crowding	93
-----	--	----

4.2	Principal Component Analysis (PCA) on tropical tree wood and leaf traits (SLA, LA, LDMC and WD): correlation circle showing the contributions of tree traits two first PCA axes used to define LES and WES respectively	95
4.3	Relationships between tree growth, estimated from the best model and the neighborhood crowding index (NCI), the wood economic spectrum (WES) and leaf economic spectrum (LES) for the emergent, canopy or sub-canopy strata	96
4.4	Relationships between tree growth, estimated from the best model and the diameter at breast height (DBH) and the Neighborhood Crowding Index (NCI) in the emergent stratum in interaction with the leaf economic spectrum (LES) in the canopy stratum	97

CHAPITRE III - Influence des fluctuations environnementales passées sur la diversité régionale actuelle d'espèces **102**

Article 5. Imprints of past habitat area reduction on extant taxonomic, functional and phylogenetic composition **105**

5.1	Expected patterns of taxonomic, phylogenetic, and trait diversity under different scenarios of habitat expansion following a period with restricted refugia	107
5.2	Eco-evolutionary dynamics of individual organisms in a habitat that experienced reduction in the past	110
5.3	Variation in simulated species richness depending on the amplitude of expansion (A) and the time (in generations) at which it occurred (t_c)	112
5.4	Principal component analysis (PCA) on the simulated species frequency spectra (SFS)	112
5.5	Variation in spectral density profile summary statistics of rerooted phylogenies (vis a vis oldest simulated MRCA) depending on the amplitude of expansion and the time (in generations) at which it occurred	113
5.6	Variation in median trait moments (mean and variance) as well as their interquartile ranges (IQR) depending on the amplitude of expansion (A)	114
5.7	Variation in median trait moments (mean and variance) as well as their interquartile ranges (IQR) along the ln-transformed principal eigenvalue (λ^*) (A) and skewness (ψ) (B), summarizing the variation in simulated spectral density profiles according to the parameters used to simulate trait evolution (β in rows and σ in columns in each individual panel)	115
	Figures Supplémentaires	123
5.S1	Variation in residuals in linear models linking phylogenetic diversity and species richness depending on the amplitude of expansion (A) and the time (in generations) at which it occurred (t_c)	123
5.S2	Variation in simulated metrics of phylogenetic diversity (A) Faith's phylogenetic diversity (PD), (B) mean pairwise distance (MPD), (C) most recent common ancestor (MRCA), and (D) mean nearest taxon distance (MNTD) depending on the amplitude of expansion (A) and the time (in generations) at which it occurred (t_c)	123

5.S3	Variation in simulated (A) species richness, (B) phylogenetic diversity (PD), (C) mean pairwise distance (MPD), (D) age of the root (in generations before present), and (E) mean nearest taxon distance (MNTD) depending on the spectral density profile ln-transformed principal eigenvalue (λ^*) and according to the time at which the expansion occurred (in generations before present)	124
Article 6. Shapes of regional trait distributions resulting from the interplay between phylogenetic imbalance and the pace of trait evolution		125
6.1	Variation of asymmetry of simulated topologies depending on different values of beta	127
6.2	Summary of variation in functional composition depending on phylogenetic tree shape and trait evolution models	129
Article 7. Diversification and divergence of rainforest woody plants in South India and Madagascar relate to geomorphological history		134
7.1	Summary of the fragmentation of Gondwana that led to Madagascar's gradual geographic isolation from India and associated biogeographic events	136
7.2	Computation of the Lineages Per Period (LPP) metric of phylogenetic diversity through time	138
7.3	Observed shared Lineage Per Period (sLPP) and unique Lineage Per Period values (α LPP) for each time period defined in Figure 1 (A) and associated effect sizes (B) in Indian and Malagasy rainforests' woody plants	140
7.4	Principal component analysis of family contributions to α LPP in India (A), Madagascar (B) and to sLPP (C)	141
	Figures Supplémentaires	146
7.S1	Distribution of forest plots available in Madagascar (A) and in Western Ghats, India (B). Plots in Madagascar cover all kinds of forests (circles), among which we will select plots occurring in rainforest context (triangles). Projection: WGS84 - World Geodetic System 1984.	146
7.S2	Contribution of families to the sLPP and α LPP values for Indian and Malagasy rainforests' woody plants. Panels represent the different metrics of phylogenetic diversity per period	146
DISCUSSION		152
D.1	Proposition d'un modèle de spéciation par différenciation fonctionnelle qui pourrait être implémenté dans <i>ecophylo</i>	158

Introduction

Introduction

Identifier les processus écologiques à l'origine de patrons de biodiversité est un objectif central de l'écologie. Les processus impliqués opèrent conjointement à une échelle écologique, en lien avec des facteurs et des dynamiques à court pas de temps opérant au niveau local, et à une échelle évolutive, sur le long terme et à un niveau régional. Cependant les deux échelles sont généralement traitées par des disciplines différentes. Comprendre comment les processus écologiques façonnent la biodiversité sur le long terme implique cependant une intégration des échelles, dans une approche dite éco-évolutive. Un point de départ pour une telle intégration a été proposé par la théorie neutre (Hubbell 2001), telle que les processus régionaux de spéciation-dérive sont couplés aux processus locaux de migration-dérive. Cette approche demeure neutre dans le sens où les espèces d'un assemblage sont équivalentes fonctionnellement, et considère un environnement de taille constante. Les fluctuations de l'environnement ont cependant marqué profondément la dynamique et la diversification des espèces sur le long terme. Un objectif central de la thèse est de comprendre comment de telles fluctuations affectent les dynamiques de spéciation, migration et de dérive et façonnent les diversités taxonomique, fonctionnelle et phylogénétique. Outre des développements méthodologiques nouveaux pour construire et analyser les dynamiques éco-évolutives, la thèse questionnera la démarche scientifique à mettre en œuvre pour évaluer de telles dynamiques. Elle traitera la question sous l'angle du problème direct, concernant la prédictions de patrons à partir des processus élémentaires, et du problème inverse, concernant l'inférence des processus à partir de patrons observés.

1 Le lien entre patrons et processus comme point de convergence des sciences de la biodiversité

1.1 Comprendre le vivant

Là où d'autres disciplines scientifiques peuvent s'intéresser à des objets très éloignés de l'expérience humaine, l'écologie est une discipline dont le sujet d'étude fait intégralement partie de notre quotidien. Nos premières expériences de l'écologie passent par l'observation de la nature qui nous entoure : des forêts, aux rivières, en passant par les abeilles et les oiseaux. Je pense qu'il y a à l'origine de tout biologiste, un souvenir d'une rencontre directe avec la nature, qui détermine aujourd'hui ce qu'il étudie en particulier. Cette familiarité que nous entretenons avec la nature ne fait pas pour autant de l'écologie une activité facile. La simple description de l'immensité de la biodiversité sur Terre est à elle seule une tâche ardue. Par ailleurs, les paradigmes que nous utilisons pour décrire et comprendre les systèmes naturels sont les nôtres, et il serait maladroit de croire que nous observons la vraie essence de la nature car elle nous est si familière (Descola 2010). Nous avons un ressenti, que nous formalisons en observations, auxquelles nous essayons de donner un sens.

L'écologie trouve ses origines dans l'histoire naturelle. Pendant une grande partie du Moyen Age, une grande attention est portée à décrire le vivant et à classifier sa variation. Cependant les préoccupations religieuses ont souvent empêché l'examen des relations de cause à effet allant au-delà des explications bibliques (Jacob 1970). Puis, les grandes expéditions du XVIIIe siècle bouleversent la vision eurocentrique et biblique de la nature. De nombreux naturalistes étaient des voyageurs. C'est leur rencontre avec des organismes et des paysages inconnus lors de leurs voyages qui leur ont permis de formuler des théories pour expliquer ces observations. Ainsi, les travaux de certains naturalistes, tels que Buffon à la fin du XVIIIe siècle, initient un changement de l'histoire naturelle; de la description de la nature en une véritable compréhension historique de la nature en tant que processus dynamique. Mais la recherche en écologie commence véritablement au XXe siècle lorsque des naturalistes comme Stephen Forbes et Eugen Warming s'interrogent sur les relations plus complexes entre les organismes, les populations et/ou espèces auxquelles ils appartiennent et leurs interactions avec l'environnement. En 1927, Charles Elton annonce dans son livre *Animal Ecology*: "*L'écologie est un nouveau nom pour un sujet très ancien. Cela signifie simplement histoire naturelle scientifique*" (Elton 2001). Ainsi, l'écologie bascule progressivement d'une discipline dont les questions étaient motivées par la description exhaustive des organismes, vers la compréhension de concepts de plus en plus abstraits ; tels que la niche, les réseaux alimentaires, les écosystèmes et, plus récemment, le lien avec les besoins et problèmes sociétaux, à travers les services écosystémiques.

Aux différentes disciplines en écologie correspondent différents niveaux d'observation de la biodiversité: biosphère, biome, paysage, écosystème, communauté, population, organisme et cellule (Figure I.1 Odum and Barrett 1971). Nous disposons aujourd'hui de plus en plus d'outils nous permettant d'observer le vivant à ces différentes échelles d'organisation. L'écologie du paysage, par exemple, a beaucoup bénéficié d'avancées technologiques comme la télédétection. Les niveaux d'organisation de cette classification sont imbriqués, c'est-à-dire que chaque niveau est constitué de groupes d'unités de niveau inférieur : les communautés sont constituées de groupes de populations qui sont elles-mêmes constituées d'individus etc. Ainsi, une communauté d'espèces peut être définie comme un ensemble d'organismes ou populations qui "*vivent dans un environnement et interagissent les uns avec les autres, formant ensemble un système vivant distinctif avec sa propre composition, structure, relations environnementales, développement et fonction.*" (Whittaker 1975). La plupart des questionnements en écologie des communautés visent à comprendre comment des espèces coexistent de façon stable, et comment celles-ci se répartissent en fonction de différents facteurs environnementaux ou historiques.

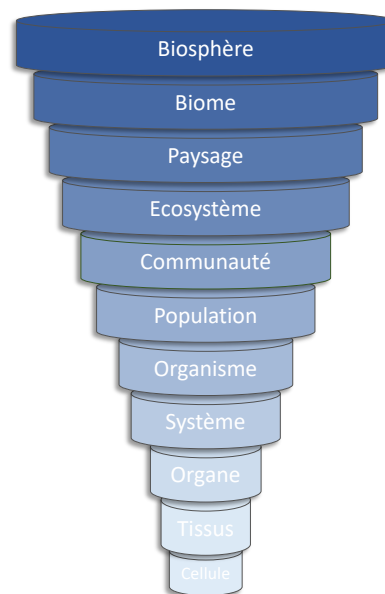


Figure I.1: Hiérarchie des différents niveaux d'organisation auxquels correspondent différentes disciplines en écologie. D'après Odum and Barrett 1971

1.1.1 De la description des patrons

Diversité taxonomique Les communautés d'espèces peuvent être classées de différentes manières, et bien que certaines soient plus faciles à délimiter que d'autres, la plupart des approches consistent à décrire divers aspects du nombre et de l'identité des espèces présentes dans une localité. Localité qu'il s'agira également de définir, là encore avec plus ou moins de difficulté. En effet, là où certains habitats forment des ensembles spatialement discrets, comme un lac par exemple, d'autres ne présentent pas de délimitation spatiale nette. Par exemple, les forêts peuvent se transformer de manière relativement imperceptible en savanes, puis en prairies, sans qu'il n'y ait de frontières claires. Toujours est-il qu'une fois cette localité définie, la première chose à faire est bien souvent d'énumérer les organismes en place, de les identifier, afin d'établir la liste d'espèces auxquels ils appartiennent, et en quelle proportion. On peut ainsi caractériser les communautés d'espèces assez instinctivement par le nombre d'espèces observées. Ce nombre, souvent appelé richesse spécifique, est une des notions les plus fondamentales de l'écologie. Mais en pratique, la richesse spécifique reste une mesure de la biodiversité assez difficile à obtenir ; d'une part, car elle nécessite de connaître les informations taxonomiques complètes de nombreux groupes d'organismes, et d'autre part, car se pose le problème pratique de savoir quel effort d'échantillonnage permet d'affirmer que toutes les espèces d'un même lieu aient été observées au moins une fois. Ceci pose évidemment problème pour les espèces rares - espèces qui selon les écosystèmes peuvent jouer un rôle particulier, voire être particulièrement fréquentes, comme c'est le cas en forêt tropicale par exemple (Hubbell 2001). Bien souvent, on se limite à évaluer la richesse de groupes taxonomiques bien identifiés, facilement échantillonnables et selon une unité d'effort répétée. Notons que les approches dites par raréfaction

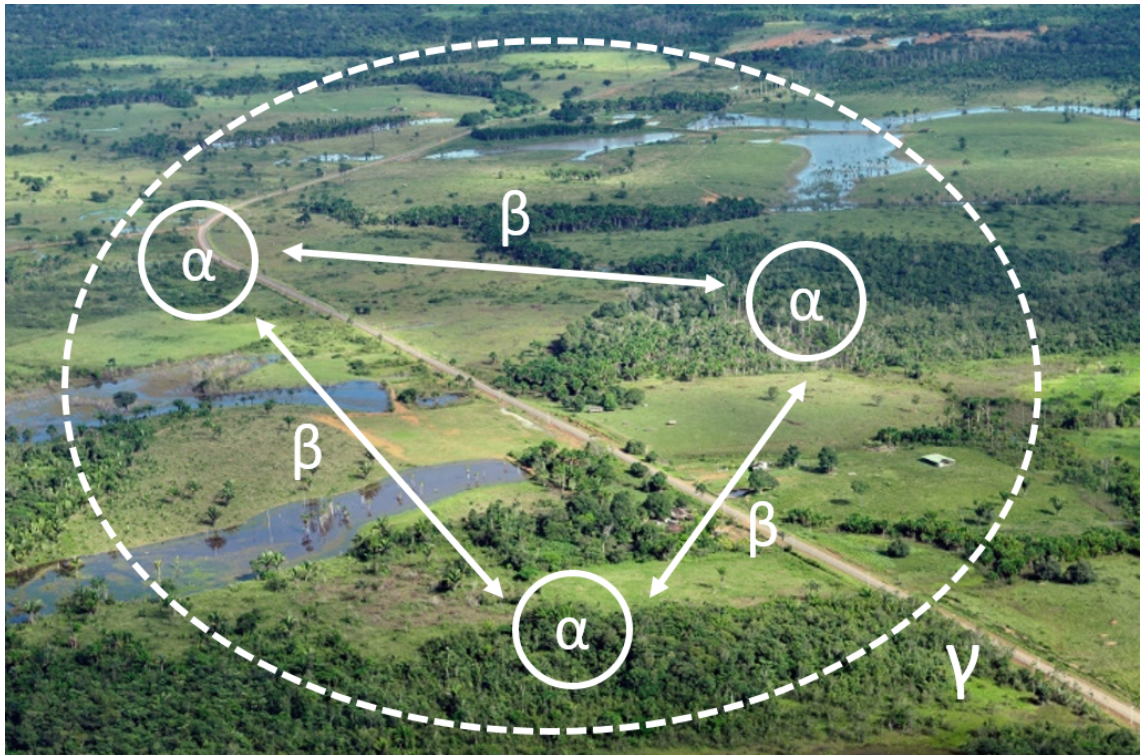


Figure I.2: Diversité α , β et γ

permettent de comparer les richesses de différentes communautés mises à l'échelle de la même quantité d'effort d'échantillonnage (Gotelli and Colwell 2001). Ainsi la richesse spécifique constitue un élément important permettant la comparaison entre communautés. En revanche, elle ne dit rien sur la fréquence relative des espèces. Différents indices existent pour rendre compte de façon synthétique des différences d'abondance entre communautés, tel que l'indice de Shannon ou de Simpson, qui permettent de prendre en compte la régularité de la distribution des abondances des espèces d'une communauté. La diversité contenue dans un seul habitat local est appelée diversité α (Whittaker 1975). A l'échelle d'une même région géographique, le changement de composition en espèces d'une localité à une autre constitue une facette supplémentaire de diversité, que l'on appelle diversité β (Figure I.2). La diversité globale observée pour un ensemble d'habitat qui compose la région est appelée diversité γ , et se compose de la somme des diversités α moyennes (dans le cas de l'indice de Simpson) et de la diversité β .

Par ailleurs, l'étude des profils de distribution des abondances relatives des espèces est une longue tradition en écologie des communautés. On représente, par exemple, les abondances relatives des espèces d'une communauté en fonction du rang qu'elles occupent : de la plus abondante à la moins abondante (Figure I.3). La similarité des formes de ce type de distribution dites de "rang-abondance", dans des contextes environnementaux différents, a motivé la recherche sur leur signification écologique. La question était alors d'obtenir des distributions similaires par des modèles statistiques sous-jacents, plus ou moins justifiés par des hypothèses biologiques sur la façon dont les espèces interagissent et aboutissent par former des communautés. Parmi les relations espèces-abondance les plus importantes ont compte la distribution en bâton rompu (ou "broken-stick"), la série géométrique et la distribution log-normale (Whittaker 1975 ; May 1975). Chaque distribution peut être dérivée en faisant des hypothèses particulières sur la façon dont les espèces se répartissent les ressources au sein d'une communauté. Par exemple, la série géométrique peut être obtenue en supposant qu'une espèce dominante représente une certaine fraction, k , du nombre total d'individus dans un échantillon, et que chaque espèce successivement moins abondante représente une fraction k du nombre restant d'individus. Bien que plus difficile à interpréter, ce type de représentation fournit une information plus complète que des mesures agrégées de diversité. Mais dans tous les cas, un élément important de la diversité des communautés reste absent: celle de la particularité des attributs des espèces.

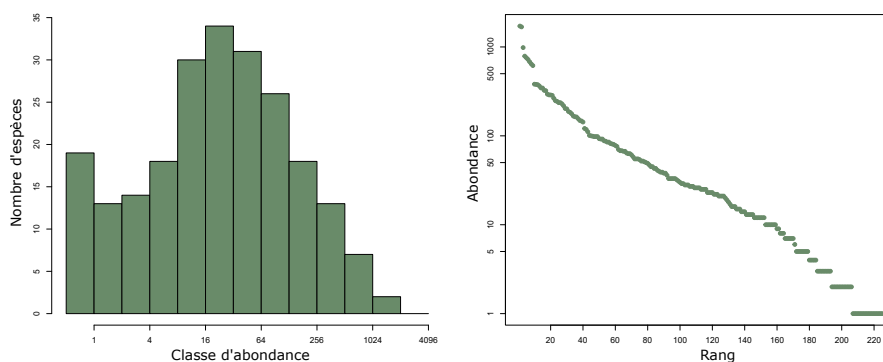


Figure I.3: Distribution d'abondance (gauche) ou de rang-abondance (droite) des espèces d'arbres tropicaux (DBH>10cm) recensées à Barro Colorado Island (Panama). DBH = Diamètre à hauteur de poitrine.

Diversité fonctionnelle Initialement la description de la biodiversité se concentrait donc sur l'énumération des espèces et sur la description de leurs abondances. Cependant on peut également s'attacher à décrire les propriétés des individus constituant ces communautés d'espèces. Ce basculement, de la description des *constituants* des communautés, à la description de leurs *propriétés*, est motivé par l'idée que les attributs biologiques et écologiques, plus que l'identité taxonomique des espèces, renseignent sur leurs adaptations (Lavorel and Garnier 2002). En écologie fonctionnelle, on s'attache à décrire les traits fonctionnels des espèces: c'est à dire, des caractéristiques morphologiques, physiologiques ou phénologiques, mesurables au niveau d'un individu, qui ont un impact sur sa performance individuelle par leurs effets sur la croissance, la reproduction ou la survie (Violle, Navas, et al. 2007). On peut ainsi caractériser une communauté par la distribution des valeurs des traits des individus qui la composent (Schleuter et al. 2010). Cette distribution peut être décrite à l'aide d'outils statistiques classiques comme sa moyenne et sa variance, ou même par des moments d'ordre supérieurs tels que la skewness ou la kurtosis (B. Enquist et al. 2015). Ces différents indices de diversité fonctionnelle permettent de caractériser la diversité des attributs biologiques des individus au sein des communautés ou encore de comparer différentes communautés. Là encore on peut distinguer la diversité fonctionnelle α des communautés, de la diversité fonctionnelle β entre communautés (Swenson 2011). Cependant, l'utilisation des traits des individus pour décrire la diversité de la communauté qu'ils composent sous-tend plusieurs contraintes. D'une part, il est préférable que la variabilité naturelle entre les valeurs de traits des individus d'une même espèce soit inférieure à celle qui existe entre les membres d'espèces différentes, surtout si l'on utilise une seule valeur de trait par espèce (Violle, B. J. Enquist, et al. 2012). Bien souvent, l'effort important que demanderait la mesure sur le terrain, de plusieurs valeurs de traits, pour tous les individus de chaque espèce présente justifie l'utilisation de valeurs moyennes par espèce, voire de valeurs mesurées dans d'autres contextes et répertoriées dans des bases de données. D'autre part, selon les indices de diversité fonctionnelle utilisés, plus le nombre d'espèces échantillonnées est grand, plus il y a de chance que la diversité fonctionnelle augmente. Enfin, les traits des espèces sont des propriétés héréditaires, qui sont le fruit de l'évolution des espèces au cours du temps. Ainsi une partie de la similarité des propriétés des espèces peut provenir d'une histoire évolutive similaire.

Diversité phylogénétique Darwin en 1859 fait la remarque suivante: *“Comme les espèces d'un même genre présentent habituellement, mais non invariablement, une certaine similitude dans leurs habitudes et leur constitution, et toujours dans leur structure, la compétition sera généralement plus sévère entre les espèces du même genre, lorsqu'elles entrent en concurrence les unes avec les autres, qu'entre les espèces de genres distincts. Nous pouvons vaguement voir pourquoi la compétition devrait être plus sévère entre des formes alliées, qui occupent presque la même place dans l'économie de la nature ; mais probablement dans aucun cas nous ne pourrions dire précisément pourquoi une espèce a été victorieuse sur une autre dans la grande bataille de la vie.”* Ce passage de l'Origine des espèces (Darwin 2011) souligne que des espèces étroitement

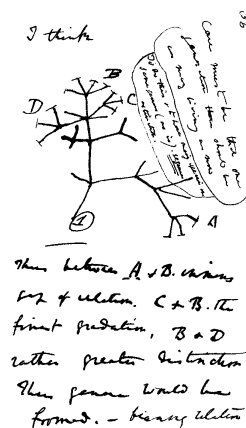


Figure I.4: La première esquisse de Darwin d'un arbre phylogénétique tirée de son First Notebook on Transmutation of Species (1837)

apparentées devraient être plus semblables les unes aux autres que les espèces plus éloignées, en raison d'un ancêtre commun partagé plus récemment. Ainsi, à la diversité basée sur l'énumération des espèces, à la description de leurs propriétés, on peut ajouter la mesure de la diversité phylogénétique visant à mesurer la part d'histoire évolutive partagée entre espèces d'une communauté.

La plupart des mesures de la diversité phylogénétique se basent sur des objets appelés phylogénies - ou arbres phylogénétiques - qui représentent des relations évolutives entre un groupe d'organismes (actuels ou éteints). Ces relations sont représentées sous la forme d'un graphe acyclique orienté composés d'arêtes, plus communément appelées "branches", reliant une série de nœuds internes et terminaux. Le nœud interne le plus ancien de la phylogénie, à partir duquel toutes les lignées descendent, est appelé "racine". La longueur des branches d'une phylogénie peut représenter un changement évolutif exprimé par une échelle de distance génétique ou temporelle (Figure I.5). Parmi les premières mesures de la diversité phylogénétique on retiendra celle de Faith, qui s'obtient en additionnant toutes les longueurs des branches reliant les espèces d'une communauté (Faith 1992). Il convient de noter d'une part que cette mesure de la diversité phylogénétique augmente lorsque le nombre d'espèces échantillonnées augmente. D'autre part, la mesure de diversité phylogénétique de Faith ne tient pas compte des différences d'abondances des espèces dans une communauté. Notons qu'il existe une version de la mesure originale de Faith pondérée par l'abondance (Barker 2002), de telle sorte que l'ajout de nouvelles espèces abondantes et anciennes maximise la diversité par rapport à des espèces rares et récentes. Encore une fois, en plus de la diversité alpha, on peut chercher à caractériser la diversité beta phylogénétique entre deux communautés par exemple en calculant l'UniFrac ("unique fraction") et pour laquelle il existe également un version pondérée par les abondances (Lozupone and Knight 2005). Il existe également d'autre mesure de la diversité

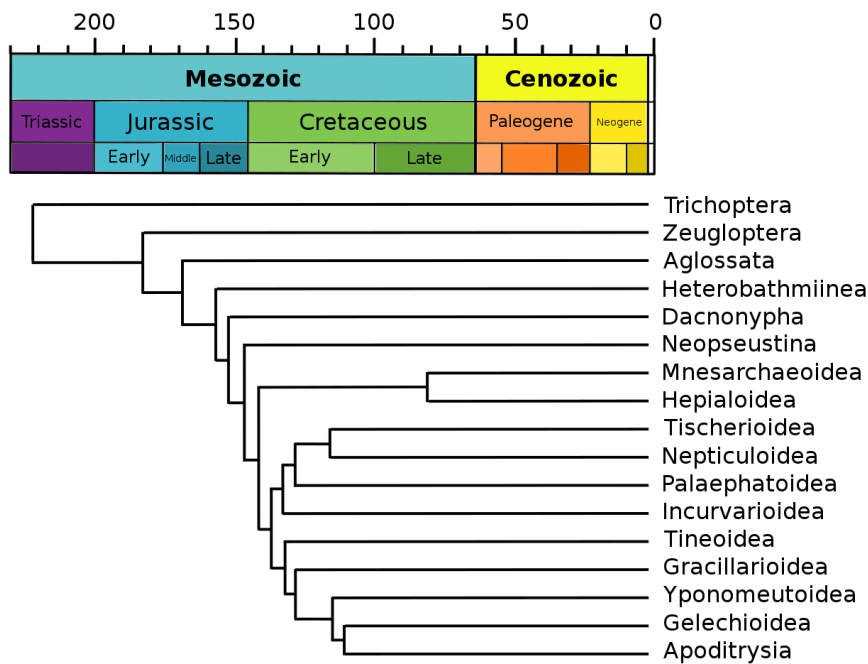


Figure I.5: Un chronogramme des lépidoptères. Dans ce type d'arbre phylogénétique, la longueur des branches est proportionnelle au temps géologique.

phylogénétique basées cette fois ci sur la matrice de distance phylogénétique ($\delta_{i,j}$). Celle-ci correspond a une matrice dont les éléments sur la diagonale sont nuls et reflètent une distance nulle entre même espèces, et les éléments hors diagonale reflètent les longueurs des branches phylogénétiques séparant deux espèces différentes. On distingue deux types de métriques: les métriques par paire (ou pairwise) et les métriques du plus proche voisin (nearest neighbor) parmi lesquelles les plus connues reste le MPD (Mean-Pairwise Distance) et le MNTD (Mean Nearest Taxon Distance), respectivement (Mazel et al. 2016).

1.1.2 A la compréhension des processus

La description des patrons de biodiversité devient particulièrement intéressante lorsque des patrons généraux émergent dans des contextes écologiques différents. Par exemple, les communautés d'espèces contiennent dans de nombreux biomes plus d'espèces rares que d'espèces abondantes (McGill et al. 2007) et ce exagérément en forêt tropicale. De la même façon qu'il existe une loi des gaz parfaits en physique, on pourrait se demander quelles sont les lois en écologie qui expliquent ces patrons répétés. Ainsi, au-delà de la description de la biodiversité, se pose plus fondamentalement la question des mécanismes à l'origine de ces observations. Comprendre comment s'organise le vivant ne passe donc pas seulement par sa description, mais par la compréhension des mécanismes élémentaires qui aboutissent à sa structuration dans l'espace et le temps. Fondamentalement, un individu membre d'une communauté peut croître, migrer, se reproduire ou mourir. La probabilité de chaque événement peut dépendre de l'espèce à laquelle l'individu appartient, de ses propriétés biologiques, et du contexte environnemental dans lequel il se place. La somme totale de ces mécanismes élémentaires pour chaque individu d'une communauté détermine la dynamique de celle-ci et s'exprime en termes de patrons observés. On peut regrouper l'ensemble de ces processus qui déterminent la structure des communautés d'espèces en deux catégories: ceux qui ajoutent de nouvelles espèces à la communauté et ceux qui en retirent. Des espèces peuvent s'ajouter à la communauté par spéciation ou immigration. Puis, au cours du temps, la sélection et la dérive façonnent les abondances relative des espèces de la communauté pouvant pousser certaines à l'extinction locale (Vellend 2010).

Migration La dispersion correspond au mouvement des organismes d'une localité à une autre. Du point de vue de la communauté, une plus grande dispersion s'accompagne d'une augmentation de la diversité locale par ajout de nouvelles espèces depuis une source de migrants (MacArthur and Wilson 1967) en supposant que les espèces puissent s'établir là où elles arrivent. La dispersion implique que les communautés ne sont pas des entités indépendantes et isolées, de telle sorte que l'on peut définir une métacommunauté comme un ensemble de communautés locales qui sont liées par la dispersion de multiples espèces (Leibold et al. 2004). Dans ce contexte, une espèce peut se maintenir régionalement si ses capacités de dispersion lui permettent d'échapper à des contraintes locales qui, autrement, provoqueraient son extinction (Figure I.6 Mouquet and Loreau 2003). Une forte immigration peut ainsi compenser une natalité locale plus faible de sorte qu'on observe des densités de populations supérieures à celles attendues pour une communauté isolée. A l'inverse, une forte émigration peut provoquer une baisse de la densité de population locale par rapport à

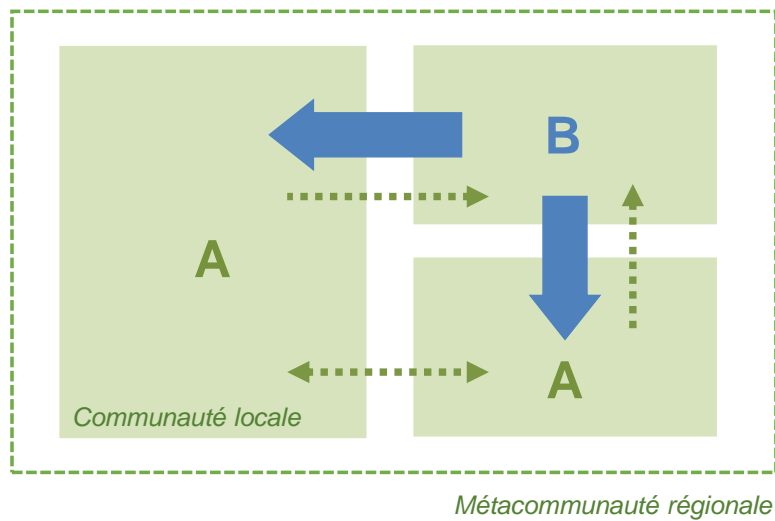


Figure I.6: Dynamique de patches de métacommunautés permettant une coexistence régionale entre une espèce compétitrice (A) et une espèce colonisatrice (B) par le biais d'un compromis entre compétition et colonisation permis par la migration. Les flèches pleines indiquent une dispersion plus importante que les flèches en pointillés.

une communauté isolée. De manière générale, la dispersion influence la similarité des communautés entre elles, ce qui a des conséquences importantes sur la diversité β (Munoz, Couteron, and Ramesh 2008).

Spéciation La spéciation est un processus évolutif fondamental à l'origine de la biodiversité. C'est par la spéciation que les espèces apparaissent pour ensuite se structurer en différentes communautés écologiques. La plupart des études en écologie des communautés excluent la question de savoir comment les espèces d'une zone donnée sont apparues en premier lieu, laissant ces questions aux domaines de la biogéographie et de la macroévolution (Ricklefs 1987 ; Brown 1995). Cela dit, pour comparer des communautés entre différentes régions - dans un contexte de métacommunauté par exemple -, il est sans doute pertinent d'inclure le contexte évolutif qui a rendu disponibles les espèces qui forment des communautés (Vellend 2010). La spéciation, étant le mécanisme générateur de nouvelles espèces, augmente la richesse spécifique de la métacommunauté. Celle-ci peut dépendre du contexte local et ainsi entraîner des différences de diversité d'une communauté à une autre, ce qui aura des répercussions sur la diversité β . Ces effets sur la diversité fonctionnelle peuvent être minimes si les nouvelles espèces ressemblent aux espèces existantes, ou à l'inverse augmenter la gamme des traits dans la métacommunauté en introduisant des espèces fonctionnellement très distinctes.

Selection Dans une communauté, la sélection est le résultat de différences de performances individuelles de ces membres, déterminées par l'espèce à laquelle ils appartiennent et la relation entre les traits qu'ils possèdent et le contexte environnemental local (Vellend 2010). En effet, les individus sont naturellement variables. Du fait de cette variabilité naturelle, certains individus seront plus susceptibles de s'éteindre dans certaines communautés, car moins adaptés à y survivre. Au contraire, certains auront plus de descendants et seront ainsi plus abondants dans la communauté. Ainsi, les abondances relatives des espèces observées sont le reflet de la sélection d'individus plus performants qui persistent dans la communauté (Chesson 2000).

Dérive Les événements de vie, mort et reproduction des individus comportent une variation aléatoire (McShea and Brandon 2010), et les modèles de dynamique démographique et de communautés peuvent alors être définis de manière probabiliste. Dans une communauté isolée et de taille finie, la disparition successive par des événements démographiques aléatoires entraîne un appauvrissement de la diversité locale - c'est ce qu'on appelle la dérive. La dérive dans une communauté locale entraîne donc une fluctuation aléatoire de l'abondance relative des espèces, indépendamment de leur identité ou de leurs caractéristiques (Vellend 2010 ; Hubbell 2001). Sans l'intervention d'autres mécanismes que la dérive, une espèce finira par devenir totalement dominante - on dira alors qu'elle est fixée - tandis que toutes les autres espèces dériveront vers l'extinction. Les chances que n'importe quelle espèce se fixe sont liées à sa fréquence initiale, et la vitesse à laquelle elle y parvient est liée négativement à la taille de la communauté. Sous l'effet de perturbation, la taille de la communauté peut fluctuer et ainsi accroître l'importance de la dérive (Nei and Tajima 1981).

1.2 Le problème inverse ou la question de l'inférence

1.2.1 Le problème direct et ses limites

A ce stade, la question qui se pose est la suivante: comment faire le lien entre les patrons de biodiversité et les processus à leur origine ? La formulation de théories nous fournit différentes hypothèses sur la façon dont sont générés certains patrons de diversité, mais la difficulté est de démontrer quelles sont les causes d'un patron observé (Quinn and Dunham 1983). La formalisation sous la forme de modèles mathématiques de ces théories permet de faire des prédictions sur la composition des communautés selon différentes valeurs de paramètres (Zurell et al. 2010). Des dispositifs expérimentaux peuvent être utilisés pour examiner le lien causal entre des déterminants écologiques et des patrons observés. Par exemple, il est possible de manipuler expérimentalement le nombre d'espèces d'une communauté ou la similarité fonctionnelle et/ou phylogénétique et d'en observer les répercussions directes sur les propriétés des écosystèmes (Tilman, Reich, and Knops 2006) ou la coexistence d'espèces (Violle, Nemergut, et al. 2011). Cependant, manipuler expérimentalement le nombre d'espèces dans des communautés vastes et reculées dont les organismes sont centenaires, comme les forêts tropicales humides, est pratiquement impossible. Concrètement, l'échelle de n'importe quel dispositif expérimental doit être adaptée à l'échelle des mécanismes hypothétiques envisagés, de sorte que si celles-ci concernent des systèmes à grande échelle la conception d'un dessin expérimental adapté devient impossible (Shipley 2016). Ainsi, la compréhension des mécanismes à l'origine de la diversité des espèces relève plus souvent d'un problème inverse, dans lequel on cherche à déterminer les processus à l'origine des patrons observés, que d'un problème direct où l'on observe les conséquences d'un processus directement sur les patrons (Tarantola 2006).

1.2.2 Les ombres au fond de la caverne

Les patrons de diversité sont, comme dans l'analogie de Platon (Platon n.d.), telles des ombres projetées sur le mur d'une caverne: des reflets indirects et imparfaits des véritables processus en jeu,

dont nous, en tant qu'observateurs, ne pouvons jamais être directement témoins. La problématique inverse, est un problème dans lequel nous essayons de déduire les paramètres régissant un système à partir des résultats qui en découlent (Tarantola 2006). Dans l'exemple de la caverne de Platon, il s'agit de savoir quels sont les processus à l'œuvre hors de la vision de l'observateur, qui sont à l'origine des ombres projetées qu'il observe. Si différents mécanismes produisent la même ombre ou si les mécanismes interagissent, on comprend comment le problème inverse peut être difficile à résoudre. En écologie, tout comme pour les ombres dans la grotte, nous devons nous fier à notre observation indirecte des systèmes naturels pour déduire les processus biologiques en jeu. Les dispositifs expérimentaux en écologie tentent de transformer le problème inverse en un problème direct, mais comme mentionné plus haut, ceux-ci ont des limites lorsque appliqués à large échelle. Une des solutions est alors pour l'observateur de simuler lui-même des ombres en fonction de ses hypothèses et les tester en comparant les résultats des simulations aux ombres réelles. Ainsi, avec des modèles mécanistes représentant directement les processus écologiques d'intérêt, nous pouvons concevoir des expériences virtuelles où les patrons de diversité sont directement observables (Zurell et al. 2010 ; Münkemüller et al. 2012).

2 De multiples processus impliquent de multiples échelles

L'une des difficultés pour la résolution du problème inverse est la question de l'échelle d'observation par rapport aux échelles auxquelles jouent les processus. Les processus et les patrons observés à une échelle peuvent être très différents de ceux observés à une autre échelle. De plus, une communauté peut être influencée par des processus opérant à plusieurs échelles simultanément (Levin 1992). Avec le concept de métacommunauté, on comprend comment les différentes communautés peuvent interagir à de multiples échelles temporelles et spatiales et comment les patrons eux aussi peuvent être décrits à différentes échelles. La nécessité d'intégrer plusieurs échelles spatiales et temporelles donne lieu à deux visions alternatives en écologie: une vision "top-down" ou descendante, dans laquelle on considère qu'une communauté locale est constituée du sous-ensemble de la diversité régionale qui s'établit et persiste dans un habitat donné, et une vision "bottom-up" ou ascendante dans laquelle l'ensemble des processus régissant la diversité de plusieurs communautés en interaction mis bout à bout fait émerger une certaine diversité régionale (Munoz and Huneman 2016). Dans les deux cas, les dynamiques locales à échelles spatiales et temporelles fines sont découplées des dynamiques régionales à larges échelles spatiales et temporelles.

2.1 Structuration de la biodiversité dans l'espace

2.1.1 Processus écologiques à l'échelle locale

Processus déterministes et dynamiques de niche Qu'est ce qui détermine la coexistence stable de multiples espèces à un endroit donné? Telle est la question fondamentale au cœur de l'étude de l'assemblage des communautés. Ceux qui se la sont posée reconnaissent que de nombreux facteurs peuvent influencer la composition en espèces d'une communauté donnée, sans qu'un seul facteur puisse fournir une explication complète des patrons observés (Schoener 1986).

Toujours est-il que bien souvent on se donne comme objectif d'établir des "règles d'assemblage", qui déterminent la co-occurrence locale de différentes espèces et qui aboutit à la biodiversité locale. Keddy 1992 propose que les capacités de dispersion et la réponse à la sélection induite par l'environnement local trient et filtrent les espèces, faisant ainsi des communautés locales des sous-ensembles non aléatoires du pool d'espèces régional (Ricklefs 2008; Lessard et al. 2012). Imaginons la succession d'étapes qui doit se produire pour qu'un individu d'une espèce donnée s'intègre dans une communauté locale. Tout d'abord, celui-ci doit arriver sur place par dispersion. Ensuite, il doit pouvoir survivre aux conditions environnementales locales et s'y reproduire avec succès. Il doit également être capable de tolérer la compétition exercée par d'autres individus avec les mêmes besoins en ressources. Chaque étape représente un filtre différent qu'un individu doit avoir passé pour faire partie de la communauté locale (Lortie et al. 2004). Ainsi cette série de filtres élimine certaines espèces tout en laissant passer d'autres. Les espèces qui possèdent les propriétés biologiques adaptées aux conditions environnementales locales "passent à travers" les filtres et les espèces inadaptées sont exclues (Figure I.7). L'existence de filtres environnementaux implique donc qu'il existe une relation entre les traits fonctionnels des espèces et les conditions environnementales locales qui déterminent l'assemblage des communautés. Par conséquent, l'étude des distributions des traits des espèces dans les communautés devrait nous permettre de caractériser l'influence des différents filtres. Le filtre d'habitat exige que seules les espèces avec les traits fonctionnels leur permettant de s'installer, survivre et se reproduire peuvent persister dans la communauté, et donc sélectionne des espèces avec des traits similaires. A l'inverse, des espèces avec des valeurs de traits plus contrastées ont plus de chances de survivre à la compétition exercée par leurs voisins. La compétition limite donc la ressemblance fonctionnelle des espèces. C'est ce qu'on appelle l'hypothèse de limite à la similarité (MacArthur and Levins 1967).

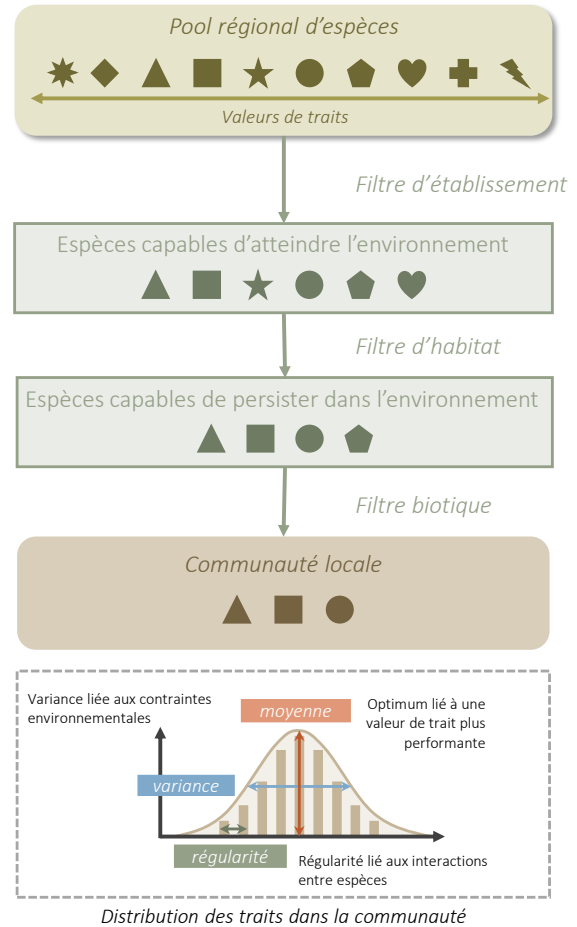


Figure I.7: Illustration traditionnelle de la série de filtres hiérarchiques qui réduisent l'ensemble des membres du pool régional aux membres d'une communauté locale.

Processus stochastiques et dynamiques neutres Une espèce peut être absente d'un relevé si ses propriétés biologiques ne sont pas adaptées à l'environnement local ou alors si celle-ci est moins compétitive que les autres espèces de la communauté. Cela dit, il peut y avoir d'autres raisons à cette absence. L'espèce peut ne pas avoir été échantillonnée localement, elle peut être présente dans la région mais s'être éteinte localement, ou ne pas avoir colonisé le milieu, elle peut connaître

un isolement biogéographique et s'éteindre régionalement, etc. Il existe intrinsèquement une forte variabilité stochastique des paramètres démographique et de colonisation, du fait du caractère discret des individus qui composent les populations, et donc les communautés. On retiendra par exemple la théorie de la biogéographie insulaire de MacArthur and Wilson 1967, qui examine les effets de la colonisation et de l'extinction sur la biodiversité des îles. Ou encore les équations de métapopulation de Levins 1968, qui décrivent la variation de l'occupation d'une population dans un réseau spatial, comme le bilan de gain par colonisation et de perte par extinction, et dont la résolution permet de connaître quelles conditions aboutissent à un équilibre stable qui permet la viabilité d'une espèce donnée dans le réseau. Mais le premier à réellement examiner les conséquences de la pure stochasticité démographique sur la structuration de la biodiversité à différentes échelles est Stephen Hubbell dans son livre *Unified Theory of Biodiversity and Biogeography* (Hubbell 2001). Il y développe la théorie neutre dans laquelle toutes les espèces d'une communauté sont équivalentes, c'est-à-dire que tous les individus vivent, meurent et dispersent en espérance de la même manière, sans effets de niches, sans compétition et sans filtre environnemental. Cette hypothèse de neutralité lui permet d'évaluer les conséquences de dynamiques stochastiques de migration et d'extinction dans des habitats de taille finie, et de prédire les propriétés émergentes des communautés en l'absence d'autres effets. En l'absence de migration, par dérive, une seule lignée se fixe et envahit la communauté. Puisque les espèces ont les mêmes capacités compétitives, l'identité de l'espèce qui survit est uniquement liée à son abondance régionale. Si de nouvelles espèces arrivent régulièrement par migration alors le temps de fixation n'est jamais atteint, de sorte que la migration compense l'effet de la dérive locale. Cet équilibre dynamique est traduit dans le modèle de Hubbell par le paramètre d'immigration I , qui représente l'équilibre entre la migration et l'effet de la taille de la communauté n qui détermine la dérive.

$$I = \frac{m}{1-m}(n-1) \quad (1)$$

Ainsi, les abondances des espèces fluctuent sous l'influence de la migration et de la dérive jusqu'à aboutir à un état stationnaire régi par le paramètre I . La dispersion compense la fixation car elle empêche les espèces de saturer leur milieu, mais si la dispersion est limitée, alors elle va surtout maintenir la différence de composition entre communautés, augmentant ainsi la diversité β (Munoz, Couteron, and Ramesh 2008)

Limites de l'hypothèse de stationnarité en écologie Bien que la plupart des travaux sur l'assemblage des communautés se soient longtemps concentrés sur les processus limitant la persistance d'espèces dans une communauté locale, découpler les échelles spatiales et temporelles locales et régionales sous-tend que la structure du pool régional a un fort impact sur la communauté locale. Des études récentes ont souligné l'influence critique de la diversité du pool d'espèces (Cornwell et al. 2014) dans la fourniture de migrants, et ont appelé à une approche de modélisation plus mécaniste qui représente directement les processus clés qui façonnent l'assemblage des communautés (Denelle, Violle, and Munoz 2019). Souvent, les efforts de modélisation font l'hypothèse que le pool d'espèces régional et les phénotypes associés représentent un instantané d'un héritage évolutif, à partir duquel les espèces s'assemblent par le biais de dynamiques locales sur des périodes récentes et

suffisamment courtes pour négliger les événements de diversification. Or, envisager le pool comme une entité fixe, c'est mettre de côté que celui-ci possède sa propre dynamique, résultant de la combinaison de processus locaux qui s'accumulent dans le temps. Dynamique qui peut être appréhendée en considérant l'apparition et la disparition des espèces qui constituent le pool, et de leurs caractéristiques biologiques à de larges échelles spatiales et temporelles (dans une perspective évolutive). L'appartenance d'une espèce au pool régional est déterminée par limites fixées par ses tolérances physiologiques, mais aussi par des facteurs historiques, et les processus évolutifs responsables de la génération d'un nombre différent d'espèces dans différents groupes taxonomiques et dans différentes régions. En effet, les espèces ne sont généralement pas présentes dans les zones bioclimatiques au-delà de leurs limites physiologiques. Cela dit, des introductions réussies d'espèces dans des zones éloignées de leur aire de répartition habituelle montrent que la contingence de la biogéographie peut exclure des groupes entiers d'espèces de certaines régions géographiques (Elton 1958). La composition des communautés actuelles reflètent l'effet des processus écologiques actuels, mais aussi un héritage biogéographique et évolutif. En effet, les communautés peuvent persister sur des milliers d'années, et leur histoire dépendra alors du temps écoulé depuis la dernière période glaciaire, de l'éloignement des refuges des différentes espèces et de la vitesse à laquelle celles-ci peuvent se déplacer dans le paysage (Ordóñez and Svenning 2015 ; Ordóñez and Svenning 2016a ; Ordóñez and Svenning 2016b).

2.1.2 Processus biogéographiques à l'échelle régionale

Pour comprendre la diversité à large échelle spatiale des pools régionaux, il est nécessaire de s'intéresser à la biogéographie. De même que l'écologie, la biogéographie historique est une science basée sur la description des faunes et des flores, ancrée dans l'histoire naturaliste issue des grandes expéditions du XVIII^e siècle (voir Figure I.8 ; Humboldt and Bonpland 1805 ; Bonpland 1807 ; De Candolle 1855). La biogéographie connaît plus tard un développement théorique, visant à formaliser et étudier avec des méthodes quantitatives les processus à l'origine des distributions à large échelle des espèces. Plus particulièrement, la biogéographie historique s'intéresse aux impacts d'événements historiques particulièrement marquants sur la distribution actuelle des espèces, tels que ceux liés à la dynamique géomorphologique des continents ou encore aux aléas climatiques. L'hypothèse au centre de la biogéographie historique est que ces événements au cours du temps ont façonné et laissé des traces sur l'aire de répartition à large échelle des espèces actuelles (Posadas, Crisci, and Katinas 2006). Les espèces ne sont pas réparties uniformément sur Terre. De plus, les aires de distribution ne sont pas spécifiques à une seule espèce: plusieurs espèces partagent la même aire de distribution de sorte qu'il existe des espèces dites « Eurasienne », « Africaine », « Paéotropicales », « Néotropicales » etc. La question se pose alors des processus à l'origine de la similarité ou de la dissimilarité des distributions géographiques de multiples espèces. Pour comprendre les distributions géographiques d'espèces actuelles il est nécessaire de coupler d'une part les mécanismes biologiques qui aboutissent à l'isolement des espèces les unes par rapport aux autres (spéciation), et d'autre part les changements géologiques de la Terre qui contraignent au cours du temps les régions que peuvent occuper les espèces. L'histoire géologique et climatique de la planète agissant tel un facteur de sélection sur l'aire de répartition des espèces.

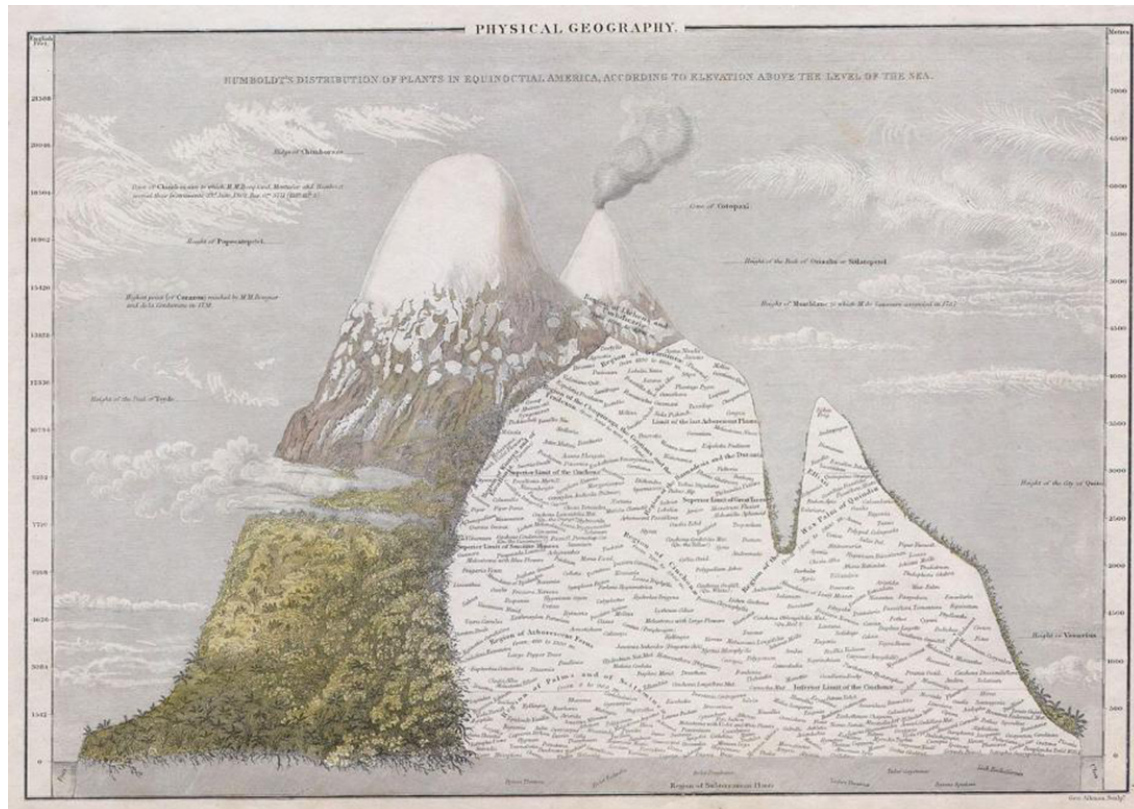


Figure 1.8: Répartition de la végétation du Mont Chimborazo en fonction de l'altitude. Cette carte constitue la pièce maîtresse de l'ouvrage fondamental de Humboldt: "Essai sur la géographie des plantes", publié en 1807, et représente également le début de la biogéographie (Bonpland 1807).

Vicariance et migration: deux processus fondamentaux en biogéographie historique Parmi les processus auxquels on s'intéresse en biogéographie, la vicariance est un modèle dans lequel une barrière géographique telle qu'une montagne ou un océan, vient fragmenter l'aire de répartition originelle d'un taxon ancestral et dans lequel des différences biologiques et écologiques s'accumulent au cours du temps jusqu'à obtenir des espèces distinctes présentes dans les régions issues de la fragmentation de l'aire initiale (Croizat 1982 ; Heads 2009). Il s'agit d'une forme de spéciation allopatrique dans laquelle l'isolement reproducteur est médié par l'apparition d'une barrière géographique entraînant l'isolement spatial et donc la différenciation de deux ou plusieurs populations. Un autre processus qui aboutit à la différenciation biogéographique des espèces est la spéciation péripatrique. Dans ce cas, un événement très peu probable de colonisation ponctuelle longue distance se produit depuis l'aire de distribution initiale de l'espèce, et aboutit par effet de fondation à l'isolement d'une population. Cette population par isolement géographique se différencie progressivement de la population initiale, jusqu'à produire une nouvelle espèce. Les espèces sont alors endémiques puisqu'elles sont chacune uniquement présente soit dans la zone géographique d'origine, soit dans la région nouvellement colonisée. Ce type de spéciation, contrairement au modèle de vicariance, met en avant le rôle de la dispersion, et en particulier à contingence de ces événements de colonisation longue distance (Croizat 2013). La biogéographie historique vise ainsi à étudier comment ces deux processus fondamentaux : vicariance et migration s'équilibrent et aboutissent à la répartition de la diversité des espèces dans l'espace. En particulier de déterminer quelles sont les espèces qui se sont différenciées par fragmentation et vicariance, et quelles sont

celles qui se sont différenciées par effet fondateur suite à un événement de dispersion rare, ainsi que les événements historiques ayant favorisé l'un ou l'autre des ces processus (Ronquist 1997 ; Yu, Harris, and He 2010).

Les grandes régions biogéographiques: héritage des biogéographes La cartographie des grandes zones biogéographiques permet de mettre en évidence des similitudes dans la répartition des faunes et des flores ainsi que d'identifier les zones d'endémisme. Ces zones biogéographiques sont les résultats des préférences environnementales des espèces en rapport avec leurs adaptations à différents climats, mais aussi de leur biogéographie, et donc lié à des événements historiques aboutissant à des zones de refuge et d'endémisme (Morrone 2008). Ainsi, les régions biogéographiques intègrent le rôle de l'histoire qui a abouti à des séparations de taxons dans des environnements similaires, et complètent l'information apporté par les zones de vie de Holdridge, qui elles ne considèrent que les préférence environnementales des espèces i.e. leurs niches fondamentales. A de larges échelles spatiales certaines limites apparaissent telle que la célèbre ligne de Wallace (Figure I.9). Celle-ci sépare les îles sub-Antarctique, la Nouvelle Zélande et la Patagonie au sud-est de l'Indonésie, Java, Bornéo, Bali, et les Philippines au nord-ouest. De chaque côté et en particulier au sud de cette frontière, ces îles abritent des faunes endémiques très similaires bien qu'elles soient très éloignées les unes des autres. On ne comprend cette ségrégation que si on s'intéresse à la distribution des masses continentales actuelles par rapport à la structure de la Pangée. En effet, l'Antarctique faisait partie du Gondwana et on observe que toutes les zones actuellement séparées étaient autrefois incluses dans une même masse continentale. La fragmentation du Gondwana et la séparation de l'Antarctique a abouti à une paléo-flore et une paléo-faune dans ces régions qui ont gardé des caractéristiques communes malgré leur séparation. Ainsi, malgré l'ancienneté de la fragmentation

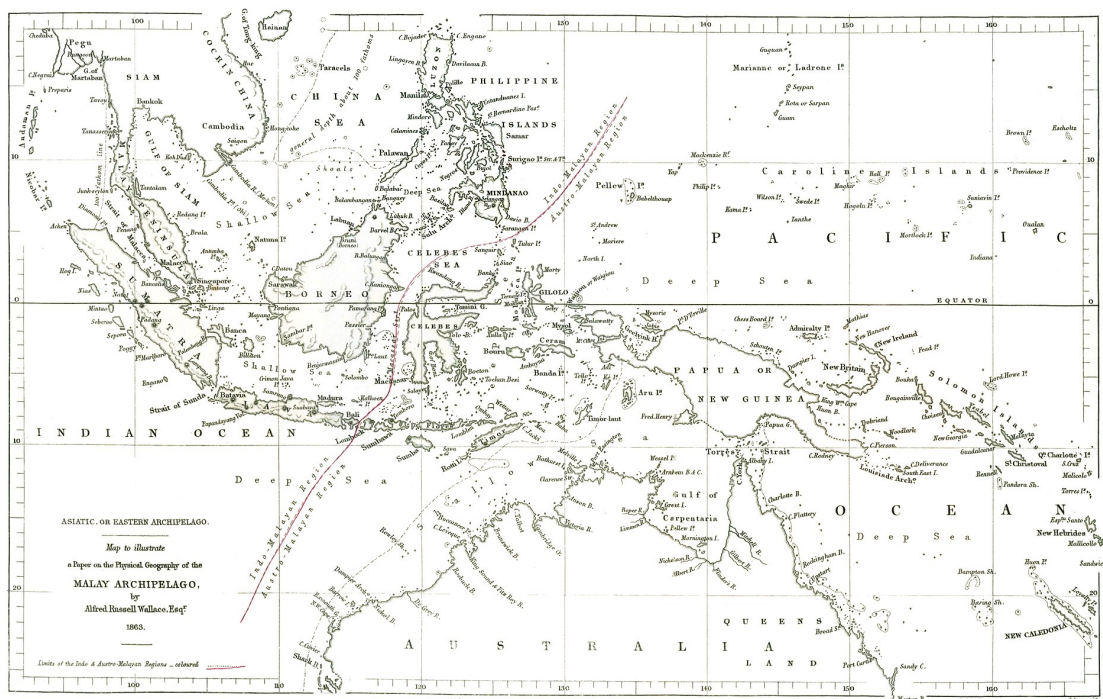


Figure I.9: Dessin original de la ligne de Wallace montrant la division entre les flores et les faunes asiatiques et australasiennes.

du Gondwana et la grande isolation actuelle des ces îles (Nouvelle Zélande, Patagonie et les îles sub-Antarctiques), leurs flores gardent des caractéristiques communes héritées de leur histoire partagée, et dont la signature de processus anciens se détectent dans la distribution de certains genres au sein des ces biotas. A l'inverse, d'autres zones témoignent de séparations anciennes, mais qui sont aujourd'hui proches géographiquement. Les zones biogéographiques de part et d'autre de la ligne de Wallace abritent des groupes d'espèces phylogénétiquement très éloignés (e.g. faunes marsupiale et placentaire). Alors qu'actuellement il n'y a pas de limite à la dispersion des espèces, leur séparation dans le passé a conduit à leur divergence. Ainsi, cette frontière biogéographique est le résultat d'une séparation ancienne et importante des biotas qui, par le mouvement des masses continentales, se sont rapprochés récemment. La ligne de Wallace est un exemple de signature biogéographique des discontinuités, entre les biotas traduisant l'ensemble d'une séquence historique de dérive des continents particulière, et un cas où les distances géographiques actuelles peuvent être décorréliées des temps de divergence des espèces (Holt et al. 2013).

Inférence de processus biogéographiques à partir des distributions d'espèces actuelles : vicariance ou dispersion ?

La dérive des continents n'ayant été découverte que bien après les observations des premiers biogéographes, la dispersion a d'abord été considérée comme l'élément fondamental pour expliquer la répartition des organismes dans les différentes régions du monde. L'hypothèse d'un centre d'origine unique et de multiples dispersions expliquant les aires de répartition discontinue des espèces a donc été soutenue par les évolutionnistes, qui au départ n'avaient pas connaissance de l'existence des mouvements continentaux importants, ni des changements des limites géographiques au cours du temps. Cette hypothèse a donc encouragé les biogéographes à trouver les chemins et l'historique de dispersion ayant permis la diversification des faunes et des flores actuelles et qui explique leur diversité actuelle. Cependant, la construction de cladogrammes et d'areogrammes - des arbres phylogénétiques qui représentent l'évolution de ces aires de répartition à la manière de caractères - et l'émergence de la biogéographie cladistique a remis en question l'idée de la dispersion comme seul processus à l'œuvre. Il n'y aurait donc pas seulement un centre et un archipel de zones colonisées, mais une diversification plus graduelle. Ainsi, il devient pertinent d'essayer d'identifier les séquences de séparation spatiale entre les faunes et les flores, et les événements qui peuvent être à l'origine de ces séparations, en évoquant soit la dispersion soit la séparation passive. L'hypothèse alternative à la dispersion: la vicariance postule donc l'existence d'un taxon ancestral cosmopolite, dont la séquence de fragmentation géographique ayant abouti aux aires de répartition des espèces actuelles doit correspondre à la séquence de diversification de ces mêmes espèces. Elle implique aussi la congruence entre espèces partageant la même structure phylogénétique, qui correspond à la séquence de fragmentation ayant influencé leur différenciation. La congruence produit des patrons de parentés phylogénétiques répétés - lié au fait qu'elles ont été contraintes par les mêmes processus de séparation spatiale - qui, s'ils sont identifiés dans plusieurs clades, permettent de mettre en évidence une origine et une séquence de fragmentation commune (et donc de vérifier l'hypothèse de vicariance). L'exemple classique qui donne raison à l'hypothèse de vicariance est l'éclatement de la Pangée qui a abouti à la différenciation des faunes et flores de l'Ancien et du Nouveau Monde. Il y a donc eu deux approches pour expliquer la répartition des espèces actuelles: dispersaliste d'une part, qui considère l'effet prépondérant de la dispersion depuis un centre d'origine, et d'autre part celle qui considère la fragmentation et la vicariance. Selon les espèces, il peut y avoir une combinaison

d'événements de vicariance et de dispersion. En particulier, quand il existe des distributions du taxon ancestral qui s'avèrent être plus petites que la distribution des taxons actuels. Dans ces cas, la séparation géographique de taxons actuels apparentés est liée à la fois à la dynamique du continent sur lequel se trouve les espèces, ainsi qu'à des événements de colonisations longues distances qui viennent étendre l'aire de répartition des espèces actuelles par rapport au taxon ancestral - c'est ce qu'on appelle la règle de progression. La règle de progression pose cependant problème lorsque des imprécisions persistent sur l'aire des espèces ancestrales. En effet, beaucoup d'informations sont nécessaires pour affirmer qu'un taxon ancestral fut cosmopolite. Résoudre la question de l'aire de répartition d'un taxon ancestral passe souvent par la découverte de fossiles. Connaître la répartition ancestrale permet de trancher entre des séquences de fragmentation et/ou de dispersion, et d'identifier la suite d'événements historiques menant à la fragmentation des aires de répartition actuelles. Aujourd'hui la datation des données moléculaires permet de trancher entre plusieurs scénarios de vicariance et/ou de dispersion selon que la date de divergence coïncide ou non avec la formation d'une barrière géographique connue (océans, chaînes de montagnes etc.). Les deux mécanismes ; vicariance d'un côté et dispersion longue distance de l'autre, ne s'excluent pas. La question est d'arriver à identifier en cas d'ambiguïté la part relative des deux processus dans la structuration de la répartition actuelle des biotas.

2.2 Diversification de la biodiversité dans le temps

La diversité des communautés d'espèces à large échelle spatiale ne peut s'appréhender qu'en considérant les processus à long terme aboutissant à la diversification des espèces. La diversification traduit l'équilibre entre le gain d'espèces par spéciation d'une part, et de la perte d'espèces par extinction d'autre part. L'étude des dynamiques de diversification est rendue particulièrement difficile par les échelles de temps considérées. En effet, les processus de spéciation et d'extinction se déroulent généralement sur une échelle de milliers à millions d'années, et nécessitent d'importantes données fossiles, bien souvent indisponibles pour de nombreux groupes d'espèces. C'est pourquoi la plupart des études en macro-évolution utilisent des phylogénies reconstruites à partir de données moléculaires, dont la comparaison avec des arbres phylogénétiques générés à partir de modèles théoriques, permettent de conclure sur les dynamiques de diversifications. De manière générale, les modèles théoriques conditionnent l'une ou l'autre des composantes de la diversification (spéciation ou extinction) à différentes grandeurs comme le temps, des facteurs abiotiques ou biotiques, ou bien les traits des espèces etc. Le but étant d'estimer dans quelles mesures ces modèles permettent de prédire les propriétés observées dans des phylogénies reconstruites (Morlon 2014).

2.2.1 Dynamiques de diversification

L'apport d'espèces par spéciation Dans la plupart des cas, on cherche à estimer à quelle fréquence de nouvelles espèces apparaissent, et si des facteurs environnementaux ou historiques peuvent expliquer comment se produit la spéciation. Dans le cas de la spéciation géographique (ou allopatrique), une barrière géographique apparaît et restreint le flux de gènes entre deux populations. Maintenues isolées pendant suffisamment de temps, celles-ci ne seront plus interfécondes et peuvent alors être

considérées comme deux espèces distinctes (Futuyma 1979). Certaines populations peuvent habiter une région qui dispose de barrières géographiques toutes faites. Bien que difficile à franchir, il arrive que des propagules traversent occasionnellement ces barrières, généralement à de larges intervalles de temps permettant la spéciation allopatrique. La région et ses barrières agissent alors comme une machine à spéciation, produisant rapidement de nouvelles espèces. C'est pourquoi les archipels et les montagnes ont tendance à augmenter les taux de spéciation. Pour les régions continues, la probabilité d'apparition d'une barrière géographique dépendra de la forme et la taille de l'aire de la population initiale, et de la longueur de la barrière. Toujours est-il qu'une fois formés, les isolats doivent diverger au point de constituer deux espèces distinctes. Un changement évolutif peut se produire par fluctuations statistiques liés à l'échantillonnage de petites populations, un phénomène appelé dérive génétique. De nombreuses mutations mineures peuvent s'accumuler régulièrement et aboutir à la différenciation des isolats. Mais ceux-ci peuvent également diverger s'ils utilisent des mécanismes différents pour résoudre des problèmes similaires, ou si des stratégies adaptatives différentes sont retenues par sélection naturelle dans des environnements différents. Dans tous les cas, les mécanismes à l'origine de la spéciation géographique sont fondamentalement les mêmes que pour la vicariance décrite précédemment (Rosenzweig 1995). Cependant, à la différence d'un modèle de spéciation allopatrique, dans la théorie issue de la biogéographie historique on ne considère pas l'isolement reproducteur primaire ou secondaire, et il arrive que des espèces vicariantes soient interfécondes et s'hybrident lorsque remises en contact, par exemple dans les jardins botaniques. Cependant, il peut y avoir spéciation sans isolement géographique. Certaines irrégularités méiotiques peuvent produire des gamètes diploïdes qui, s'ils se combinent, produisent une nouvelle espèce avec plus de chromosomes que chacun de ses parents (R. C. Jackson 1976). Certains taxons sont des candidats idéaux pour la polyploïdie, comme les plantes qui contiennent souvent des espèces capables d'autofécondation. Par rapport à la spéciation géographique, la polyploïdie ne nécessite pas la formation d'une barrière et se produit relativement instantanément. De plus, la spéciation peut se produire au sein de la même région géographique si une opportunité écologique nouvelle conduit à la scission sympatrique de deux populations, l'une utilisant l'opportunité originelle, l'autre la nouvelle opportunité exploitée. Dans ce cas, les taux de spéciation dépendent des opportunités écologiques, tandis que les taux de spéciation géographique dépendent des possibilités d'isolement. Les différents modes de spéciation offrent différents types de prédictions. Ainsi, savoir quels modes produisent les espèces, et dans quels contextes, permet d'expliquer les différences de patrons de diversité régionale.

La perte d'espèces par extinction La force antagoniste à la spéciation est l'extinction. On doit l'idée qu'une espèce puisse s'éteindre à l'anatomiste Georges Cuvier en 1796 et de l'étude plus générale de fossiles (Faria 2013). Cela dit, on a longtemps considéré l'extinction comme un processus graduel jusqu'à la découverte de preuves de la disparition soudaine des dinosaures non aviaires il y a 66 millions d'années. Malheureusement, on ne manque pas d'exemples récents qui illustrent comment une espèce entière peut complètement disparaître en très peu de temps. L'impact actuel des activités humaines entraîne encore plus d'extinctions à mesure que les habitats des organismes disparaissent, ou sont perturbés par l'introduction de nouvelles espèces invasives. Mais les espèces peuvent aussi s'éteindre sans raison prévisible. Des catastrophes naturelles peuvent balayer des populations entières. Leur patrimoine génétique peut aléatoirement dériver vers des combinaisons de gènes

inadaptés à l'environnement local. L'ensemble des conditions environnementales nécessaires à la survie de l'espèce peut également disparaître par malchance, auquel cas les espèces aussi disparaîtront. Et bien qu'une grande part de l'extinction est le fruit du hasard, certaines caractéristiques des espèces les rendent plus susceptibles de s'éteindre. L'idée que des petites populations sont plus à risque de s'éteindre fait assez largement consensus. Si chaque individu d'une population a une probabilité p de mourir, alors la totalité d'une population de taille N disparaîtra en un temps proportionnel à p^N . Ainsi, plus N est grand, moins l'espèce est susceptible de s'éteindre. De plus, la taille d'une population détermine la gravité de la dérive génétique, qui modifie les fréquences des gènes de génération en génération, possiblement vers des combinaisons inadaptées.

2.2.2 Temporalité du signal: importance des dynamiques hors équilibres

Le déséquilibre de la nature Les communautés sont loin d'être des entités statiques. Cela dit, l'idée selon laquelle celles-ci atteignent un état idéal et permanent auquel elles reviennent si elles sont perturbées, persiste. Mais même si un point d'équilibre stable existe, il n'est pas garanti que les communautés l'aient atteint ni qu'elles l'atteignent jamais. Par exemple, on sait que les perturbations empêchent certaines communautés d'atteindre l'équilibre impliqué par le principe d'exclusion compétitive. Ce principe aussi connu comme le principe de Gause (Gause 1934), veut que dans le cas où deux espèces auraient strictement les mêmes exigences environnementales, alors la moins compétitive des deux espèces s'éteindra sous la pression exercée par l'espèce compétitive. Or, si le milieu est régulièrement perturbé, alors l'espèce moins compétitive peut survivre si elle s'en sort mieux que sa concurrente dans les conditions perturbées. Si des perturbations se produisent suffisamment régulièrement, elles déplacent sans arrêt l'état du système, qui se maintient dans un état qui autrement serait instable. Autrement dit, si l'environnement extérieur change plus fréquemment qu'il ne faut pour atteindre l'équilibre, celui-ci ne sera jamais atteint. Imaginons maintenant que la réponse des communautés à un changement de l'environnement soit particulièrement lente, par exemple car les espèces qui les composent ont des cycles de vie qui s'étendent sur des centaines d'années. Dans ce cas des événements historiques très anciens peuvent avoir entraîné des dynamiques transitoires (S. T. Jackson and Sax 2010 ; Huth et al. 2015 ; Blanchard and Munoz 2022). Ainsi, des patrons actuels de diversité peuvent résulter de l'effet durable de conditions environnementales passées.

Importance des dynamiques transitoires: exemple des refuges Fondamentalement c'est l'existence de dynamiques transitoires qui permet d'inférer des événements passés à partir de l'observation de la diversité actuelle. Lors de dynamiques transitoires, l'équilibre entre de nouvelles conditions extérieures et la configuration d'un système n'est pas atteint. C'est ce délai de réponse entre la diversité des espèces actuelles et leur dynamique éco-évolutive qui rend la signature d'un processus observable dans le présent, et qui rend possible la reconstitution des événements passés. Par exemple, une partie de la diversité actuelle en Europe ne peut se comprendre qu'en prenant en compte l'effet des variations climatiques importantes des épisodes glaciaires et interglaciaires du Quaternaire (Abellán and Svenning 2014). Lors des périodes où la calotte glaciaire était largement répartie sur l'Europe médiane, certaines zones en périphérie et au sud de l'Europe ont servi de refuges pour des espèces tempérées et subtropicales qui y ont survécu, avant de progressivement

coloniser de nouveaux habitats lorsque les conditions climatiques furent à nouveau plus favorables. La nature transitoire des dynamiques éco-évolutives couplée au caractère contingent des événements climatiques et géologiques contribuent à la présence de refuges. La phylogéographie permet de reconstituer les routes et les chemins de migration post-glaciaire, mais on constate que de nombreuses espèces ayant trouvé refuge dans ces régions n'ont depuis pas encore atteint les limites d'occupation prédites en se fiant à leurs préférences environnementales.

Dans certains cas, ces espèces ne recoloniseront jamais leur habitat d'origine et resteront des reliques de ces événements climatiques importants. C'est le cas de *Carex maritima* (Figure I.10), qui lors des dernières glaciations a colonisé une large partie de l'Europe depuis les côtes Scandinaves, et qui occupe actuellement un isolat dans les Alpes en réponse au réchauffement récent du climat, et qui tend même à disparaître sous l'effet du changement climatique (Escudero et al. 2010).

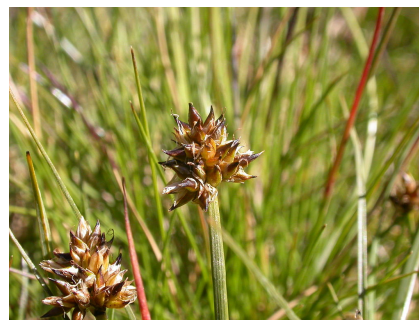


Figure I.10: *Carex maritima*. Zermatt, Riffelsee (Kanton Wallis, Suisse), 2700m

La question de l'échelle La biodiversité se décrit à différents niveaux d'organisation et selon différentes facettes, et est façonnée par des processus opérant à différentes échelles de temps. De ce fait, tous les aspects de la biodiversité ne sont donc pas nécessairement à l'équilibre simultanément. Avec le concept de métacommunauté, on voit bien comment les dynamiques internes des communautés peuvent présenter des propriétés d'équilibre différentes de celles des dynamiques de dispersion reliant les communautés entre elles. Dans un environnement hétérogène notamment, la dispersion peut favoriser la stabilité régionale de communautés locales asynchrones soumises à une dynamique instable (Briggs and Hoopes 2004). Par ailleurs, puisque l'on peut décrire les communautés selon leur diversité taxonomique, fonctionnelle ou phylogénétique, l'équilibre peut très bien dépendre de l'aspect de la diversité que l'on considère. Par exemple, il est très difficile de prédire l'identité des espèces des communautés planctoniques, du fait de l'émergence de dynamiques chaotiques, alors que considérer la diversité fonctionnelle de telles communautés, comme leur biomasse algale, permet de faire émerger des patrons réguliers et prévisibles (Scheffer et al. 2003).

3 Vers l'intégration des processus à de multiples échelles spatio-temporelles dans un cadre unifié

Les processus impliqués aux échelles locales et régionales s'influencent mutuellement. La diversité de différents pools régionaux dépend de la capacité des espèces à traverser les frontières biogéographiques et à migrer vers une région (Pennington and Dick 2004), et les abondances régionales des espèces ainsi que leurs attributs biologiques dépendent de l'équilibre entre la spéciation et l'extinction. La compréhension des dynamiques éco-évolutives nécessite donc d'intégrer l'interaction de mécanismes agissant à de multiples échelles spatiales et temporelles : d'un côté, pour prendre en compte

l'influence de processus à court terme, mais répétés dans le temps, sur les trajectoires évolutives des espèces, et de l'autre, sur l'impact de cet héritage évolutif sur les assemblages locaux.

3.1 La théorie neutre comme première intention

3.1.1 Hypothèses et prédictions de la théorie neutre

Comment faire le lien entre les processus écologiques qui façonnent la structure locale des communautés, et ceux qui façonnent sur le long terme les espèces rendues disponibles par la diversité des processus évolutifs? Pour répondre à cette question, la théorie neutre (Hubbell 2001) commence par faire un certain nombre d'hypothèses. La première, est l'hypothèse de neutralité qui postule que tous les individus d'une communauté sont équivalents en termes de survie, de reproduction et de migration. La spéciation est également neutre. Des individus variants apparaissent aléatoirement et leur descendance constitue immédiatement une nouvelle espèce à part entière. Autrement dit, la théorie neutre propose un modèle d'assemblage des communautés sans différence de niche, sans filtre environnemental et sans sélection naturelle. De plus, le nombre d'individus que contient la communauté est constant. Cette capacité biotique est à la fois limitée et saturée, de sorte que lorsqu'un individu meurt il est instantanément remplacé, soit par un descendant de la même communauté, soit par immigration depuis une autre communauté de la métacommunauté. Enfin, cette capacité du milieu est constante au cours du temps. Tout ceci contraint la dynamique locale des espèces et aboutit à une structuration des abondances des espèces, qui dépend de l'immigration, et de la taille des communautés déterminant la dérive. L'ensemble de ces communautés reliées entre elles par migration constitue une métacommunauté, dans laquelle les espèces apparaissent par spéciation et disparaissent par extinction. L'équilibre à l'échelle de la métacommunauté entre extinction et spéciation détermine la diversité régionale, et est décrit par le paramètre θ . Ainsi avec seulement deux paramètres ; I qui représente l'équilibre entre migration et dérive, et θ celui entre la spéciation et l'extinction, le modèle de la théorie neutre permet de faire des prédictions sur les abondances relatives des espèces, en l'absence de différences de niche. L'hypothèse de neutralité en fait un modèle fondamentalement peu réaliste, mais l'importance accordée aux dynamiques stochastiques à de multiples échelles le rend suffisamment complexe pour avoir des propriétés non-triviales. Certaines distributions d'abondance sont mieux prédites avec le modèle neutre qu'avec les lois d'échantillonnage classique, comme la distribution log-normale, la série géométrique ou la "broken-stick". C'est le cas des forêts tropicales pour lesquelles le modèle neutre permet d'expliquer le grand nombre d'espèces rares dans ces milieux, comme résultant des fluctuations aléatoires des abondances des espèces et la dispersion limitée (Hubbell 2001). D'autres études ont depuis montré que le modèle de la théorie neutre prédit étonnamment bien les distributions d'abondance (Hubbell 2001 ; Etienne 2005) et les courbes de aire-espèces (Rosindell and Cornell 2007 ; Rosindell and Cornell 2009) de nombreux systèmes allant des forêts tropicales aux communautés aquatiques et microbiennes (Volkov et al. 2007 ; Muneeppeerakul et al. 2008).

3.1.2 Unification des échelles écologiques et évolutive

En somme, la théorie neutre permet de prendre en compte l'effet des aléas de la colonisation et de dérive dans des communautés locales, et de la diversification régionale. La combinaison de

ces différents effets aléatoires structurent la diversité neutre, à la fois locale et régionale, via les processus écologiques locaux et les processus évolutifs régionaux. Dans la théorie neutre, il y a donc un couplage entre les dynamiques démographiques locales et les dynamiques régionales de diversification. Bien que ce soit dans un contexte neutre, en ce sens, la théorie neutre réconcilie les échelles écologiques et évolutives. Ainsi, parmi les nombreux modèles de diversification, les modèles basés sur la théorie neutre ont longtemps été parmi les seuls utilisant la dynamique des individus pour obtenir des phylogénies prédites, qui peuvent être ajustées à des données empiriques (Morlon 2014). C'est pourquoi la théorie neutre constitue un bon point de départ pour l'étude des dynamiques éco-évolutives.

3.2 Limites de la théorie neutre

3.2.1 Limites de la stationnarité temporelle de l'environnement

La théorie neutre de la biodiversité fait plusieurs hypothèses limitantes hormis celle de neutralité. La première: il n'y a pas de fluctuations de la taille des communautés au cours du temps. Ainsi, l'approche éco-évolutive de la théorie neutre suppose un état d'équilibre entre la biodiversité et la configuration actuelle de l'habitat. Et bien que de telles fluctuations dans les tailles des populations aient été prises en compte dans les modèles phylogéographiques portant sur l'étude de la diversité génétique (Avice 2009), cela reste rare dans les modèles biogéographiques multi-espèces (Overcast, Emerson, and Hickerson 2019 ; Overcast, Ruffley, et al. 2021). Il est pourtant essentiel de pouvoir intégrer l'impact des variations de la configuration de l'environnement passé sur la diversité actuelle. En effet, le temps de génération assez long de certains organismes peuvent entraîner des dynamiques transitoires, auquel cas les dynamiques démographiques des espèces seront sensibles à des fluctuations environnementales, même anciennes. De manière générale, les fluctuations environnementales peuvent façonner la trajectoire évolutive des organismes en impactant la coexistence des espèces, les opportunités de migration, les risques d'extinctions et les voies de spéciation au cours du temps.

3.2.2 Limites de la spéciation ponctuelle

Un autre aspect limitant de la théorie de Hubbell concerne la spéciation ponctuelle. Cette vision de la spéciation comme un événement instantané, engendre plusieurs prédictions peu réalistes comme l'apparition trop fréquente de nouvelles espèces, une durée de vie moyenne des espèces trop courte et l'excès d'espèces ne présentant qu'un ou deux individus (Ricklefs 2003 ; Nee 2005 ; Ricklefs 2006). Le fait est que la spéciation est un processus qui prend du temps. Ce constat a encouragé le développement de nouveaux modèles de spéciation dite "prolongée". Dans ce cas on ne considère plus la mutation ponctuelle comme un événement instantané mais comme l'initiation d'un processus de spéciation long qui ne produit une nouvelle espèce qu'après l'écoulement d'une période de transition (Rosindell, Cornell, et al. 2010).

3.3 Développer de nouveaux modèles éco-évolutifs hérités de la théorie neutre

Le couplage de différents processus à différentes échelles spatiales et temporelles, permis par le cadre de modélisation de la théorie neutre, en fait néanmoins une base intéressante pour de nouveaux

modèles éco-évolutifs. L'enjeu étant de relâcher l'hypothèse de stationnarité de l'environnement ; en permettant des changements de la configuration de l'environnement par vicariance, ainsi que la fluctuation de la taille des communautés, et d'en examiner les effets sur différents aspects de la biodiversité.

3.3.1 Sur les propriétés du coalescent

Il existe au moins deux façons de se représenter les dynamiques démographiques d'une population. La première, et probablement la plus intuitive, consiste à suivre le devenir de chaque individu. Rappelons qu'ils peuvent migrer, mourir ou se reproduire auquel cas on s'intéresse au devenir de leur descendance et ainsi de suite. Si l'on associe des probabilités à chacun de ces événements, on peut alors conclure sur la composition de la population après un certain nombre de générations. Toutefois, il y a de fortes chances que des individus soient morts sans avoir eu de descendance à la fin de la simulation, et donc d'utiliser du temps de calcul inutilement. Mais on peut également se représenter le processus dans le sens inverse. Imaginons une collection d'individus dans le présent, à la génération précédente, deux d'entre eux, pris au hasard, peuvent descendre du même ancêtre, avec une probabilité d'autant plus grande que la population est petite. Dans ce cas, on dira que les lignées issus de ces deux individus coalescent. Avec la théorie du coalescent, on peut ainsi retracer l'ascendance de n individus (qui sera alors composée d'exactly $n-1$ événements de coalescence) dans laquelle des paires d'individus se rejoignent en remontant dans le temps (Wakeley 2009). La probabilité de coalescence à chaque génération étant liée à la taille de la population, la généalogie ainsi obtenue traduit directement les effets d'une taille de population contrainte, sur la démographie des individus qui la composent au cours du temps (Figure I.11). Si l'on considère que l'environnement physique contraint le nombre d'individus que peut contenir une population, alors ce type de représentation est un bon moyen d'appréhender l'effet des fluctuations environnementales passées sur la démographie d'une collection d'individus (Etienne 2005). Par ailleurs, la modélisation par coalescence permet d'appréhender l'effet de la dérive sur une communauté d'espèces. En effet, en l'absence d'autres effets, la fixation d'une seule espèce (dont l'identité est déterminée par hasard sous l'hypothèse de neutralité) se traduit par une généalogie aboutissant à un ancêtre commun unique. La longueur des branches de cette généalogie dépendra alors de la taille de la communauté - qui détermine la force de la dérive. Cependant la migration compense l'appauvrissement dû à la dérive, de sorte que tous les membres de la communauté ne partagent pas forcément le même ancêtre, puisqu'il peut s'agir d'un migrant. Un autre avantage non négligeable de la modélisation par coalescence vient du fait que seules les lignées avec des descendants actuels sont représentées, ce qui en fait une alternative beaucoup plus rapide aux méthodes dites « forward in time ».

3.3.2 Réconcilier la niche et la neutralité (*ecolottery*)

Toutefois, la modélisation par coalescence repose initialement sur l'hypothèse de neutralité, du fait qu'aucun ancêtre n'est plus susceptible qu'un autre d'engendrer une descendance. Cela dit, en conditionnant l'établissement et la survie des descendants dans la généalogie, au passage d'un filtre environnemental, la modélisation par coalescence peut être adaptée, pour à la fois refléter l'effet

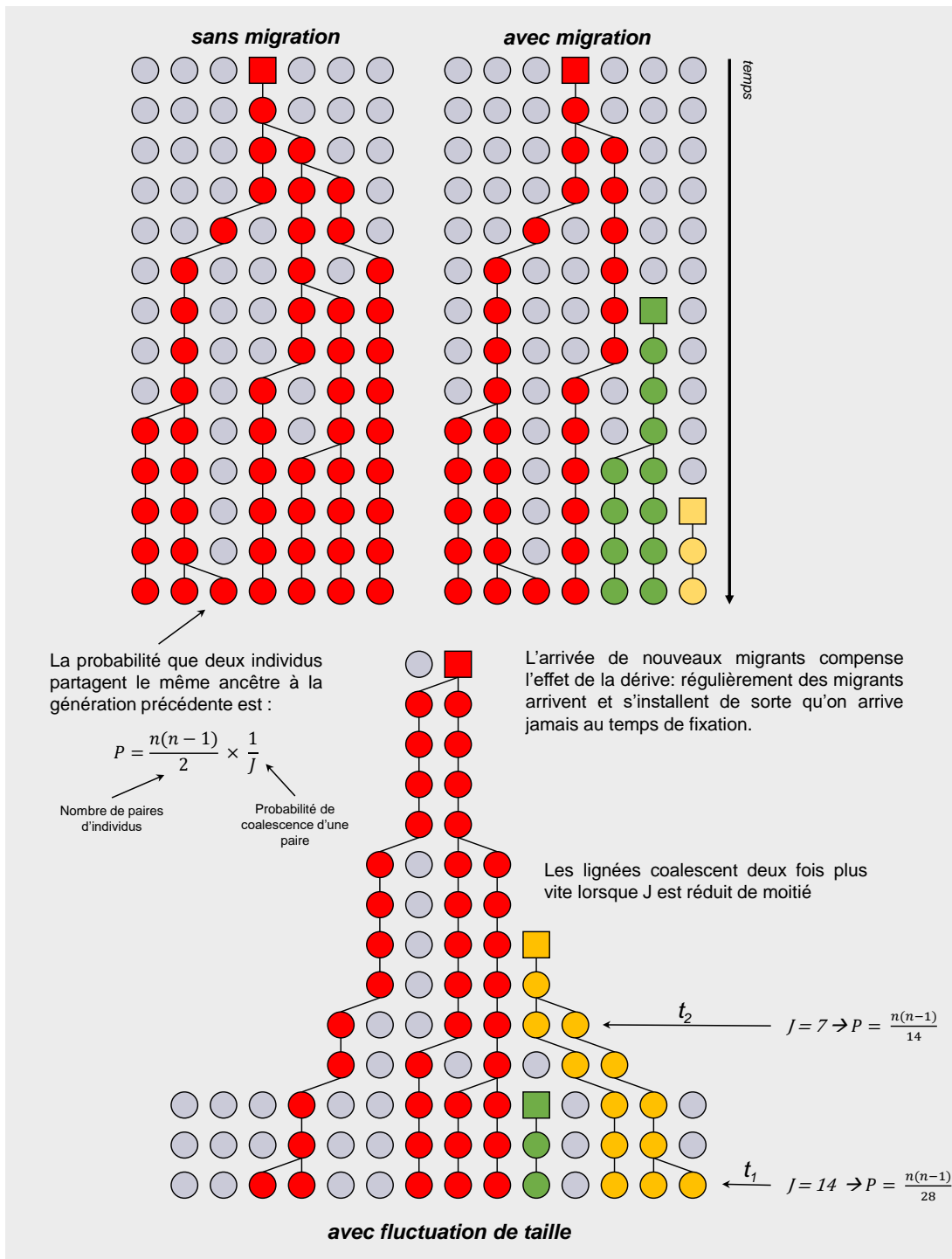


Figure I.11: Sur les propriétés du coalescent

de la dérive neutre et des différences de niches (*ecolottery*, Munoz, Grenié, et al. 2018). Dans ce cas, la probabilité qu'un ancêtre migrant de la généalogie soit tiré depuis le pool régional, dépend directement de l'adéquation de sa valeur de trait au filtre environnemental - qui détermine quelles valeurs de traits garantissent un établissement optimal. Ainsi, dans une perspective où il n'y a pas

de limite stricte entre les effets des dynamiques neutres et ceux des dynamiques de niche, mais plutôt un continuum de leur importance relative (Gravel et al. 2006), ce type de méthode permet de situer une communauté le long de ce gradient.

3.3.3 Intégrer le rôle de l'histoire (*ecophylo*)

Cependant, cette approche fait l'hypothèse d'un pool régional stationnaire et n'intègre donc pas comment les dynamiques écologiques et évolutives s'influencent mutuellement. Grâce à la modélisation par coalescence, il est possible d'obtenir des généalogies dont la topologie dépend de changements dans la configuration de l'environnement physique, contraignant la taille d'une population à différents moments dans le passé (Figure I.11). Ainsi il est possible d'utiliser cette approche pour représenter l'influence de fluctuations environnementales sur l'écologie d'individus au cours du temps. Reste la question de l'évolution des espèces. Le passage d'une généalogie d'individus à une communauté d'espèces passe par la considération des dynamiques de diversification. Considérons deux individus pris au hasard dans la généalogie. Si un grand nombre de génération les séparent alors il est probable que des différences génétiques se soient accumulées et qu'ils ne soient plus interféconds. Ainsi, la spéciation résulte naturellement de l'accumulation de mutations ponctuelles au cours du temps (Manceau, Lambert, and Morlon 2015). En fusionnant les branches de la généalogie en fonction des mutations accumulées par les individus qui en descendent, il est possible de générer des communautés d'espèces actuelles dont les dynamiques éco-évolutives résultent de fluctuations environnementales passées.

4 Comprendre et prédire l'impact de fluctuations environnementales à de multiples échelles spatiales et temporelles

Puisque les processus biologiques qui influencent la diversité des communautés sont multiples et interagissent à différentes échelles spatiales et temporelles, l'approche de modélisation des dynamiques éco-évolutives se doit de faire un effort explicite d'intégration des processus à différentes échelles. Ce travail de thèse repose sur la volonté de représenter explicitement les processus microscopiques - i.e. le comportement des individus - afin de prédire les propriétés émergentes des communautés d'espèces.

Dans cette logique, un premier objectif de la thèse fut de proposer une nouvelle approche de modélisation éco-évolutive, basée sur la simulation des dynamiques des individus, pour prédire les trajectoires évolutives des espèces dans un contexte de fluctuations environnementales. Inspiré de la théorie neutre, le modèle développé lors de la thèse utilise une approche par coalescence pour explicitement représenter, sous la forme d'une généalogie, l'effet d'un environnement fluctuant sur la démographie des individus. Le couplage entre démographie des individus et diversification des espèces est permis par un modèle de spéciation prolongée, et permet d'obtenir des prédictions sur différentes facettes de la biodiversité actuelle en fonction de changements anciens.

Dans un second temps, ce travail s'est attaché à l'étude de l'influence jointe des processus déterministes et stochastiques dans l'assemblage des communautés à l'échelle locale. De même, l'utilisation d'une approche par simulation explicite des processus permet d'inférer - lorsqu'il n'y a pas de limite stricte entre les effets des processus neutres et ceux des dynamiques de niche, mais plutôt un continuum de leur importance relative - où se placent les communautés le long de ce gradient, en fonction du contexte environnemental. Ainsi, inférer la contribution relative de processus qui s'opposent, à partir de l'observation des patrons de diversité, constitue un second objectif de cette thèse.

Enfin, un dernier chapitre de cette thèse est consacré à l'évaluation de la signature sur le long terme de processus historiques sur la biodiversité actuelle à l'échelle régionale. Dans une perspective où les communautés d'espèces ne sont pas à l'équilibre avec les conditions environnementales actuelles, mais restent influencées par des événements historiques anciens, du fait de dynamiques internes lentes, l'émergence de dynamiques transitoires permet de détecter l'influence de tels événements. C'est pourquoi il fut nécessaire de savoir quelles peuvent être les conséquences d'événements historiques sur les patrons de diversité, et si ceux-ci peuvent en porter la signature. L'utilisation d'expériences virtuelles relevant du problème direct peuvent ainsi bénéficier la résolution du problème inverse.

Références

- Abellán, P. and Svenning, J.-C. (2014). “Refugia within refugia—patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula”. In: *Biological Journal of the Linnean Society* vol. 113, no. 1. ISBN: 0024-4066 Publisher: Oxford University Press, pp. 13–28.
- Avise, J. C. (2009). “Phylogeography: retrospect and prospect”. In: *Journal of Biogeography* vol. 36, pp. 3–15.
- Barker, G. M. (2002). “Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation”. In: *Biological Journal of the Linnean Society* vol. 76, no. 2. ISBN: 1095-8312 Publisher: Oxford University Press, pp. 165–194.
- Blanchard, G. and Munoz, F. (2022). “Causes and Consequences of Extinction Debts: Perspectives for Historical Ecology and Biological Conservation”. In: *Historical Ecology: Learning from the Past to Understand the Present and Forecast the Future of Ecosystems*. Publisher: Wiley Online Library, pp. 273–284.
- Bonpland, A. (1807). *Essai Sur La Géographie Des Plantes, Accompagné D’Un Tableau Physique Des Régions Équinoxiales: Fondé sur des mesures exécutées, depuis le dixième degré de latitude boréale jusqu’au dixième degré de latitude australe, pendant les années 1799, 1800, 1801, 1802 et 1803*. Vol. 5. Schoell.
- Briggs, C. J. and Hoopes, M. F. (2004). “Stabilizing effects in spatial parasitoid–host and predator–prey models: a review”. In: *Theoretical population biology* vol. 65, no. 3. ISBN: 0040-5809 Publisher: Elsevier, pp. 299–315.
- Brown, J. H. (1995). *Macroecology*. University of Chicago Press.
- Chesson, P. (2000). “Mechanisms of maintenance of species diversity”. In: *Annual Review of Ecology and Systematics* vol. 31, pp. 343–366.
- Cornwell, W. K. et al. (2014). “Functional distinctiveness of major plant lineages”. In: *Journal of Ecology* vol. 102, pp. 345–356.
- Croizat, L. (2013). *Manual of phytogeography: an account of plant-dispersal throughout the world*. Springer.
- Croizat, L. (1982). “Vicariance/vicariism, panbiogeography, vicariance biogeography, etc.: A clarification”. In: *Systematic Zoology*. ISBN: 0039-7989 Publisher: JSTOR, pp. 291–304.
- Darwin, C. (2011). *The origin of species*. William Collins. Collins Classics.
- De Candolle, A. (1855). *Géographie botanique raisonnée ou exposition des faits principaux et des lois concernant la distribution géographique des plantes de l’époque actuelle*. Vol. 2.
- Denelle, P., Violle, C., and Munoz, F. (2019). “Distinguishing the signatures of local environmental filtering and regional trait range limits in the study of trait–environment relationships”. In: *Oikos*.
- Descola, P. (2010). “L’écologie des autres: L’anthropologie et la question de la nature”. In: *L’écologie des autres*. Publisher: Editions Quae, pp. 1–112.
- Elton, C. S. (1958). *The ecology of invasions by animals and plants*. University of Chicago Press.
- (2001). *Animal ecology*. University of Chicago Press.

- Enquist, B. et al. (2015). “Scaling from traits to ecosystems: Developing a general Trait Driver Theory via integrating trait-based and metabolic scaling theories”. In: *Advances in Ecological Research* vol. 52.
- Escudero, M. et al. (2010). “Bipolar disjunctions in *Carex*: long-distance dispersal, vicariance, or parallel evolution?” In: *Flora-Morphology, Distribution, Functional Ecology of Plants* vol. 205, no. 2. ISBN: 0367-2530 Publisher: Elsevier, pp. 118–127.
- Etienne, R. S. (2005). “A new sampling formula for neutral biodiversity”. In: *Ecology Letters* vol. 8, pp. 253–260.
- Faith, D. P. (1992). “Conservation evaluation and phylogenetic diversity”. In: *Biological conservation* vol. 61, no. 1, pp. 1–10.
- Faria, F. (2013). “Georges Cuvier et le premier paradigme de la paléontologie”. In: *Revue de Paléobiologie, Genève* vol. 32, no. 2, pp. 297–302.
- Futuyma, D. J. (1979). *Evolutionary Biology*. Sunderland, Mass: Sinauer Associates.
- Gause, G. F. (1934). “The struggle for existence Williams and Wilkins”. In: *Baltimore, Maryland*.
- Gotelli, N. J. and Colwell, R. K. (2001). “Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness”. In: *Ecology Letters* vol. 4, pp. 379–391.
- Gravel, D. et al. (2006). “Reconciling niche and neutrality: the continuum hypothesis”. In: *Ecology Letters* vol. 9, pp. 399–409.
- Heads, M. (2009). “Darwin’s changing views on evolution: from centres of origin and teleology to vicariance and incomplete lineage sorting”. In: *Journal of Biogeography* vol. 36, no. 6, pp. 1018–1026.
- Holt, B. G. et al. (2013). “An update of Wallace’s zoogeographic regions of the world”. In: *Science* vol. 339, no. 6115. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 74–78.
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton and Oxford: Princeton University Press.
- Humboldt, A. v. and Bonpland, A. (1805). “Géographie des Plantes Equinoxiales. Tableau physique des Andes et Pays voisins. Dressé d’après des Observations & des Mesures prises sur les Lieux depuis le 10e degré de latitude boréale jusqu’au 10e de latitude australe en 1799, 1800, 1801, 1802 et 1803”. In: Publisher: Humboldt, Alexander von.
- Huth, G. et al. (2015). “Long-distance rescue and slow extinction dynamics govern multiscale metapopulations”. In: *The American Naturalist* vol. 186, no. 4. ISBN: 0003-0147 Publisher: University of Chicago Press Chicago, IL, pp. 460–469.
- Jackson, R. C. (1976). “Evolution and systematic significance of polyploidy”. In: *Annual Review of Ecology and Systematics*. ISBN: 0066-4162 Publisher: JSTOR, pp. 209–234.
- Jackson, S. T. and Sax, D. F. (2010). “Balancing biodiversity in a changing environment: extinction debt, immigration credit and species turnover”. In: *Trends in ecology & evolution* vol. 25, no. 3. ISBN: 0169-5347 Publisher: Elsevier, pp. 153–160.
- Jacob, F. (1970). “La logique du vivant”. In: Publisher: Gallimard.
- Keddy, P. A. (1992). “Assembly and response rules: two goals for predictive community ecology”. In: *Journal of Vegetation Science* vol. 3, no. 2, pp. 157–164.

- Lavorel, S. and Garnier, É. (2002). “Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail”. In: *Functional ecology* vol. 16, no. 5. ISBN: 0269-8463 Publisher: Wiley Online Library, pp. 545–556.
- Leibold, M. A. et al. (2004). “The metacommunity concept: a framework for multi-scale community ecology”. en. In: *Ecology Letters* vol. 7, no. 7, pp. 601–613.
- Lessard, J.-P. et al. (2012). “Inferring local ecological processes amid species pool influences”. In: *Trends Ecol Evol* vol. 27, pp. 600–607.
- Levin, S. A. (1992). “The problem of pattern and scale in ecology”. In: *Ecology* vol. 73, pp. 1943–1967.
- Levins, R. (1968). *Evolution in changing environments: some theoretical explorations*. Princeton University Press.
- Lortie, C. J. et al. (2004). “Rethinking plant community theory”. In: *Oikos* vol. 107, no. 2, pp. 433–438.
- Lozupone, C. and Knight, R. (2005). “UniFrac: a new phylogenetic method for comparing microbial communities”. In: *Applied and environmental microbiology* vol. 71, no. 12. ISBN: 0099-2240 Publisher: Am Soc Microbiol, pp. 8228–8235.
- MacArthur, R. and Levins, R. (1967). “The limiting similarity, convergence, and divergence of coexisting species”. In: *The american naturalist* vol. 101, no. 921. ISBN: 0003-0147 Publisher: Science Press, pp. 377–385.
- MacArthur, R. and Wilson, E. (1967). *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press.
- Manceau, M., Lambert, A., and Morlon, H. (2015). “Phylogenies support out-of-equilibrium models of biodiversity”. In: *Ecology Letters* vol. 18, pp. 347–356.
- May, R. M. (1975). “Patterns of species abundance and diversity”. In: *Ecology and evolution of communities*. Publisher: Harvard University Press, pp. 81–120.
- Mazel, F. et al. (2016). “Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics”. In: *Ecography* vol. 39, no. 10. ISBN: 0906-7590 Publisher: Wiley Online Library, pp. 913–920.
- McGill, B. J. et al. (2007). “Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework”. In: *Ecology letters* vol. 10, no. 10. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 995–1015.
- McShea, D. W. and Brandon, R. N. (2010). *Biology’s first law: the tendency for diversity and complexity to increase in evolutionary systems*. University of Chicago Press.
- Morlon, H. (2014). “Phylogenetic approaches for studying diversification”. en. In: *Ecology Letters* vol. 17, no. 4. Ed. by Mooers, A., pp. 508–525.
- Morrone, J. J. (2008). *Evolutionary biogeography: an integrative approach with case studies*. Columbia University Press.
- Mouquet, N. and Loreau, M. (2003). “Community patterns in source-sink metacommunities”. 0003-0147. In: *American Naturalist* vol. 162, pp. 544–557.
- Muneepeerakul, R. et al. (2008). “Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin”. In: *Nature* vol. 453, pp. 220–222.

- Munoz, F., Couteron, P., and Ramesh, B. R. (2008). “Beta diversity in spatially implicit neutral models: a new way to assess species migration”. In: *The American Naturalist* vol. 172, no. 1. ISBN: 0003-0147 Publisher: The University of Chicago Press, pp. 116–127.
- Munoz, F., Grenié, M., et al. (2018). “ecolottery : Simulating and assessing community assembly with environmental filtering and neutral dynamics in R”. en. In: *Methods in Ecology and Evolution* vol. 9, no. 3. Ed. by Chisholm, R., pp. 693–703.
- Munoz, F. and Huneman, P. (2016). “From the neutral theory to a comprehensive and multiscale theory of ecological equivalence”. In: *The Quarterly Review of Biology* vol. 91, no. 3. ISBN: 0033-5770 Publisher: University of Chicago Press Chicago, IL, pp. 321–342.
- Münkemüller, T. et al. (2012). “From diversity indices to community assembly processes: a test with simulated data”. In: *Ecography* vol. 35, no. 5. ISBN: 0906-7590 Publisher: Wiley Online Library, pp. 468–480.
- Nee, S. (2005). “The neutral theory of biodiversity: do the numbers add up?” In: *Functional Ecology* vol. 19, no. 1. ISBN: 0269-8463 Publisher: JSTOR, pp. 173–176.
- Nei, M. and Tajima, F. (1981). “Genetic drift and estimation of effective population size”. In: *Genetics* vol. 98, no. 3. ISBN: 1943-2631 Publisher: Oxford University Press, pp. 625–640.
- Odum, E. P. and Barrett, G. W. (1971). *Fundamentals of ecology*. Vol. 3. Saunders Philadelphia.
- Ordonez, A. and Svenning, J.-C. (2015). “Geographic patterns in functional diversity deficits are linked to glacial-interglacial climate stability and accessibility”. In: *Global Ecology and Biogeography* vol. 24, no. 7. ISBN: 1466-822X Publisher: Wiley Online Library, pp. 826–837.
- (2016a). “Functional diversity of North American broad-leaved trees is codetermined by past and current environmental factors”. In: *Ecosphere* vol. 7, no. 2.
- (2016b). “Strong paleoclimatic legacies in current plant functional diversity patterns across Europe”. In: *Ecology and Evolution* vol. 6, pp. 3405–3416.
- Overcast, I., Emerson, B. C., and Hickerson, M. J. (2019). “An integrated model of population genetics and community ecology”. en. In: *Journal of Biogeography* vol. 46, no. 4, pp. 816–829.
- Overcast, I., Ruffley, M., et al. (2021). “A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities”. In: *Molecular ecology resources* vol. 21, no. 8. ISBN: 1755-098X Publisher: Wiley Online Library, pp. 2782–2800.
- Pennington, R. T. and Dick, C. W. (2004). “The role of immigrants in the assembly of the South American rainforest tree flora”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 359, pp. 1611–1622.
- Platon (n.d.). *La république*. Les Belles Lettres. Collection des universités de France.
- Posadas, P., Crisci, J. V., and Katinas, L. (2006). “Historical biogeography: a review of its basic concepts and critical issues”. In: *Journal of Arid Environments* vol. 66, no. 3. ISBN: 0140-1963 Publisher: Elsevier, pp. 389–403.
- Quinn, J. F. and Dunham, A. E. (1983). “On hypothesis testing in ecology and evolution”. In: *The American Naturalist* vol. 122, no. 5. ISBN: 0003-0147 Publisher: University of Chicago Press, pp. 602–617.
- Ricklefs, R. (1987). “Community diversity: relative roles of local and regional processes”. In: *Science* vol. 235, no. 4785. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 167–171.

- (2003). “A comment on Hubbell’s zero-sum ecological drift model”. In: *Oikos* vol. 100, pp. 185–192.
- (2006). “The unified neutral theory of biodiversity: do the numbers add up?” en. In: *Ecology* vol. 87, no. 6, pp. 1424–1431.
- (2008). “Disintegration of the ecological community: American Society of Naturalists Sewall Wright award winner address”. In: *The American Naturalist* vol. 172, no. 6, pp. 741–750.
- Ronquist, F. (1997). “Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography”. In: *Systematic biology* vol. 46, no. 1. ISBN: 1076-836X Publisher: Society of Systematic Biologists, pp. 195–203.
- Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge university press.
- Rosindell, J. and Cornell, S. J. (2007). “Species–area relationships from a spatially explicit neutral model in an infinite landscape”. In: *Ecology letters* vol. 10, no. 7. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 586–595.
- (2009). “Species–area curves, neutral models, and long-distance dispersal”. In: *Ecology* vol. 90, no. 7. ISBN: 1939-9170 Publisher: Wiley Online Library, pp. 1743–1750.
- Rosindell, J., Cornell, S. J., et al. (2010). “Protracted speciation revitalizes the neutral theory of biodiversity”. In: *Ecology Letters* vol. 13, pp. 716–727.
- Scheffer, M. et al. (2003). “Why plankton communities have no equilibrium: solutions to the paradox”. In: *Hydrobiologia* vol. 491, no. 1. ISBN: 1573-5117 Publisher: Springer, pp. 9–18.
- Schleuter, D. et al. (2010). “A user’s guide to functional diversity indices”. In: *Ecological monographs* vol. 80, no. 3. ISBN: 1557-7015 Publisher: Wiley Online Library, pp. 469–484.
- Schoener, T. W. (1986). “Mechanistic approaches to community ecology: a new reductionism”. In: *American Zoologist* vol. 26, no. 1. ISBN: 0003-1569 Publisher: Oxford University Press UK, pp. 81–106.
- Shipley, B. (2016). *Cause and correlation in biology: a user’s guide to path analysis, structural equations and causal inference with R*. Cambridge University Press.
- Swenson, N. G. (2011). “Phylogenetic Beta Diversity Metrics, Trait Evolution and Inferring the Functional Beta Diversity of Communities”. In: *Plos One* vol. 6, e21264.
- Tarantola, A. (2006). “Popper, Bayes and the inverse problem”. In: *Nature physics* vol. 2, no. 8, p. 492.
- Tilman, D., Reich, P. B., and Knops, J. M. H. (2006). “Biodiversity and ecosystem stability in a decade- long grassland experiment”. en. In: vol. 441, p. 4.
- Vellend, M. (2010). “Conceptual synthesis in community ecology”. In: *The Quarterly review of biology* vol. 85, no. 2, pp. 183–206.
- Violle, C., Navas, M. L., et al. (2007). “Let the concept of trait be functional!” In: *Oikos* vol. 116, pp. 882–892.
- Violle, C., Enquist, B. J., et al. (2012). “The return of the variance: intraspecific variability in community ecology”. en. In: *Trends in Ecology & Evolution* vol. 27, no. 4, pp. 244–252.
- Violle, C., Nemergut, D. R., et al. (2011). “Phylogenetic limiting similarity and competitive exclusion: Phylogenetic relatedness and competition”. en. In: *Ecology Letters* vol. 14, no. 8, pp. 782–787.
- Volkov, I. et al. (2007). “Patterns of relative species abundance in rainforests and coral reefs”. In: *Nature* vol. 450, no. 7166. ISBN: 1476-4687 Publisher: Nature Publishing Group, pp. 45–49.
- Wakeley, J. (2009). “Coalescent theory”. In: *Roberts & Company*.

- Whittaker, R. H. (1975). *Communities and ecosystems*. New York: MacMillan Publishing.
- Yu, Y., Harris, A. J., and He, X. (2010). “S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories”. In: *Molecular phylogenetics and Evolution* vol. 56, no. 2. ISBN: 1055-7903 Publisher: Elsevier, pp. 848–850.
- Zurell, D. et al. (2010). “The virtual ecologist approach: simulating data and observers”. en. In: *Oikos* vol. 119, no. 4, pp. 622–635.

Chapitre I

**Développements algorithmiques et
statistiques pour l'étude des
dynamiques éco-évolutives**

Tandis que l'écologie s'intéresse aux dynamiques à court terme aboutissant à la coexistence d'espèces en fonction de leurs adaptations présentes à l'environnement local, l'évolution se préoccupe de l'apparition et la disparition d'espèces et de leurs caractéristiques biologiques à de larges échelles spatiales et temporelles. Bien que longtemps l'objet de disciplines distinctes, l'intégration de ces deux types de processus devient nécessaire pour appréhender certaines questions biologiques, notamment celle de l'influence d'événements passés sur les dynamiques transitoires lente de certaines communautés. Ce travail de thèse se base essentiellement sur une analogie entre les mécanismes élémentaires qui structurent la diversité des génotypes et ceux qui structurent la diversité spécifique de communautés d'espèces. Par conséquent, les outils qui sont traditionnellement utilisés pour appréhender la diversité des génotypes peuvent l'être pour comprendre la diversité des espèces.

Ce chapitre synthétise le développement d'outils nouveaux pour l'étude des dynamiques éco-évolutives et pour la résolution plus générale du problème inverse. L'Article 1 présente le modèle développé lors de la thèse et la librairie python associée: `ecophylo`. Le modèle repose sur la théorie du coalescent pour simuler une ascendance partagée d'individus cooccurrents, sous l'influence de fluctuations démographiques passées dues, par exemple, à des fluctuations de l'habitat, à la fragmentation et/ou à des événements de migration entre des zones distinctes. Des mutations se produisent au fil du temps dans la généalogie, et les génotypes divergents représentent des espèces distinctes. Le logiciel comprend des outils permettant de simuler un grand nombre de communautés selon différentes gammes de paramètres afin de générer diverses statistiques de diversité taxonomiques et phylogénétiques. Cet article propose également une exploration du comportement du modèle en fonction de différentes dynamiques démographiques et de modalités de spéciation. De cette façon, il est possible d'examiner les conséquences de fluctuations de l'environnement passées sur les patrons de diversité qui en résultent. Cette logique relève du problème direct puisqu'elle prédit les propriétés de communautés virtuelles en fonction de paramètres du modèle. Cependant, dans la plupart des cas en écologie et en biogéographie, on ne cherche pas tant à modéliser et à prédire des patrons de diversité virtuels mais plutôt à déduire quels processus ont façonné nos observations. Ainsi, si les modèles écologiques peuvent être utiles pour prédire les patrons de biodiversité à partir des valeurs des paramètres d'un modèle (problème direct), ils peuvent être insuffisants pour déduire la nature des processus écologiques à partir de l'observation des patrons (problème inverse). Dans l'Article 2 nous soutenons que pour résoudre le problème inverse, nous devons d'abord considérer le problème direct, en proposant que de nombreuses expériences virtuelles soient utilisées pour déduire, par comparaison, comment les patrons de biodiversité observés sont façonnés, et ainsi résoudre le problème inverse. Dans cet article d'opinion, nous appelons à reconnaître l'influence des connaissances a priori sur notre capacité à tirer des conclusions et encourageons donc la quantification de la façon dont même le meilleur modèle peut être erroné afin d'améliorer nos choix de modélisation. Les nouveaux outils algorithmiques et approches méthodologiques présentés dans ce chapitre permettent ainsi d'aboutir à des conclusions plus robustes concernant l'influence des fluctuations environnementales passées sur les dynamiques éco-évolutives, et leurs impacts sur les patrons de diversité - en particulier si et quand différents processus interagissent conjointement.

Article 1

ecophylo: Simulating and assessing eco-evolutionary dynamics under past environmental changes in Python.

Elizabeth Barthélemy¹, Maxime Jaunatre¹, François Munoz¹

Submitted after revision to *Plos Computational Biology*.

Abstract

A major challenge is to bridge the gap between ecological and evolutionary models, in order to address eco-evolutionary dynamics. Specifically, past environmental changes have determined changing regimes of population and migration dynamics, yet existing approaches do not allow investigating the long-term influence of such historical changes on extant biodiversity patterns. We introduce the Python package *ecophylo* dedicated to coalescent-based simulation of neutral eco-evolutionary dynamics. The model rests on coalescent theory to simulate a shared ancestry of co-occurring individuals, under the influence of past demographic fluctuations due to, for example, habitat fluctuations, fragmentation and/or migration events among separate areas. Mutations occur over time in the genealogy, and divergent genotypes represent distinct extant species. The package includes tools to simulate large numbers of datasets and associated summary statistics, so that diverse patterns of taxonomic and phylogenetic compositions can be generated. We used these options to explore how past demographic fluctuations in habitat size and different speciation dynamics affect extant observable patterns using *in silico* experiments. We then assessed whether and how to recognize the influence of alternative models and infer parameter values from a set of diversity metrics using misclassification and cross-validation analyses. The source code is readily available on Github and easily installable via pip. We included the scripts used for the simulations which can be used as templates for user wishing to conduct similar experiments using the model.

¹LIPhy Université Grenoble Alpes, Grenoble, France

1 Introduction

Observed species distributions and biodiversity patterns are shaped by current ecological processes but also reflect the influence of past evolutionary and biogeographic dynamics (Svenning et al. 2015). For instance, past environmental variations (especially climate) influence habitat area and thus the population dynamics of multiple species within. In the case where these fluctuations occur rapidly compared to the expected time-to-equilibrium of speciation, migration and drift dynamics (for instance with long-lived organisms with slow population dynamics), current biodiversity patterns should retain the signature of past environmental fluctuations (Barthelemy et al. 2021). Hence, a fundamental goal of biogeography is to understand how changes in suitable environmental conditions due to past climatic and geomorphologic history have shaped biodiversity patterns observed at present (Bennett 1990).

Eco-evolutionary approaches try to reconcile short term ecological assembly and long term evolutionary dynamics which requires integrating on one hand the historical biogeography of species and on the other hand the dynamics of individuals forming populations in fluctuating environments (Vellend 2010). By failing to integrate population dynamics within species, traditional approaches (e.g. Pigot et al. 2010 ; Tello and Stevens 2012) cannot acknowledge how speciation, migration and drift together drive the variation in species abundances (as envisioned in Hubbell 2001, see Morlon 2014) under the influence of historical variations in environmental area and configuration. Such variation has been considered in phylogeographic models of genetic diversity with changing population size (Avice 2009), but seldom in multispecies biogeographic models of biodiversity dynamics (Barthelemy et al. 2021 ; Overcast, Ruffley, et al. 2021). Here we propose a novel eco-evolutionary, multispecies model, named *ecophylo*, to bridge the gap.

Hubbell's neutral theory of biogeography (NTB) has underlined the role of migration, speciation and drift in shaping patterns of taxonomic diversity in space and time (Hubbell 2001) but makes two limiting assumptions: (i) it assumes constancy of the size of a regional biota (called metacommunity) to predict regional composition under a speciation-drift equilibrium (Etienne et al. 2006 ; Jabot, Etienne, and Chave 2008) and (ii) species arise instantaneously following a single mutation process (but see Rosindell et al. 2010). Hence, most approaches investigating how intertwined ecological and evolutionary drivers shape the relative abundances of species have assumed an equilibrium state of biodiversity with present habitat availability and configuration. Yet, by allowing metacommunity size to vary at specific ages of the past we can relax the NTB's constancy assumption and account for vicariance and/or distinct demographic histories between metacommunities.

Here we propose a novel neutral coalescent-based approach in which we consider how assembly dynamics in temporally changing environments (ecological perspective) affect diversification over a long-term (evolutionary perspective). We introduce a new Python library dedicated to the extensive simulation of multispecies eco-evolutionary dynamics: *ecophylo*. We expose the logic and advantages of the approach to examine how past multi-species demographic, migration, and diversification dynamics have shaped extant patterns of taxonomic and phylogenetic diversity. We show how the library can be used to design *in silico* experiments (Zurell et al. 2010), and explore the influence

of eco-evolutionary parameters on diversity patterns. The package includes options to simulate large numbers of datasets over broad ranges of parameters and scenarios of past demographic events and fluctuations. We use these to assess the long-lasting influence of past changes in habitat size and of different speciation dynamics on extant observable patterns. Then we assess whether and how to recognize the influence of alternative models and infer parameter values from a set of diversity metrics (Csilléry, M. G. B. Blum, et al. 2010). We discuss the originality and novelty of the approach for inference of eco-evolutionary dynamics.

2 Design and Implementation

2.1 Coalescent-based simulation of assemblage genealogies

The core model rests on the fundamental hypothesis that stochastic population dynamics rule the patterns of relative species abundances in a given habitat over time, depending on both extant and past metacommunity sizes. The past metacommunity size at time t is henceforth denoted $Ne(t)$. Following the standard coalescent model (Hudson 1983), we assume that all individuals in a multi-species assemblage can reproduce freely (panmixia assumption), with equal fitness (neutral assumption). The dynamics of the assemblage can be traced through the structure of its genealogy, i.e. by tracing the shared co-ancestry of extant individuals backwards in time until a single common ancestor is found (Kingman 1982). Coalescence is an event such as two lineages at generation t share the same ancestor at generation $t - 1$, which defines a bifurcating node in the genealogy of individuals. Generations in the model are discrete and non-overlapping, and there are no polytomies in the genealogy. When the number of sampled individuals n at present is such as $n \ll Ne(0)$, the distribution of coalescence times can be approximated as an exponential distribution with parameter λ proportional to $1/(2Ne(t))$ (Wakeley 2009). Therefore, the distribution depends on the fluctuation of $Ne(t)$, thereby on the history of regional habitat e.g. throughout paleoenvironmentally distinct periods.

In addition, the genealogy depends on spatial structuring and migration dynamics across distinct assemblages. The msprime coalescent simulator (Kelleher and Lohse 2020) used in *ecophylo* can represent distinct metacommunities exchanging migrants at given rates supplied in a migration matrix. In a coalescent approach, individuals moving from metacommunity j to k backwards in time actually correspond to individuals migrating from metacommunity k to j forwards in time. Thus the migration rates in the migration matrix represent the proportion of immigrants received by a given metacommunity at each generation. The migration matrix can be parameterized to encompass a wide range of migration scenarios, and can also depend on time, so as to represent changing migration dynamics during biogeographic history. By combining past fluctuations of Ne and migration scenarios across metacommunities, the coalescent-based modeling approach can account for a variety of dispersal and vicariance events hypothesized to have shaped the biogeographic histories of extant metacommunities. The topology of the resulting genealogy then reflects the combined influence of migration and ecological drift in a finite fluctuating environment with possibly differing demographic histories linked by migration (Figure 1.1).

1. *ecophylo*: Simulating and assessing eco-evolutionary dynamics under past environmental changes in Python.

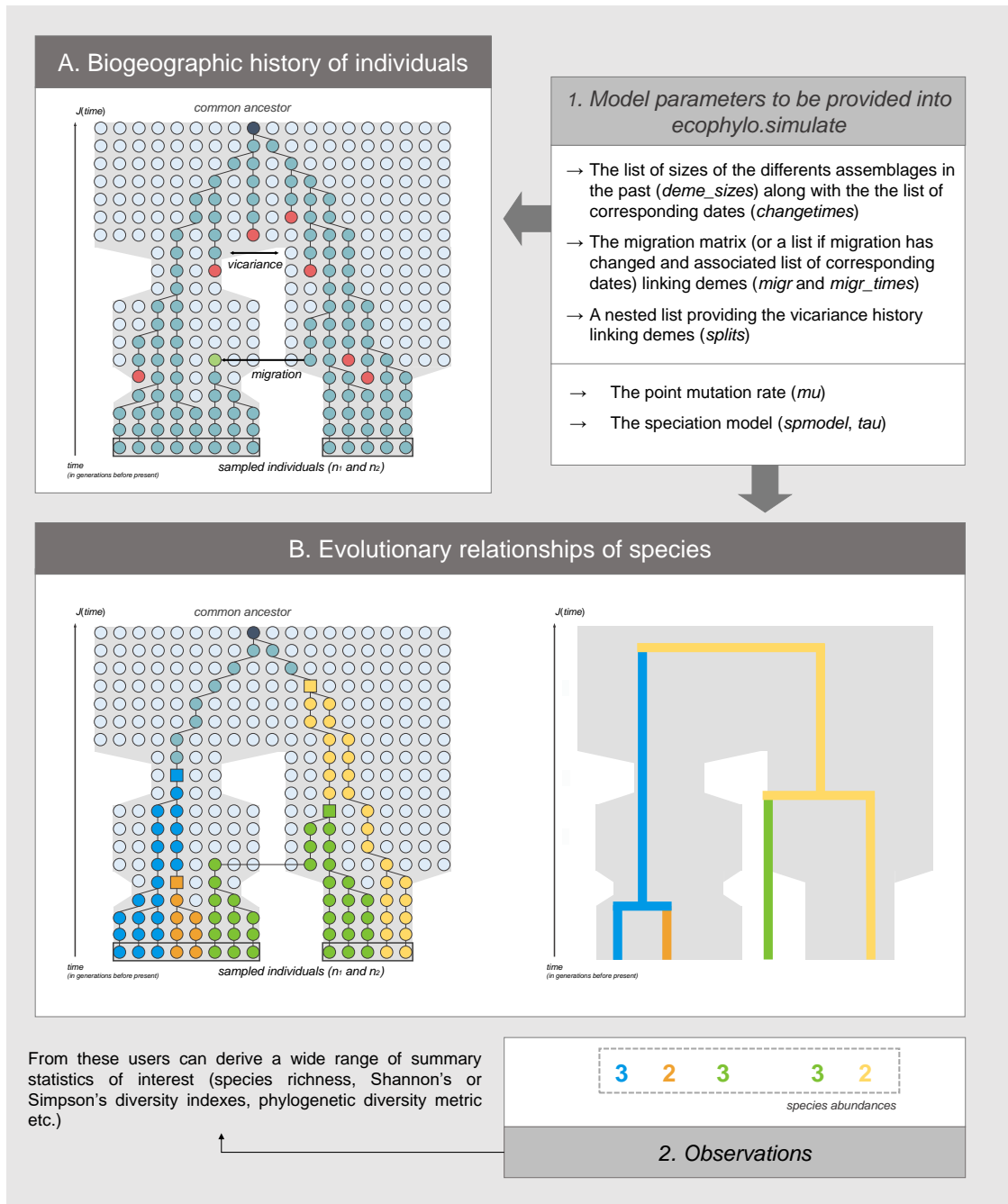


Figure 1.1: Conceptual diagram illustrating the primary components and associated parameters to the eco-evolutionary model *ecophylo*. Panel A shows how the model first reconstructs the biogeographic history of individuals having experienced past habitat fluctuations and/or fragmentation as well as migration. Grey shaded areas represent the size of suitable habitat conditioning the number of individuals over time. The shared ancestry of sampled individuals from each metacommunity is reconstructed using coalescent theory without assuming to which species they belong to until a common ancestor is found. Note that lineages culminating in local lineages extinctions (red circles) are depicted on the diagram but are not explicitly represented in the coalescent process (thus saving computation time). Panel B depicts how the model then determines the relations between species by sprinkling mutation events along the branches of the genealogy (squares) and then deriving the phylogenetic relationships amongst individuals and their abundances are obtained by merging paraphyletic clades into single species depending on the topology of the genealogy. Panel 1 gives details on the parameters that should be provided when using the *ecophylo* library to simulate such scenarios, while panel 2 shows the type of outputs that can be obtained from the simulations.

2.2 Phylogenetic reconstruction and species abundances

First, the algorithm simulates the genealogies of individuals in an assemblage experiencing past demographic fluctuations and/or linked by vicariance or migration events using the msprime coalescent simulator (Kelleher and Lohse 2020), without allotting a species label to each individual. Second, mutation events are sprinkled over the branches of simulated genealogies depending on branch lengths, so that the number of mutations over a branch follows a Poisson distribution with parameters $\mu \cdot B$, where μ is the point mutation rate and B is the length of the branch (in number of generations). The descendants of a branch with at least one mutation define a genetically distinct clade. Since an extant species should be a monophyletic genetic clade distinct from other species, all paraphyletic clades of haplotypes at present are merged to related clades to form a single monophyletic species (Manceau, Lambert, and Morlon 2015). Furthermore, in the NTB point mutations instantaneously give rise to a new species, in which case all branches descending from a mutation event are merged to form species. Yet, there is a general consensus that the mechanisms behind speciation are more complex and should at the very least spread out over more than a single generation. Thus considering speciation as a drawn out protracted process, whereby groups of genetically distinct individuals should only be recognized as new species after a given period of τ generations (Rosindell et al. 2010) meets more realistic expectations regarding the speciation process. Therefore, only monophyletic lineages with distinct genotypes and older than τ generations were considered a distinct species (Manceau, Lambert, and Morlon 2015). Both options (protracted vs. instantaneous) speciation models are implemented in the *ecophylo* library and can be used to derive the phylogenetic relationships among individuals as well as the number of individuals descending from a speciation event in the genealogy, which defined the species abundance in the sample at present (Figure 1.1). The *ecophylo* package essentially articulates itself around the *ecophylo.simulate* function that implements the core simulation algorithm presented above. It yields a phylogeny of species belonging to an observed assemblage (in Newick format) along with corresponding species abundances, for a combination of eco-evolutionary parameter values depending on past biogeographic history. Example scripts are provided in the supplementary materials, allowing users to simulate phylogenies by detailing the demographic history by not only considering demographic fluctuations but also including multiple metacommunities that can be linked by migration of varying intensity and/or originate from the split of an ancestral metacommunity.

3 Results

Whether to conduct virtual experiments examining how different eco-evolutionary scenarios have shaped patterns of diversity (see Barthelemy et al. 2021 for an illustration of the software) or to carry out inference (Csilléry, M. G. B. Blum, et al. 2010), *ecophylo* allows simulating large datasets over wide ranges of eco-evolutionary parameters. Specifically, the *ecophylo.dosimul* function can be used to call the underlying simulation algorithm mentioned above, and to retrieve summary statistics generated for different parameter values drawn from specified distributions. The parameter format makes it fairly straightforward to simulate large datasets over specified ranges of parameter values and is exemplified in the supplementary materials.

1. ecophylo: Simulating and assessing eco-evolutionary dynamics under past environmental changes in Python.

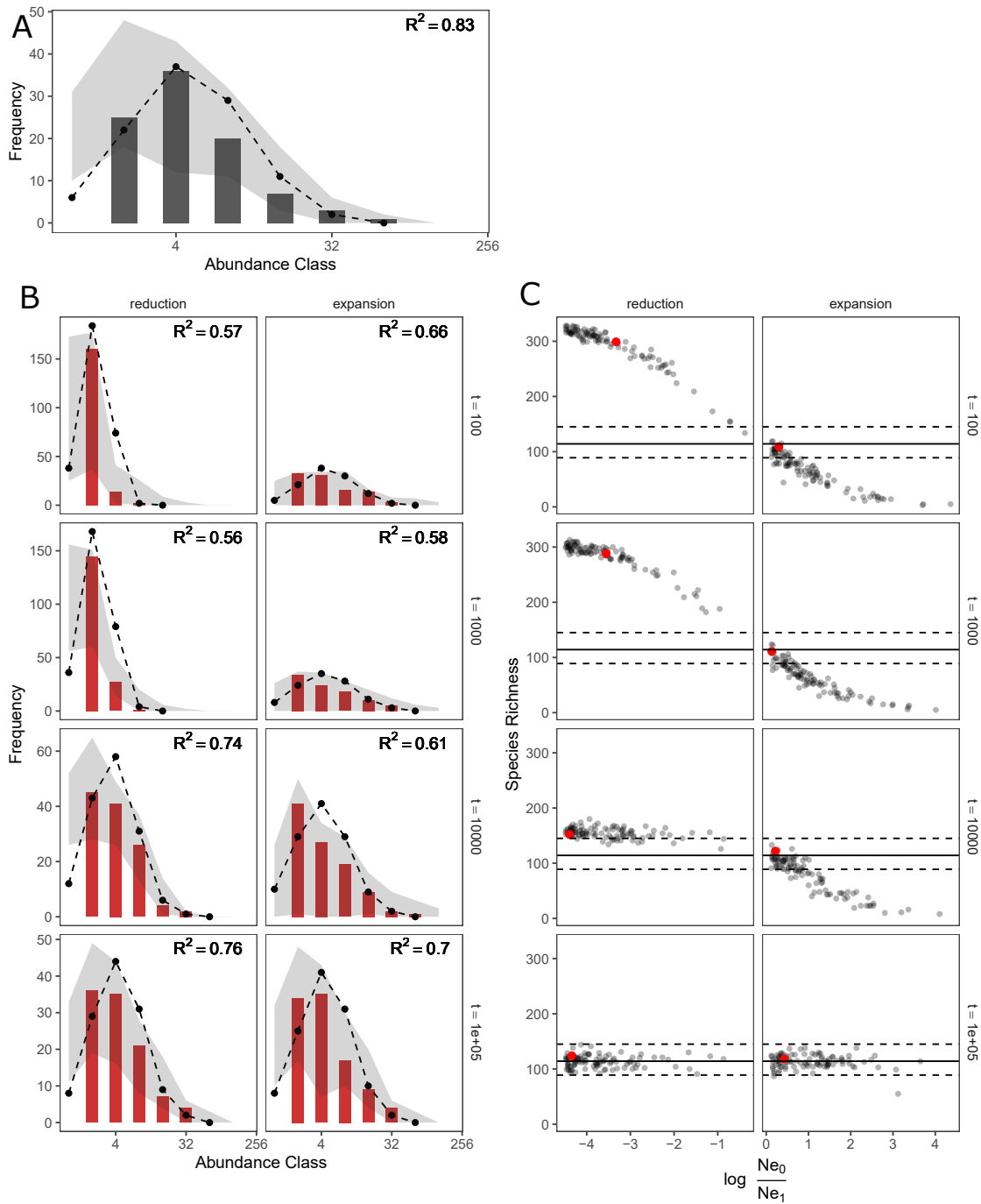


Figure 1.2: Variation in simulated taxonomic diversity. The left column shows the variation in simulated species abundance distributions (SAD) with regard to the time (in generations before present) at which the assemblage changes in size and the type of demographic fluctuation with (A) no demographic change (constant size model) or (B) with a fixed current assemblage size (N_{e0}) but a fluctuating past size (N_{e1}). Shaded grey areas correspond to the variation in abundance class frequency over all simulations whereas the red histograms show an example SAD. The associated fit to a log-normal distribution is shown by the dashed line and the associated R^2 given in the top right corner of each panel. (C) Variation in simulated species richness with regard to the time t (in generations before present) at which the assemblage changes in size and the log-transformed ratio between current and past assemblage sizes (N_{e0} and N_{e1} respectively). Negative values indicate past assemblage reduction, and positive values assemblage expansion. Mean species richness simulated for a constant size model is represented by horizontal lines while dashed lines represent the minimum and maximum simulated species richness. Points in red represent the species richness of the simulated communities whose SADs are illustrated in the left column. For each parameter combination we simulated 100 samples, each containing 500 individuals.

3.1 Examining the effect of past demographic fluctuations

We demonstrated the influence of the model's parameters with regard to past demographic expansion or contraction on species diversity, in terms of species richness or number of species by abundance class, by comparing these patterns to those obtained with a constant size model for different parameter combinations. Parameters related to the evolutionary aspect of the model were fixed ($\mu = 10^{-4}$ and $\tau = 0$). The extant metacommunity size ($Ne_0 = Ne(0)$) was set to 9×10^4 , and either did not vary in the past (constant size model) or was changed to a value drawn in a uniform prior distribution bounded between 9.9×10^5 and 8×10^6 (with subsequent demographic reduction) or 8×10^2 and 8×10^4 (with subsequent demographic expansion) (parameter $Ne_1 = Ne(t_1)$) at a given time t_1 (either 10^2 , 10^3 , 10^4 or 10^5 generations before present). For each parameter combination we simulated 100 samples, each containing 500 individuals. We characterized past demographic fluctuations in terms of: (i) the time (in generations before present) at which the assemblage changes in size and (ii) the ratio between current and past assemblage size (Ne_0/Ne_1). The ratio was log transformed so that negative values meant demographic reduction while positive values meant demographic expansion.

Simulated species richness decreased as the size of the assemblage in the past (Ne_1) was small when compared to that of the assemblage today (Ne_0) (Figure 1.2C) and did so faster when the subsequent expansion was more recent. Also, the decrease in richness particularly concerned rarest species (Figure 1.2B). Increased ecological drift in smaller assemblages led to the loss of species (particularly rare species), which was still observable when the time since expansion began was insufficient to reach a new speciation/drift equilibrium. However, when the change from Ne_1 to Ne_0 occurred before 104 generations ago, there was no difference in species richness between the constant size model and that with a change in assemblage size. Thus, maximum species richness (Figure 1.2C) and a greater number of rare species (Figure 1.2B) were obtained when the size of the assemblage in the past was the largest compared to Ne_0 and when the change in size was recent enough that species diversity still reflected past environmental conditions. This is due to the time elapsed since the change not being sufficient to attain an equilibrium with the current habitat configuration, i.e. extant species diversity still reflects the imprint of past habitat area. Also, the remnant influence of smaller past assemblage sizes (with subsequent expansion) is greater than that left when the assemblage was larger (with subsequent reduction).

3.2 Examining the effect of the pace of evolution

We then examined the influence of parameters controlling the evolutionary component of the *ecophylo* model, i.e. the point mutation rate (μ) and protractedness (τ) for constant size models of metacommunity size ($Ne(t) = Ne$). We tested 4 values of τ (0, 50, 500 or 1000) and 3 values of Ne (8×10^4 , 9×10^5 and 8×10^7). For each parameter combination 100 samples were simulated, each containing 500 individuals, while μ was drawn from a uniform distribution bound between 0.01 and 0.01. We characterized species diversity in terms of either species richness or number of species by abundance class (Figure 3).

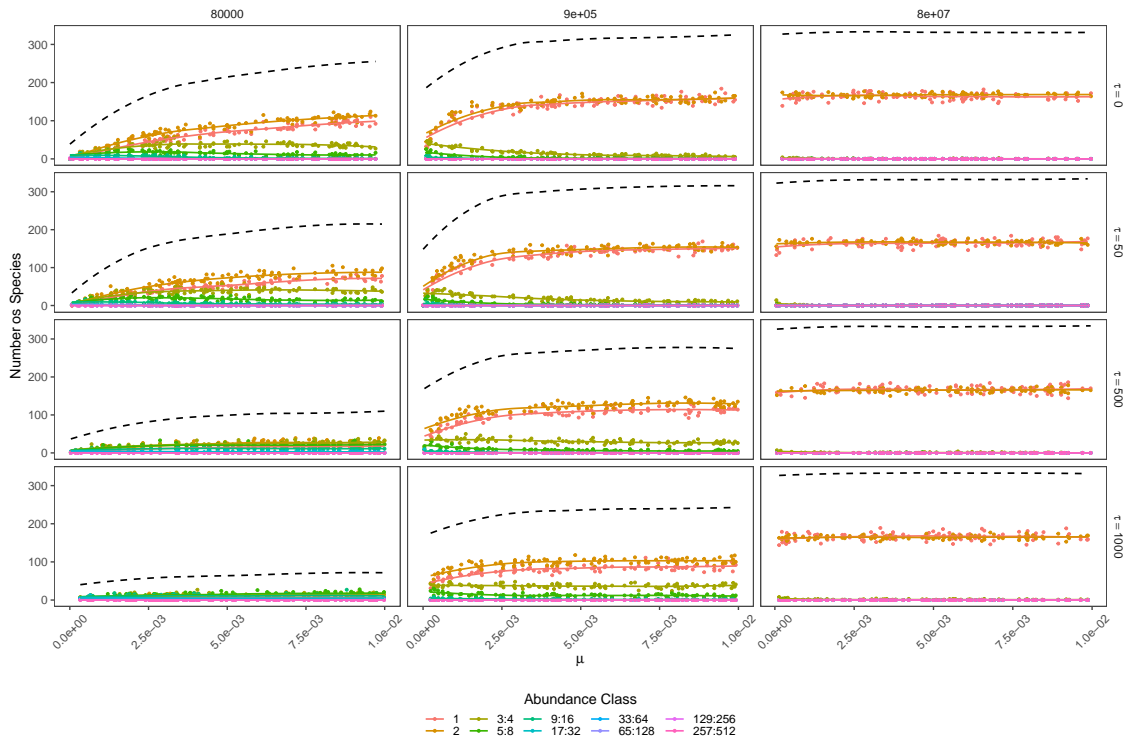


Figure 1.3: Variation in the simulated number of species with regard to the point mutation rate (μ), protractedness (τ) and the size of the assemblage (N_e). The dashed line represents the total species richness while color indicate the number of species with a given number of individuals (abundance class).

Maximum species richness could not exceed the sample size (500 individuals), and reached different maximum values depending on the protractedness parameter (τ) and the size of the metacommunity (N_e) (Figure 1.3). Furthermore, the number of rare species (species with less than 2 individuals) dropped when τ increased. With a larger mutation rate more species were produced because the odds of introducing a variant individual in the genealogy increased, yet these chances decreased when a certain number of generations must pass to consider the descendants of these variants as new species. With larger assemblages, the probability of a coalescence event decreases, making for longer branches in the genealogy and thus more chances that individuals vary to the point of producing new species. On the other hand, in smaller assemblages, more individuals can coalesce at a time earlier than τ generations, so that species richness decreases under the protracted speciation process.

All simulations can be reproduced using the Python scripts provided in the supplementary materials.

3.3 Inferring eco-evolutionary mechanisms

The *ecophylo* library can be used as described in the previous section to conduct virtual experiments, i.e by examining the effect of the model's parameters relating to past environmental fluctuations and migration dynamics on a variety of diversity patterns. Yet another (and perhaps more important) motivation behind *ecophylo* is to conduct inference. Using the simulations produced by *ecophylo* to retrieve parameter values yielding diversity patterns which have come closest to one's actual data

can achieve two things: (i) comparing and selecting models with different scenarios of biogeographic change and (ii) estimating their parameters values. We, therefore, assessed whether a limited set of biodiversity patterns could be used to distinguish between two simple models (misclassification analysis) and how well we could estimate these models' parameter values (cross-validation).

Two sets of 5000 simulations were produced (each containing 21,454 individuals, as in the classical Barro Colorado forest plots (Condit 1995) for which τ was fixed to 0 and the mutation rate was sampled in a uniform prior distribution bounded between 10^{-3} and 10^{-9} . In the first model the assemblage size did not fluctuate over time and the single associated parameter (Ne_0) was sampled from a uniform prior distribution varying between 30,000 and 10^9 . In the second model, the size of the assemblage changed 30,000 generations before present and the associated parameters (Ne_0 and Ne_1) were sampled from a uniform prior distribution varying between 30,000 and 10^9 (reduction scenarios). The results of the simulations were summarized by means of species richness, Shannon's and Simpson's diversity indexes (Oksanen et al. 2007) and by the number of species by octave of abundance. The metrics constituted the summary statistics used for the misclassification and cross-validation analyses.

Misclassification errors between the two considered models were computed by assigning 100 randomly selected simulations to the model with the highest posterior probability given a tolerance of 5% (Csilléry, François, and M. G. Blum 2012). Here we use the existing simulations to determine whether an ABC-based method of model selection can correctly distinguish between models. Some of the simulated summary statistics are used as observations and assigned to model with the highest posterior probability given the rest of the simulated statistics. This procedure is then repeated 100 times. Our results indicated that the simulated outcomes of the two models were clearly distinguishable given the considered summary statistics as the constant size model was correctly identified in 85% of cases and the model with a change in past assemblage size in 75% of cases.

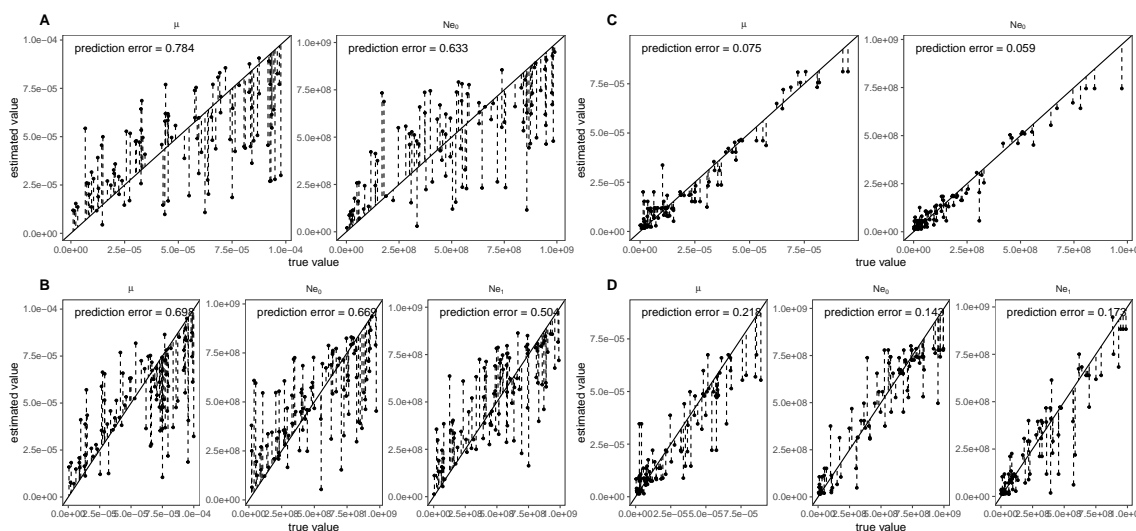


Figure 1.4: Cross-validation analysis with dimension reduction of simulated communities using the constant size model (A,C) and that with a change in size occurring 30,000 generations ago (B,D) based on either species richness, Shannon's and Simpson's diversity indexes and the number of species by octave of abundance class (A,B) or the same statistics transformed into the PCAIV components trained on the 100 closest simulations (C,D). Vertical dashed lines represent the misprediction errors for each of the 100 considered pseudo-observations and given a tolerance threshold of 1%.

Then, for each model we considered a random subset of 100 simulations as pseudo-observations and used the rest of the simulations to retrieve the estimated parameter values (leave-one-out cross validation) (Csilléry, François, and M. G. Blum 2012). The estimated values were then compared to actual simulation values to gauge the accuracy of parameter estimation given the chosen summary statistics (Figure 1.4). The results of the cross validations showed that given the set of chosen summary statistics (and the tolerance threshold of 1%), the simulations produced using *ecophylo* made it possible to estimate both considered models' parameters, i.e. the mutation rate and the size of the assemblage in the past, values reasonably well (Figure 1.4). Yet, since ABC can be sensitive to the curse of dimensionality (M. G. Blum et al. 2013), we also considered an alternative protocol for cross validations. We performed a dimension reduction of the summary statistics using a principal component analysis with instrumental variables (PCAIV) which allows to compute linear combinations of the summary statistics that explain the most variance in the parameter distribution (Lebreton et al. 1991). We trained the PCAIV using the 10% of simulated summary statistics (taxonomic diversity indices, richness and species abundances) closest to the observed statistics. In a second step, we estimated the posterior parameter distribution in ABC using the observed and all the original simulated statistics transformed into the PCAIV components as summary statistics using a simple rejection method with a tolerance of 5%. Cross validation results using this method (Figure 1.4) showed that we were able to reduce prediction errors by a factor ranging between 6 and 10 depending on the model.

4 Availability and future directions

ecophylo bridges the gap between ecological models representing population and assembly dynamics, and evolutionary models representing the appearance and disappearance of taxa over time. Thanks to great flexibility in parametrization and simplicity in output format, we feel that this freely available Python library meets the need of ecologists and biogeographers developing novel eco-evolutionary models to explain or predict biodiversity patterns. The source code is available on Github and the library easily installable via pip.

4.1 Modeling out-of-equilibrium diversification dynamics in a fluctuating environment

Studies in ecology often neglect the influence of evolutionary events, by customarily considering the regional species pool from which communities assemble via rapid local dynamics as a static snapshot of evolutionary history (Lessard et al. 2012). However, recent studies have called for a fuller integration of biogeographic history into community ecology (Overcast, Emerson, and Hickerson 2019 ; Overcast, Ruffley, et al. 2021 ;Swenson 2019), as well as for a more mechanistic understanding and modeling of key processes shaping the composition of species pools (Denelle, Violle, and Munoz 2019). With the novel simulation algorithm proposed in the *ecophylo* package, we explicitly represent how elementary ecological mechanisms such as ecological drift and migration

(Vellend 2010) in a finite, temporally variable environment shape evolutionary trajectories and result in extant patterns of phylogenetic and taxonomic diversity.

First, our model allows relaxing the assumption of constancy in metacommunity size of the NTB (Hubbell 2001), and thus can account for the vicariance and distinct demographic histories of metacommunities with fluctuating sizes over time. This feature is an essential advantage of *ecophylo* to understand and model the influence of past environmental changes on biogeographic dynamics and biodiversity patterns (Barthelemy et al. 2021). Also, modeling speciation as a gradual process by providing a way to delimit extant taxa (Manceau, Lambert, and Morlon 2015) meets more realistic expectations than considering speciation as an instantaneous event (Rosindell et al. 2010). Second, the simulation algorithm proposed in the *ecophylo* package by integrating the dynamics of individuals in a fluctuating environment and the speciation process, allows extending the coalescent modeling framework (Kingman 1982 ; Hudson 1983) to multiple species assemblages (Manceau, Lambert, and Morlon 2015). Thus, *ecophylo* can be used in the context of comparative phylogeography (Avisé 2009) to investigate the remnant influence of past biogeographic events on several co-occurring taxa (Barthelemy et al. 2021).

We have shown how the *ecophylo* package makes it possible to generate patterns of diversity along broad ranges of parameter values reflecting contrasted scenarios of past demographic fluctuations as well as different speciation dynamics. *ecophylo* can be used to similarly investigate how migration and/or habitat fragmentation history have affected multi-species assemblages. We have also demonstrated that our model can generate meaningful patterns of diversity in terms of species richness and relative abundance. For instance, the number of rare species dropped when including protractedness to the speciation process, thus endorsing the often raised criticism of the NTB's point speciation process amounting to an unrealistic excess of rare species (Rosindell et al. 2010 ; Hubbell 2015). Our model makes it possible to generate patterns of diversity which are out-of-equilibrium with current environmental conditions, an idea which has gained in momentum (Jackson and Sax 2010 ; Ordonez and Svenning 2016) and for which integrative methods are still needed. Thus, *ecophylo* is useful to explore the remnant influence of past environmental changes on current diversity patterns (Barthelemy et al. 2021).

4.2 Inference of eco-evolutionary dynamics from observed biodiversity patterns

An overarching goal in both ecology and evolution is to infer the mechanisms responsible for shaping diversity patterns from the limited information they contain at present (McGill 2019 ; Stouffer 2019). By adopting a mechanistic modeling approach of eco-evolutionary processes, we can first address whether a set of given metrics is sensitive to the mechanisms we chose to investigate by designing virtual experiments. We can thus compare large virtual datasets generated over a wide range of parameters, which represent a wide range of biogeographic scenarios we wish to investigate, and then determine which subset of metrics, if any, can keep track of these mechanisms of interest (Barthelemy et al. 2021). For instance, in the case where confounding extant distribution patterns can arise from either vicariance or migration or a combination of both processes, *ecophylo* can be used to examine patterns of phylogenetic and/or taxonomic diversity simulated over a broad

spectrum of scenarios combining migration and vicariance events. It is then possible to assess whether the chosen set of diversity metrics allows distinguishing between the influence of migration versus vicariance (Hastie, Tibshirani, and Friedman 2009; Csilléry, François, and M. G. Blum 2012), using misclassification and/or cross validation analyses as exemplified here.

Second, we can use simulated datasets to carry out inference, that is, retracing the most likely past demographic history of a metacommunity, based on the comparison of these datasets to actual extant biodiversity patterns (Csilléry, M. G. B. Blum, et al. 2010). While most existing approaches estimate an effective constant size as the harmonic mean of the per-generation metacommunity size (Nee, 2005), explicitly modeling its temporal variation may prove useful to investigate the relationship between current diversity patterns and past environment history. Yet this adds considerable complexity to the model, not to mention added constraints to the speciation process (for instance with protraction or the strict respect of monophyly). With Approximate Bayesian Computation (ABC) methods we can avoid computing the explicit likelihood of the model by approximating it through the relative proximity between the true dataset and datasets simulated given a model (Beaumont 2010 ; Jabot and Chave 2011 ; Csilléry, François, and M. G. Blum 2012). Here, we not only show that we can correctly recognize simulations issuing from different biogeographic models generated by *ecophylo* (model misclassification), but also that these models' parameters values could accurately be estimated using ABC methods (cross-validation). Our results thus warrant the use of *ecophylo* to conduct inference on actual observed data.

4.3 Limitations and future perspectives

The baseline eco-evolutionary model provided in *ecophylo* assumes neutrality. That is, species are evenly subject to fluctuations in the size of the assemblage they form with no regards to how individual fitness differences may influence their probability to persist in a given environment or migrate. Increasing the complexity of the model would require accounting for the ways in which deterministic processes such as ecological biotic and abiotic filtering, jointly with evolutionary natural selection, can affect diversification and assemblage dynamics over time (Ewing and Hermisson 2010 ; Ovaskainen et al. 2011 ; Karhunen et al. 2013 ; Manceau, Lambert, and Morlon 2016). Yet, because of the implications simulating all these elementary mechanisms for all individuals forward in time would have on computation times, we have focused on a coalescent framework (Munoz et al. 2018). Previous individual-based simulations of eco-evolutionary dynamics depending on birth, death, mutation, and dispersal have been designed to depend on an environmental raster determining environmental filtering and dispersal pathways over time (Rangel et al. 2018 ; Garwood, Spencer, and Sutton 2019) However, as mentioned in Rangel et al. 2018, these approaches are well suited to simulate and explore some theoretical scenarios, but are not designed for inference from real data. Likewise, coalescent-based simulation of extant community composition depending on migration, drift and speciation events have been proposed and can consider a spatially explicit fluctuating landscape (Thompson, Chisholm, and Rosindell 2020). Yet in order to use such models for parameter inference, users must run simulations by drawing parameters in prior distributions, which in itself is possible but may for some users deter their use. Conversely, *ecophylo* provides

quick simulation of extant community composition for broad ranges of eco-evolutionary parameters. Thanks to the computation power and rapidity of its backward-in-time algorithm, *ecophylo* then allows massive simulation and inference of parameter values from real data.

Later versions of the *ecophylo* package should gain in complexity by drawing from recent developments in coalescent simulation algorithms implementing spatially explicit methods (Rosindell et al., 2008; Kelleher, Etheridge, and Barton 2014), natural selection (Shlyakhter, Sabeti, and Schaffner 2014) or individual/species interactions (Lepers et al. 2021). Yet, for the time being, *ecophylo* in its initial version, with the flexibility in parameterization of simulations along with in the choice of summary statistics for either virtual experiments or ABC analyses on observed data, can be used as a neutral baseline to investigate how elementary mechanisms of mutation, drift and migration in a finite fluctuating environment influence evolutionary trajectories and thus shape patterns of diversity.

Data availability. The source code for the *ecophylo* package is available on Github (<https://github.com/thegreatlizzyator/ecophylo>) and can be installed via pip.

Author contributions. EB conceived the study and built the basic architecture of the *ecophylo* package. EB and MJ did substantial work on adding functionalities, testing and cleaning code. All the authors contributed to setting up the framework and to write the manuscript.

References

- Avice, J. C. (2009). “Phylogeography: retrospect and prospect”. In: *Journal of Biogeography* vol. 36, pp. 3–15.
- Barthelemy, E. et al. (2021). “Imprints of past habitat area reduction on extant taxonomic, functional and phylogenetic composition”. In: *Frontiers in Ecology and Evolution* vol. 9. ISBN: 2296-701X Publisher: Frontiers, p. 332.
- Beaumont, M. A. (2010). “Approximate Bayesian Computation in Evolution and Ecology”. In: *Annual Review of Ecology and Systematics* vol. 41, pp. 379–406.
- Bennett, K. D. (1990). “Milankovitch cycles and their effects on species in ecological and evolutionary time”. In: *Paleobiology*. ISBN: 0094-8373 Publisher: JSTOR, pp. 11–21.
- Blum, M. G. et al. (2013). “A comparative review of dimension reduction methods in approximate Bayesian computation”. In: *Statistical Science* vol. 28, no. 2. ISBN: 0883-4237 Publisher: Institute of Mathematical Statistics, pp. 189–208.
- Condit, R. (1995). “Research in large, long-term tropical forest plots”. In: *Trends in Ecology & Evolution* vol. 10, no. 1. ISBN: 0169-5347 Publisher: Elsevier, pp. 18–22.
- Csilléry, K., Blum, M. G. B., et al. (2010). “Approximate Bayesian computation (ABC) in practice”. In: *Trends in Ecology & Evolution* vol. 25, pp. 410–418.
- Csilléry, K., François, O., and Blum, M. G. (2012). “abc: an R package for Approximate Bayesian Computation (ABC)”. In: *Methods in Ecology and Evolution* vol. 3, pp. 475–479.
- Denelle, P., Violle, C., and Munoz, F. (2019). “Distinguishing the signatures of local environmental filtering and regional trait range limits in the study of trait–environment relationships”. In: *Oikos*.

- Etienne, R. S. et al. (2006). “Comment on” Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot”. In: *Science* vol. 311, no. 5761. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 610–610.
- Ewing, G. and Hermisson, J. (2010). “MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus”. In: *Bioinformatics* vol. 26, pp. 2064–2065.
- Garwood, R. J., Spencer, A. R., and Sutton, M. D. (2019). “RE voSim: Organism-level simulation of macro and microevolution”. In: *Palaeontology* vol. 62, no. 3. ISBN: 0031-0239 Publisher: Wiley Online Library, pp. 339–355.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton and Oxford: Princeton University Press.
- Hubbell, S. P. (2015). “Estimating the global number of tropical tree species, and Fisher’s paradox”. In: *Proceedings of the National Academy of Sciences* vol. 112, no. 24. ISBN: 0027-8424 Publisher: National Acad Sciences, pp. 7343–7344.
- Hudson, R. R. (1983). “Properties of a neutral allele model with intragenic recombination”. In: *Theoretical population biology* vol. 23, no. 2. ISBN: 0040-5809 Publisher: Elsevier, pp. 183–201.
- Jabot, F. and Chave, J. (2011). “Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference”. In: *The American Naturalist* vol. 178, no. 2. ISBN: 0003-0147 Publisher: University of Chicago Press Chicago, IL, E37–E47.
- Jabot, F., Etienne, R. S., and Chave, J. (2008). “Reconciling neutral community models and environmental filtering: theory and an empirical test”. In: *Oikos* vol. 117, pp. 1308–1320.
- Jackson, S. T. and Sax, D. F. (2010). “Balancing biodiversity in a changing environment: extinction debt, immigration credit and species turnover”. In: *Trends in ecology & evolution* vol. 25, no. 3. ISBN: 0169-5347 Publisher: Elsevier, pp. 153–160.
- Karhunen, M. et al. (2013). “DRIFTSEL: an R package for detecting signals of natural selection in quantitative traits”. In: *Molecular Ecology Resources* vol. 13, no. 4, pp. 746–754.
- Kelleher, J., Etheridge, A. M., and Barton, N. H. (2014). “Coalescent simulation in continuous space: Algorithms for large neighbourhood size”. In: *Theoretical Population Biology* vol. 95, pp. 13–23.
- Kelleher, J. and Lohse, K. (2020). “Coalescent simulation with msprime”. In: *Statistical Population Genomics*. Humana, New York, NY, pp. 191–230.
- Kingman, J. F. C. (1982). “On the genealogy of large populations”. In: *Journal of Applied Probability*, pp. 27–43.
- Lebreton, J.-D. et al. (1991). “Principal component and correspondence analyses with respect to instrumental variables: an overview of their role in studies of structure-activity and species-environment relationships”. In: *Applied multivariate analysis in SAR and environmental studies*. Publisher: Springer, pp. 85–114.
- Lepers, C. et al. (2021). “Inference with selection, varying population size, and evolving population structure: application of ABC to a forward–backward coalescent process with interactions”. In: *Heredity* vol. 126, no. 2. ISBN: 1365-2540 Publisher: Nature Publishing Group, pp. 335–350.

- Lessard, J.-P. et al. (2012). “Strong influence of regional species pools on continent-wide structuring of local communities”. In: *Proceedings of the Royal Society of London B: Biological Sciences* vol. 279, pp. 266–274.
- Manceau, M., Lambert, A., and Morlon, H. (2015). “Phylogenies support out-of-equilibrium models of biodiversity”. In: *Ecology Letters* vol. 18, pp. 347–356.
- Manceau, M., Lambert, A., and Morlon, H. (2016). “A Unifying Comparative Phylogenetic Framework Including Traits Coevolving Across Interacting Lineages”. In: *Systematic Biology* vol. 66, no. 4, pp. 551–568.
- McGill, B. J. (2019). “The what, how and why of doing macroecology”. In: *Global Ecology and Biogeography* vol. 28, no. 1. ISBN: 1466-822X Publisher: Wiley Online Library, pp. 6–17.
- Morlon, H. (2014). “Phylogenetic approaches for studying diversification”. en. In: *Ecology Letters* vol. 17, no. 4. Ed. by Mooers, A., pp. 508–525.
- Munoz, F. et al. (2018). “ecolottery : Simulating and assessing community assembly with environmental filtering and neutral dynamics in R”. en. In: *Methods in Ecology and Evolution* vol. 9, no. 3. Ed. by Chisholm, R., pp. 693–703.
- Oksanen, J. et al. (2007). “The vegan package”. In: *Community ecology package* vol. 10, pp. 631–637.
- Ordóñez, A. and Svenning, J.-C. (2016). “Strong paleoclimatic legacies in current plant functional diversity patterns across Europe”. In: *Ecology and Evolution* vol. 6, pp. 3405–3416.
- Ovaskainen, O. et al. (2011). “A New Method to Uncover Signatures of Divergent and Stabilizing Selection in Quantitative Traits”. en. In: *Genetics* vol. 189, no. 2, pp. 621–632.
- Overcast, I., Emerson, B. C., and Hickerson, M. J. (2019). “An integrated model of population genetics and community ecology”. en. In: *Journal of Biogeography* vol. 46, no. 4, pp. 816–829.
- Overcast, I., Ruffley, M., et al. (2021). “A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities”. In: *Molecular ecology resources* vol. 21, no. 8. ISBN: 1755-098X Publisher: Wiley Online Library, pp. 2782–2800.
- Pigot, A. L. et al. (2010). “The shape and temporal dynamics of phylogenetic trees arising from geographic speciation”. In: *Systematic biology* vol. 59, no. 6. ISBN: 1076-836X Publisher: Oxford University Press, pp. 660–673.
- Rangel, T. F. et al. (2018). “Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves”. en. In: *Science* vol. 361, no. 6399, eaar5452.
- Rosindell, J. et al. (2010). “Protracted speciation revitalizes the neutral theory of biodiversity”. In: *Ecology Letters* vol. 13, pp. 716–727.
- Shlyakhter, I., Sabeti, P. C., and Schaffner, S. F. (2014). “Cosi2: an efficient simulator of exact and approximate coalescent with selection”. en. In: *Bioinformatics* vol. 30, no. 23, pp. 3427–3429.
- Stouffer, D. B. (2019). “All ecological models are wrong, but some are useful”. en. In: *Journal of Animal Ecology* vol. 88, no. 2. Ed. by Gill, J., pp. 192–195.
- Svenning, J.-C. et al. (2015). “The Influence of Paleoclimate on Present-Day Patterns in Biodiversity and Ecosystems”. In: *Annual Review of Ecology, Evolution, and Systematics* vol. 46, null.
- Swenson, N. G. (2019). *Phylogenetic Ecology: A History, Critique, and Remodeling*. en. University of Chicago Press.
- Tello, J. S. and Stevens, R. D. (2012). “Can stochastic geographical evolution re-create macroecological richness–environment correlations?” In: *Global Ecology and Biogeography* vol. 21, no. 2. ISBN: 1466-822X Publisher: Wiley Online Library, pp. 212–223.

1. ecophylo: Simulating and assessing eco-evolutionary dynamics under past environmental changes in Python.

- Thompson, S. E., Chisholm, R. A., and Rosindell, J. (2020). “pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity”. In: *Methods in Ecology and Evolution* vol. 11, no. 10. ISBN: 2041-210X Publisher: Wiley Online Library, pp. 1237–1246.
- Vellend, M. (2010). “Conceptual synthesis in community ecology”. In: *The Quarterly review of biology* vol. 85, no. 2, pp. 183–206.
- Wakeley, J. (2009). “Coalescent theory”. In: *Roberts & Company*.
- Zurell, D. et al. (2010). “The virtual ecologist approach: simulating data and observers”. en. In: *Oikos* vol. 119, no. 4, pp. 622–635.

Appendix 1.A Python scripts

This script runs the simulations used for the following analyses. If users wish to save these simulations, a specified file name should be provided in the "file_name" argument of the ecophylo.dosimul function

1. an exploration of the influence of past demographic fluctuations on diversity patterns. In this case the size parameters related to the evolutive aspect of the model were fixed ($\mu = 10^{-4}$ and $\tau = 0$). The extant assemblage size (N_{e0}) was set to 9×10^4 , and either did not vary in the past (constant size model) or was changed to a value drawn in a uniform prior distribution bounded between 9.9×10^5 and 8×10^6 (demographic reduction) or 8×10^2 and 8×10^4 (demographic expansion) (parameter N_{e1}) at a given time t (either 10^2 , 10^3 , 10^4 or 10^5 generations before present).
2. an exploration of the influence of the pace of evolution. Here, we tested 4 values of τ (0, 50, 500 or 1000) and 3 values of N_e (8×10^3 , 9×10^4 and 8×10^6). For each of these parameter combinations 100 communities were simulated, each containing 500 individuals for which μ was drawn from a uniform distribution bound between 0.01 and 10^{-10} .
3. an analysis of the ability to distinguish between a constant size model and a model in which the size of the assemblage changes at a given date. Two sets of 5000 simulations are produced (each containing 21,454 individuals) for which τ is fixed to 0 and the mutation rate is sampled in a uniform prior distribution bounded between 10^{-3} and 10^{-9} . In the first model the assemblage size did not fluctuate over time and the single associated parameter (N_{e0}) was sampled from a uniform prior distribution varying between 30,000 and 108. In the second model, the size of the assemblage changed 30 000 generations before present and the associated parameters (N_{e0} and N_{e1}) were sampled from a uniform prior distribution varying between 30,000 and 10^8 .

1. ecophylo: Simulating and assessing eco-evolutionary dynamics under past environmental changes in Python.

```
1 import ecophylo
2
3 # (i) an exploration of the influence of past demographic fluctuations on
4 # diversity patterns
5
6 n = 500
7 mu = 1e-4
8 nsim = 100
9
10 times = [1e2, 1e3, 1e4, 1e5]
11
12 for i in range(len(times)):
13     ds = [[9e5, [99e4,8e7, "uniform"]]] # reduction
14     changetimes = [[0,times[i]]]
15     ecophylo.dosimuls(nsim = nsim,
16                       samples = n,
17                       deme_sizes= ds,
18                       mu = mu,
19                       changetimes= changetimes,
20                       output=['Params','Sumstat','Trees'])
21
22 for i in range(len(times)):
23     ds = [[9e5, [8e3,8e5, "uniform"]]] # expansion
24     changetimes = [[0,times[i]]]
25     ecophylo.dosimuls(nsim = nsim,
26                       samples = n,
27                       deme_sizes= ds,
28                       mu = mu,
29                       changetimes= changetimes,
30                       output=['Params','Sumstat','Trees'])
31
32 # constant
33 ecophylo.dosimuls(nsim = nsim,
34                   samples = n,
35                   deme_sizes= [[9e5]] ,
36                   mu = mu,
37                   changetimes= [[0]],
38                   output=['Params','Sumstat','Trees'])
39
40 # (ii) an exploration of the influence of the pace of evolution
41
42 mu = [1e-2, 1e-10, "uniform"]
43
44 changetimes = [[0]]
45 nsim = 100
46 tau = [0,50,5e2,1e3]
47 sizes = [8e4, 9e5, 8e7]
48
49 for j in range(len(sizes)):
50     deme_sizes = [[ sizes[j] ]]
51     for i in range(len(tau)):
52         out = ecophylo.dosimuls(nsim = nsim,
53                                 samples = n,
```

```
54         deme_sizes= deme_sizes,
55         mu = mu,
56         tau = tau[i],
57         changetimes= changetimes,
58         output=['Params','Sumstat','Trees'])
59
60
61 # (iii) an analysis of the ability to distinguish between a constant
62 # size model and a model in which the size of the assemblage changes
63 # at a given date
64
65 n = 21454
66 mu = [1e-4, 1e-10, "uniform"]
67 deme_sizes= [[[3e4, 1e9,"uniform"]]]
68 changetimes = [[0]]
69 totnsim = 5000
70
71
72 ecophylo.dosimuls(nsim = totnsim,
73                 samples = n,
74                 deme_sizes= deme_sizes,
75                 mu = mu,
76                 changetimes= changetimes,
77                 output=['Params','Sumstat','Trees'])
78
79 deme_sizes= [[[3e5, 1e9,"uniform"], [3e4,1e9, "uniform"]]] # reduction
80 changetimes = [[0,3e4]]
81
82 ecophylo.dosimuls(nsim = totnsim,
83                 samples = n,
84                 deme_sizes= deme_sizes,
85                 mu = mu,
86                 changetimes= changetimes,
87                 output=['Params','Sumstat','Trees'])
```


Article 2

From biodiversity patterns to processes and back again: solving the inverse problem in ecology

Elizabeth Barthélemy¹, Pierre Denelle^{2,3}, Grégoire Blanchard^{4,5},
Cyrille Violle³, François Munoz¹

Abstract

An essential goal in ecology is to faithfully identify the contributions of ecological processes from the limited information enclosed in biodiversity patterns. While, ecological models can be useful to predict biodiversity patterns from a model's parameter values (direct problem), they can be insufficient to deduce the nature of ecological processes from the observation of patterns (inverse problem). Experimental designs attempt to consider the causal relationship between patterns and process as a direct problem, rather than an inverse one, and while field experiments are impracticable in many systems, we assert that we can reap similar benefits from virtual experiments. We argue that in order to solve the inverse problem, we must consider the direct problem first, by proposing that many virtual experiments be used to deduce, by comparison, how observed biodiversity patterns are shaped, and thus solve the inverse problem. We call for recognizing the influence a priori knowledge holds on our ability to draw conclusions. We thus encourage quantifying how even the best model can be wrong to improve our modelling choices. Doing so will allow ecologists to reach more robust conclusions regarding the relationship between pattern and process - especially if and when processes jointly interact in opposition.

¹LIPhy, Université Grenoble Alpes, Grenoble, France

²Department of Biodiversity, Macroecology and Biogeography, University of Goettingen, Goettingen, Germany

³CEFE, Université de Montpellier, CNRS, EPHE, IRD, Université Paul Valéry, Montpellier, France

⁴Equipe Sol & Vegetation (SolVeg), Institut Agronomique neo-Caledonien, Nouméa, New Caledonia

⁵AMAP, IRD, CIRAD, CNRS, INRAE. Université de Montpellier, Montpellier, France

1 Introduction – solving the inverse problem in ecology

Understanding the mechanisms by which observed biodiversity patterns emerge is a key objective in ecology and evolution. Biodiversity studies recurrently disagree on the validity of numerous and at times opposing theories (Adler, HilleRisLambers, and Levine 2007 ; McGill 2010). For instance, some theories define how the persistence and coexistence of organisms in populations and communities is determined by their ability to establish, survive and reproduce in a given environment (Cadotte and Tucker 2017), while others emphasize the role of random fluctuations in their relative abundances regardless of organisms' individual performances (Hubbell 2001). Ecological theories are useful as they provide different hypotheses as to how our observation of biodiversity patterns come to be. Yet, where these might fall short is in providing causal relationships between process and pattern (Quinn and Dunham 1983).

Mathematical models are routinely derived from ecological theory to address the direct problem, that is, they can predict community composition given a set of parameter values (Zurell et al. 2010), for example based on certain properties of species affecting their local persistence (Munoz, Grenié, et al. 2018). However, ecologists are not so much interested in modelling and predicting virtual diversity patterns. Rather, what they are interested in is to use these theories and models as a lens to scrutinize biodiversity patterns, in order to deduce which processes have shaped our observations. Yet, models do not reveal the mechanisms responsible for shaping biodiversity in the same way that a microscope reveals the presence of microorganisms to an observer. These patterns are, as in the analogy drawn by Plato, like shadows cast on a cave wall: indirect and imperfect reflections of the true processes at play, which we, as observers, cannot directly witness. Despite being meant for a different purpose, this allegory perfectly sums up what is, in essence, an inverse problem, and one especially central to how we study ecological systems. An inverse problem is one in which we attempt to infer the parameters ruling a system from its resulting outcomes (Tarantola 2006). Thus, most concerns in ecology boil down to tackling the inverse problem of deducing the nature of ecological processes from the limited observation of diversity patterns (McGill 2019 ; Stouffer 2019). And although it is quite straightforward to characterize shadows by projecting light on an object (direct problem), it is far more difficult to deduce the nature of the object from the projected shadows (inverse problem).

Multiple issues stem from considering the inverse problem in ecology. Distinguishing between underlying processes is near impossible when these occur simultaneously or if there are strong interactions between underlying causes (Quinn and Dunham 1983 ; Chave, Muller-Landau, and Levin 2002). And while, randomized experiments are often sanctioned to examine causal effects, these can be insufficient to identify a mechanism (Imai et al. 2011) and are often impractical for most biological systems (Stouffer 2019). Because, the scale of the experimental device must be relevant to that of the hypotheses of interest, if these concern large-scale systems random experiments become quite impractical (Shipley 2016). For instance, it possible to experimentally manipulate levels of biodiversity – whether it be the number of species or their functional and/or phylogenetic similarity – to observe the direct repercussions on ecosystem properties (Tilman, Reich, and Knops 2006) or mechanisms of coexistence (Violle, Nemergut, et al. 2011). However, doing so for extensive and

remote communities with long-lived organisms, e.g. tropical rainforests, is near impossible. Hence, for many systems it is inconceivable to directly witness how diversity patterns are shaped, but much like with the shadows in the cave, we have to rely on our indirect observation of natural systems to infer ecological processes at play. This task is challenging, as there is no universal law relating biodiversity patterns to processes (Lawton 1999), because ambiguous patterns may lead to misinterpretations (Chave, Muller-Landau, and Levin 2002 ; Yanco et al. 2020) and because scale matters in ecology (Levin 1992). With experimentation, the objective is to consider the relationship between patterns and process as a direct problem, rather than an inverse one, because it makes causality easier to consider, and while field experiments have material limitations in time and space we argue that these can be overcome with virtual experiments (Münkemüller et al. 2012).

Thus, while models are useful for tackling the direct problem, they can be insufficient to solve the inverse problem. And although such fundamental duality has often been addressed in physics for instance (Kirsch 2011 ; Aster, Borchers, and Thurber 2018), less attention has been devoted to theoretical and methodological implications of dealing with the inverse problem when analyzing biodiversity patterns. We argue that in order to solve the inverse problem in ecology, we need to work on the direct problem first. This means defining models of biodiversity dynamics with clear mechanistic rules and well-defined assumptions. Because a major issue with correlative or regressive methods is that they can fail to disambiguate the complex and confounding effects of entangled processes on biodiversity patterns (Chave, Muller-Landau, and Levin 2002 ; Mayfield and Levine 2010). Meanwhile, permutation-based null models are routinely used to assess the contribution of non-neutral processes, namely, traits are assumed to be under environmental filtering if observed trait-based diversity statistics deviate from those obtained when randomizing species functional traits (Bernard-Verdier et al. 2012) or community composition (Gotelli 2000). However, such approaches do not explicitly address the nature and the shape of trait-based processes constraining functional composition, nor do they acknowledge the influence of stochastic neutral dynamics (Gotelli and McGill 2006 ; Bausman 2018). With mechanistic models directly representing ecological processes of interest we can actually design virtual experiments whereby diversity patterns are directly observable (Zurell et al. 2010 ; Münkemüller et al. 2012).

Ultimately, we propose that the outcomes of many virtual experiments be used to deduce by comparison how observed biodiversity patterns are shaped (Csilléry et al. 2010). In other words, we can simulate shadows on the cave wall based on our hypotheses and test them by comparing the outcomes of our simulations to the actual shadows. However, to conclude on the drivers of observed biodiversity patterns, we first need to know how the patterns change when the drivers change i.e. consider the direct problem. This means we must first consider the direct problem which we argue can be done in a virtual experimental context. Confounding effects can be identified by virtually examining how similar patterns emerge from different mechanisms. This allows us to identify signatures unique to the mechanisms we are interested in uncovering in real data. We can then make the comparison between observed diversity patterns and those we have explicitly simulated in order to infer which of our mechanistic models and their parameters are likely to have shaped our observations. Thus, it is crucial that we be able to draw conclusions from this comparison.

2 The curse of conditionality

From this premise, testing the respective influences of elementary processes and estimating their parameter values require designing appropriate statistical comparisons of observed and simulated patterns. To this end, Approximate Bayesian Calculation (ABC) methods, while now widely used in population genetics and phylogeography, are still little used to address biodiversity dynamics in ecology (Csilléry et al. 2010 ; Beaumont 2010 ; Ruffley et al. 2019). These methods repeatedly compare large numbers of datasets simulated under a hypothesized model (by sampling parameter values from a prior probability distribution) to the observed data, by means of summary statistics (Csilléry et al. 2010 ; Sunnåker et al. 2013). Sampled parameter values are then accepted or rejected based on the distance between the simulated and observed summary statistics, thus producing a distribution of most likely parameter values related to observed data under the hypothesized model. Thus, it is possible to consider statistical inference for models based on complex algorithms mimicking elementary mechanisms, by bypassing using a likelihood function – which is either impossible or too costly to evaluate (Hartig et al. 2011). Also, in communities where there is no strict limit between competing processes, these methods allow us to statistically compare the behavior of diversity metrics simulated over gradually changing assembly scenarios. The problem is that a virtual experiment cannot show the outcome of every imaginable model given every single parameter combination, we must choose what to simulate and which mechanisms that means we are testing. The important implication here is that conclusions are necessarily contingent upon the architecture and assumptions of modelling choices. In other words, statistically comparing our observations to simulations will only disclose the best amongst the simulated options, and not whether the simulations and thus our models are in fact pertinent. The experimental design therefore explicitly determines the limits of validity and of interpretation of the results (Hickerson 2014). This is what we call “the curse of conditionality”.

This issue is not exclusive to simulation-based approaches. The use of species pools is quite routine in ecology (Karger, Cord, et al. 2016 ; Karger, Wüest, et al. 2019). It is a practical way to represent assembly processes with a top-down scheme, in which species present in a regional pool can disperse, be filtered depending on local environment, and interact with other species to yield the observed composition (Lortie et al. 2004). Yet, species pools are defined a priori. When we perform permutation-based null models we are still testing the hypothesis of non-random assembly against an a priori defined pool of potential immigrant species (but see Karger, Cord, et al. 2016 ; Karger, Wüest, et al. 2019). It is important to acknowledge this because the definition of the species pool can have repercussions on how we interpret the resulting diversity patterns (Denelle, Violle, and Munoz 2019). For instance, inferring environmental filtering parameters and migration rates from a large spatial scale pool across a broad abiotic conditions will emphasize the influence of species fundamental niches and of their limited dispersal abilities depending on landscape structure (Karger, Cord, et al. 2016). Conversely, inferring the same parameters from a species pool restricted to the composition of adjacent communities will emphasize the role of finer-scale biotic interactions in shaping community assembly (Perronne et al. 2017) or may lessen the importance of environmental filtering if it has already been taken into account in the definition of the species pool. Hence, when we draw conclusions regarding community assembly, these can only ever be conditional to the

species pool we have ourselves designed.

3 Drawing more robust conclusions from virtual experiments

Since conclusions are conditional to the assumptions of the analyst regarding the competing models and the parameter space in which they are defined, it is essential to question how and which summary statistics are simulated. First, the way we define prior distributions is essential. Defining uninformative priors runs the risk of simulating highly unrealistic scenarios that can blur and bias the statistical analysis of the resulting posterior distributions. There is also a chance of increasing the overlap between simulated assembly scenarios which can mislead our ability to discriminate between models i.e. model selection. Restricting priors based on independent knowledge may also be biased and induce misleading conclusions either by ignoring possible outcomes that may be a good fit to the data or by introducing erroneous information into the model (Betini, Avgar, and Fryxell 2017).

Second, that diversity metrics should carry the signature of underlying processes is important because inference is directly based on the information withheld in the derived summary statistics (Grimm et al. 2010 ; Hartig et al. 2011). This is when considering the direct problem becomes especially useful. Benchmarking the chosen statistics through intensive simulations for different assembly models over broad ranges of parameter values can help characterize how diversity metrics are affected by ecological processes and how we choose to simulate them (Zurell et al. 2010). Once we know whether or not a set of diversity metrics carries the distinguishable signature of an ecological process of interest, can these be used to solve the inverse problem.

Model design through the definition of priors and choice of summary statistics will influence our ability to both discriminate between models, i.e model selection (Johnson and Omland 2004), and perform parameter estimation. If the simulated metrics do not significantly differ between models, either because they do not carry the signature of the simulated processes or because of overlapping priors, model selection will essentially boil down to the flip of a coin. Fortunately, it is possible to gauge whether the variation of summary statistics across simulations allows distinguishing between the proposed assembly models i.e. misclassification and appraise the robustness of parameter estimation i.e. cross-validation (Csilléry et al. 2010). Both approaches rely on the generation of artificial datasets from the simulated parameter values and the subsequent appraisal of how well a given ABC method either assigns the simulations to the correct model (model selection) or recovers the true parameter values (parameter estimation) respectively (Sunnåker et al. 2013). Systematically performing misclassification and cross-validations will reveal whether the variability of the simulated statistics suffices to carry out robust statistical inference and support conclusions.

4 All models are wrong, all of which are useful

However, the results of the ABC analysis rest on the summary statistics that have come the closest to the observation, without any guarantee that they are, in fact, actually close to observed diversity patterns. So even though a set of parameter values can better fit an observed diversity pattern for a given model of community assembly, it does not mean that the fit is perfect, or even good for that matter – just better. Thus, it is useful to know why a model is wrong.

Even a high goodness-of-fit cannot guarantee that a model is true, however, its “badness-of-fit” can tell us which of its aspects are poorly predicted, and thereby help us iteratively define more relevant alternative models. Because the design of the virtual experiment conditions the interpretation of the results, it is essential that we acquit ourselves, as far as possible, from cognitive biases that limit our ability to consider and test alternative hypotheses. There lies the risk of focusing our attention on a narrow set of favored theories (Elliott and Brook 2007 ; Betini, Avgar, and Fryxell 2017) which ABC methods have great potential to blindly confirm. Thus it is important to consider competing models which allow us to devise multiple hypotheses, test their validity by comparison to observed patterns and finally re-evaluate them to refine and successively improve modelling choices (Platt 1964 ; Betini, Avgar, and Fryxell 2017 ; Yanco et al. 2020).

To this end, posterior predictive checks can be used to test for discrepancies between the posterior predictions derived from the model and the actual data (Berkhof, Van Mechelen, and Hoijsink 2000 ; Conn et al. 2018). These dissimilarities can be appraised in two ways: quantitatively or qualitatively either by comparing the simulated posterior model predictions to the observed data through test statistics – which can differ from those used for the ABC scheme – or graphically appraising their resemblance. For instance, we can identify which species are significantly over or under-represented based on the comparison of their observed and predicted relative abundances. A systematic misrepresentation of a specific group of species can indicate that an assembly process has been overlooked in model design. It is then essential to assess which aspects of diversity have contributed to fitting the model if not just to understand why a model’s posterior predictions may not benefit the observed data. If only the accepted summary statistics related to a certain facet of diversity are in fact close to the observations, then the issuing model may only be suited to correctly predict this aspect of diversity. In which case predictive checks relative to other metrics may fall short. Systematically checking for such statistics behavior at each iteration can benefit the choice of more relevant summary statistics together with improving how mechanisms are simulated. Thus, repeatedly improving the choice of diversity metrics and assembly models in order to gain a firmer understanding of the relation between pattern and process.

5 Resolving the conundrum of neutral and non-neutral processes

Two distinct families of theories have been developed to predict the distribution and abundance of species, which have often been debated in opposition (Leibold 2008). Specifically, neutral theory emphasizes the role of random fluctuations in their relative abundances regardless of organisms’

individual performances. Thus, if species' ecological differences do not influence their fitness and only stochastic dynamics play, community assembly mainly results from the balance between dispersal and demography and yield neutral distributions of local and regional relative abundances (Hubbell 2001). Conversely, niche theory emphasizes that species' biological attributes in relation to the local biotic and abiotic environment primarily drive community composition (Cadotte and Tucker 2017). The concept of environmental filtering specifically represents how local environmental conditions act as a filter that only allows species with a particular set of traits to establish and develop. Community composition should then reflect the way fitness differences arising from variations in functional traits affect abundance dynamics and biodiversity patterns (Shipley, Vile, and É. Garnier 2006 ; McGill et al. 2006 ; Violle, Navas, et al. 2007), environmental filtering thereby generating a determined relationship between traits and local abundances. Despite these opposing mechanisms possibly producing confounding patterns, especially on species abundances (Bell 2001 ; Hubbell 2001), there is an emerging consensus that the presence and abundance of species in assemblages jointly depend on species adaptations and competitive abilities in a local environment, on their dispersal abilities and on demographic stochasticity (Leibold and McPeck 2006 ; Gravel et al. 2006 ; Shipley, Paine, and Baraloto 2012). The issue, moving forward, rather than opposing both theories and debating whether one prevails over the other, lies in characterizing these components' relative contributions. We argue that from a perspective in which there is no strict limit between neutral and niche influences but rather a continuum of their relative importance (Gravel et al. 2006), a simulation-based approach allows explicitly addressing where a given community is located along such gradient.

To this end, different metrics of diversity can be considered based on how these are hypothesized to respond to either neutral or niche-based processes. For instance, trait information can be used at species or individual levels, which will determine whether it is possible to detect and characterize assembly processes playing in intraspecific dynamics (Violle, Enquist, et al. 2012). The community-weighted mean of the trait distribution can deviate from a reference pool due to local filtering (Loranger et al. 2018), and vary across communities reflecting how local optima change with abiotic conditions (E. Garnier et al. 2004 ; Denelle, Violle, and Munoz 2019). Furthermore, other trait distribution moments convey the signatures of other aspects of environmental filtering (Loranger et al. 2018). Conversely, neutral drift dynamics are known to influence patterns of taxonomic diversity regardless of functional trait values, so that summary statistics based on taxonomic composition should help deciphering the contribution of stochastic neutral dynamics (Munoz and Coueron 2012 ; Munoz, Grenié, et al. 2018). Thus, the design of community sampling and trait measurement along with the choice of statistics will determine how to detect the influence of assembly processes and to infer their parameter values. Again, these choices can be cautioned by first considering the direct problem, and verifying in a controlled setting whether the chosen metrics carry sufficient signatures of the simulated processes to carry out robust inferences.

From there, we can consider different - and complementary - ways to address the relative contributions of neutral and niche-based dynamics on community assembly using simulations (Munoz, Grenié, et al. 2018). We can simulate communities in which neutral drift is estimated by a dispersal limitation parameter or directly by the number of immigrants colonizing local communities

2. From biodiversity patterns to processes and back again: solving the inverse problem in ecology

from an a priori defined species pool. When environmental filtering plays, establishment success can depend on how the ecological properties or trait values of immigrants allow greater performance in the local environment, for instance as a Gaussian function around some local optimum (stabilizing filtering, Rolhauser and Pucheta 2017). Thus, we can simulate patterns of community composition over a broad range of scenarios with varying relative contributions of niche-based or neutral processes. At this point, a misclassification analysis will assess whether the selected summary statistics allow distinguishing the influence of niche-based against neutral processes (Chave, Muller-Landau, and Levin 2002). Then, model selection can be used to compare a purely neutral model, in which assembly does not depend on any species difference, to a model in which the local performance further depends on species attributes. Furthermore, the extent to which environmental filtering constrains community composition also depends on the strength of the filter, for instance, the standard deviation in the case of Gaussian filtering. A very broad filter implies that community assembly is less constrained by trait values (weaker filtering), suggesting either neutral assembly or that unmeasured traits are actually filtered. In addition, lower estimated migration rates mean that local drift is less counterbalanced by immigration, denoting enhanced influence of stochastic demographic dynamics on community assembly. These options are not alternative ways to test neutrality but provide complementary insights on the extent to which niche-based dynamics influence community assembly.

We can then determine the influence of environmental variation directly on the simulated processes in one of two ways. First, model selection can be used to distinguish different hypotheses regarding assembly mechanisms - for instance to examine the absence or shape of an environmental filter driving communities in contrasted environments. Second, it is possible to relate parameters of trait-based filtering or neutral drift dynamics to environmental factors through post-hoc analysis. Possible applications include analyzing how estimated optimal trait values vary along environmental gradients, and relating the variation in migration rate to habitat configuration reflecting the influence of dispersal limitation and extinction risk in fragmented habitat (Blanchard, Birnbaum, and Munoz 2020).

6 Conclusion

Most ecological concerns fall under solving the inverse problem of deducing the nature of ecological processes from the limited observation of diversity patterns. And, while experimental designs attempt to uncover causal relationships between pattern and process, in remote ecosystems with slow dynamics experimentation is impossible or at best impractical. With virtual experiments not only can we bypass these technical difficulties but we can also virtually examine how similar patterns emerge from different mechanisms and thus identify confounding effects. We have argued that in order to solve the inverse problem, we need to consider the direct problem first, which for many systems can mainly - if not only - be achieved virtually. Thus, we have advocated for (i) defining ecological models with clear mechanistic rules and well-defined assumptions, (ii) questioning and systematically verifying these assumptions when comparing the results of many virtual experiments to our actual observations so as to avoid blindly validating preconceived assumptions, and (iii)

quantifying how even the best model can be wrong to then iteratively improve our modelling choices. We believe that doing so will allow ecologists to reach more robust conclusions regarding the relationship between patterns and the processes that have shaped them - especially if and when these processes jointly interact in opposition, for instance niche-based and neutral mechanisms ruling community assembly.

Author contributions. EB and FM conceived the ideas and methodology; EB led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Funding. This work was supported by the GdR ecostat.

References

- Adler, P., HilleRisLambers, J., and Levine, J. M. (2007). “A niche for neutrality”. In: *Ecology Letters* vol. 10, pp. 95–104.
- Aster, R. C., Borchers, B., and Thurber, C. H. (2018). *Parameter estimation and inverse problems*. Elsevier.
- Bausman, W. (2018). “Modeling: Neutral, Null, and Baseline”. In.
- Beaumont, M. A. (2010). “Approximate Bayesian Computation in Evolution and Ecology”. In: *Annual Review of Ecology and Systematics* vol. 41, pp. 379–406.
- Bell, G. (2001). “Neutral macroecology”. In: *Science* vol. 293, no. 5539. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 2413–2418.
- Berkhof, J., Van Mechelen, I., and Hooijink, H. (2000). “Posterior predictive checks: Principles and discussion”. In: *Computational Statistics* vol. 15, no. 3, pp. 337–354.
- Bernard-Verdier, M. et al. (2012). “Community assembly along a soil depth gradient: contrasting patterns of plant trait convergence and divergence in a Mediterranean rangeland”. In: *Journal of Ecology* vol. 100, pp. 1422–1433.
- Betini, G. S., Avgar, T., and Fryxell, J. M. (2017). “Why are we not evaluating multiple competing hypotheses in ecology and evolution?” In: *Royal Society Open Science* vol. 4.
- Blanchard, G., Birnbaum, P., and Munoz, F. (2020). “Extinction–immigration dynamics lag behind environmental filtering in shaping the composition of tropical dry forests within a changing landscape”. In: *Ecography*.
- Cadotte, M. W. and Tucker, C. M. (2017). “Should Environmental Filtering be Abandoned?” In: *Trends in Ecology & Evolution*.
- Chave, J., Muller-Landau, H. C., and Levin, S. A. (2002). “Comparing classical community models: theoretical consequences for patterns of diversity”. In: *The American Naturalist* vol. 159, no. 1, pp. 1–23.
- Conn, P. B. et al. (2018). “A guide to Bayesian model checking for ecologists”. en. In: *Ecological Monographs* vol. 88, no. 4, pp. 526–542.
- Csilléry, K. et al. (2010). “Approximate Bayesian computation (ABC) in practice”. In: *Trends in Ecology & Evolution* vol. 25, pp. 410–418.
- Denelle, P., Violle, C., and Munoz, F. (2019). “Distinguishing the signatures of local environmental filtering and regional trait range limits in the study of trait–environment relationships”. In: *Oikos*.

2. From biodiversity patterns to processes and back again: solving the inverse problem in ecology

- Elliott, L. P. and Brook, B. W. (2007). “Revisiting Chamberlin: Multiple Working Hypotheses for the 21st Century”. en. In: *BioScience* vol. 57, no. 7, pp. 608–614.
- Garnier, E. et al. (2004). “Plant functional markers capture ecosystem properties during secondary succession”. In: *Ecology* vol. 85, no. 9, pp. 2630–2637.
- Gotelli, N. J. (2000). “Null model analysis of species co-occurrence patterns”. In: *Ecology* vol. 81, pp. 2606–2621.
- Gotelli, N. J. and McGill, B. J. (2006). “Null Versus Neutral Models: What’s The Difference?” en. In: *Ecography* vol. 29, no. 5, pp. 793–800.
- Gravel, D. et al. (2006). “Reconciling niche and neutrality: the continuum hypothesis”. In: *Ecology Letters* vol. 9, pp. 399–409.
- Grimm, V. et al. (2010). “The ODD protocol: a review and first update”. In: *Ecological modelling* vol. 221, no. 23, pp. 2760–2768.
- Hartig, F. et al. (2011). “Statistical inference for stochastic simulation models – theory and application”. In: *Ecology Letters* vol. 14, pp. 816–827.
- Hickerson, M. J. (2014). “All models are wrong”. In: *Molecular Ecology* vol. 23, pp. 2887–2889.
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton and Oxford: Princeton University Press.
- Imai, K. et al. (2011). “Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies”. en. In: *American Political Science Review* vol. 105, no. 4, pp. 765–789.
- Johnson, J. B. and Omland, K. S. (2004). “Model selection in ecology and evolution”. en. In: *Trends in Ecology & Evolution* vol. 19, no. 2, pp. 101–108.
- Karger, D. N., Cord, A. F., et al. (2016). “Delineating probabilistic species pools in ecology and biogeography”. In: *Global Ecology and Biogeography*, n/a–n/a.
- Karger, D. N., Wüest, R. O., et al. (2019). “Disentangling the drivers of local species richness using probabilistic species pools”. en. In: *Journal of Biogeography*.
- Kirsch, A. (2011). *An introduction to the mathematical theory of inverse problems*. Vol. 120. Springer Science & Business Media.
- Lawton, J. H. (1999). “Are there general laws in ecology?” English. In: *Oikos* vol. 84, pp. 177–192.
- Leibold, M. A. (2008). “Return of the niche”. In: *Nature* vol. 454, no. 7200. ISBN: 1476-4687 Publisher: Nature Publishing Group, pp. 39–41.
- Leibold, M. A. and McPeck, M. A. (2006). “Coexistence of the niche and neutral perspectives in community ecology”. In: *Ecology* vol. 87, no. 6, pp. 1399–1410.
- Levin, S. A. (1992). “The problem of pattern and scale in ecology”. In: *Ecology* vol. 73, pp. 1943–1967.
- Loranger, J. et al. (2018). “What makes trait–abundance relationships when both environmental filtering and stochastic neutral dynamics are at play?” In: *Oikos* vol. 127, no. 12, pp. 1735–1745.
- Lortie, C. J. et al. (2004). “Rethinking plant community theory”. In: *Oikos* vol. 107, no. 2, pp. 433–438.
- Mayfield, M. M. and Levine, J. M. (2010). “Opposing effects of competitive exclusion on the phylogenetic structure of communities”. In: *Ecology Letters* vol. 13, pp. 1085–1093.
- McGill, B. J. et al. (2006). “Rebuilding community ecology from functional traits”. In: *Trends in Ecology & Evolution* vol. 21, no. 4, pp. 178–185.

- McGill, B. J. (2010). “Towards a unification of unified theories of biodiversity”. In: *Ecology Letters* vol. 13, pp. 627–642.
- (2019). “The what, how and why of doing macroecology”. In: *Global Ecology and Biogeography* vol. 28, no. 1. ISBN: 1466-822X Publisher: Wiley Online Library, pp. 6–17.
- Munoz, F. and Couteron, P. (2012). “Estimating immigration in neutral communities: theoretical and practical insights into the sampling properties”. In: *Methods in Ecology and Evolution* vol. 3, no. 1. ISBN: 2041-210X Publisher: Wiley Online Library, pp. 152–161.
- Munoz, F., Grenié, M., et al. (2018). “ecolottery : Simulating and assessing community assembly with environmental filtering and neutral dynamics in R”. en. In: *Methods in Ecology and Evolution* vol. 9, no. 3. Ed. by Chisholm, R., pp. 693–703.
- Münkemüller, T. et al. (2012). “From diversity indices to community assembly processes: a test with simulated data”. In: *Ecography* vol. 35, no. 5. ISBN: 0906-7590 Publisher: Wiley Online Library, pp. 468–480.
- Perronne, R. et al. (2017). “How to design trait-based analyses of community assembly mechanisms: insights and guidelines from a literature review”. In: *Perspectives in Plant Ecology, Evolution and Systematics* vol. 25, pp. 29–44.
- Platt, J. R. (1964). “Strong inference”. In: *science* vol. 146, no. 3642. ISBN: 0036-8075 Publisher: JSTOR, pp. 347–353.
- Quinn, J. F. and Dunham, A. E. (1983). “On hypothesis testing in ecology and evolution”. In: *The American Naturalist* vol. 122, no. 5. ISBN: 0003-0147 Publisher: University of Chicago Press, pp. 602–617.
- Rolhauser, A. G. and Pucheta, E. (2017). “Directional, stabilizing, and disruptive trait selection as alternative mechanisms for plant community assembly”. In: *Ecology* vol. 98, no. 3, pp. 668–677.
- Ruffley, M. et al. (2019). “Identifying models of trait-mediated community assembly using random forests and approximate Bayesian computation”. In: *Ecology and Evolution*.
- Shipley, B. (2016). *Cause and correlation in biology: a user’s guide to path analysis, structural equations and causal inference with R*. Cambridge University Press.
- Shipley, B., Paine, C. E. T., and Baraloto, C. (2012). “Quantifying the importance of local niche-based and stochastic processes to tropical tree community assembly”. In: *Ecology* vol. 93, pp. 760–769.
- Shipley, B., Vile, D., and Garnier, É. (2006). “From plant traits to plant communities: a statistical mechanistic approach to biodiversity”. In: *science* vol. 314, no. 5800, pp. 812–814.
- Stouffer, D. B. (2019). “All ecological models are wrong, but some are useful”. en. In: *Journal of Animal Ecology* vol. 88, no. 2. Ed. by Gill, J., pp. 192–195.
- Sunnåker, M. et al. (2013). “Approximate Bayesian Computation”. In: *PLOS Computational Biology* vol. 9, no. 1, e1002803.
- Tarantola, A. (2006). “Popper, Bayes and the inverse problem”. In: *Nature physics* vol. 2, no. 8, p. 492.
- Tilman, D., Reich, P. B., and Knops, J. M. H. (2006). “Biodiversity and ecosystem stability in a decade-long grassland experiment”. en. In: vol. 441, p. 4.
- Violle, C., Navas, M. L., et al. (2007). “Let the concept of trait be functional!” In: *Oikos* vol. 116, pp. 882–892.

2. From biodiversity patterns to processes and back again: solving the inverse problem in ecology

- Violle, C., Enquist, B. J., et al. (2012). “The return of the variance: intraspecific variability in community ecology”. en. In: *Trends in Ecology & Evolution* vol. 27, no. 4, pp. 244–252.
- Violle, C., Nemergut, D. R., et al. (2011). “Phylogenetic limiting similarity and competitive exclusion: Phylogenetic relatedness and competition”. en. In: *Ecology Letters* vol. 14, no. 8, pp. 782–787.
- Yanco, S. W. et al. (2020). “A modern method of multiple working hypotheses to improve inference in ecology”. en. In: *Royal Society Open Science* vol. 7, no. 6, p. 200231.
- Zurell, D. et al. (2010). “The virtual ecologist approach: simulating data and observers”. en. In: *Oikos* vol. 119, no. 4, pp. 622–635.

Chapitre II

Influence des processus écologiques sur la diversité locale des communautés d'espèces

A l'échelle locale, deux familles distinctes de théories ont été développées pour prédire la distribution et l'abondance des espèces; théories qui ont souvent été mises en opposition. D'un côté la théorie neutre met l'accent sur le rôle de fluctuations stochastiques dans les abondances relatives des espèces de sorte que l'assemblage des communautés résulte principalement d'un équilibre entre dispersion et dynamique démographique. De l'autre, la théorie de la niche souligne que les attributs biologiques des espèces en relation avec l'environnement biotique et abiotique local déterminent principalement la composition de la communauté. Le concept de filtre environnemental représente comment les conditions abiotiques locales excluent de la communauté les espèces qui ne possèdent pas la combinaison de traits leur permettant de s'installer, persister et s'y reproduire avec succès. De ce fait, la composition de la communauté devrait alors refléter la manière dont ces différences d'aptitude découlant des variations de traits fonctionnels affectent la dynamique des espèces. Cela dit, plutôt que d'opposer les deux théories et de débattre de la prédominance de l'une sur l'autre, il conviendra de caractériser les contributions relatives de ces composantes. Dans une perspective où il n'y a pas de limite stricte entre les influences neutres et de niche mais plutôt un continuum de leur importance relative, une approche basée sur la simulation permet d'aborder explicitement où se situe une communauté donnée le long de ce gradient. Comme évoqué dans [Article 2](#), la simulation explicite de patron de diversité permet de comparer un modèle purement neutre, dans lequel l'assemblage ne dépend d'aucune différence entre espèces, à un modèle dans lequel la performance locale dépend davantage des attributs des espèces. D'une part, une analyse des erreurs de classification permettra de déterminer si un ensemble de métriques de diversité permettent de distinguer l'influence des processus de niche de celle des processus neutres. D'autre part, l'estimation des paramètres liés à la forme du filtre, par exemple, l'écart type dans le cas du filtrage gaussien, permet d'établir dans quelle mesure le filtre environnemental contraint la composition de la communauté. Un filtre très large implique que l'assemblage de la communauté est moins contraint par les valeurs des traits (filtrage plus faible), ce qui suggère soit un assemblage neutre.

L'[Article 3](#) propose d'évaluer les contributions jointes des dynamiques déterministes de niche et des dynamiques stochastiques neutre sur l'assemblage d'adventices. Les adventices de culture représentent un cas d'étude intéressant pour étudier les contributions relatives des dynamiques neutres et non-neutre dans l'assemblage des communautés. D'une part, les pratiques agricoles (bien souvent en monoculture) sont conçues pour maximiser le rendement des espèces cultivées et ciblent souvent les adventices par le biais de stratégies de désherbage. Ces méthodes de gestion constituent une contrainte majeure à la persistance des adventices. D'autre part, en raison des perturbations fréquentes, les adventices de cultures forment des petites populations sujettes à des taux de mortalité élevés et à une forte variabilité démographique. Du point de vue des adventices, les paysages arables en monoculture constituent un ensemble d'environnements perturbés qui forment une mosaïque d'habitat hétérogène à différents stades de succession. En raison de cette importante variabilité de la dispersion et de la démographie des espèces d'adventices, il est fort probable que les dynamiques neutres jouent un rôle important dans l'assemblage de ces communautés. L'[Article 3](#) montre que l'influence des dynamiques de niche sur l'assemblage des communautés d'adventices est plus importante au centre des parcelles. Ces conclusions concordent avec l'hypothèse selon laquelle les pratiques de gestion contribuent principalement au filtrage basé sur les traits, et que ces pratiques ont une plus grande influence au centre qu'en périphérie des parcelles agricoles.

2. From biodiversity patterns to processes and back again: solving the inverse problem in ecology

Cette étude montre également une plus grande importante dispersion limitée dans l'assemblage des communautés d'adventices au centre des parcelles, alors que l'assemblage des communautés en marge des champs était le plus souvent neutre. Dans le cas où un filtre environnemental contraint l'établissement des adventices, celles-ci sont filtrées pour ressembler à l'espèce cultivée avec laquelle elles coexistent. Par conséquent, l'hypothèse de filtre environnemental, plus que celle de limite de la similarité avec la culture, a affecté l'assemblage des communautés d'adventices de culture. Cela-dit, les interactions biotiques, et notamment la compétition, peuvent contraindre la croissance des espèces et donc leur appartenance à la communauté locale. Dans le cas des forêts tropicales, se pose la question de pourquoi autant d'espèces peuvent coexister dans des conditions environnementales relativement homogènes sans exclusion compétitive et sans domination. Si la théorie neutre permet d'expliquer certaines classes d'abondance par des modèles d'assemblage des communautés basée sur la dynamiques démographique des individus qui la composent sans intervention de différences biologiques, d'autres études évoquent le rôle des mécanismes de stabilisation tels que la partition d'habitat, des effets de densité-dépendance et un recrutement limité dans ces communautés. L'utilisation d'un indice de voisinage dans l'Article 4, permet d'évaluer grâce à des modèles bayésiens multiniveaux l'effet du voisinage sur l'accroissement des arbres tropicaux d'Uppangala en Inde. En prenant en compte la stratification verticale de cette parcelle forestière, cet article met en évidence un effet négatif du voisinage sur la croissance des arbres et donc une forme de compétition qui dépend de la topographie de la parcelle. Ainsi, la croissance des arbres tropicaux dépendait d'une combinaison de caractéristiques topographiques et fonctionnelles inhérentes à leur voisinage proche, qui permettait de détourner les interactions entre individus en termes d'effets de compétition pour les nutriments, l'eau, l'espace et la lumière.

Article 3

Dispersal limitation and environmental filtering interplay in shaping the assembly of weed plant communities

Elizabeth Barthélemy¹, Sabrina Gaba^{2,3,4}, François Munoz¹

in prep.

Abstract

Weed communities found in arable fields undergo strong environmental constraints related to management practices, make up spatially isolated communities, are subject to strong interactions with crops and to dispersal dynamics as well as demographic stochasticity. Niche theory, specifically the concept of environmental filtering dictates that only weeds with an appropriate set of traits allowing them to establish, grow, survive and reproduce, should persist. On the other hand, important demographic fluctuations leading to neutral dynamics should influence community composition independently from their trait values.

Using a simulation-based, mechanistic model of community assembly, we performed Approximate Bayesian Computation (ABC) to infer the relative contribution of neutral and environmental filtering on the composition of weed communities. In this model, environmental filtering weights the probability of immigration success depending on trait values, while neutral stochastic drift further shapes the variation in abundance of established immigrants. We hypothesized that environmental filtering could affect 6 functional traits: height, seed mass and Specific Leaf Area (SLA) as well as Ellenberg N, F and L indices.

We found that environmental filtering influenced the assembly of communities located at the center of arable fields more often than those located in field margins, consistently with the hypothesis that management practices mostly contribute to trait-based filtering, and that such practices have a greater influence at the center than at the interface of cultivated fields.

¹LIPhy, Université Grenoble Alpes, Grenoble, France

²LTSER Zone Atelier “Plaine & Val de Sèvre”, CNRS, Villiers-en-Bois, France

³Centre d’Études Biologiques de Chizé, USC Agripop, INRAE, Villiers-en-Bois, France

⁴Centre d’Études Biologiques de Chizé, UMR 7372, CNRS — Université de La Rochelle, Villiers-en-Bois, France

1 Introduction

Agrosystems represent a range of habitats, each defined by its own set of environmental constraints (sowing date, tillage, rotation, fertilization), and in which management practices are designed to maximize the yield of crop species at the expense of weed plant communities. In addition, agrosystems are embedded in a spatial context (hedges, neighboring fields, or pathways) leading to high fragmentation of weed communities and, because of frequent disturbance, weed populations are prone to high mortality rates and significant demographic variability (Navas 1991). Demographic stochasticity coupled with dispersal limitation should entail neutral dynamics (Hubbell 2001) of weed communities, while strong environmental filtering due to the management practices should conversely entail niche-based dynamics (Booth and Swanton 2002). Arable landscapes thus make up a set of disturbed environments of varying spatial extent, resulting in a heterogeneous mosaic of weed communities subject to strong stochastic dynamics and undergoing strong environmental filtering. Here, we address how the interplay of neutral and niche-based processes jointly shape weed community composition in a mosaic of agricultural fields.

The ability of weed species to occupy agricultural fields depends on the adequacy of their biological traits determining their dispersion, establishment and survival abilities in each community (Jabot, Etienne, and Chave 2008 ; Bourgeois, Munoz, et al. 2019). Thus, only weeds with an appropriate set of characteristics allowing them to settle, grow, survive and reproduce alongside a particular crop, should persist (Booth and Swanton 2002). In this case, functional traits reflect the ability of plants to establish, survive and reproduce in a given environment (Violle et al. 2007), so that the diversity of functional trait values in weed communities should reflect the influence of underlying ecological processes (McGill et al. 2006 ; Enquist et al. 2015). Specifically, the concept of environmental filtering represents how local environmental conditions act as a filter that only allows species with particular sets of traits to establish and develop, excluding all others (Bazzaz 1991 ; Woodward and Diament 1991). Species possessing less adapted characteristics can be, therefore, filtered out from the community. For instance, management practices, such as herbicide input (Ryan et al. 2010) or cover-crop mixtures (Smith et al. 2015), were shown to act as a filter on weed community composition, while non-managerial environmental filters included soil type (basic clay vs. acidic sandy soil, see Fried, Norton, and Reboud 2008). In fact, an examination of weed traits (Bourgeois, Munoz, et al. 2019), showed that it is the specific set of traits that allow such species to tolerate the ecological filters associated with arable fields, such as regular soil disturbances and fertilization, that make them explicitly “weeds”. Yet biotic and abiotic factors restricting the establishment of weeds are likely to vary over space and time. For instance, in the center of fields - where competition and management practices are harshest on weeds - weed assemblages have been shown to depend on crop type (Bourgeois, Gaba, et al. 2020) or crop height, a proxy of crop competition (Berquer, Martin, and Gaba 2021). In field margins, on the other hand, weed community assembly was influenced by spatial dispersal (Berquer, Martin, and Gaba 2021). Weed communities in field margins usually harbor more species and thus serve as sources for local dispersal (Bourgeois 2020 ; *source-sink dynamics* see Mouquet and Loreau 2003). Landscape heterogeneity in cultivated areas with more crop edges contributes to greater weed diversity, underlying the importance of dispersal (Gaba et al. 2010). Because of their intermediate location - weed communities in field margins are thus better

connected to potential sources of immigration. Meanwhile, the role of neutral processes ruling the assembly of weed communities has already been stressed (Perronne, Le Corre, et al. 2015), but the combined influence of these dynamics and trait-based ecological dynamics remain to be addressed.

Here we devise a model of community dynamics including the influence of both neutral dynamics and environmental filtering (Munoz, Grenié, et al. 2018). The probability of successful immigration from the species pool is weighed by an environmental filtering function depending on trait values (niche-based component), after which the variation in abundance of established immigrants depends on the balance between immigration and neutral stochastic drift (neutral component). By simulating communities over a large array of neutral to non-neutral scenarios and comparing simulated outcomes to our observations of weed communities, we performed Approximate Bayesian Computation (ABC) (Csilléry, M. G. B. Blum, et al. 2010) to infer which processes are most involved in shaping observed diversity patterns in weed communities.

2 Materials and Methods

2.1 Study site and sampling

The study site, covering 435km^2 , is named Zone Atelier Plaine & Val de Sèvre (ZA PVS, Bretagnolle et al. 2018). Located in the Poitou-Charentes region, in the south of Deux-Sèvres district (46.23°N , 0.41°W) in France, the ZA PVS is an intensive farming area including 11 000 fields, with mainly winter cereals. The region has a warm temperate climate, with 820mm mean annual precipitation, and poor alkaline soils with low water retention. Most common crops include wheat (33.8%), corn (9.6%), sunflower (10.4%), oilseed rape (8.3%), pea (2%) and meadows (13.5%), including both permanent grasslands and temporary hay (such as alfalfa). For more detailed information about the study site, see Bretagnolle et al. 2018. From 2005, weed presence/absence per m^2 in each quadrat (Gaba et al. 2010 ; Henckel et al. 2015) has been recorded each year between April to July for 100 to 250 fields of major crops and their margins, including at least 100 winter wheat fields per year.

2.2 Functional traits

We selected 6 traits among which response traits involved in the L-H-S strategy scheme (Westoby 1998): height, Specific Leaf Area (SLA) and seed mass; each being essential to the following aspect of plant development: competition with neighboring plants for light, light use efficiency and dispersal. High nutrient and low light conditions under crop canopies may additionally filter weeds depending on their resource-use strategy and tolerance. Hence, we considered Ellenberg N, L and F to account for weed preferences to nitrogen, light and soil moisture rich environments, respectively. We retrieved L-H-S functional trait values from the BiolFlor (Kühn, Durka, and Klotz 2004), Ecoflora (Fitter and Peat 1994), and LEDA databases (Knevel et al. 2003), whereas we extracted Ellenberg indices from Baseflor database (Julve 1998).

3. Dispersal limitation and environmental filtering interplay in shaping the assembly of weed plant communities

Trait	Unit	Function	Missing Data	Ref.
SLA	$m^2.kg^{-1}$	Ressource aquisition	23%	Kühn, Durka, and Klotz 2004 ; Fitter and Peat 1994 ; Knevel et al. 2003
Seed mass	g	Dispersal ; seedling competition	21.2%	Kühn, Durka, and Klotz 2004 ; Fitter and Peat 1994 ; Knevel et al. 2003
Height	m	Competition for light	21.4%	Kühn, Durka, and Klotz 2004 ; Fitter and Peat 1994 ; Knevel et al. 2003
Ellenberg L	-	Light preferences for growth	3.1%	Julve 1998
Ellenberg N	-	Nitrogen preferences for growth	3.1%	Julve 1998
Ellenberg F	-	Drought resistance	3.2%	Julve 1998

Table 3.1: Description of functional traits, related plant functions based on the literature and average relative missing trait information in the regional species pool

2.3 Delimitating local weed communities and defining the regional species pool

A regional species pool was defined as all weed occurrences that were recorded throughout the study site in 2016, 2017, 2018 and 2019. We defined weed assemblages as regrouping weeds occurring in field with identical crops as well as according to their position relative to the field: either in the center or at the field margin (where there is no tillage). To this end, we regrouped field based on a k-mean clustering analysis of their location based on the longitudinal and latitudinal coordinates ($n=15$). Then, a weed community was defined as all weeds occurring (i) in a specific group of fields, (ii) alongside a specific crop and (iii) given a certain position.

2.4 Choice of summary statistics

Simulated communities were compared to observed data, on the basis of 8 summary statistics meant to capture both functional and taxonomic patterns of diversity - the 4 first moments of the trait distribution (Enquist et al. 2015) as well as the 4 first Hill numbers of taxonomic composition (Chao, Chiu, and Jost 2014). Based on the assumption that stochastic processes should primarily influence species abundances and taxonomic diversity (Hubbell 2001 ; Munoz, Couteron, et al. 2007), while niche-based processes should be better detected with functional diversity metrics (McGill et al. 2006), we expected statistics related to both aspects to be key in uncovering both neutral and non-neutral dynamics in community assembly.

2.5 Estimating parameters of neutral and niche-based dynamics in weed communities

We used the ecolottery package in R language to simulate communities (i) receiving migrants from a reference external pool, (ii) subject to environmental filtering determining the establishment of migrants and the persistence of their descendants according to their functional traits, and (iii) subject to demographic stochasticity (Munoz et al., 2018). We compared two models of community assembly. In the first model, stabilizing environmental filtering was defined as a Gaussian distribution centered on an optimal trait value ($topt$), with a standard deviation (σ)

representing the strength of environmental filtering, i.e. the performance of species away from topt decreases quicker if s is small (Munoz, Grenié, et al. 2018). We simulated communities with topt values falling within the observed range of trait values in the species pool, and σ ranging between 1 and half the range of all trait values in the species pool. The second model was a neutral model, without trait-based filtering, in which all immigrant ancestors established in the community with equal probabilities - regardless of individual trait values - community assembly resulting only in the balance between demography and migration (at rate m).

We simulated communities over a wide range of migration and environmental filter parameter values and compared them to observed communities. We performed Approximate Bayesian Computation (ABC, Csilléry, M. G. B. Blum, et al. 2010 ; Csilléry, François, and M. G. Blum 2012) to infer parameter values and their plausible intervals for the observed communities. For each model, and for each trait, we produced 10,000 simulations. The sampled parameter values were accepted or rejected based on the distance between the simulated and observed summary statistics, given a tolerance threshold of 1%, thus producing a posterior distribution of most likely parameter values. First, we performed model selection between the purely neutral assembly model and that with additional environmental filtering, by computing the posterior model probabilities as the proportion of accepted simulations given each model. We selected a given model if that probability was greater than 70%. Then, given the selected model of assembly for each assemblage, we summarized the corresponding posterior parameter distributions by their mean. We compared these estimates between assemblages and according to position in field: whether at the center or at the field margin. We tested parameter variability among communities using an ANOVA to test whether the crop type and position (center vs. margin) accounted for differences in estimated parameter values.

2.6 Model misclassification and cross validation

Model selection with ABC is only relevant if it is able, given the existing simulations, to distinguish between the proposed models. Model misclassification was performed using the abc package (Csilléry, François, and M. G. Blum 2012) whereby a set of summary statistics is selected and classified according to all the remaining simulations and average the posterior probabilities over many simulations. Results are usually given as a matrix that encompasses the average posterior probabilities for a given model. Model selection was thus only undertaken for communities for which model selection was deemed satisfactory, i.e. when models were correctly identified in more than 70% of cases.

The accuracy of parameter estimation was appraised by cross-validation. A subset of 100 simulations were randomly selected and their parameters estimated by ABC so that the true parameter value for each simulation can be compared to its estimated value, for a given tolerance rate, thereby indicating the reliability of the parameter estimations for each model. Parameter estimation was thus only undertaken for communities for which the estimation error was no greater than 0.6.

3 Results

3.1 Selecting a model of weed community assembly

Model misclassification (Figure 3.S1) showed that for most traits, the chosen ABC scheme was able to correctly discriminate between the neutral assembly model and that with additional environmental filtering. The exception was for the assembly model with environmental filtering affecting weed seed mass, most likely because our dataset included numerous missing entries for this particular trait. For this reason, we discarded the assembly model affecting seed mass from model selection and further parameter estimation. Nevertheless, communities for which model selection was deemed unsatisfactory, i.e. when models were correctly identified in less than 70% of cases, were removed from the rest of the study (model selection and posterior parameter estimation).

The results of model selection (Figure 3.1) revealed a greater influence of environmental filtering in communities occurring in field centers rather than at the margin, alongside most crops and for most traits. The assembly model with the environmental filtering component was selected more often when it affected height, SLA and Ellenberg N, especially in grasslands, wheat and corn fields.

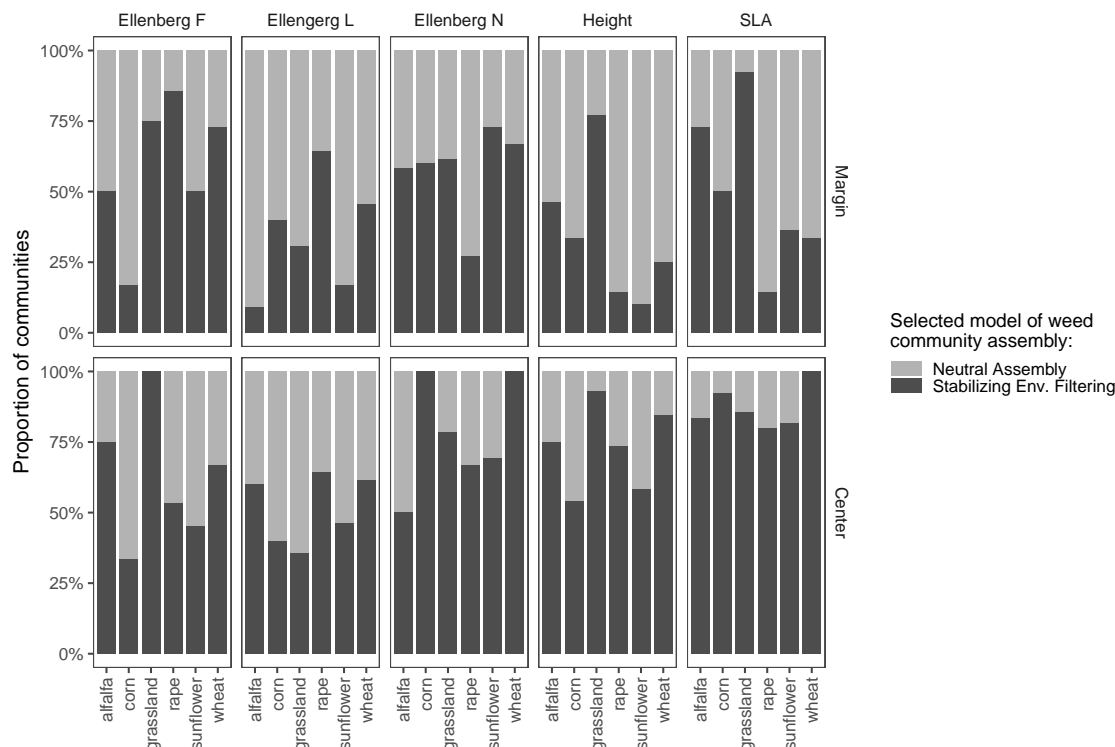


Figure 3.1: Model selection. Proportion of communities that were assigned to either a neutral model of community assembly (light gray) versus a model with additional stabilizing environmental filtering (dark gray) depending on the considered trait, on the field crop and on whether communities were located in the field center or at the margin.

3.2 Parameters of neutral and niche-based assembly in weed communities

Parameter estimation in communities for which the assembly model with a filtering component was selected showed significant effects of the type of crop and position relative to the field on the shape of the filter. In particular, the type of crop influenced the optimal trait value selected by the stabilizing filter (parameter t_{opt} , Figure 3.3), while the position – whether at the field center or at the edge – impacted the shape of the filter (parameter σ). For instance, the filter affecting weed communities growing alongside corn, selected species with more acquisitive strategies i.e. taller with larger SLA values and with larger Ellenberg N and F values. In the case of grasslands and alfalfa fields the filter tended to select species with lower SLA and Ellenberg N values (Figure 3.3). Meanwhile, for all crop types, the filter was narrower (smaller estimated σ) for communities occurring in the field center compared to those at the edges (Figure 3.4). These results were most significant for filters affecting SLA and height. Meanwhile, the migration parameter - which determines along with the size of weed communities the balance between immigration and neutral demographic drift - was lower (more limited dispersal) in field centers than at the margins (Figure 3.2).

4 Discussion

Using a simulation-based mechanistic model of community assembly (Munoz, Grenié, et al. 2018), we inferred using Approximate Bayesian Computation (ABC) the relative contribution of neutral and environmental filtering influencing the composition of weed communities. We found that environmental filtering influenced the assembly of communities located at the center of arable fields more often than those located in field margins, consistently with the hypothesis that management practices mostly contribute to trait-based filtering, and that such practices have a greater influence at the center than at the interface of cultivated fields (Berquer, Martin, and Gaba 2021). The misclassification analysis showed that we could distinguish the imprint of neutral dynamics and environmental filtering on taxonomic and functional diversity with ABC analysis. Hence, despite low individual counts and species richness, community composition proved informative enough to discern between models. Because misclassification results were highly discriminant, these results did not reflect lack of detection of environmental filtering, but a greater influence of stochastic dynamics. Yet the misclassification analysis showed that in the case of seed mass, we could not discriminate between assembly models efficiently, plausibly because of lack of trait data. However, this was not an issue for height and SLA, for which trait data was missing in comparable proportions.

We also found a greater influence of dispersal limitation in the assembly of weed communities in field centers in conjunction with environmental filtering, while the assembly of communities in field margins was most often purely neutral. Neutral theory underlines the influence of demographic stochasticity on assembly, especially when immigration is low and does not counterbalance local extinction (e.g., tropical forests, Hubbell 2001). In the context of weed communities, relaxed environmental constraints found at the interface may enhance the influence of demographic fluctuations, however, given their location these communities are also better connected to potential sources of immigration nearby in the landscape. In field centers, there was both stronger trait-based

3. Dispersal limitation and environmental filtering interplay in shaping the assembly of weed plant communities

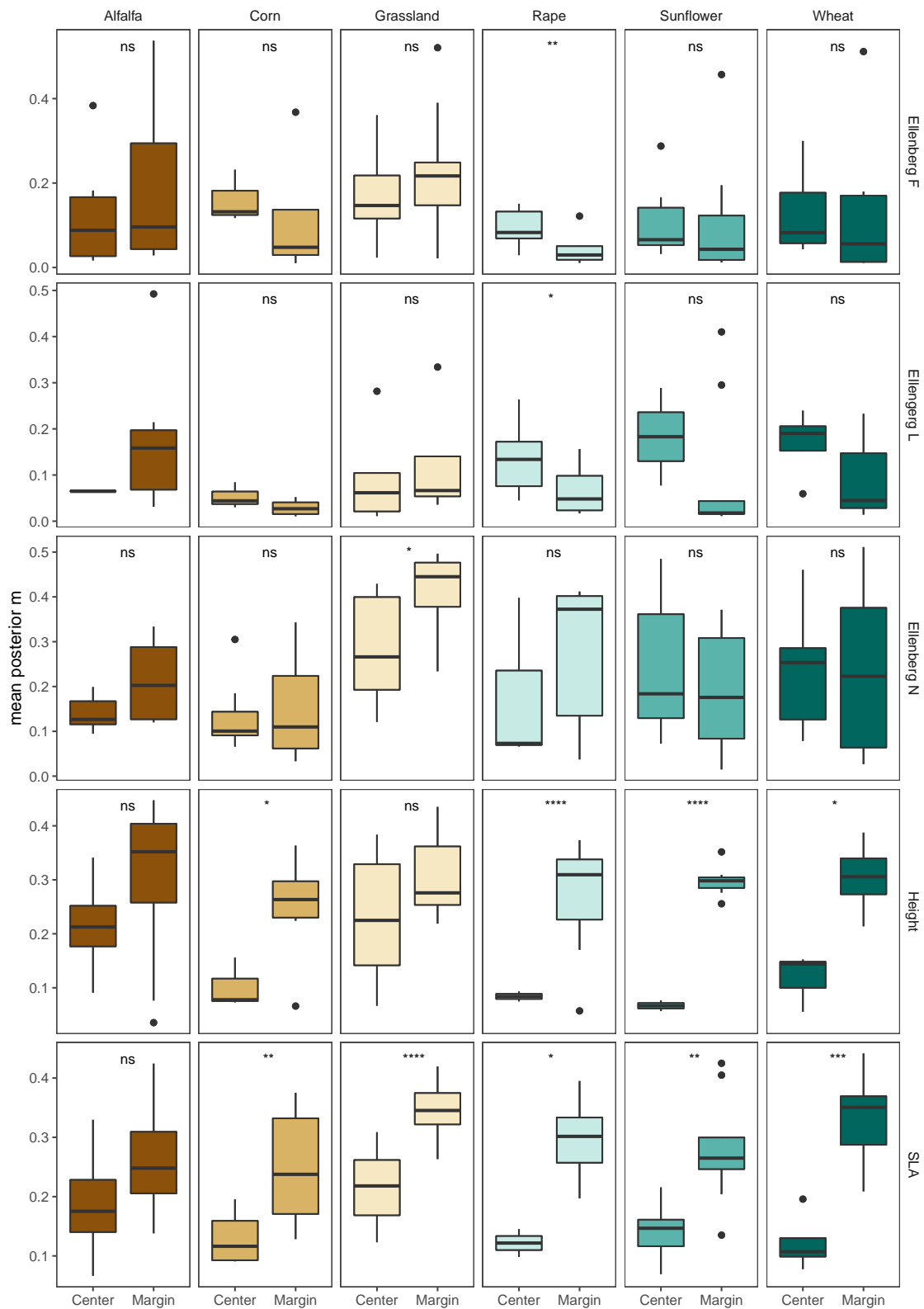


Figure 3.2: Mean posterior migration rate m of the weed assembly model (with additional environmental filtering) depending on different traits (Height, Specific Leaf Area, and Ellenberg F, L and N) and according to the type to the type of crop and position relative to the field: either in the center or at the field margin (where there is no tillage)

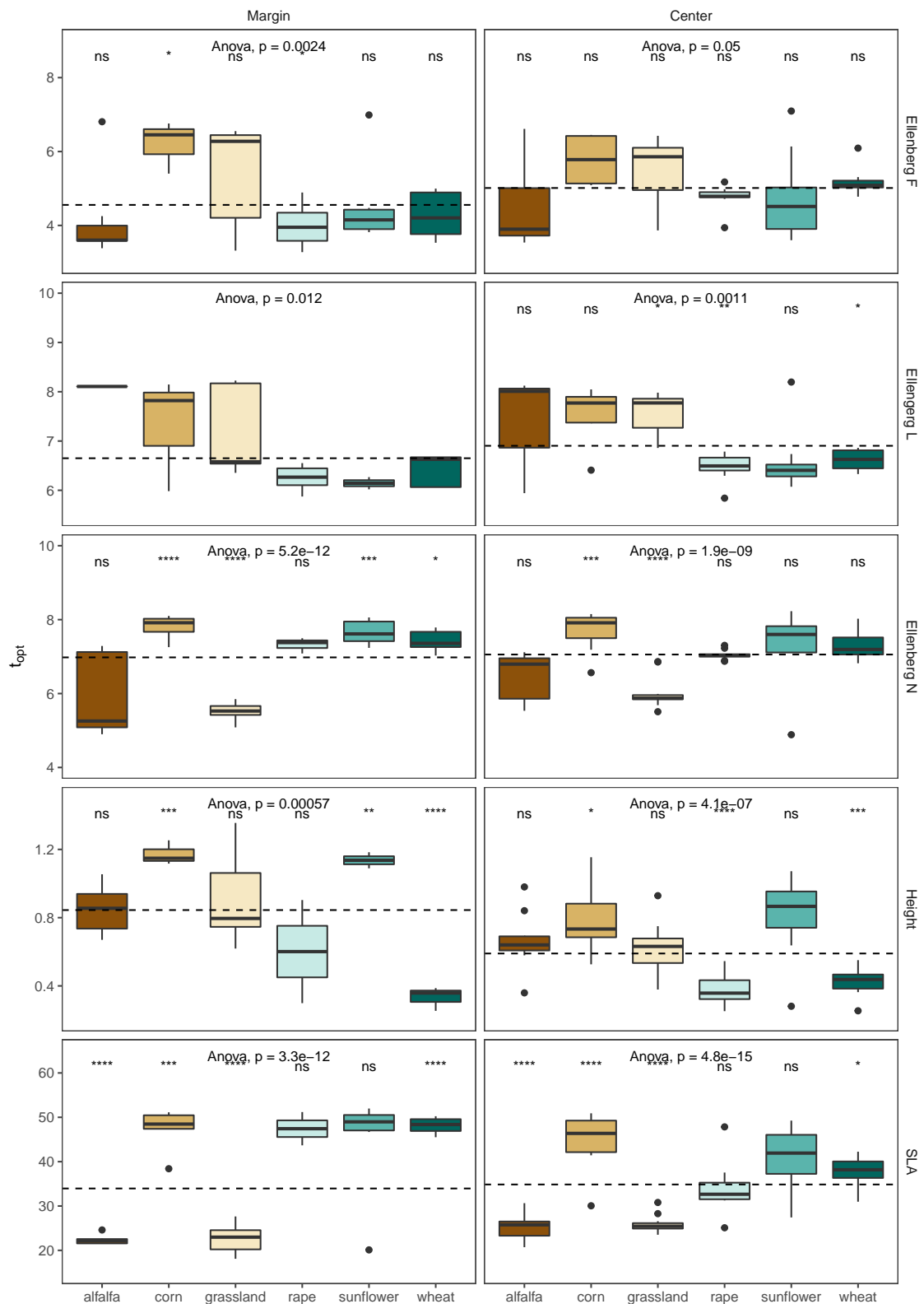


Figure 3.3: Mean posterior optimal trait value (t_{opt}) of the environmental filtering component of the weed assembly model depending on different traits (Height, Specific Leaf Area, and Ellenberg F, L and N) and according to the type to the type of crop and position relative to the field: either in the center or at the field margin (where there is no tillage)

3. Dispersal limitation and environmental filtering interplay in shaping the assembly of weed plant communities

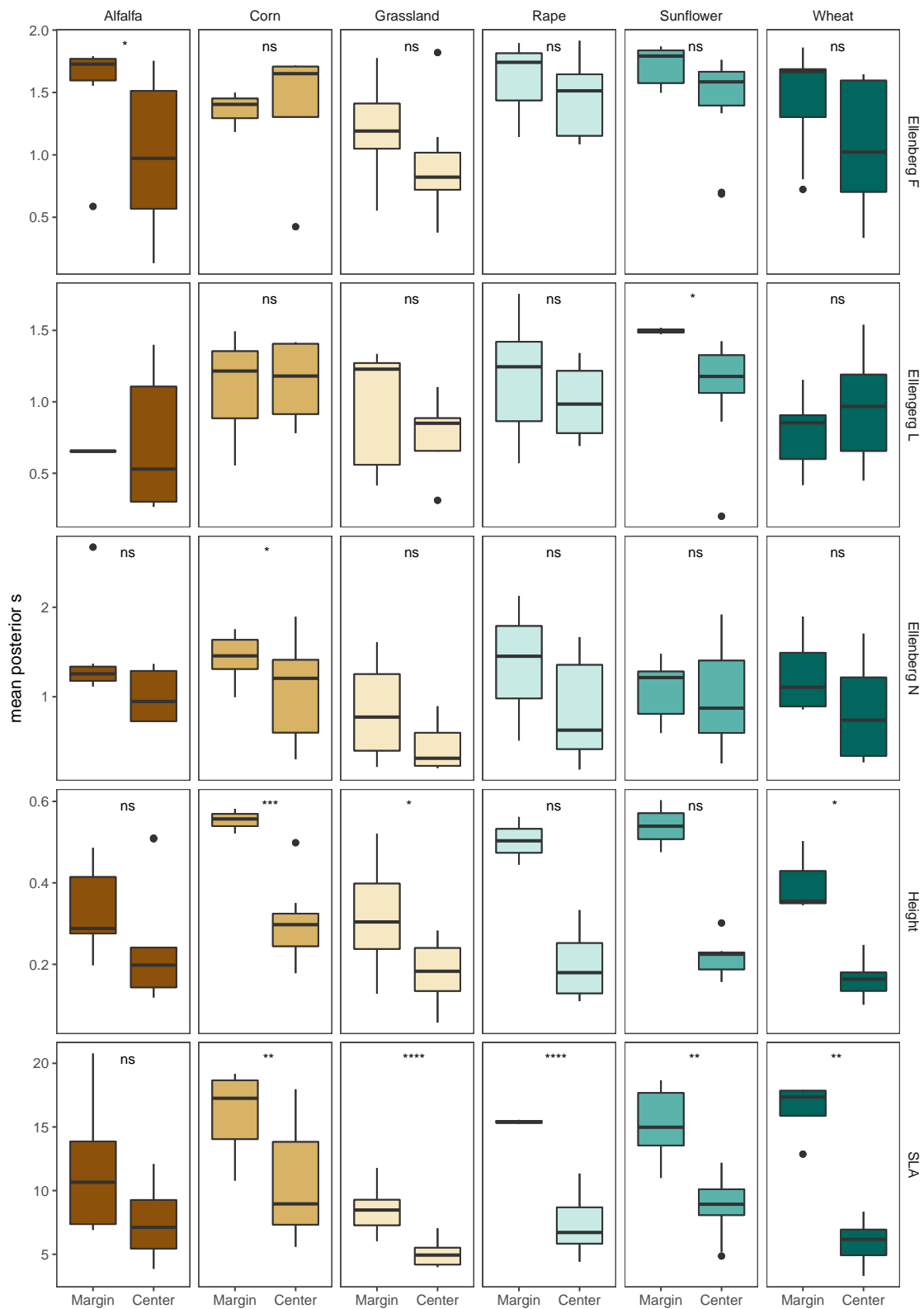


Figure 3.4: Mean posterior σ of the environmental filtering component of the weed assembly model depending on different traits (Height, Specific Leaf Area, and Ellenberg F, L and N) and according to the type to the type of crop and position relative to the field: either in the center or at the field margin (where there is no tillage)

filtering and less accessibility to immigrants. Therefore, our results support the hypothesis that dispersal limitation (Bourgeois, Gaba, et al. 2020) increases from the margin to the center, but does not necessarily entail prevailing neutral dynamics over environmental filtering at the center. However, by focusing on environmental filtering in a spatially implicit manner, we do not account for landscape complexity especially at the field edges which has shown to play an important part in weed community assembly (Gaba et al. 2010 ; Bourgeois, Gaba, et al. 2020), and will no doubt have repercussions on dispersal limitations.

Two of the traits from the LHS (leaf-height-seed) model (Westoby 1998), which considers specific leaf area (SLA), species height and seed mass as three primary components summarizing plant strategies and the influence of selective forces, were most often significantly filtered according to the crop alongside which weed communities occurred. However, we assumed that crop type was a proxy for both environmental constraints and competition with the crop (Berquer, Martin, and Gaba 2021). We also considered that typical management practices (fertilization, irrigation, sowing, tillage and harvest dates etc.) for each crop type were being upheld without actual in situ measurements. Making this assumption fails to take into account other important environmental gradients within and/or between fields, such as soil types for instance (Fried, Norton, and Reboud 2008). Since we do not have specific data regarding management practices in each individual field, whether filtering according to crop type occurred because of the direct interaction between the crop species and the weeds or rather because of the agricultural practices associated with specific crops, cannot be said. Yet, our results indicate that species in weed communities were filtered to resemble the crop they coexisted with. For instance, the filter restraining the establishment of weeds alongside corn, selected taller species with more exploitative leaves. Thus, functional convergence more than limiting similarity with the crop affected weed community assembly.

However, these results are contingent upon our own definition of the species pool used in this study. Inferring environmental filtering parameters and migration rates from a large-scale pool covering a broad spectrum of abiotic conditions, as we have done here, will likely emphasize the influence of species niche differences and of their limited dispersal abilities depending on landscape structure (Karger et al. 2016). Had we inferred the same parameters from a species pool restricted to the composition of adjacent communities we may have detected the role of finer-scale biotic interactions in shaping community assembly (Bernard-Verdier et al. 2012 ; Perronne, Munoz, et al. 2017). Furthermore, we did not explicitly account for field history which, because of crop rotation and fallow lands, can have repercussions on seed banks and thus impact relative abundances of species in the pool (Leck, Parker, and Simpson 1989 ; Cardina, Herms, and Doohan 2002 ; Borgy et al. 2015 ; Bohan et al. 2011).

Finally, a major challenge in community ecology is to assess the relative influence of neutral and niche-based processes, but available approaches mostly consider neutral and niche theories separately. Here we propose a methodological framework to assess their joint influence. More than permutation-based null models, which do not explicitly represent neutral processes when defining their null hypothesis, this approach can quantify the contribution of neutral and niche-based processes to community assembly – specifically through model selection and parameter estimation. The generalization of such approaches (see Blanchard, Birnbaum, and Munoz 2020, for another

3. Dispersal limitation and environmental filtering interplay in shaping the assembly of weed plant communities

example) may help uncover the relative entwined contributions of what are still considered to be opposing mechanisms.

Author contributions. All authors contributed to design the study. EB performed the analyses. EB wrote the first draft of the manuscript. All authors contributed revisions to the manuscript and gave final approval for publication.

References

- Bazzaz, F. A. (1991). "Habitat selection in plants". In: *The American Naturalist* vol. 137. ISBN: 0003-0147 Publisher: University of Chicago Press, S116–S130.
- Bernard-Verdier, M. et al. (2012). "Community assembly along a soil depth gradient: contrasting patterns of plant trait convergence and divergence in a Mediterranean rangeland". In: *Journal of Ecology* vol. 100, pp. 1422–1433.
- Berquer, A., Martin, O., and Gaba, S. (2021). "Landscape Is the Main Driver of Weed Assemblages in Field Margins but Is Outperformed by Crop Competition in Field Cores". In: *Plants* vol. 10, no. 10. ISBN: 2223-7747 Publisher: MDPI, p. 2131.
- Blanchard, G., Birnbaum, P., and Munoz, F. (2020). "Extinction–immigration dynamics lag behind environmental filtering in shaping the composition of tropical dry forests within a changing landscape". In: *Ecography*.
- Bohan, D. A. et al. (2011). "Modelling rotations: can crop sequences explain arable weed seedbank abundance?" In: *Weed Research* vol. 51, no. 4. ISBN: 0043-1737 Publisher: Wiley Online Library, pp. 422–432.
- Booth, B. D. and Swanton, C. J. (2002). "Assembly theory applied to weed communities". In: *Weed Science* vol. 50, no. 1. ISBN: 0043-1745 Publisher: Cambridge University Press, pp. 2–13.
- Borgy, B. et al. (2015). "Dynamics of weeds in the soil seed bank: a hidden Markov model to estimate life history traits from standing plant time series". In: *PloS one* vol. 10, no. 10. ISBN: 1932-6203 Publisher: Public Library of Science San Francisco, CA USA, e0139278.
- Bourgeois, B., Gaba, S., et al. (2020). "Weed diversity is driven by complex interplay between multi-scale dispersal and local filtering". In: *Proceedings of the Royal Society B* vol. 287, no. 1930. ISBN: 0962-8452 Publisher: The Royal Society, p. 20201118.
- Bourgeois, B., Munoz, F., et al. (2019). "What makes a weed a weed? A large-scale evaluation of arable weeds through a functional lens". In: *American Journal of Botany* vol. 106, no. 1. ISBN: 0002-9122 Publisher: Wiley Online Library, pp. 90–100.
- Bretagnolle, V. et al. (2018). "Description of long-term monitoring of farmland biodiversity in a LTSER". In: *Data in brief* vol. 19. ISBN: 2352-3409 Publisher: Elsevier, pp. 1310–1313.
- Cardina, J., Herms, C. P., and Doohan, D. J. (2002). "Crop rotation and tillage system effects on weed seedbanks". In: *Weed Science* vol. 50, no. 4, pp. 448–460.
- Chao, A., Chiu, C.-H., and Jost, L. (2014). "Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers". In: *Annual review of ecology, evolution, and systematics* vol. 45. ISBN: 1543-592X Publisher: Annual Reviews, pp. 297–324.

- Csilléry, K., Blum, M. G. B., et al. (2010). “Approximate Bayesian computation (ABC) in practice”. In: *Trends in Ecology & Evolution* vol. 25, pp. 410–418.
- Csilléry, K., François, O., and Blum, M. G. (2012). “abc: an R package for Approximate Bayesian Computation (ABC)”. In: *Methods in Ecology and Evolution* vol. 3, pp. 475–479.
- Enquist, B. et al. (2015). “Scaling from traits to ecosystems: Developing a general Trait Driver Theory via integrating trait-based and metabolic scaling theories”. In: *Advances in Ecological Research* vol. 52.
- Fitter, A. H. and Peat, H. J. (1994). “The ecological flora database”. In: *Journal of ecology*. ISBN: 0022-0477 Publisher: JSTOR, pp. 415–425.
- Fried, G., Norton, L. R., and Reboud, X. (2008). “Environmental and management factors determining weed species composition and diversity in France”. In: *Agriculture, ecosystems & environment* vol. 128, no. 1-2. ISBN: 0167-8809 Publisher: Elsevier, pp. 68–76.
- Gaba, S. et al. (2010). “Weed species richness in winter wheat increases with landscape heterogeneity”. In: *Agriculture, Ecosystems & Environment* vol. 138, no. 3-4. ISBN: 0167-8809 Publisher: Elsevier, pp. 318–323.
- Henckel, L. et al. (2015). “Organic fields sustain weed metacommunity dynamics in farmland landscapes”. In: *Proceedings of the Royal Society B: Biological Sciences* vol. 282, no. 1808. ISBN: 0962-8452 Publisher: The Royal Society, p. 20150002.
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton and Oxford: Princeton University Press.
- Jabot, F., Etienne, R. S., and Chave, J. (2008). “Reconciling neutral community models and environmental filtering: theory and an empirical test”. In: *Oikos* vol. 117, pp. 1308–1320.
- Julve, P. H. (1998). “Baseflor. Index botanique, écologique et chorologique de la flore de France”. In: *Institut Catholique de Lille, Lille*.
- Karger, D. N. et al. (2016). “Delineating probabilistic species pools in ecology and biogeography”. In: *Global Ecology and Biogeography*, n/a–n/a.
- Knevel, I. C. et al. (2003). “Life-history traits of the Northwest European flora: the LEDA database”. In: *Journal of Vegetation Science* vol. 14, no. 4. ISBN: 1100-9233 Publisher: Wiley Online Library, pp. 611–614.
- Kühn, I., Durka, W., and Klotz, S. (2004). “BiolFlor: a new plant-trait database as a tool for plant invasion ecology”. In: *Diversity and Distributions* vol. 10, no. 5/6. ISBN: 1366-9516 Publisher: JSTOR, pp. 363–365.
- Leck, M. A., Parker, V. T., and Simpson, R. L. (1989). *Ecology of soil seed banks*. en. San Diego: Academic Press.
- McGill, B. J. et al. (2006). “Rebuilding community ecology from functional traits”. In: *Trends in Ecology & Evolution* vol. 21, no. 4, pp. 178–185.
- Mouquet, N. and Loreau, M. (2003). “Community patterns in source-sink metacommunities”. 0003-0147. In: *American Naturalist* vol. 162, pp. 544–557.
- Munoz, F., Couteron, P., et al. (2007). “Estimating parameters of neutral communities: from one single large to several small samples”. In: *Ecology* vol. 88, no. 10. ISBN: 1939-9170 Publisher: Wiley Online Library, pp. 2482–2488.

3. Dispersal limitation and environmental filtering interplay in shaping the assembly of weed plant communities

- Munoz, F., Grenié, M., et al. (2018). “ecolottery : Simulating and assessing community assembly with environmental filtering and neutral dynamics in R”. en. In: *Methods in Ecology and Evolution* vol. 9, no. 3. Ed. by Chisholm, R., pp. 693–703.
- Navas, M.-L. (1991). “Using plant population biology in weed research: a strategy to improve weed management”. In: *Weed Research* vol. 31, no. 4. ISBN: 0043-1737 Publisher: Wiley Online Library, pp. 171–179.
- Perronne, R., Le Corre, V., et al. (2015). “Stochastic processes and crop types shape weed community assembly in arable fields”. In: *Journal of Vegetation Science* vol. 26, no. 2. ISBN: 1100-9233 Publisher: Wiley Online Library, pp. 348–359.
- Perronne, R., Munoz, F., et al. (2017). “How to design trait-based analyses of community assembly mechanisms: insights and guidelines from a literature review”. In: *Perspectives in Plant Ecology, Evolution and Systematics* vol. 25, pp. 29–44.
- Ryan, M. R. et al. (2010). “Management filters and species traits: weed community assembly in long-term organic and conventional systems”. In: *Weed Science* vol. 58, no. 3. ISBN: 0043-1745 Publisher: Cambridge University Press, pp. 265–277.
- Smith, R. G. et al. (2015). “Cover-crop species as distinct biotic filters in weed community assembly”. In: *Weed Science* vol. 63, no. 1. ISBN: 0043-1745 Publisher: Cambridge University Press, pp. 282–295.
- Violle, C. et al. (2007). “Let the concept of trait be functional!” In: *Oikos* vol. 116, pp. 882–892.
- Westoby, M. (1998). “A leaf-height-seed (LHS) plant ecology strategy scheme”. In: *Plant and soil* vol. 199, no. 2. ISBN: 1573-5036 Publisher: Springer, pp. 213–227.
- Woodward, F. I. and Diament, A. D. (1991). “Functional approaches to predicting the ecological effects of global change”. In: *Functional Ecology*. ISBN: 0269-8463 Publisher: JSTOR, pp. 202–212.

Appendix 3.A Supplementary Figures

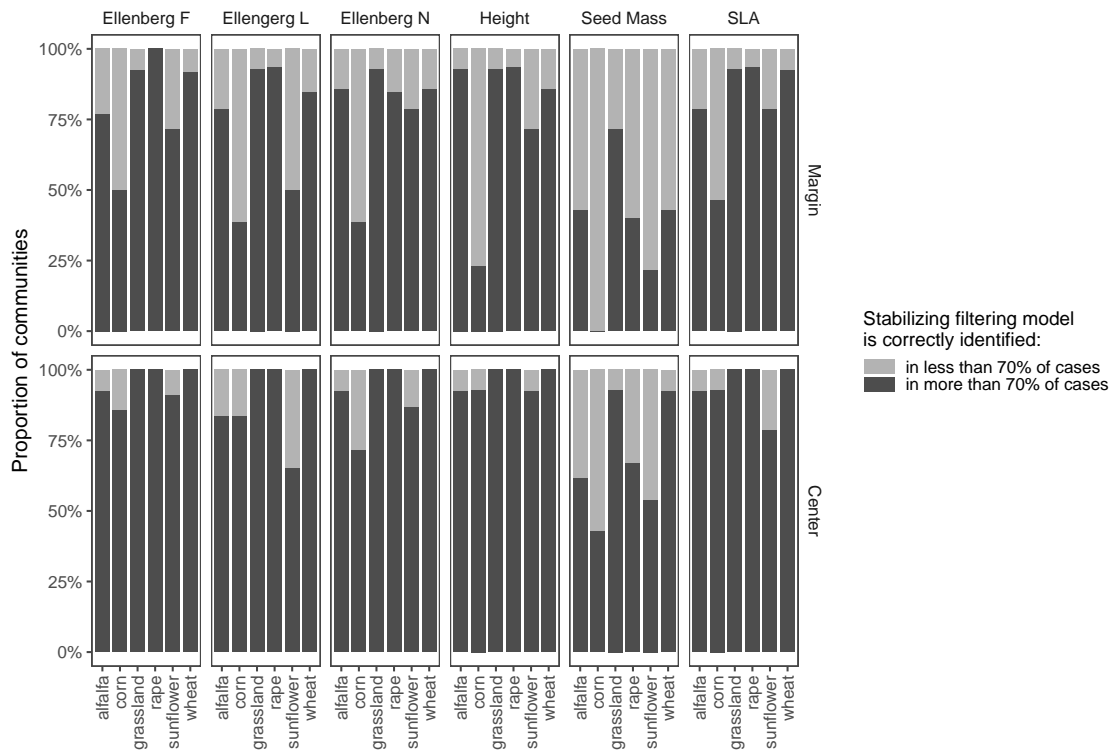


Figure 3.S1: Misclassification analysis. For each trait, and given each crop type and relative position in the field, dark grey bars represent the proportion of weed communities for which the environmental filtering assembly model is deemed satisfactory (i.e. correctly identified in more than 70% of cases).

Article 4

The spatial occupancy and functional similarity of neighboring trees determine tree growth in a tropical rainforest

Geoffray Reboud¹, Elizabeth Barthélemy¹, François Munoz¹

in prep.

Abstract

Tropical rainforests represent a very rich and diverse environment that provides a subject for the study of tree interactions, which are still poorly understood. Based on a neighborhood index as well as trait and topography data we used multilevel Bayesian models to investigate effect of tree proximity on a vital index, increment, for trees surveyed in the Uppangala permanent plots in India. These models allow the integration of parameters at the population and group level. We show that traits have little effect in isolation, but rather in interaction with topography and/or neighborhood. Despite a rather homogeneous occupation of the plot, the neighborhood index showed negative effects regardless of the structural level of the trees. Thus the growth of the trees depended on a combination of topographical and functional features inherent to their close vicinity, which made detouring the interactions between individuals in terms of competition effects for nutrients, water, space and light possible.

¹LIPhy Université Grenoble Alpes, Grenoble, France

1 Introduction

Tropical rainforests represent a flagship case to investigate the role of ecological processes in shaping biodiversity (Whitmore 1990). These forests include many species of trees, yet each of them occurs at low population density. A long-standing conundrum regards why so many species can coexist in relatively homogeneous environmental conditions without competitive exclusion and dominance. Much debate has then focused on whether and how the patterns of species abundances can be explained by models of community assembly and population dynamics (Hubbell 2001). The measured patterns can concern species identity and relative abundance in sampling units (taxonomic diversity), or the diversity of functional attributes of coexisting trees (functional diversity, e.g. Baraloto et al. 2010). Despite there having been much effort in this direction, there is no consensus on which ecological processes shape the observed patterns. Recently though, some works have promoted measuring and analyzing more directly the parameters of tree and population dynamics, in terms of variation of individual numbers, tree growth, fecundity and establishment (a.k.a. vital rates), in relationship to the putative drivers. Although such an approach should help resolve uncertainty on the nature of dynamics underlying rainforest hyperdiversity, it has seldom been applied to tropical forests till now. Here, we investigate a key vital rate in forest dynamics, i.e., tree diameter growth, and question whether and how it depends on competition with neighboring individuals.

Competition is a central process in ecology due to the limited resources on which coexisting individuals must thrive to complete their life cycle. It is, therefore, a tenet of ecological theory, since the seminal works of Gause on competitive exclusion (1934) (Kneitel 2019), of MacArthur and Levins on the coexistence of competing species (MacArthur and Wilson 1967, and of Tilman (Tilman 1994) on the way resource depletion and resource use jointly drive coexistence. Whether competition is direct (e.g., vertical competition for light) or indirect (e.g., use of shared resources), it basically affects individuals that are close enough in space to interact in this way. From this premise, several works have considered evaluating the nature and the intensity of competition in the neighborhood of individuals. For instance, the Neighborhood Crowding Index (NCI) encapsulates the spatial occupancy of trees in the neighborhood of a focal tree (Fortunel et al. 2016 ; Kunstler, Lavergne, et al. 2012). In terms of vital rates decomposing the parameters of an individual life cycle, a way to assess and validate the role of competition is to assess how vital rates change in response to variation in NCI (Laughlin et al. 2020; Pélissier and J.-p. Pascal 2000). A commonly addressed vital rate in forest ecology is tree growth (basically in terms of diameter increment).

A specific feature of forests, especially tropical rainforests, is vertical stratification (Richards, 1979). This phenomenon represents the distribution of the trees across vertical layers. During its lifetime a tree can belong successively to different vertical strata, and ultimately reaches a specific stratum when at the adult stage. We can thus define potential strata of trees that define their adult habit, e.g., emergent, canopy, subcanopy and understorey tree species (Figure 4.1). Each stratum represents distinct environmental constraints and conditions. Light availability sharply decreases from the canopy down to understorey. Hydraulic constraints conversely increase from the understorey up to the canopy. Therefore, we expect that competition for light and resources do

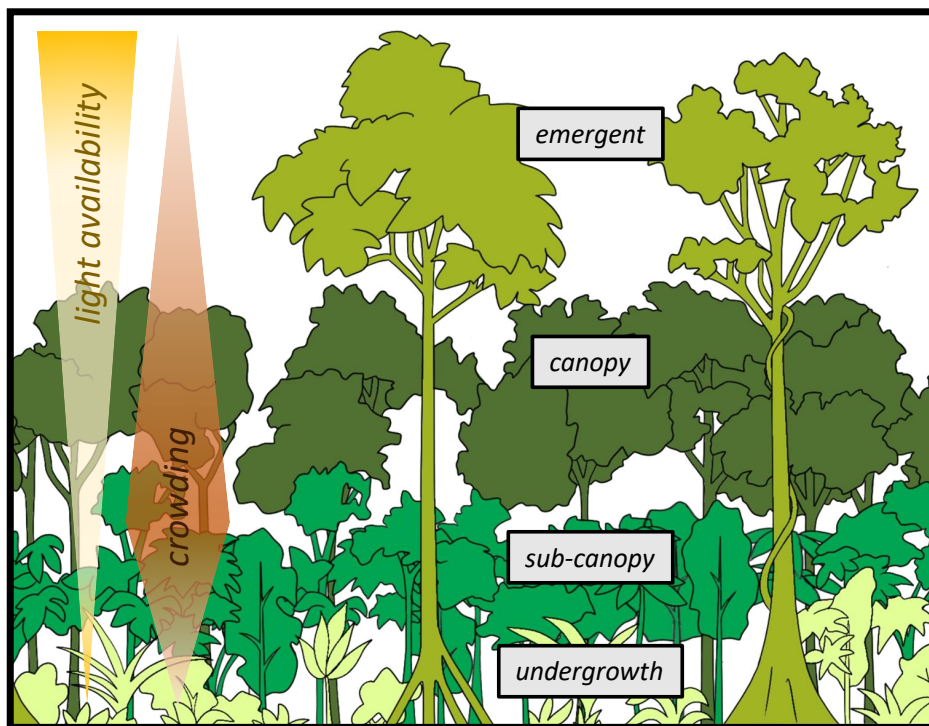


Figure 4.1: Diagram of the different tropical tree strata and associated constraints in terms of light availability and crowding

not play the same way across strata (Schmitt et al. 2021). In addition, the potential stratum of a given species represents a feature of its fundamental niche in terms of its ability to cope with competition during lifetime (Kunstler, Falster, et al. 2016). We thus hypothesize here that the role of competition on tree growth will vary depending on the potential stratum of tree species.

The Western Ghats in South India represent a 1000 km long mountain chain facing the South West Coast. Due to its steep topography, it is subject to an intensive monsoon regime allowing the presence of wet evergreen forests throughout the range. During recent decades, most of the surface of the rainforests has been deforested in lowland and only a few, protected remnants can be found. One of the remnants is the Uppangala permanent forest plots in Karnataka State. For more than 30 years, all tree individuals in the forests have been monitored and their diameter growth recorded. Uppangala thereby represents an ideal case to address the spatial structure of trees and the role of competition on temporal dynamics of a tropical rainforest.

We hypothesized that the composition of the neighborhood can affect tree growth in two main respects. First, competition for light is a key constraint on tree development, growth and survival. The spatial occupancy of trees in the neighborhood likely affects the access to light of a focal tree: the larger they are, the more competition they should exert (Figure 4.1). Therefore, we expect that tree growth should decrease when the density and basal area of trees in the neighborhood increase. Furthermore, as light access depends on the tree height, the potential stratum of individuals should affect the competitive impact of the neighborhood, thus we expect that trees belonging to each strata behave differently towards competition. Second, competition for resources in a broader sense depends on a set of functional traits driving resource uptake and conservation. We expect that

4. The spatial occupancy and functional similarity of neighboring trees determine tree growth in a tropical rainforest

the more similar trees are to a focal tree in terms of these functional traits, the more competition there should be for resources overall among them. We expect the leaf economics spectrum (LES) to influence especially the light uptake (Chen et al. 2020), the wood economics spectrum (WES) to affect the growth depending on the strata and represent the strategic trade-offs (Chave et al. 2009) and both LES and WES to be representative of the resources allocation of trees (Laughlin et al. 2020). Using long-term monitoring and spatial cartography of trees in a lowland tropical rainforest in South India, we measured tree growth, characterized their neighborhood, and tested our two main hypotheses with these data.

2 Material and Methods

2.1 Study site and sampling

The Uppangala study site is located within a 28-ha station in the Kadamakal Reserve Forest which is part of the Pushpagiri Wildlife Sanctuary in Karnataka state, in the Western Ghats of India. A total of 5.07 ha of undisturbed, old-growth wet evergreen Dipterocarp forest were censused every 3 to 5 years over a 20 year period to include 3870 trees with a minimum girth at breast height (gbh) of 30 cm having been identified, mapped and fitted with stainless steel dendrometer bands for accurate growth monitoring. The complete sampling scheme of the Uppangala forest plot is described in (Pélissier, J.-P. Pascal, et al. 2011).

2.2 Neighborhood Crowding Index (NCI)

The NCI was computed as described in formula [ref] where n is the number of neighbors and s_i their index, k the focal tree for which the NCI is computed, DBH_i is the DBH of the i th tree and $i:k$ represents either the physical distance or the functional distance between the i th tree and the focal tree. We computed the NCIs [ref eq] by first dividing the original NCI by the proportion of neighboring trees for which DBH was available and second dividing by the proportion of empty area surrounding each focal tree. The code in R language for the NCIs is available in the supplementary materials. This allowed us to take into account the bias caused by edge effects due to the somewhat irregular plot shape and where therefore, considered this corrected version of the original NCI as our final and most robust index, as it should better represent the reality of the competition between the censused trees.

2.3 Bayesian multilevel model implementation

We used two variables in our models that stand for wood economics spectrum (WES) and leaf economics spectrum (LES). They ensue from the scores on the 2 first axes of the PCA derived from all available trait data. WES main components are WD (wood density) and LA (leaf area), LES main components are SLA (specific leaf area) and LDMC (leaf dry matter content).

We designed bayesian multilevel models as exemplified in (Fortunel et al. 2016), using the ‘brms’ R package. The models included both 23 predictors with 5 main and 12 interaction fixed effects, plus 6 random effects. The predictors were the DBH (Diameter at Breast Height), NCI (Neighborhood Crowding Index), Altitude, values of traits and strategies for leaves and wood (WES and LES), interactions between the two last and all the other and also topographical variables as Slope, Curvature, altitude and SW (SouthWesterness), random effects of year, individual and species, and the random effects of species on altitude, DBH and NCI.

Each species belonged to one of four potential strata (Figure 4.1) representing the structural level of the forest that adult individuals could reach. We built one BRM model for each stratum, hence yielding a total of 4 distinct models.

All the analyses have been performed with R version 4.1.1. (R_Development_Core_Team 2010)

3 Results

The first axis of the PCA captured 36.4% of the total variation in tree traits and was essentially carried by the leaf dry matter content (LDMC) and specific leaf area (SLA). Scores along this axis were assigned to species and represented their leaf economic strategy further referred to as LES (Figure 4.2). The second axis of the PCA accounted for 32.5% of tree trait variation. Wood density (WD) and leaf area (LA) contributed most to the second axis. Species’ scores along the second axis of the PCA thus reflected their preference along the wood economic spectrum (WES) (Figure 4.2).

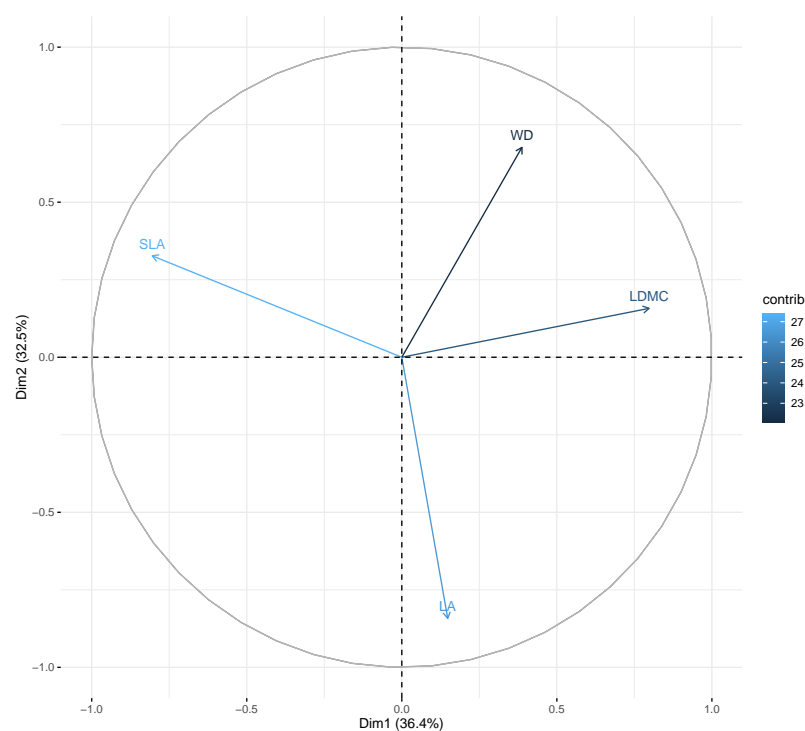


Figure 4.2: Principal Component Analysis (PCA) on tropical tree wood and leaf traits (SLA, LA, LDMC and WD): correlation circle showing the contributions of tree traits two first PCA axes used to define LES and WES respectively

4. The spatial occupancy and functional similarity of neighboring trees determine tree growth in a tropical rainforest

Of the 4 models each representing a different strata, all but the understory stratum yielded significant results. For all 3 strata the NCI had a negative effect on tree growth (Figure 4.3). DBH had a negative effect on tree growth only for subcanopy stratum and WES a positive effect for canopy stratum (Figure 4.3). Alone LES did not have a significant effect, however in interaction with the DBH it had a negative effect on the growth of trees in the emergent stratum while in the canopy stratum it had a negative effect in interaction with the NCI (Figure 4.4). The interaction between SW (South-Westerness) and LES had a negative effect on tree growth for all 3 strata. Finally, in the sub-canopy stratum, the DBH had a negative effect on tree growth.

Strata	Emergent	Canopy	Sub-canopy
positive effects	LES:Curvature	WES WES:SW	
negative effects	NCI LES:SW DBH:LES	NCI LES:SW WES:Curvature Alt:WES LES:Slope NCI:LES	NCI DBH LES:SW

Table 4.1: Significant effects for the model with the spatial NCI with edge effects correction.

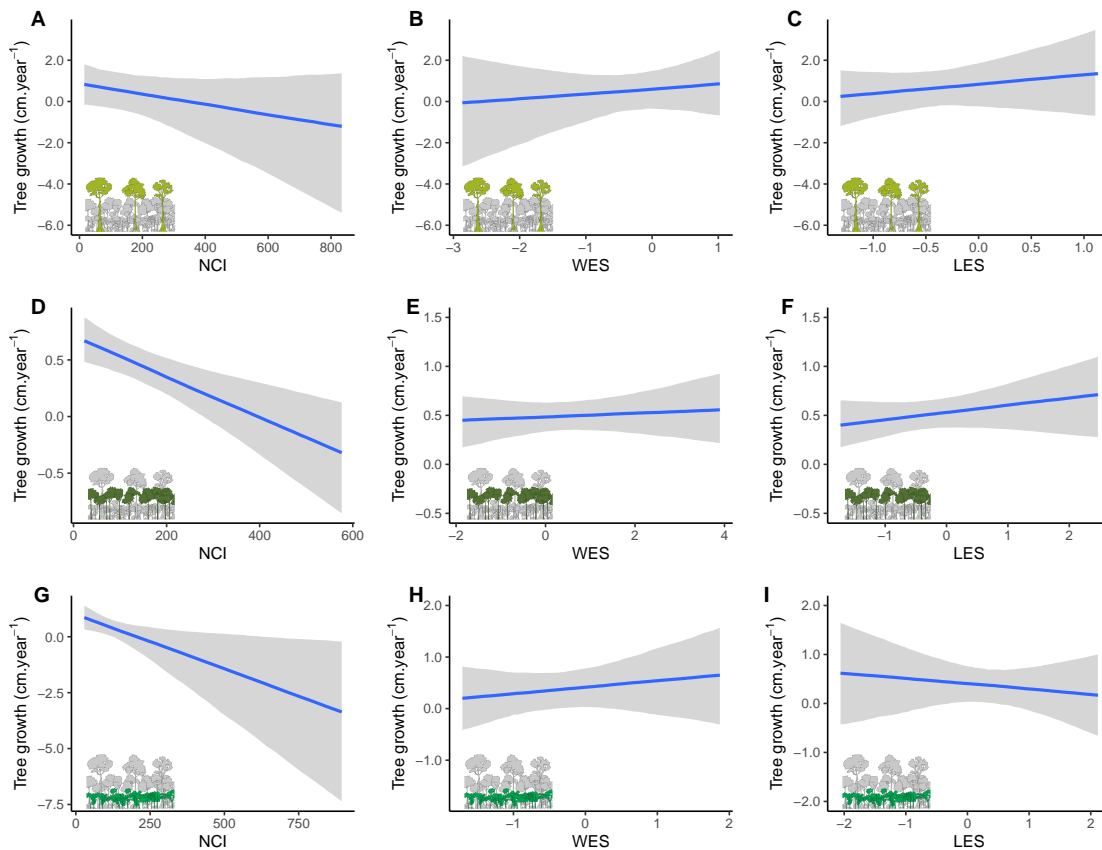


Figure 4.3: Relationships between tree growth, estimated from the best model and the neighborhood crowding index (NCI) (left), the wood economic spectrum (WES) (center) and leaf economic spectrum (LES) (right) for the emergent (A,B,C), canopy (D,E,F) or sub-canopy strata (G,H,I)

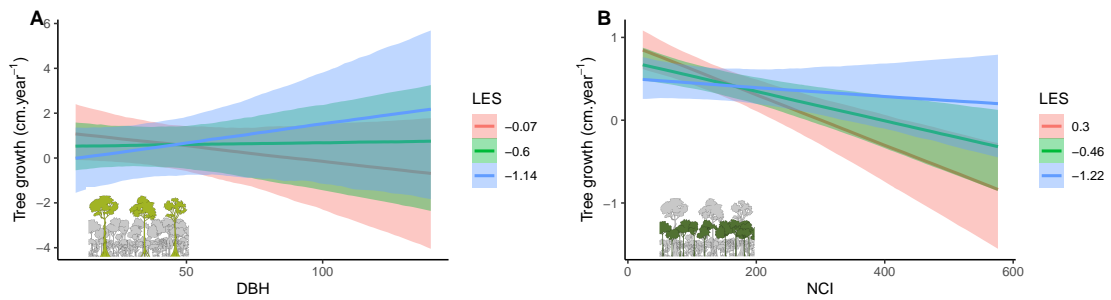


Figure 4.4: Relationships between tree growth, estimated from the best model and the diameter at breast height (DBH) (A) and the Neighborhood Crowding Index (NCI) in the emergent stratum (B) in interaction with the leaf economic spectrum (LES) in the canopy stratum. Larger LES values represent trees with more exploitative leaf strategies while smaller values reflect more conservative strategies

4 Discussion

Using BRMS models with neighboring indexes we showed that for most tropical rainforest strata competition ensued by neighboring trees negatively impact tree growth. In the emergent stratum, the growth of trees exhibiting exploitative strategies (of leaf traits) depended on their size with larger trees exhibiting lower growth than smaller ones with similar strategies. Our results thus show that acquisitive leaves consist in a more effective strategy for smaller trees rather than for large ones in the emergent stratum. Moreover, in the canopy, trees with more neighbors incidentally had greater tree growth when their leaves allowed greater photosynthetic yields. Therefore trees which highly invest in acquisitive leaves suffer more from the cost of this investment under the competing effect of neighboring trees. These same trees showed maximum growth when they had fewer neighbors and thus experienced less competition. On the other hand, the lack of investment in acquisitive leaves did not interact with the competition exerted by neighboring trees. Ultimately, there was an effect of the resource allocation of trees on their growth however only in interaction with the competition exerted by neighbors.

Canopy tree investment in wood tissues for sap conduction had a positive effect on their growth. Trees that invested in less durable wood tissues with high sap conduction exhibited greater tree growth. Yet, in interaction with topographical variables such as curvature this effect was lessened. Thus, on terrain which retains less nutrients, trees which have invested in sap conduction are less at an advantage because of the lack of nutrients inherent to the lixiviation of this topography (Quinto-Mosquera and Moreno-Hurtado 2016).

Our results showed that interactions in our models were significant but not very strong in comparison with the parameters that they depended on. For instance, for emergent trees, our results show a significant effect of the LES and Curvature interaction, but it is not strong in regard to the global effect of LES on the girth increase. We would have expected the Curvature to impact more importantly growth in combination with LES due to nutrients requirements of trees with productive strategy (Asner et al. 2016).

Because we grouped trees in their respective strata, especially their potential strata may have

4. The spatial occupancy and functional similarity of neighboring trees determine tree growth in a tropical rainforest

affected our results especially regarding competition. For instance we considered the resource allocation strategies of trees which had not yet reached their stratum. It is safe to say that the effect of competition should differ as trees cross the strata separating them from their realized stratum. However measuring the height of each individual in a dense tropical forest context remains a challenge for tropical ecologists and as more data become available, this type of study will benefit from using actual tree height. Nevertheless we were able to draw conclusions from potential strata which in itself is promising.

Author contributions. All authors contributed to design the study. GR performed the analysis and wrote the first draft of the manuscript. All authors contributed revisions to the manuscript and gave final approval for publication.

References

- Asner, G. P. et al. (2016). “Large-scale climatic and geophysical controls on the leaf economics spectrum”. In: *Proceedings of the National Academy of Sciences* vol. 113, no. 28. ISBN: 0027-8424 Publisher: National Acad Sciences, E4043–E4051.
- Baraloto, C. et al. (2010). “Decoupled leaf and stem economics in rain forest trees”. In: *Ecology letters* vol. 13, no. 11. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 1338–1347.
- Chave, J. et al. (2009). “Towards a worldwide wood economics spectrum”. In: *Ecology letters* vol. 12, no. 4. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 351–366.
- Chen, X. et al. (2020). “The leaf economics spectrum constrains phenotypic plasticity across a light gradient”. In: *Frontiers in plant science* vol. 11. ISBN: 1664-462X Publisher: Frontiers Media SA, p. 735.
- Fortunel, C. et al. (2016). “Functional trait differences influence neighbourhood interactions in a hyperdiverse Amazonian forest.” In: ISBN: 1461-023X.
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton and Oxford: Princeton University Press.
- Kneitel, J. M. (2019). “Gause’s competitive exclusion principle”. In: *Encyclopedia of ecology*. Sacramento State, pp. 110–113.
- Kunstler, G., Falster, D., et al. (2016). “Plant functional traits have globally consistent effects on competition”. In: *Nature* vol. 529, no. 7585. ISBN: 1476-4687 Publisher: Nature Publishing Group, pp. 204–207.
- Kunstler, G., Lavergne, S., et al. (2012). “Competitive interactions between forest trees are driven by species’ trait hierarchy, not phylogenetic or functional similarity: implications for forest community assembly”. In: *Ecology letters* vol. 15, no. 8. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 831–840.
- Laughlin, D. C. et al. (2020). “The net effect of functional traits on fitness”. In: *Trends in Ecology & Evolution* vol. 35, no. 11. ISBN: 0169-5347 Publisher: Elsevier, pp. 1037–1047.
- MacArthur, R. and Wilson, E. (1967). *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press.

-
- Pélissier, R. and Pascal, J.-p. (2000). “Two-year tree growth patterns investigated from monthly girth records using dendrometer bands in a wet evergreen forest in India”. In: *Journal of Tropical Ecology* vol. 16, no. 3. ISBN: 1469-7831 Publisher: Cambridge University Press, pp. 429–446.
- Pélissier, R., Pascal, J.-P., et al. (2011). “Tree demography in an undisturbed Dipterocarp permanent sample plot at Uppangala, Western Ghats of India.” In: *Ecology* vol. 92, no. 6. ISBN: 0012-9658 Publisher: Ecological Society of America.
- Quinto-Mosquera, H. and Moreno-Hurtado, F. (2016). “Precipitation effects on soil characteristics in tropical rain forests of the Chocó biogeographical region”. In: *Revista Facultad Nacional de Agronomía Medellín* vol. 69, no. 1. ISBN: 0304-2847 Publisher: Universidad Nacional de Colombia, pp. 7813–7823.
- R_Development_Core_Team (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Schmitt, S. et al. (2021). “Canopy and understorey tree guilds respond differently to the environment in an Indian rain forest”. In: *Journal of Vegetation Science* vol. 32, no. 5. ISBN: 1100-9233 Publisher: Wiley Online Library, e13075.
- Tilman, D. (1994). “Competition and biodiversity in spatially structured habitats”. In: *Ecology* vol. 75, no. 1. ISBN: 1939-9170 Publisher: Wiley Online Library, pp. 2–16.
- Whitmore, T. C. (1990). *An introduction to tropical rain forests*. Clarendon Press.

Chapitre III

Influence des fluctuations environnementales passées sur la diversité régionale actuelle d'espèces

La diversité des forêts tropicales se distingue par leur nombre extraordinaire d'espèces et en particulier d'espèces rares, de sorte qu'il est communément admis que dans ces forêts la rareté est commune. Mais cette incroyable biodiversité remet surtout en question l'idée assez fondamentale en écologie des communautés selon laquelle les espèces ne peuvent coexister dans des communautés que si elles utilisent des niches différentes. En effet, les différences de niches peuvent être insuffisantes dans ces communautés pour en expliquer la diversité spécifique. Ce sont les prédictions de la théorie neutre qui ont notamment permis d'expliquer cette abondance d'espèces rares comme le fruit de la dispersion limitée dans des populations de petite taille soumises à la dérive. Cependant, il se peut que des fluctuations de la superficie de la forêt tropicale au fil du temps aient affecté la dynamique des populations et les variations d'abondances des espèces d'arbres. Or le modèle neutre en écologie ne prend pas en compte l'influence d'un environnement variable. Des études paléoclimatiques et paléoécologiques ont montré que les changements climatiques passés ont généré une alternance de périodes plus sèches et plus humides qui ont fortement affecté la superficie des forêts tropicales humides dans le monde. Des études antérieures ont montré que ces changements d'habitat passés peuvent expliquer la biodiversité de la faune des forêts pluviales actuelles. Dans le cas des forêts tropicales, on pense aux oscillations climatiques du Quaternaire, et plus récemment les activités humaines, qui peuvent avoir impacté la disponibilité d'habitats favorables à la persistance de certaines espèces et ainsi fortement contraindre leur répartition à des refuges lors des périodes de contraction. Du fait du temps de génération long des arbres tropicaux (environ 60 ans), il se peut que leur démographie soit sensible à des fluctuations environnementales qui, à notre échelle d'observation, nous paraissent très anciennes. Sur un plus long terme, la fragmentation du Gondwana a une importance toute particulière pour la biogéographie des forêts tropicales, car celles-ci se répartissent dans des régions biogéographiques autrefois réunies au sein de ce même supercontinent. Ainsi les massifs actuels de forêts tropicales ont une histoire commune dont les débuts précèdent l'éclatement du Gondwana il y a environ 160 millions d'années. La séparation du continent africain puis de la plaque Indo-Malgache du reste du Gondwana a pu favoriser l'isolation et la divergence des lignées d'arbres et donc entraîner de la vicariance, et également affecter les possibilités d'échanges par migration au cours du temps. Il y a environ 100 à 80 millions d'années, l'Inde et Madagascar formaient une seule et même île localisée à environ 30°S dont la séparation environ 20 millions d'années plus tard entama l'orogénèse de l'Himalaya lors de la rencontre de l'Inde avec la plaque Laurasienne. L'histoire biogéographique des forêts tropicales fut ainsi marquée par la fragmentation progressive du Gondwana qui n'a abouti à la configuration actuelle qu'à la fin de l'ère Tertiaire (il y a environ 10 millions d'années) en se poursuivant au cours du temps sur les différents fragments et au rythme des possibles migrations transocéaniques.

La question qui motive ce chapitre est la suivante: quelles signatures ce genre d'événements biogéographiques majeurs ont-ils laissé sur la biodiversité actuelle? Dans un premier temps, il fut nécessaire de caractériser comment la présence de refuges a pu influencer la dynamique des communautés ainsi que les conséquences sur la diversification des espèces, au moyen d'une expérience théorique. L'Article 5 utilise ainsi la simulation de dynamiques éco-évolutives neutres basées sur la coalescence (*ecophylo*) pour évaluer comment une contraction de taille des communautés peut influencer la diversité des espèces actuelles en fonction des processus de dérive démographique, de mutation et de spéciation. Cet article examine différents patrons de diversité taxonomique,

4. Influence des fluctuations environnementales passées sur la diversité régionale actuelle d'espèces

phylogénétique et fonctionnelle et met en évidence des profondeurs temporelles différentes de l'influence des changements environnementaux passés selon l'aspect de la diversité que l'on considère. En effet, les abondances relatives existantes reflètent l'influence durable des refuges plus récents sur la dynamique démographique, tandis que la composition phylogénétique reflète l'influence des changements d'habitat plus anciens. La diversité fonctionnelle quant à elle dépend de l'interaction entre la dynamique de diversification et l'évolution des traits, offrant de nouvelles possibilités pour déduire conjointement l'adaptation actuelle des traits et leur dynamique d'évolution passée. Une autre étude montre par ailleurs que la diversité fonctionnelle actuelle porte la signature conjointe des processus évolutifs générant l'asymétrie phylogénétique et du rythme d'évolution des traits. L'Article 6 met en évidence que dans le cas où la majorité de l'évolution des traits a lieu dans les branches profondes de l'arbre, la distribution des traits en termes de variance, d'asymétrie et d'aplatissement conserve une signature claire témoignant de la présence de lignées anciennes qui se sont très peu diversifiées et qui persistent dans les assemblages actuels. L'intérêt de cette démarche par expérience virtuelle consiste à évaluer la capacité de modèles à simuler des patrons de diversité réalistes sur lesquelles baser l'inférence de différents scénarios de changements passés en comparant les patrons simulés aux données observées.

Enfin, l'Article 7 propose une nouvelle méthode pour analyser les lignées uniques ou partagées entre plusieurs régions, lors de périodes évolutives définies sur la base d'événements marquants de l'histoire géologique de Madagascar et de l'Inde. La proportion de lignées partagées entre Madagascar et l'Inde diminue progressivement suite aux séparations successives et à l'isolation de Madagascar de la plaque indo-malgache. Les lignées propres à l'Inde sont plus nombreuses dans les périodes les plus récentes, alors que les lignées propres à Madagascar étaient significativement plus présentes dans les périodes plus anciennes. Ainsi l'histoire biogéographique de ces deux régions se reflète dans la structure de la phylogénie des espèces actuelles.

Article 5

Imprints of past habitat area reduction on extant taxonomic, functional and phylogenetic composition

Elizabeth Barthélemy¹, Maxime Jaunatre², Claire Fortunel², François Munoz¹

published in *Frontiers in Ecology and Evolution*.

Abstract

Past environmental changes have shaped the evolutionary and ecological diversity of extant organisms. Specifically, climatic fluctuations have made environmental conditions alternatively common or rare over time. Accordingly, most taxa have undergone restriction of their distribution to local refugia during habitat contraction, from which they could expand when suitable habitat became more common. Assessing how past restrictions in refugia have shaped species distributions and genetic diversity has motivated much research in evolutionary biology and biogeography. But there is still lack of clear synthesis on whether and how the taxonomic, functional and phylogenetic composition of extant multispecies assemblages retains the imprint of past restriction in refugia. We devised an original eco-evolutionary model to investigate the temporal dynamics of a regional species pool inhabiting a given habitat today, and which have experienced habitat reduction in the past. The model includes three components: (i) a demographic component driving stochastic changes in population sizes and extinctions due to habitat availability, (ii) a mutation and speciation component representing how divergent genotypes emerge and define new species over time, and (iii) a trait evolution component representing how trait values have changed across descendants over time. We used this model to simulate dynamics of multispecies assemblages that occupied a restricted refugia in the past and could expand their distribution subsequently. We characterized the past restriction in refugia in terms of two parameters representing the ending time of past refugia, and the extent of habitat restriction in the refugia. We characterized extant patterns of taxonomic, functional and phylogenetic diversity depending on these parameters. We found that extant relative abundances reflect the lasting influence of more recent refugia on demographic dynamics, while phylogenetic composition reflects the influence of more ancient habitat change. Extant functional diversity depends on the interplay between diversification dynamics and trait evolution, offering new options to jointly infer current trait adaptation and past trait evolution dynamics.

¹LIPhy, Université Grenoble Alpes, Grenoble, France

²AMAP (botAnique et Modélisation de l'Architecture des Plantes et des végétations), CIRAD (Centre de coopération internationale en recherche agronomique pour le développement), CNRS (Centre national de la recherche scientifique), INRAE (Institut national de recherche pour l'agriculture, l'alimentation et l'environnement), IRD (Institut de Recherche pour le Développement), Université de Montpellier, Montpellier, France

1 Introduction

Alternating periods of contraction and expansion of suitable environmental conditions should affect demographic and speciation-extinction dynamics in a given habitat over time. Hence, a fundamental goal of biogeography is to better understand how changes in suitable environmental conditions due to past climatic and geomorphologic history have shaped current biodiversity patterns (Bennett 1990). Much emphasis has been put on the impact of dramatic habitat restriction, whereby species could only persist in reduced areas called refugia (e.g. Weiss and Ferrand 2007). Most studies so far have focused on how substantial habitat restriction in glacial refugia have shaped the genetic diversity within organisms at present (Avice 2000). Conversely, we still poorly understand how past restrictions in refugia have influenced extant patterns for other dimensions of diversity in multi-species assemblages, a.k.a., taxonomic, phylogenetic and functional diversity.

Here, considering the diversity of organisms making a biota in specific environmental conditions (habitat) of a given region, we examine how habitat contraction in refugia and subsequent expansion can affect these three different facets of biodiversity. Firstly, taxonomic diversity, based on the number and relative abundances of species in a local assemblage, depends on species demography (Hubbell 2001). Changes in the area of available habitat should affect stochastic demographic dynamics, yielding greater abundance variation among species and extinction risk in smaller refugia (Figure 5.1). Specifically, we expect that the influence of ecological drift be greater in smaller assemblages and thus shape the relative abundances of species (Vellend 2010 ; Blanchard, Birnbaum, and Munoz 2020). Thus, we expect demographic fluctuations to happen quickly over the timescale of habitat changes, so that extant taxonomic diversity should be more sensitive to recent than to ancient refugia. Secondly, phylogenetic diversity represents how closely-related species are in a habitat according to their evolutionary history. We hypothesize that a larger biota in a larger habitat can undergo more or less frequent speciation or extinction events, respectively, and thus greater net diversification (Figure 5.1). Because diversification occurs at a slower pace than demographic dynamics, we expect phylogenetic diversity at present to keep a signal of ancient refugia. A greater number of old lineages could persist larger ancient refugia and have descendants in current assemblages (Bose et al. 2019). Conversely if the refuge was particularly small relative to the current assemblage size and expansion has only occurred recently, the loss of many lineages in refugia will have caused the most ancient common ancestor to be more recent (Figure 5.1). Indeed, the probability of sharing a common ancestor should increase in small assemblages (Patterson 2005). Furthermore, the variation of trait attributes among extant organisms depends on how trait values have evolved across taxa in the past (Pearse et al. 2019). Specifically, assuming that species sharing greater common history have more similar phenotypes, we can account for the tempo of trait evolution to predict the functional composition of assemblages having undergone a common demographic history that shaped long-term diversification events (Revell, Harmon, and Collar 2008). The speed of trait evolution could indeed have changed and depended on when and how adaptive radiations happened and shaped trait diversification over time in a given habitat (Bartish et al. 2016 ; Harmon and Open Textbook Library 2019). Therefore, extant patterns of trait diversity should differ depending on the pace and the timing of trait evolution. The extent of trait variation in descendants should depend on both trait evolution rate and the size of the

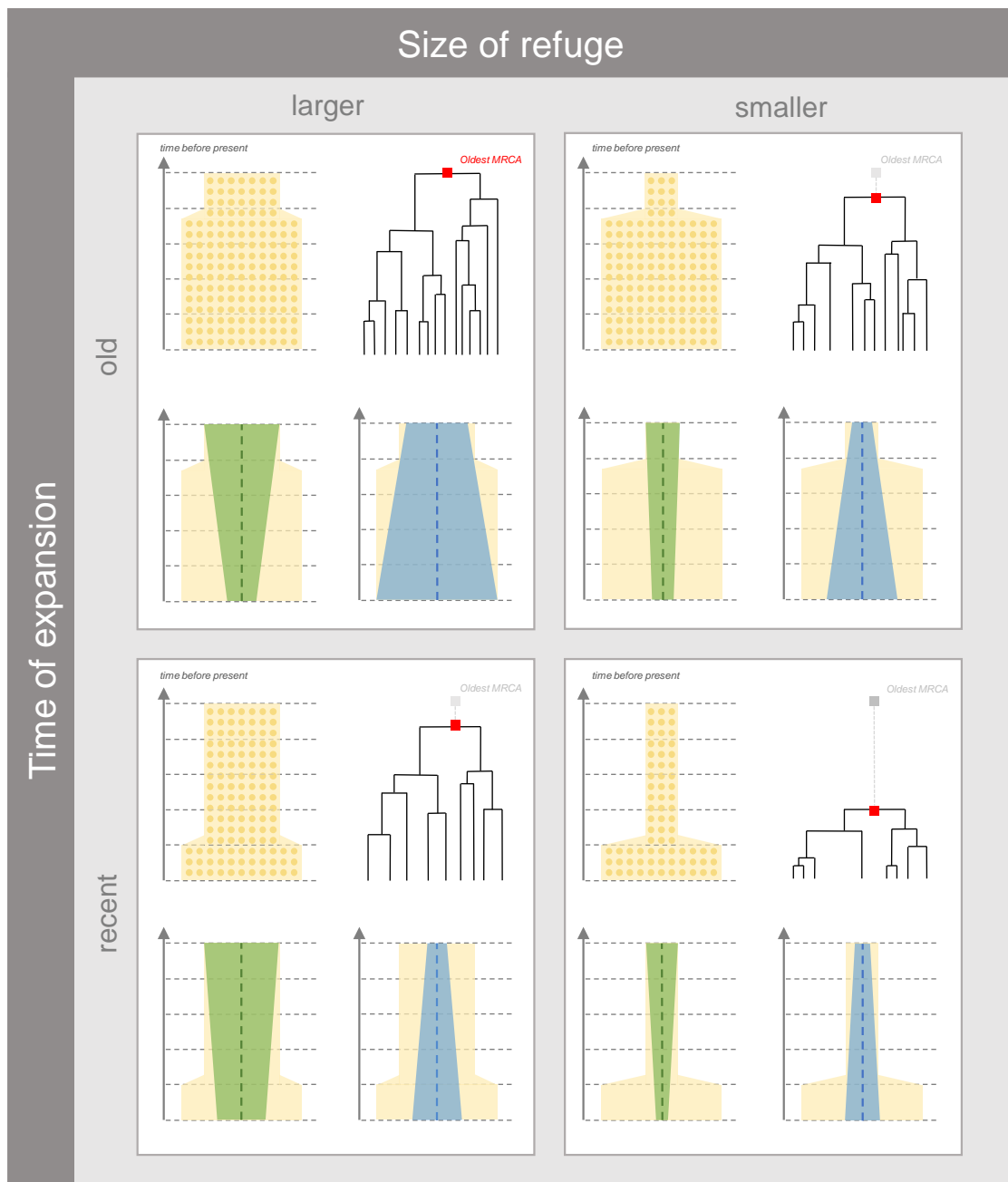


Figure 5.1: Expected patterns of taxonomic, phylogenetic, and trait diversity under different scenarios of habitat expansion following a period with restricted refugia. Yellow shaded areas represent the size of suitable habitat conditioning the number of individuals over time (represented as circles). The size of the assemblage instantaneously increases at a given time (parameter t_c) so as to represent habitat expansion (parameter A). Current patterns of diversity will depend on how restrictive past habitat sizes were relative to present (left vs. right panels) and how long assemblages have spent in refugia (top vs. bottom panels). In larger refugia lower extinction rates allow more species in older lineages to persist in extant assemblages, while in smaller refugia more species will be lost to the increased influence of drift. Functional diversity will depend on the pace and timing of trait evolution in relation to diversification events. Trait evolution can deviate from the ancestral trait values (vertical dashed lines), and can experience exponential deceleration (Early Burst—EB represented in green) or acceleration (Late Burst—LB represented in blue), the resulting patterns of functional diversity will depend of whether acceleration or deceleration of trait evolution happens in or out of refugia. For example, if greater trait evolution occurs during strong habitat restriction, trait diversity among descendant lineages would be reduced.

biota over time (Figure 5.1). Specifically, if greater trait evolution occurs during strong habitat restriction, trait diversity among descendant lineages would be reduced.

The taxonomic, phylogenetic and functional facets of biodiversity are thus influenced by the interplay of historical variation in habitat area and past evolutionary dynamics. These three facets thereby convey complementary insights into population dynamics, evolutionary diversification, and trait evolution. To understand how past contraction in refugia and subsequent expansion of a habitat can affect taxonomic, phylogenetic and functional diversity, we propose an eco-evolutionary model explicitly representing the basic processes at play. First, we devise a genealogical model of population dynamics depending on the temporal variation in habitat area (Figure 5.2a). The model reconstructs past co-ancestry in the genealogy of extant individuals, using coalescent theory (Wakeley 2009). We represent habitat restriction in past refugia with two parameters: one representing the ending time of refugia with subsequent and instantaneous expansion until present, and the other the extent of habitat restriction in the refugia (ratio of refugial area compared to present). Second, we model evolutionary diversification and phylogenetic relationships based on the genealogical structure. We group extant individuals into distinct species under a model of point mutations and protracted speciation of monophyletic genetic groups (Manceau, Lambert, and Morlon 2015) (Figure 5.2b&c). Third, we model alternative scenarios of trait evolutions in the phylogeny, yielding extant trait composition. To represent different historical trajectories in trait evolution, we define the rate of trait variation in descendants along the phylogeny as either constant (Brownian motion), greater in older time (early burst), or greater in recent time (late burst). Different paces of trait evolution can be considered to reflect evolutive radiations concurring or not with lineages diversification brought about by demographic expansions (Figure 5.2c).

Here, we design a virtual experiment to examine how increases in favorable habitats following refugia influence taxonomic, phylogenetic and functional diversity. To this end, we explicitly simulate the demography of assemblages, the emergence of new species, their phylogenetic relationships and associated trait values.

2 Material and methods

2.1 Simulating phylogenies and trait values in extant assemblages

In our model, fluctuations in the relative species abundances in a given habitat are driven by neutral drift dynamics, depending on the global size of the assemblage (Hubbell 2001). Regardless of which species these individuals belong to, the dynamics of the assemblage they form can be represented by coalescence, i.e. by tracing the shared co-ancestry of extant individuals backwards in time until a single common ancestor is found (Kingman 1982). Only the ancestry of individuals observed at present is traced back, so that coalescent methods do not require simulating lineages with no extant descendants and are thereby much faster than their forward-in-time alternatives (Munoz et al. 2018). Thus, without assuming to which species each individual belongs, we simulated the genealogies of individuals in an assemblage with constant size, and modelled the effect of habitat

expansion on coalescence events following a period with restricted refugia (Figure 5.2a) (Hudson 2002 ; Kelleher and Lohse 2020). We modulated both the amplitude of habitat expansion (A) and the time elapsed since expansion began (t_c), so as to simulate genealogies covering a wide range of refugia age and habitat expansion scenarios (Table 5.1).

We sprinkled speciation events over the simulated genealogies (Figure 5.2b). The probability that a mutation occurred on a specific branch, leading to a new variant individual, followed a Poisson distribution with parameter $\mu \times B$, where μ is the point mutation rate (Table 5.1) and B the length of the branch. Descendant lineages of mutated individuals formed new clades that, if paraphyletic, were merged to form species (Manceau, Lambert, and Morlon 2015), thus monophyletic lineages with distinct genotypes and older than two generations were considered a new species (Figure 5.2c). We derived the number of individuals descending from a speciation event which gave us both the relative abundances of species and their phylogenetic relations to one another. Note that the point mutation rate is fixed for all simulations and thus the time scale (in generations before present) should be considered relative to μ . Additionally, any effect of expansions following periods with restricted refugia will depend on how parameter μ and the date of expansion are related. Indeed, if too few diversification events should happen during later expansion phases, as they may if μ is particularly small, we expect there to be little if not any signature in extant functional and phylogenetic diversity.

We simulated alternative models of trait evolution on the simulated phylogenies (Figure 5.2d), using the mvMORPH package in R (Clavel, Escarguel, and Merceron 2015). We considered three models of trait evolution from an ancestral value set to 0 (Table 5.1). In the first model, traits evolved from their ancestral state following a pure Brownian motion (BM), with a variance parameter σ^* set either to 0.1, 0.2 or 0.4 to represent varying speed of trait evolution. In the second model, trait evolution exponentially decelerated with time ($\beta = -0.01$) following an Early Burst (EB). The third model represented Late Burst (LB) trait evolution, with exponential acceleration ($\beta = 0.001$).

We simulated 1000 phylogenies for every past change date (5000 in total), in which the model parameter values varied uniformly within above-mentioned limits (Table 1). We summarized the amplitude of the demographic expansion by parameter A , which reflects how much smaller the refuge was in the past before the favorable habitat expanded to its current and fixed size. Then, trait evolution was simulated with 100 replicates along each phylogeny according to the 9 parameter combinations given in Table 5.1.

2.2 Characterizing extant taxonomic, functional and phylogenetic diversity

We investigated the influence of the parameters of past expansion (A, t_c) on three different facets of extant biodiversity. We characterized taxonomic diversity by means of (i) species richness, and (ii) species frequencies spectra (SFS). The SFS represented the number of species by octaves of abundance class, and we performed a principal components analysis (PCA) (Lê, Josse, and Husson 2008) to summarize the variation of SFS across simulations. Similarly to using allelic frequency spectra in population genetics to infer population history (Marth et al. 2004), we expected the SFS

5. Imprints of past habitat area reduction on extant taxonomic, functional and phylogenetic composition

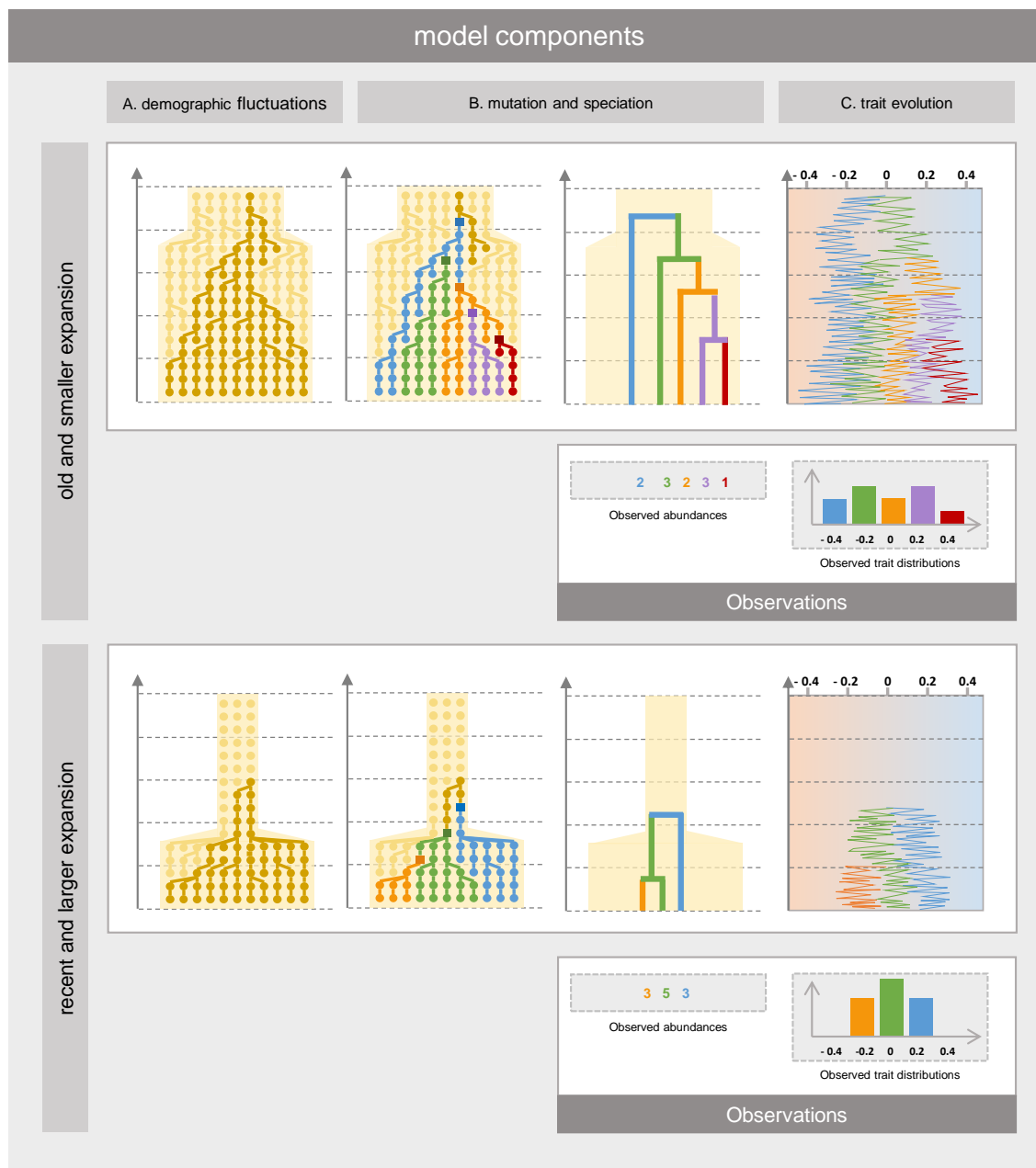


Figure 5.2: Eco-evolutionary dynamics of individual organisms in a habitat that experienced reduction in the past. The model relies on three components. (A) The population component reconstructs the effect of habitat expansion on the past co-ancestry in the genealogy of extant individuals, using coalescent theory without assuming to which species they belong to. (B) The evolutionary component represents how divergent genotypes have made new species over time. It is modeled by first sprinkling mutation events over the simulated genealogy conditionally its branch lengths. Then, the phylogenetic relationships amongst individuals and their abundances are obtained by merging paraphyletic clades into single species depending on the topology of the genealogy. (C) The trait evolution component is modeled using Brownian motion (BM), Early Burst (EB) or Late Burst (LB) along the simulated phylogeny and represents how species niches may have changed over time and at which pace. We represent two extreme scenarios: small demographic expansion (relative to refugia) that occurred long ago (top) and important habitat expansion that occurred recently (bottom) and how these have led to different patterns of taxonomic, phylogenetic, and functional diversity.

to retain the signature of past demographic fluctuations (here, periods of habitat restriction) that caused more important drift dynamics affecting species relative abundances in smaller assemblages.

Model components	Parameter	Prior	Description
Demographic fluctuations	s	500	(Assemblage) Sample size: number of individuals observed at present in a region of size Ne
	Ne	10,000	Regional assemblage or pool size at present
	P	(1,000 5,000)	Regional assemblage or pool size in past refugia
	tc	100, 500, 1500, 2000, 3000	Beginning date of expansions (in generations before present), after which the regional assemblage size changed from P to Ne
	A = Ne/P	(2,10)	Amplitude of expansion
Mutation and speciation	mu	0.001	Non-synonymous point mutation rate. The distribution of distinct genotypes allow defining monophyletic clades as distinct species
Trait evolution	sigma	0.1, 0.2, 0.4	Standard-deviation of Brownian motion
	beta	-0.01, 0, 01	Exponent of the exponential component in trait evolution: if negative, Early-Burst (EB), if positive Late-Burst (LB), if 0 pure Brownian motion (BM)

Table 5.1: Model parameters and their prior distributions used to simulate species assemblages, the phylogenetic relationships, and trait values. Whenever the prior is a single value, the parameter value is fixed for all simulations. Bracketed values correspond to the upper and lower bounds of a uniform distribution from which the according parameter value is drawn for the simulations.

We used complementary metrics of phylogenetic diversity to characterize the evolutionary history of extant species of the assemblage. We quantified (i) Phylogenetic Diversity (PD), (ii) Mean pairwise distance (MPD) and (iii) Mean Nearest Taxon Distance (MNTD) using the picante package (Kembel et al. 2010). MPD is expected to be more sensitive to ancient and MNTD to recent diversification events (Mazel et al. 2016). Additionally, we used the approach described in (Lewitus and Morlon 2016) implemented in the RPANDA package Morlon et al. 2015) to derive the ln-transformed principal eigenvalue (λ^*) and skewness (ψ) of the Laplacian spectral density profiles of phylogenies. These summary statistics quantified phylogenetic expansion and stem-to-tip distribution of branching events respectively. Taxonomic and phylogenetic metrics of diversity were related to parameters of demographic expansion (parameters A and t_c) using linear regression models. To evaluate the relative effect of species richness from those of the parameters used to simulated demographic expansions, we used linear models between (i) the residuals of a linear model relating phylogenetic metrics to species richness and (ii) the parameters of demographic expansion. Linear models were fitted using the generalized least squares (GLS) method from the nlme package in R. In the case of PD and the MRCA, allowing the variance to increase with species richness yielded better AIC values when compared to simple linear regressions.

Finally, functional diversity was characterized by the first two moments of the simulated trait distributions (mean and variance) (Fortunel et al. 2014) given each phylogeny and given each trait evolution model parameter combination σ and β .

All simulations were conducted using Python version 3.7.4 (Python Software Foundation) and subsequent analyses performed using the R (3.6.3) statistical platform (R_Development_Core_Team 2010).

3 Results

3.1 Taxonomic diversity

Simulated species richness decreased when the refugia was more restricted compared to current assemblage size, and it did so even faster for more recent changes (Figure 5.3). We obtained maximum species richness for oldest demographic expansion (i.e., 3000 generations ago). When the refugia was more ancient, the restrictedness of the refugia did no longer influence extant species richness.

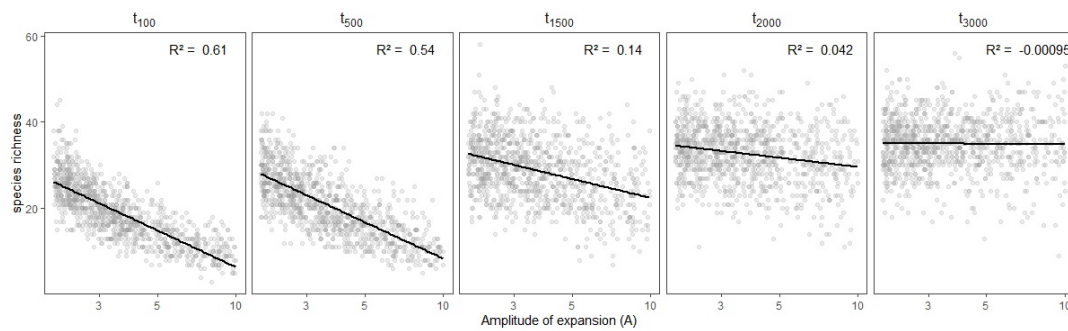


Figure 5.3: Variation in simulated species richness depending on the amplitude of expansion (A) and the time (in generations) at which it occurred (t_c). Linear regressions are shown in black with the associated R^2 statistic in the upper-right side of each panel.

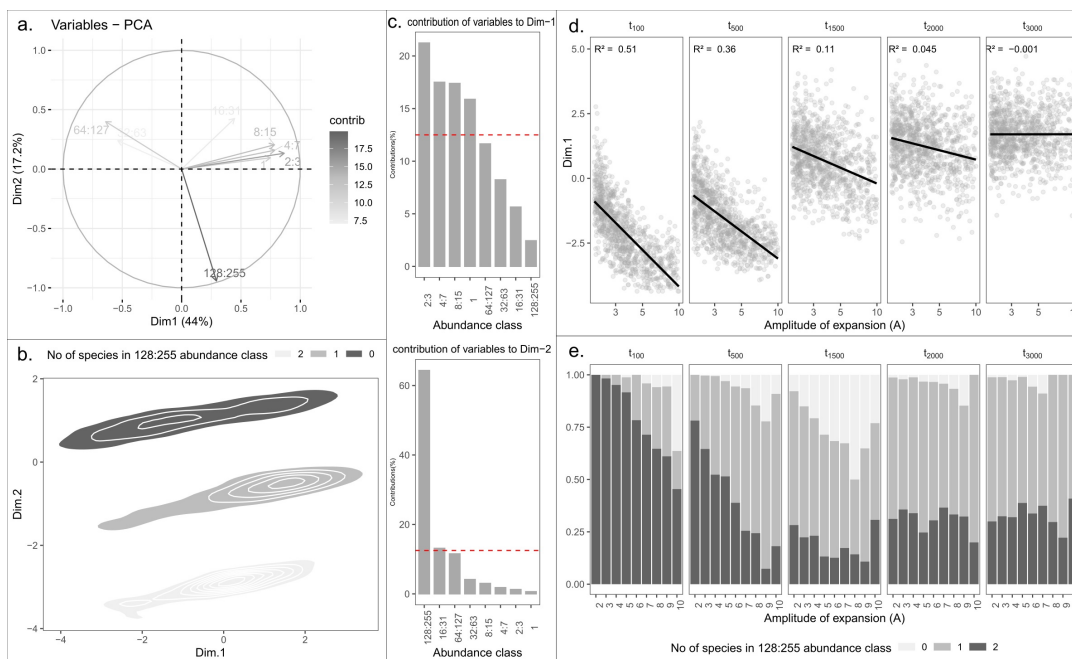


Figure 5.4: Principal component analysis (PCA) on the simulated species frequency spectra (SFS). (A) Correlation circle showing the contributions of SFS abundance classes to the two first PCA axes. (B) Individual PCA scores projected onto the two first dimensions of the PCA and clustered according to the number of species in the most abundant abundance class. (C) Contributions of SFS abundances classes to the first (top) and second (bottom) axes of the PCA. (D) Variation in PCA-scores along the first dimension depending on the amplitude of expansion (parameter A) and the time (in generations) at which expansion occurred (parameter t_c). Linear regressions are shown in black with the associated R^2 statistic in the upper-left side of each panel. (E) Variation in the number of points with, respectively, none, one, or two hyperabundant species depending on the amplitude of expansion (parameter A) and the time (in generations) at which expansion occurred (parameter t_c).

The first PCA axis captured 44% of SFS variation and opposed the classes of lowest abundance to classes of intermediate abundance. PCA scores along the first axis were strongly related to species richness ($R^2 = 0.88$), and likewise decreased with the amplitude of expansion (Figure 5.4d). Yet, depending on how ancient the end of refugia was, the amplitude of subsequent expansion did not affect the abundance classes equivalently. Positive PCA scores along the first axis indicated assemblages with many rare species (Figure 5.4a). Such assemblages were mostly simulated when expansions occurred more than 1500 generations ago (Figure 5.4d). Negative PCA scores along the first axis indicated assemblages with greater frequency of common species which, conversely, happened mostly when expansions occurred more recently (less than 500 generations ago). The second axis captured 17.2% of SFS variance and was mostly carried by the largest abundance class, representing assemblages with 0, 1 or 2 hyperabundant species (Figure 4b,c). For changes predating 2000 generations, the number of abundant species increased with the amplitude of expansion and was maximal for 1500 generations (Figure 5.4e).

3.2 Phylogenetic diversity

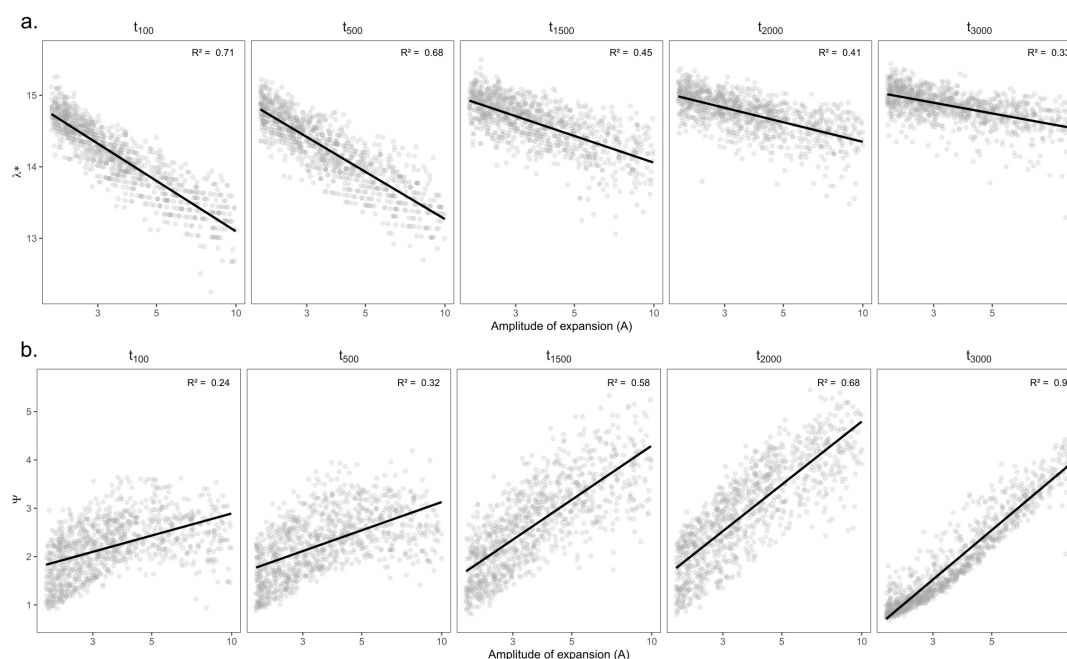


Figure 5.5: Variation in spectral density profile summary statistics: (A) In-transformed principal eigenvalue (λ^*) and (B) skewness (ψ), of rooted phylogenies (vis a vis oldest simulated MRCA) depending on the amplitude of expansion and the time (in generations) at which it occurred. Linear regressions are shown in black with the associated R^2 statistic in the upper-left side of each panel.

We found that phylogenetic diversity (λ^* , Faith's PD, MPD and the age of the MRCA) decreased overall with the amplitude of expansion - i.e when the refuge relative to present was smaller, and did so more rapidly for recent change (see Fig. Figure 5.S1 and Figure 5.5). The In-transformed principal eigenvalue λ^* correlated positively to PD, MRCA, MPD and species richness (supplementary materials S2), and thus larger λ^* values were indicative of lengthy older rooted phylogenies with many lineages and greater overall phylogenetic diversity (Lewitus and Morlon 2016). We found

5. Imprints of past habitat area reduction on extant taxonomic, functional and phylogenetic composition

strong linear relationships between (i) the residuals of the regressions between phylogenetic metrics and species richness and (ii) the parameters of demographic expansions (all R^2 greater than 0.68; Figure 5.S2). This suggests that despite the influence of species richness on phylogenetic metrics of diversity, the latter still carried the signature of past demographic expansions. On the other hand, MNTD did not significantly vary along parameters of demographic change (Figure 5.S1), even when accounting for the influence of species richness. The density profile skewness ψ was always positive, which indicated stem-to-tip asymmetry favoring terminal branching events. ψ values increased with the amplitude of expansion (i.e when the size of refugia decreased relative to present) and reached higher values when expansions occurred long ago (Figure 5.5b).

3.3 Functional composition

Trait mean - as expected - averaged around the ancestral trait value and trait variance increased with σ^* (Figure 5.6 and Figure 5.7). Yet, depending on the trait evolution model (here, EB, BM or LB, based on the sign of β^*) simulated trait values deviated more or less importantly from the ancestral value from one simulation to the next (larger IQR) and their variance decreased depending on the amplitude of expansion (Figure 5.6). More interestingly, the trait IQRs and variance increased depending on λ^* (Figure 5.7a), for the BM or LB trait evolution models. In the case of an EB trait evolution model, the IQR and trait variance did not depend on phylogenies' spectral density metrics, but instead increased depending on σ^* . For the BM and LB trait evolution models, the IQR and trait variance increased both depending on σ^* and λ^* , more so in the case of a LB. Simulated trait IQRs and variance also increased, with similar trends, for phylogenies with stem-to-tip asymmetry favoring terminal branch lengths.

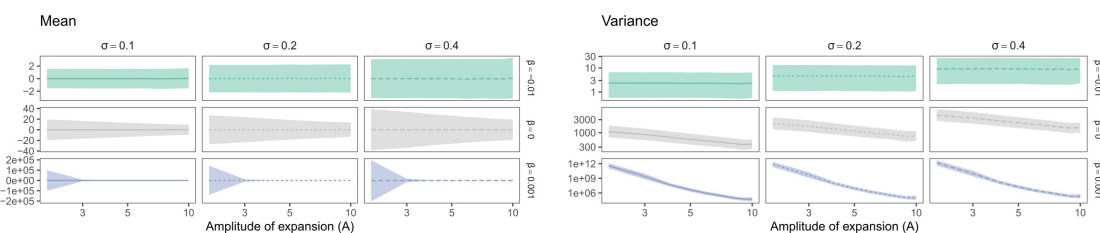


Figure 5.6: Variation in median trait moments (mean and variance) as well as their interquartile ranges (IQR) depending on the amplitude of expansion (A). Trait mean and variance did not vary depending on the time at which expansion occurred (not shown). Species richness is fixed for all simulations ($S = 31$). Brownian Motion is represented in gray, while Early and Late Burst are shown, respectively, in green or blue.

4 Discussion

By simulating the genealogies of individuals in an extant assemblage, for a wide range of scenarios of past restriction in refugia and subsequent expansion we found that the resulting patterns of taxonomic, phylogenetic and functional diversity in extant assemblages, reflect how past refugia have influenced population dynamics, evolutionary diversification, and trait evolution until present.

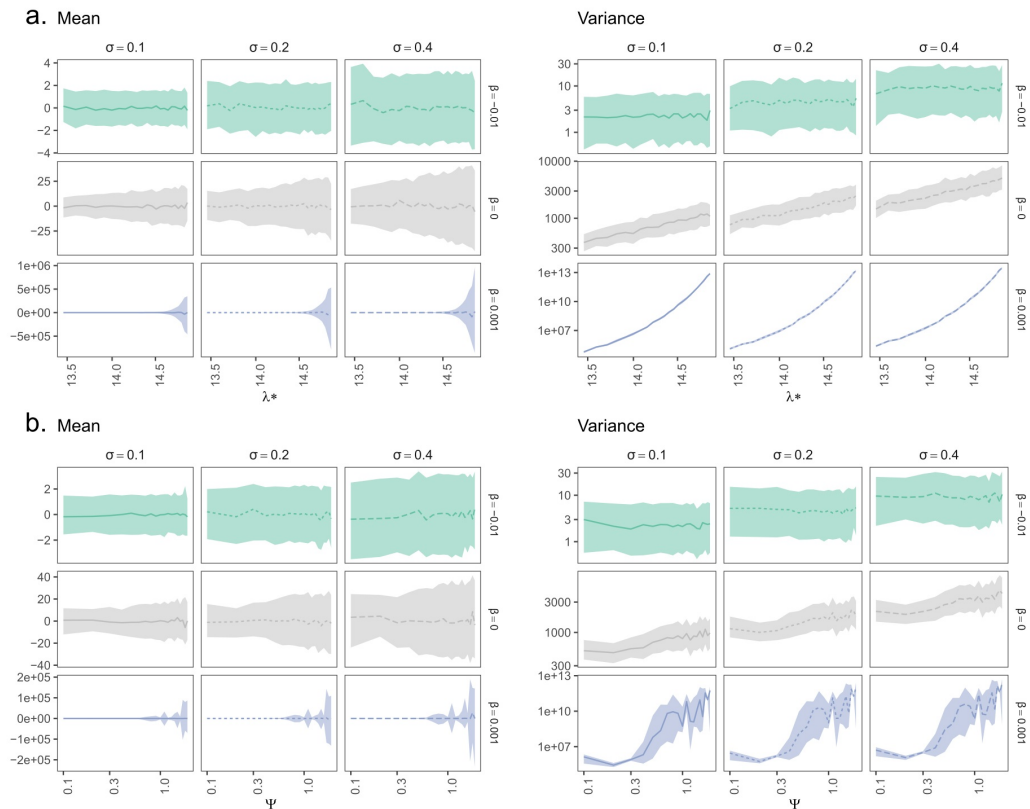


Figure 5.7: Variation in median trait moments (mean and variance) as well as their interquartile ranges (IQR) along the ln-transformed principal eigenvalue (λ^*) (A) and skewness (ψ) (B), summarizing the variation in simulated spectral density profiles according to the parameters used to simulate trait evolution (β in rows and σ in columns in each individual panel). The lines represent the median over 100 replications of trait evolution models and are dashed according to sigma values, while the shaded area represents the replicates' interquartile ranges (IQR). Species richness is fixed for all simulations ($S = 31$). Brownian Motion is represented in grey, while Early and Late Burst are shown respectively in green or blue.

4.1 Relative species abundances keep track of (more) recent habitat changes

When the size of favorable habitat in refugia was smaller relative to present size (i.e. ampler habitat expansion afterward), our results suggest that more species were lost over time due to greater extinction and ecological drift in reduced assemblages (Vellend 2010). This phenomenon amplified as the time spent in refugia increased (i.e. when the subsequent habitat expansion was recent), and mostly affected rare species. Reciprocally, species richness and the number of rare species were highest when expansion occurred long enough ago that specific diversity could subsequently reach a new equilibrium, i.e. when it no longer depended on the amplitude of expansion. Hence, our results suggest that extant taxonomic diversity, such as species richness, is influenced by past demographic events having affected whole assemblages. More specifically, for recent habitat changes, taxonomic diversity kept track of the influence of restriction in more recent refugia. Motivated by analogous uses in population genetics, the Species Frequency Spectra (SFS), a multispecies extension of allele frequency spectra used in population genetics (Fisher 1930), was especially useful to detect the influence of more recent demographic expansions. Because SFS provide more detailed and unabridged information regarding the variation in species abundance in an assemblage, these allowed us to detect the specific signature carried by specific classes of abundance, in particular that of rare species.

4.2 Phylogenetic diversity keeps track of (more) ancient habitat changes

Overall, phylogenetic diversity decreased as past assemblages experienced reduced-sized refugia and subsequent habitat expansion. Metrics of phylogenetic diversity (PD, MPD, λ^*) reflected the higher extinction of lineages in smaller refugia. The effect of past habitat restriction was especially well captured by metrics of phylogenetic diversity even when expansions were recent. MNTD failed to capture the signature of more recent expansions (Mazel et al. 2016). Nevertheless, the signature of assemblage-wide expansions was still perceptible in other phylogenetic diversity metrics (λ^* , PD, MPD) despite very ancient and small changes in the past.

There has been a growing interest in how network science can provide novel ways to characterize the shape of phylogenetic trees (Chindelevitch et al. 2019). We found that the shape of a phylogeny, and more specifically its stem-to-tip asymmetry increased - indicating more frequent terminal branching events (Pennell, Sarver, and Harmon 2012).

Hence, using different metrics of phylogenetic diversity and tree shape provide insights into how past habitat restrictions have impacted the diversification of species. Phylogenetic diversity provided a complementary insight to taxonomic diversity as taxonomic metrics mostly reflected the effect of past assemblage-wide demographic expansions while phylogenetic metrics reflect the effect of long-term speciation/extinction dynamics. Yet, species richness was influenced by global diversification, hence using phylogenetic diversity metrics which did not depend on species richness by capturing tree-shape such as the skewness of the spectral density profiles.

Phylogenetic diversity metrics carried the signature of assemblage-wide past demographic changes. Even if expansions occurred long ago, extant patterns of phylogenetic diversity still carried the signature of the loss of past lineages, unlike metrics of taxonomic diversity.

4.3 Functional composition reflects the interplay between trait evolution and past diversification in phylogenies

Simulating trait evolution along phylogenies obtained for different expansion scenarios, trait values deviated from the ancestral value depending on the pace of trait evolution with regards to overall phylogenetic diversity. All replicates remained centered on the ancestral trait value. In the case of a Brownian (BM) or late burst (LB) trait evolution model, we showed that the extent by which traits deviated from the ancestral value depended on phylogenetic diversity (λ^*), which in turn depended on parameters of past demographic changes. Our results show how patterns of functional composition change with regards to parameters of demographic expansion: past refugia dynamics have driven the levels of diversification of lineages and thereby the opportunity for traits to deviate from an ancestral trait value. In larger past refugia, increased diversification amounting to higher phylogenetic diversity allowed lineages to deviate more from the ancestral trait value. However, in the case of an early burst (EB), most trait evolution happened in smaller assemblages with fewer lineages, thus trait mean and variance remained small and unaffected by subsequent habitat expansion. Thus, the shape of a phylogeny interplays with trait evolution to shape extant functional

composition (Tucker, Davies, et al. 2018 ; Pearse et al. 2019). Specifically, phylogenetic asymmetry can be related to asymmetry in evolutionary models (Kirpatrick and Slatkin 1993 ; Blum and François 2006). In this study, using a metric of stem-to-tip asymmetry (Lewitus and Morlon 2016), we show that a substantial increase in trait variance can be due to a combined effect of increased terminal branching events and the acceleration of trait evolution (in the case of a LB model).

Yet additional metrics capturing other aspects of phylogenetic tree shape may have interesting applications to understand how regional trait diversity emerges (Pavoine, Baguette, and Bonsall 2010 ; Tucker, Cadotte, et al. 2017). For instance, if a phylogeny is particularly unbalanced, i.e. if branch lengths are unevenly distributed in the phylogenetic tree - then trait evolution may introduce multi-modal trait distributions. The emergence of functional outliers should then depend on the pace of trait evolution relative to the depth of diverging branching events (Violle et al. 2017). Reciprocally, we can expect trait distributions to reflect evolutive convergence if the distributions of branch-lengths are especially regular. Thus, specific signatures in other trait distribution moments such as the skewness or kurtosis (Gross et al. 2017) may result from the interplay between trait evolution and phylogenetic tree shape (Tucker, Davies, et al. 2018).

Past environmental conditions may have constrained assemblages to refugia and removed lineages with ill-adapted sets of traits in a non-random manner (i.e natural selection) and affect extant patterns of functional diversity (Svenning 2003 ; Ordoñez and Svenning 2015). Here, we focus on the patterns of functional diversity emerging from the combination of lineage diversification (under the influence of expansion dynamics) and the pace of trait evolution. These types of approaches can reflect radiations which may or may not be adaptive (Harmon and Open Textbook Library 2019 ; Czekanski-Moir and Rundell 2019). Our approach does not explicitly consider differential trait diversification between species, as may occur in the case of niche differentiation for instance (Dayan and Simberloff 2005). A promising perspective would be to add a supplementary component to our model to represent more explicitly how natural selection affects the diversification of species niches depending on their adequacy to fluctuating environments (Ewing and Hermisson 2010 ; Ovaskainen et al. 2011 ; Karhunen et al. 2013 ; Manceau, Lambert, and Morlon 2016).

4.4 Process-based eco-evolutionary modelling

In this study, we devised an original modeling framework to address how past refugia size and age can affect extant patterns of taxonomic, phylogenetic and functional diversity. Typically, phylogeographic studies focus on how current patterns of genetic diversity within species can retain the influence of past fragmentation and isolation of populations in refugia on intraspecific genetic diversity at present (Avice 2000). Conversely, comparative phylogeographic approaches aim to grasp congruent or differing influence of refugia on several co-occurring taxa (Arbogast and Kenagy 2001 ; Swenson 2019 ; Overcast, Emerson, and Hickerson 2019 ; Overcast, Ruffley, et al. 2021). Because smaller refugia entail greater extinction risk and fewer opportunities for speciation, the diversification of taxa over time should be shaped by changes in habitat availability. Here we confirm these hypotheses, by proposing an approach in which mechanisms of assemblage-wide dynamics of species abundances and trait values are explicitly simulated depending on changing

habitat availability.

By directly representing the consequences of assemblage-wide environmental fluctuations over a wide range of parameter values relating to refugia size and age, we can address the broader question of whether any set of chosen metrics carry the signature of such processes. In fact, this approach echoes a more general challenge in ecology: can a snapshot of diversity patterns unambiguously reflect the signatures of many processes acting simultaneously at different spatial and temporal scales (Levin 1992 ; Pearse et al. 2019). Relating extant patterns of diversity to the processes having hypothetically shaped them is an overarching goal in many fields of ecology and evolution. Examining how patterns relate to processes in a virtual setting can allow us to evaluate to what extent we can in turn recover parameter values of refugia dynamics from the observation of extant diversity patterns (Nielsen and Beaumont 2009 ; Beaumont 2010). Approximate Bayesian Computation (ABC) approaches for instance, repeatedly compare large numbers of simulated datasets to observed data, by means of summary statistics (Csilléry et al. 2010 ; Sunnåker et al. 2013). Thus, it is important that the simulated diversity metrics carry the signature of underlying processes, as we have shown here, because inference is directly based on the information withheld in the derived summary statistics (Grimm et al. 2010 ; Hartig et al. 2011). From here, we could now compare observed patterns of taxonomic, phylogenetic and functional diversity to those simulated using our process-based eco-evolutionary model in order to infer whether and how past expansion dynamics have shaped actual observations of biodiversity.

4.5 Perspectives - From a regional pool to local communities: integrating the influence of assembly processes on local community composition

In recent years, there has been a renewed interest in better integrating evolutionary history and community ecology (Swenson 2019 ; Overcast, Emerson, and Hickerson 2019 ; Overcast, Ruffley, et al. 2021). Yet, studies of the assembly of communities often assume that the regional species pool (and associated phenotypes) represents a snapshot of an evolutionary heritage from which species assemble through local dynamics over recent and short enough time periods that diversification events can be ignored. Such approaches consider community assembly as a top-down process, whereby species present in a regional pool go through a series of filters (dispersal limitation, environmental filtering, biotic interactions) to assemble in local communities (Lortie et al. 2004). In this perspective, the species pool is stable, ignoring how the pool species composition is shaped by biogeographic history, colonization events or the tempo of trait evolution (Tucker, Davies, et al. 2018 ; Swenson 2019). However, as we have highlighted above, it is possible to explicitly represent how such processes together generate species composition and traits at a regional level. Recent studies have underlined the critical influence of the diversity of the species pool in providing immigrants (Cornell and Harrison 2014) and called for a more mechanistic modelling approach that directly represents key processes shaping community assembly (Denelle, Violle, and Munoz 2019). The difficulty lies in the definition of the species pool as a fixed phenomenological object (Lessard, Borregaard, et al. 2012 ; Lessard, Belmaker, et al. 2012). Depending on how it is delineated, from species' fundamental niches and their limited dispersal abilities in a broad-scale landscape (Karger

et al. 2016) to finer-scale biotic interactions in shaping community assembly (Perronne et al. 2017), our perspective on the community level can vary. Here we show how local demographic processes influence the taxonomic, phylogenetic and functional competition of the species pool under the influence of long-term environmental fluctuations. Yet we could extend our framework to consider how these same large-scale biogeographic historical processes feed back into community assembly at a local scale. This could be done for instance by integrating a subsequent step to our modeling framework in which species can potentially be filtered based on their traits (Munoz et al. 2018).

Author contributions. All authors contributed to design the study. EB and MJ performed the analyses. EB wrote the first draft of the manuscript. All authors contributed revisions to the manuscript and gave final approval for publication.

Funding. MJ was supported by the Institut de Recherche pour le Développement. This work has benefited from an ‘Investissements d’Avenir’ grant managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX- 25-01).

References

- Arbogast, B. S. and Kenagy, G. J. (2001). “Comparative phylogeography as an integrative approach to historical biogeography”. In: *Journal of Biogeography* vol. 28, no. 7, pp. 819–825.
- Avice, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard university press.
- Bartish, I. V. et al. (2016). “Different habitats within a region contain evolutionary heritage from different epochs depending on the abiotic environment”. In: *Global Ecology and Biogeography*, n/a–n/a.
- Beaumont, M. A. (2010). “Approximate Bayesian Computation in Evolution and Ecology”. In: *Annual Review of Ecology and Systematics* vol. 41, pp. 379–406.
- Bennett, K. D. (1990). “Milankovitch cycles and their effects on species in ecological and evolutionary time”. In: *Paleobiology*. ISBN: 0094-8373 Publisher: JSTOR, pp. 11–21.
- Blanchard, G., Birnbaum, P., and Munoz, F. (2020). “Extinction–immigration dynamics lag behind environmental filtering in shaping the composition of tropical dry forests within a changing landscape”. In: *Ecography*.
- Blum, M. G. and François, O. (2006). “Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance”. In: *Systematic Biology* vol. 55, no. 4. ISBN: 1076-836X Publisher: Society of Systematic Zoology, pp. 685–691.
- Bose, R. et al. (2019). “Phylogenetic diversity in the Western Ghats biodiversity hotspot reflects environmental filtering and past niche diversification of trees”. en. In: *Journal of Biogeography* vol. 46, no. 1, pp. 145–157.
- Chindelevitch, L. et al. (2019). “Network science inspires novel tree shape statistics”. In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory, p. 608646.
- Clavel, J., Escarguel, G., and Merceron, G. (2015). “mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data”. In: *Methods in Ecology and Evolution* vol. 6, no. 11. ISBN: 2041-210X Publisher: Wiley Online Library, pp. 1311–1319.

- Cornell, H. V. and Harrison, S. P. (2014). “What Are Species Pools and When Are They Important?” In: *Annual Review of Ecology, Evolution, and Systematics* vol. 45, pp. 45–67.
- Csilléry, K. et al. (2010). “Approximate Bayesian computation (ABC) in practice”. In: *Trends in Ecology & Evolution* vol. 25, pp. 410–418.
- Czekanski-Moir, J. E. and Rundell, R. J. (2019). “The Ecology of Nonecological Speciation and Nonadaptive Radiations”. In: *Trends in Ecology & Evolution* vol. 34, no. 5, pp. 400–415.
- Dayan, T. and Simberloff, D. (2005). “Ecological and community-wide character displacement: the next generation”. In: *Ecology Letters* vol. 8, pp. 875–894.
- Denelle, P., Violle, C., and Munoz, F. (2019). “Distinguishing the signatures of local environmental filtering and regional trait range limits in the study of trait–environment relationships”. In: *Oikos*.
- Ewing, G. and Hermisson, J. (2010). “MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus”. In: *Bioinformatics* vol. 26, pp. 2064–2065.
- Fisher, R. A. (1930). “The Distribution of Gene Ratios for Rare Mutations”. In: *Proceedings of the Royal Society of Edinburgh* vol. 50, pp. 205–220.
- Fortunel, C. et al. (2014). “Environmental factors predict community functional composition in Amazonian forests”. en. In: *Journal of Ecology* vol. 102, no. 1. Ed. by De Deyn, G., pp. 145–155.
- Grimm, V. et al. (2010). “The ODD protocol: a review and first update”. In: *Ecological modelling* vol. 221, no. 23, pp. 2760–2768.
- Gross, N. et al. (2017). “Functional trait diversity maximizes ecosystem multifunctionality”. In: *Nature ecology & evolution* vol. 1, no. 5. ISBN: 2397-334X Publisher: Nature Publishing Group, pp. 1–9.
- Harmon, L. J. and Open Textbook Library (2019). *Phylogenetic Comparative Methods: learning from trees*. en. OCLC: 1151100908. CreateSpace Independent Publishing.
- Hartig, F. et al. (2011). “Statistical inference for stochastic simulation models – theory and application”. In: *Ecology Letters* vol. 14, pp. 816–827.
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton and Oxford: Princeton University Press.
- Hudson, R. R. (2002). “Generating samples under a Wright–Fisher neutral model of genetic variation”. In: *Bioinformatics* vol. 18, no. 2, pp. 337–338.
- Karger, D. N. et al. (2016). “Delineating probabilistic species pools in ecology and biogeography”. In: *Global Ecology and Biogeography*, n/a–n/a.
- Karhunen, M. et al. (2013). “DRIFTSEL: an R package for detecting signals of natural selection in quantitative traits”. In: *Molecular Ecology Resources* vol. 13, no. 4, pp. 746–754.
- Kelleher, J. and Lohse, K. (2020). “Coalescent simulation with msprime”. In: *Statistical Population Genomics*. Humana, New York, NY, pp. 191–230.
- Kembel, S. W. et al. (2010). “Picante: R tools for integrating phylogenies and ecology”. In: *Bioinformatics* vol. 26, pp. 1463–1464.
- Kingman, J. F. C. (1982). “On the genealogy of large populations”. In: *Journal of Applied Probability*, pp. 27–43.
- Kirxpatrick, M. and Slatkin, M. (1993). “Searching for evolutionary patterns in the shape of a phylogenetic tree”. In: *Evolution* vol. 47, no. 4. ISBN: 0014-3820 Publisher: Wiley Online Library, pp. 1171–1181.

- Lê, S., Josse, J., and Husson, F. (2008). “FactoMineR: an R package for multivariate analysis”. In: *Journal of statistical software* vol. 25, no. 1. ISBN: 1548-7660 Publisher: Los Angeles, pp. 1–18.
- Lessard, J.-P., Belmaker, J., et al. (2012). “Inferring local ecological processes amid species pool influences”. In: *Trends Ecol Evol* vol. 27, pp. 600–607.
- Lessard, J.-P., Borregaard, M. K., et al. (2012). “Strong influence of regional species pools on continent-wide structuring of local communities”. In: *Proceedings of the Royal Society of London B: Biological Sciences* vol. 279, pp. 266–274.
- Levin, S. A. (1992). “The problem of pattern and scale in ecology”. In: *Ecology* vol. 73, pp. 1943–1967.
- Lewitus, E. and Morlon, H. (2016). “Characterizing and comparing phylogenies from their Laplacian spectrum”. In: *Systematic biology* vol. 65, no. 3. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 495–507.
- Lortie, C. J. et al. (2004). “Rethinking plant community theory”. In: *Oikos* vol. 107, no. 2, pp. 433–438.
- Manceau, M., Lambert, A., and Morlon, H. (2015). “Phylogenies support out-of-equilibrium models of biodiversity”. In: *Ecology Letters* vol. 18, pp. 347–356.
- Manceau, M., Lambert, A., and Morlon, H. (2016). “A Unifying Comparative Phylogenetic Framework Including Traits Coevolving Across Interacting Lineages”. In: *Systematic Biology* vol. 66, no. 4, pp. 551–568.
- Marth, G. T. et al. (2004). “The Allele Frequency Spectrum in Genome-Wide Human Variation Data Reveals Signals of Differential Demographic History in Three Large World Populations”. en. In: *Genetics* vol. 166, no. 1, pp. 351–372.
- Mazel, F. et al. (2016). “Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics”. In: *Ecography* vol. 39, no. 10. ISBN: 0906-7590 Publisher: Wiley Online Library, pp. 913–920.
- Morlon, H. et al. (2015). “RPANDA: an R package for macroevolutionary analyses on phylogenetic trees”. In: *Methods in Ecology and Evolution*, n/a–n/a.
- Munoz, F. et al. (2018). “ecolottery : Simulating and assessing community assembly with environmental filtering and neutral dynamics in R”. en. In: *Methods in Ecology and Evolution* vol. 9, no. 3. Ed. by Chisholm, R., pp. 693–703.
- Nielsen, R. and Beaumont, M. A. (2009). “Statistical inferences in phylogeography”. en. In: *Molecular Ecology* vol. 18, no. 6, pp. 1034–1047.
- Ordóñez, A. and Svenning, J.-C. (2015). “Geographic patterns in functional diversity deficits are linked to glacial-interglacial climate stability and accessibility”. In: *Global Ecology and Biogeography* vol. 24, no. 7. ISBN: 1466-822X Publisher: Wiley Online Library, pp. 826–837.
- Ovaskainen, O. et al. (2011). “A New Method to Uncover Signatures of Divergent and Stabilizing Selection in Quantitative Traits”. en. In: *Genetics* vol. 189, no. 2, pp. 621–632.
- Overcast, I., Emerson, B. C., and Hickerson, M. J. (2019). “An integrated model of population genetics and community ecology”. en. In: *Journal of Biogeography* vol. 46, no. 4, pp. 816–829.
- Overcast, I., Ruffley, M., et al. (2021). “A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities”. In: *Molecular ecology resources* vol. 21, no. 8. ISBN: 1755-098X Publisher: Wiley Online Library, pp. 2782–2800.

- Patterson, N. J. (2005). “How Old Is the Most Recent Ancestor of Two Copies of an Allele?” en. In: *Genetics* vol. 169, no. 2, pp. 1093–1104.
- Pavoine, S., Baguette, M., and Bonsall, M. B. (2010). “Decomposition of trait diversity among the nodes of a phylogenetic tree”. en. In: *Ecological Monographs* vol. 80, no. 3, p. 23.
- Pearse, W. D. et al. (2019). “The interaction of phylogeny and community structure: Linking the community composition and trait evolution of clades”. In: *Global Ecology and Biogeography* vol. 28, no. 10, pp. 1499–1511.
- Pennell, M. W., Sarver, B. A. J., and Harmon, L. J. (2012). “Trees of Unusual Size: Biased Inference of Early Bursts from Large Molecular Phylogenies”. In: *PLOS ONE* vol. 7, no. 9, e43348.
- Perronne, R. et al. (2017). “How to design trait-based analyses of community assembly mechanisms: insights and guidelines from a literature review”. In: *Perspectives in Plant Ecology, Evolution and Systematics* vol. 25, pp. 29–44.
- R_Development_Core_Team (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Revell, L. J., Harmon, L. J., and Collar, D. C. (2008). “Phylogenetic Signal, Evolutionary Process, and Rate”. In: *Systematic Biology* vol. 57, pp. 591–601.
- Sunnåker, M. et al. (2013). “Approximate Bayesian Computation”. In: *PLOS Computational Biology* vol. 9, no. 1, e1002803.
- Svenning, J.-C. (2003). “Deterministic Plio-Pleistocene extinctions in the European cool-temperate tree flora”. In: *Ecology Letters* vol. 6, pp. 646–653.
- Swenson, N. G. (2019). *Phylogenetic Ecology: A History, Critique, and Remodeling*. en. University of Chicago Press.
- Tucker, C. M., Cadotte, M. W., et al. (2017). “A guide to phylogenetic metrics for conservation, community ecology and macroecology: A guide to phylogenetic metrics for ecology”. en. In: *Biological Reviews* vol. 92, no. 2, pp. 698–715.
- Tucker, C. M., Davies, T. J., et al. (2018). “On the relationship between phylogenetic diversity and trait diversity”. en. In: *Ecology* vol. 99, no. 6, pp. 1473–1479.
- Vellend, M. (2010). “Conceptual synthesis in community ecology”. In: *The Quarterly review of biology* vol. 85, no. 2, pp. 183–206.
- Violle, C. et al. (2017). “Functional rarity: the ecology of outliers”. In: *Trends in Ecology and Evolution* vol. 32, no. 5, pp. 356–367.
- Wakeley, J. (2009). “Coalescent theory”. In: *Roberts & Company*.
- Weiss, S. and Ferrand, N. (2007). *Phylogeography of southern European refugia*. Springer.

Appendix 5.A Supplementary Figures

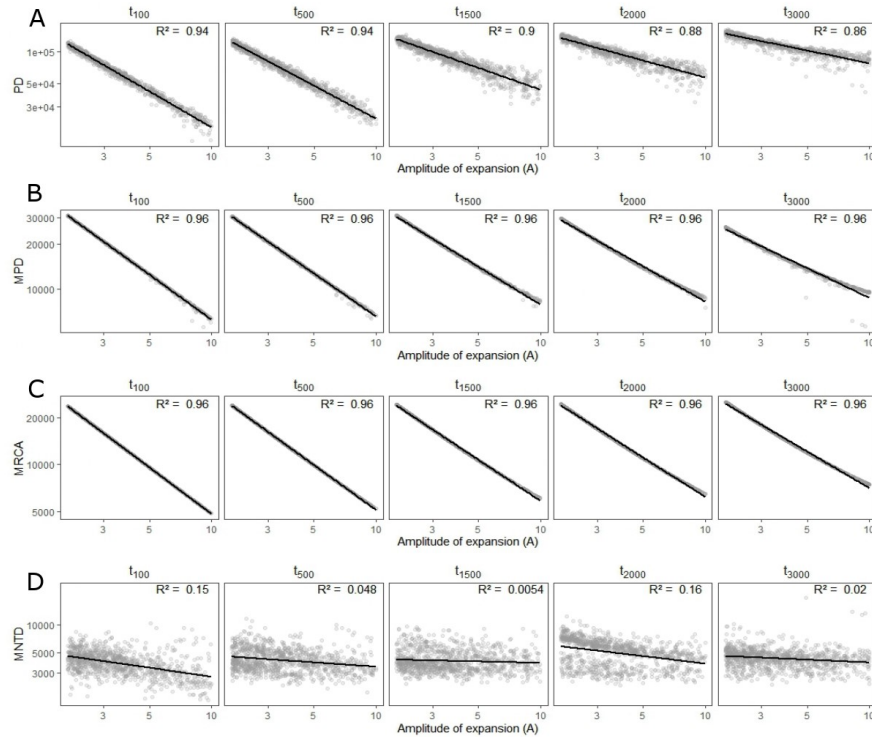


Figure 5.S1: Variation in residuals in linear models linking phylogenetic diversity and species richness depending on the amplitude of expansion (A) and the time (in generations) at which it occurred (t_c). (a) Faith's phylogenetic diversity (PD), (b) mean pairwise distance (MPD), (c) most recent common ancestor (MRCA) and (d) mean nearest taxon distance.

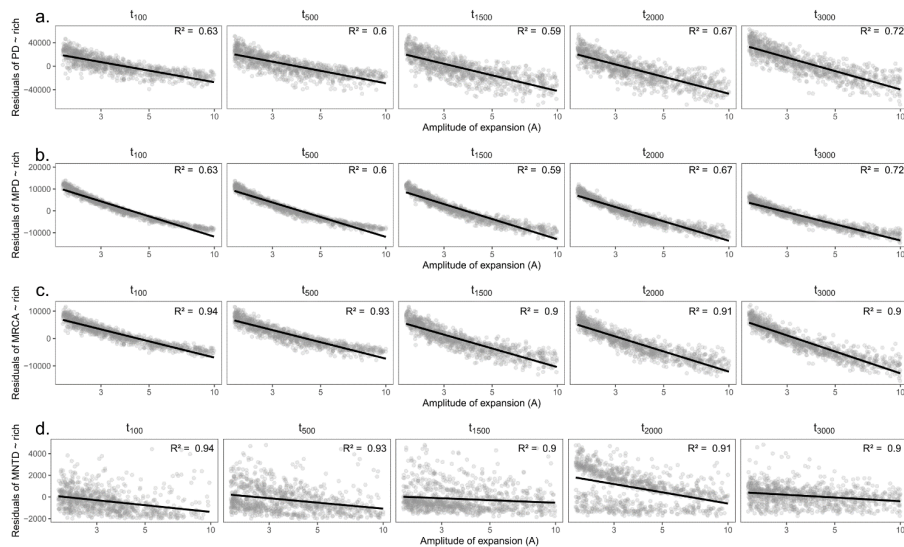


Figure 5.S2: Variation in simulated metrics of phylogenetic diversity (A) Faith's phylogenetic diversity (PD), (B) mean pairwise distance (MPD), (C) most recent common ancestor (MRCA), and (D) mean nearest taxon distance (MNTD) depending on the amplitude of expansion (A) and the time (in generations) at which it occurred (t_c). Linear regressions are shown in black with the associated R^2 statistic in the upper-right side of each panel.

5. Imprints of past habitat area reduction on extant taxonomic, functional and phylogenetic composition

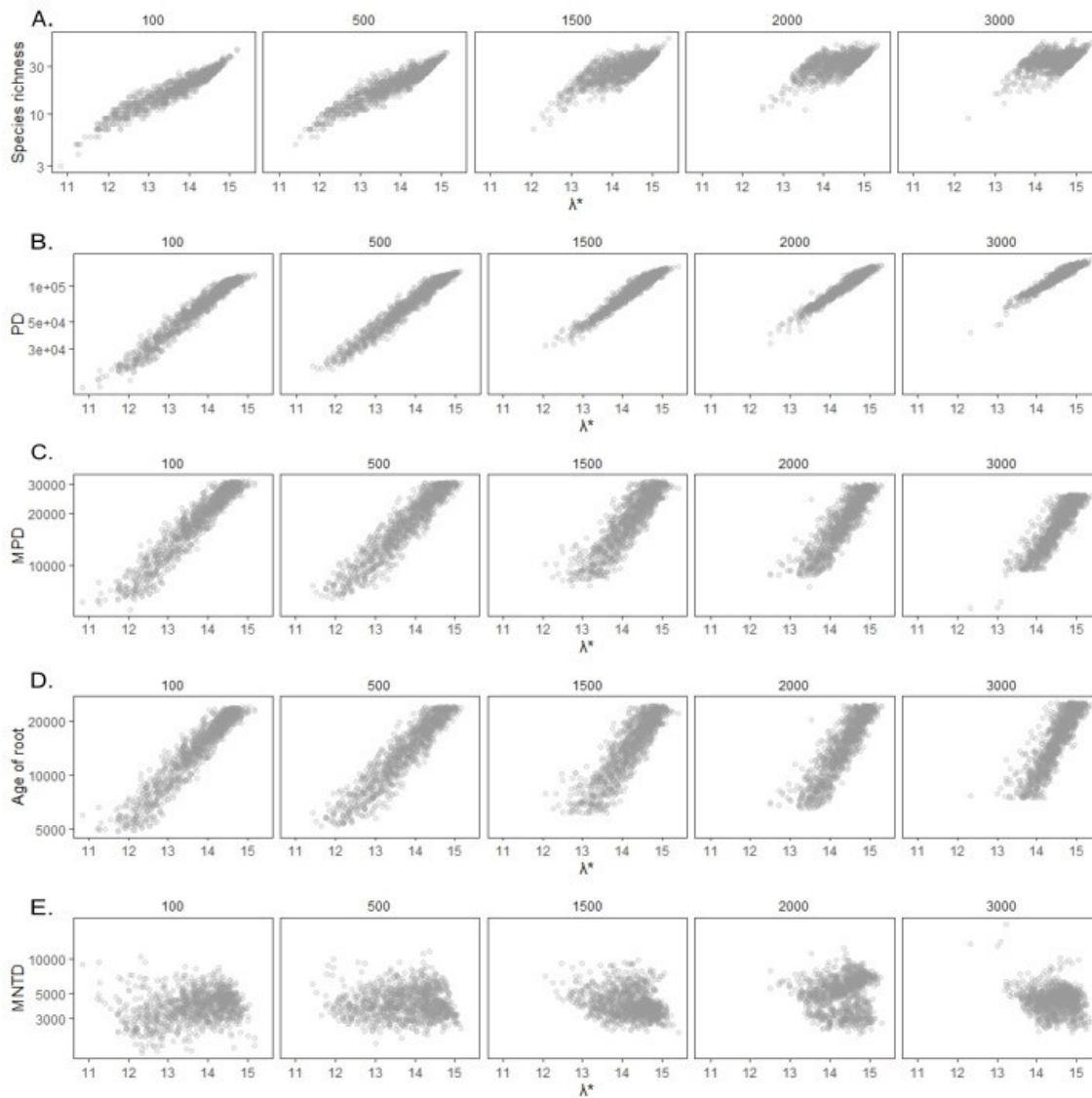


Figure 5.S3: Variation in simulated (A) species richness, (B) phylogenetic diversity (PD), (C) mean pairwise distance (MPD), (D) age of the root (in generations before present), and (E) mean nearest taxon distance (MNTD) depending on the spectral density profile ln-transformed principal eigenvalue (λ^*) and according to the time at which the expansion occurred (in generations before present)

Article 6

Shapes of regional trait distributions resulting from the interplay between phylogenetic imbalance and the pace of trait evolution

Elizabeth Barthélemy¹, Ana Pichelin¹, François Munoz¹

submitted to *Journal of Theoretical Ecology*.

Abstract

A key challenge in community ecology is to identify and describe the processes responsible for observed biodiversity patterns. Although often considered independent of evolutionary processes, the taxonomic and functional diversity of regional species pools is nevertheless shaped by processes that determine how species emerge or disappear through speciation/extinction dynamics and how their traits evolve over time. Recent studies, have since highlighted the limitations of considering these pools as fixed and called for more explicit modeling of the mechanisms that drive species pools dynamics. We devised a virtual experiment in which phylogenies are generated given different levels of imbalance representing variations in diversification rates between lineages. The imbalance of the phylogenies is then characterized and linked to functional diversity patterns according to different models of trait evolution that reflect more or less ancient episodes of evolutionary radiation. We show that current functional diversity carries the joint signature of both evolutionary processes generating phylogenetic asymmetry and the tempo of trait evolution. Our results show that in the case where most of trait evolution takes place in the deep branches of the tree, the distribution of traits in terms of variance, skewness and kurtosis retains a clear signature testifying to the presence of ancient lineages that have diversified very little, and that persist in the current assemblages.

¹LIPhy Université Grenoble Alpes, Grenoble, France

1 Introduction

One of the major challenges in community ecology is to characterize precisely the mechanisms responsible for the currently observed taxonomic, functional and phylogenetic diversity of organisms (Mittelbach and McGill 2019 ; Morin 2009). These different facets of biodiversity are shaped at the local scale as a result of abiotic constraints or biotic interactions limiting the dispersal, survival and reproduction of species in a finite environment (Vellend 2016). Yet these also depend on evolutionary processes that have shaped, over longer periods of time, the diversity of species that are regionally available as well as their characteristics or traits (Ricklefs 2008). Studying the distribution and composition of functional traits within extant communities thus reflects both the role of current ecological processes and the outcome of evolutionary dynamics determining regional diversity (Costa-Pereira et al. 2019 ; Enquist et al. 2015). However, how the evolutionary dynamics shape the functional composition of extant assemblages remains little addressed (Barthelemy et al. 2021).

One way to consider the diversification of species over time is by means of phylogenetic approaches (Morlon 2014) because we expect the topology of extant phylogenetic trees to retain the signal of the evolutionary processes that led to their construction. Specifically, one often studied feature of empirical phylogenies is the nature and dynamics of tree imbalance, which can result from a variation in speciation and extinction rates (Figure 6.1a) (Agapow and Purvis 2002). A balanced tree suggests a uniform distribution of branch lengths across the tree (Figure 6.1c), and can occur under a uniform distribution of diversification events over its evolutionary history. Thus, the study of the imbalance of phylogenetic trees is of particular interest to understand macro-evolutionary dynamics deviating from such uniformity (Heard 1996).

Niche conservatism makes the assumption that more related species should resemble each other more than distantly related ones (Pearse et al. 2019). This can result from trait evolution along a phylogeny, such as Brownian trait evolution (Tucker et al. 2018), so that trait evolution interplays with diversification dynamics along the phylogeny to shape extant functional composition. Differences in the evolution of species traits in terms of tempo (e.g. radiation) or in terms of mode (e.g. selection) can thereby entail different signatures in regional functional composition.

The first moments of the functional trait composition, i.e., mean, variance, skewness, and kurtosis convey complementary insights into ecological assembly rules at a local scale (Enquist et al. 2015 ; Gross, Le Bagousse-Pinguet, et al. 2021), but we posit here that the interplay of diversification dynamics shaping phylogenies and trait evolution should also affect the facets of extant functional composition. Specifically, the skewness measures the asymmetry of a distribution and increases in absolute value when few individuals have extreme traits relative to the rest of the distribution. While functional skewness can reflect rapid environmental change in ecological dynamics, we hypothesize here that trait evolution on an imbalanced tree can also yield functional skewness at present. Excess kurtosis is a measure of the contribution of the tails of a distribution relative to its center and is therefore sensitive to the presence of functional outliers. An excess of positive kurtosis has been shown to reflect low trait diversity while conversely negative kurtosis indicates high diversity (Gross,

Le Bagousse-Pinguet, et al. 2021 ; Gross, Bagousse-Pinguet, et al. 2017). Here we hypothesize that kurtosis can depend on the tempo of trait evolution in a phylogeny.

We developed a virtual experiment to investigate how different facets of functional composition of extant organisms depend on the shape of simulated phylogenies varying in terms of imbalance, and to the tempo of trait evolution on the phylogenies. We show that extant patterns of functional skewness and kurtosis can reveal the interplay of these dynamics, which opens new prospects for a more integrative and eco-evolutionary understanding of the functional composition of extant assemblages.

2 Material and methods

2.1 Simulating phylogenies and characterizing their imbalance

Phylogenetic trees were generated from the α - β_s splitting model described in (Maliot, Gascuel, and Lambert 2018). The parameters α and β_s of the model represent the relationship between a clade's species richness and its age ($\alpha \in]-\infty, +\infty[$) and tree skewness ($\beta_s \in]-2; +\infty[$), respectively. When $\beta_s < 0$, tree skewness increases, while trees tend to be increasingly balanced for $\beta_s > 0$ (Figure 6.1). Setting $\alpha = 1$ and $\beta_s = 0$ provides the standard Yule model of diversification (Figure 6.1b). The latter is often used as a null model for the shape of trees where each branch has an equal probability of splitting (Blum and François 2006 ; Paradis, Claude, and Strimmer 2004).

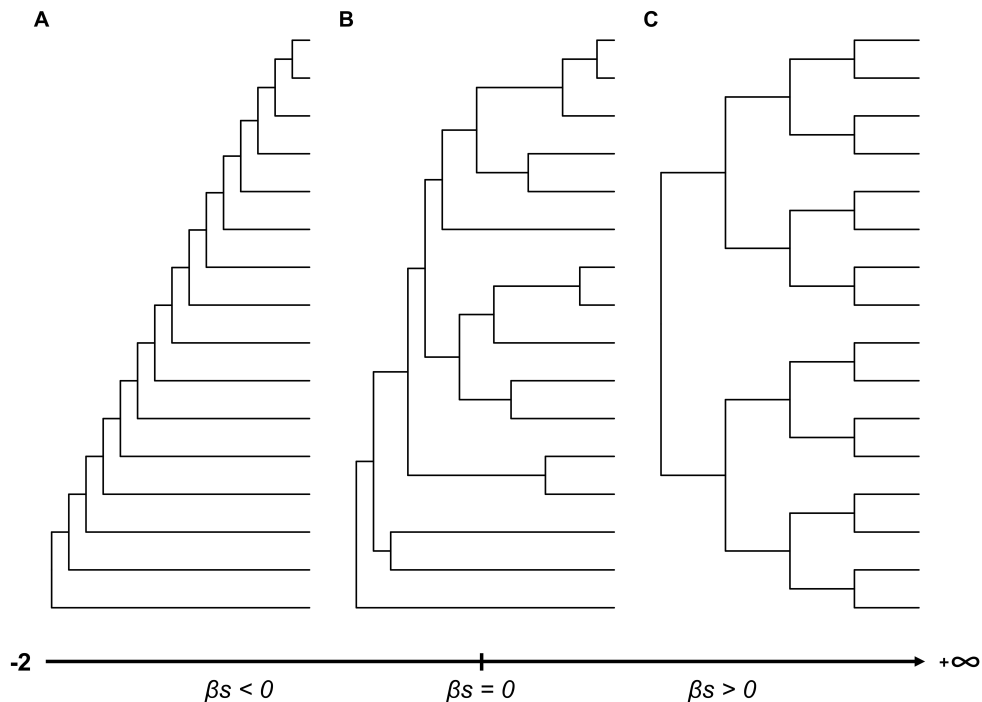


Figure 6.1: Variation of asymmetry of simulated topologies depending on different values of beta. (A) Asymmetric tree ($I_c = 1$). (B) Yule tree ($0 < I_c < 1$). (C) Balanced tree ($I_c = 0$).

We simulated 500 phylogenies using the `simulate_tree` function in the `apTreeshape` package in R (Bortolussi et al. 2006 ; Maliet, Gascuel, and Lambert 2018). Specific richness was set to 100, $\alpha = 1$, and β s varied uniformly between -2 and 0 to generate a range of unbalanced to balanced trees.

2.2 Simulating functional trait distributions

We simulated different models of trait evolution on the simulated phylogenies, depending on the parameter β , the exponential rate of trait evolution, using the `mvMORPH` package (Clavel, Escarguel, and Merceron 2015). In the first model, trait evolution follows Brownian motion ($\beta = 0$). In the second Early Burst model, trait evolution decreases exponentially with time ($\beta < 0$, here fixed at -0.1). In the Late Burst model, evolution accelerates exponentially with time ($\beta > 0$, here fixed at 0.1). For each model, the trait evolved from an ancestral state θ equal to 0 with different evolutionary rates estimated by σ (standard deviation of Brownian motion) set to either 0.2, 0.5, or 0.8. For each phylogeny, trait evolution was simulated in 100 replicates for each combination of β and σ (9 in total).

We characterized the functional composition of extant assemblages in terms of the second, third, and fourth moment of the trait distribution, i.e., variance, skewness, and kurtosis, respectively. We calculated skewness (ψ) and excess kurtosis (γ^2) with the `e1071` package (Meyer et al., 2014). We averaged each moment and calculated its variance over all 100 replicates for each combination of σ and β . Finally, we evaluated the relationship between functional diversity and tree imbalance in terms of parameter β_s .

All analyses were performed in R language using RStudio (4.0.5) (R_Development_Core_Team 2010)

3 Results

Under an EB model of trait evolution, the kurtosis of the resulting trait distributions decreased with imbalance parameter β_s (Figure 6.2b), i.e. when phylogenies were increasingly balanced. In the case of BM and LB evolution models, simulated kurtosis did not depend on phylogenetic imbalance and reached similar values approaching the expected kurtosis of a normal distribution i.e zero. Positive kurtosis values were obtained in unbalanced phylogenies only under a EB trait evolution model. Furthermore, the skewness of the simulated trait distributions did not depend on parameter β_s under LB and BM trait evolution while, in the case of an EB, trait distributions were more skewed when phylogenies were more unbalanced (Figure 6.2d) and became progressively less so as parameter β_s increased (and phylogenies became more balanced). Also, for all trait evolution models, trait skewness and kurtosis did not vary according to parameter σ , while trait variance increased with σ for all models. In the case of BM and EB trait variance also increased with parameter β_s - that is when phylogenies were more balanced (Figure 6.2e). Simulated trait variance was greatest under the LB trait evolution model with the largest σ value ($\sigma = 0.8$).

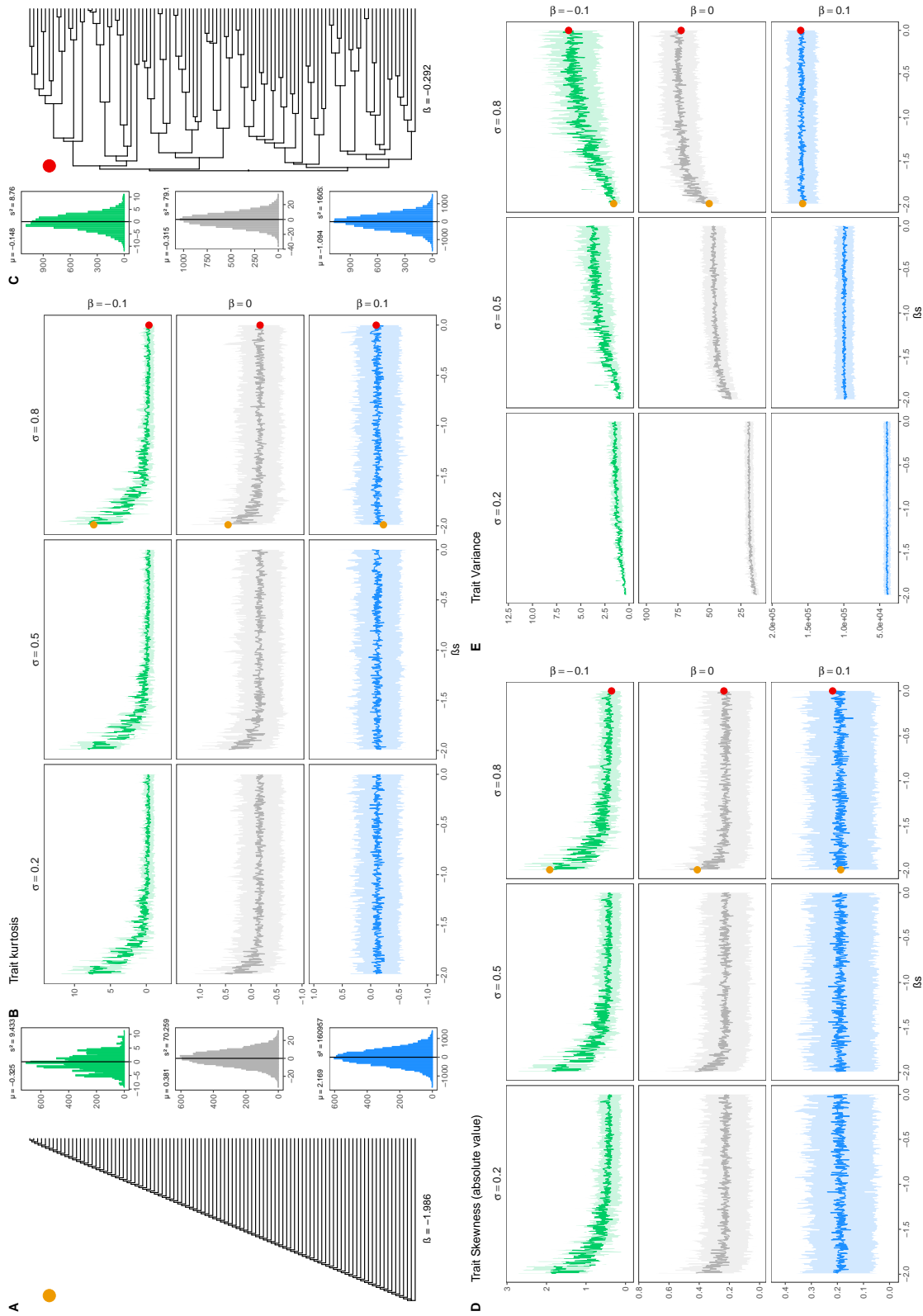


Figure 6.2: Summary of variation in functional composition depending on phylogenetic tree shape and trait evolution models. Variation in mean simulated kurtosis (B), variance (D), and skewness (E) according to β_s depending on the parameters used to simulate trait evolution (β in rows and σ in columns in each individual panel). The lines represent the mean over 100 replications of trait evolution models, while the shaded area represents the replicates' standard deviation. Brown Motion trait evolution is represented in grey, while Early and Late Burst are shown in green or blue, respectively. Example trait distributions associated with a very balanced (in red) or a very unbalanced phylogeny (in orange) under different trait evolution models ($\sigma = 0.8$) are given in panels (A) and (C) respectively. Vertical lines in the histograms represent the ancestral trait value. Species richness is fixed for all simulations ($S = 100$).

4 Discussion

In this study we developed a virtual experiment to assess the joint influence of evolutionary dynamics of taxon diversification and trait evolution on extant functional composition. Our results show that the distribution of extant functional traits carries the signature of (i) evolutionary processes generating phylogenetic imbalance and (ii) the tempo of trait evolution. In particular, we have shown that the shape of the distribution of traits, in terms of variance, skewness and kurtosis, depends on the mechanisms determining the shape of the phylogeny and notably its imbalance.

4.1 Evolutionary legacies in extant functional composition

We found that the functional composition characterized by the shape of trait distributions results from the interplay of phylogenetic imbalance, generated by a skewed diversification model, and the tempo of trait evolution. In particular, our results show that in the case where the bulk of trait evolution occurs in the deep branches of the tree (Early Burst), the trait distribution retains a clear signature of tree imbalance due to the fact that it reflects the presence of ancient lineages that have since diversified very little, and that persist in current assemblages (museum effect described in (Bartish et al. 2016)). Trait evolution from an ancestral value and thus the variance of the resulting distribution, can be limited by the few early branching events in tree topology (Maliet, Gascuel, and Lambert 2018 ; Nee and May 1997), even more so when trait evolution is concentrated in a small number of deep nodes. This results in a decrease in the variance of the trait (in the case of an EB) as most of the diversification takes place in the older lineages which when persisting in current assemblages, strongly influence the distribution of functional traits by yielding especially peaked distributions (reflected in the kurtosis). Similarly, the kurtosis can increase when and if the persistence of old trait values appearing at deep nodes, lead to the appearance of a peaky distribution with functional outliers (to which the kurtosis of any distribution is particularly sensitive). Thus, we highlight that some patterns that reflect a form of regional functional sparsity could reflect a "museum" effect (Bartish et al. 2016), in the context of a virtual experiment.

Much emphasis has been put in previous works on the effects of trait evolution on the variance of resulting trait distribution, so that metrics of phylogenetic signal (Blomberg and Garland 2002) and of Phylogenetic Independent Contrasts (Felsenstein 1985) are all based on trait variances and covariances. We here considered two other metrics of the shape of trait distribution, skewness and kurtosis. We have shown that these metrics provide further insights on how variation in tempos of trait evolution coupled with tree imbalance affect trait distributions of extant taxa. Therefore, as shown recently for the inference of assembly processes from community composition (Gross, Le Bagousse-Pinguet, et al. 2021), we advocate that variance, skewness and kurtosis provide complementary cues for the inference of diversification processes and trait evolution along phylogenies.

4.2 Lessons for community ecology

Many studies in community ecology use regional species pools to represent how regionally present species may immigrate, be filtered by local environmental conditions, and interact with other species to yield local community composition (Lessard et al. 2012 ; Lortie et al. 2004). In most cases, this is considered not only independent of the evolutionary and biogeographic processes that shaped it over the long term (Cornell and Harrison 2014 ; Swenson 2019 ; Tucker et al. 2018) but can be defined using uniform distributions of trait values and/or species abundances, or as the sum of all regionally sampled communities. We argue that the design of species pools then reflects community contexts rather than the long-term evolutionary legacies, and thus can fail to account for the influence of evolutionary dynamics maintaining functional outliers in a given region. Therefore, the mechanisms generating regional phylogenetic and functional diversity are not trivial and a too agnostic definition of species pool can in some cases bias our conclusions (Denelle, Violle, and Munoz 2019 ; Spasojevic et al. 2018).

In addition, our results indicate that in cases where ancient lineages persist in current assemblages, phylogenetic relatedness is not necessarily a good proxy for trait variance (Swenson 2019). This challenges the use of phylogenetic diversity as a proxy of trait variance in ecological assemblages (Webb et al. 2002). Conversely, our results indicate that patterns of trait skewness and kurtosis related to the topology of phylogenies inform on heterogeneity in diversification dynamics and in trait evolution at the origin of extant functional distribution.

Understanding functional and phylogenetic diversity across multiple scales - from a local to regional perspective (Ricklefs 2008) - thereby calls for adopting eco-evolutionary thinking and integrated modeling frameworks (Swenson 2019).

4.3 Limits and Perspectives

Accurately characterizing the mechanisms responsible for the current diversity of species assemblages in all its aspects is a critical issue in community ecology. Examining how functional diversity patterns reflect lineage diversification processes and trait evolution in a virtual experiment allows us to see if these patterns actually capture the signature and to identify potential confounding effects (Zurell et al. 2010). The virtual experiment approach, which aims to represent the direct consequence of the simulated mechanisms, allows us to evaluate to what extent we are able to infer the parameters of such patterns on a selection of statistics of interest (Beaumont 2010). Future research can build on these results to evaluate, in particular, the signature carried by "museum" lineages in real data. That said, in order to abstract from the influence of richness on the metrics measuring phylogenetic imbalance (which would have limited our ability to compare simulated phylogenies) we fixed the richness for all simulations. If we were to use these results to study observed patterns in real communities (Folk, Siniscalchi, and Soltis 2020), the potential influence of species richness on phylogenetic diversity metrics should be kept in mind (Cadotte et al. 2010).

Author contributions. All authors contributed to design the study. AP and EB performed the analyses. AP wrote the first draft of the manuscript. All authors contributed revisions to the manuscript and gave final approval for publication.

References

- Agapow, P.-M. and Purvis, A. (2002). “Power of Eight Tree Shape Statistics to Detect Nonrandom Diversification: A Comparison by Simulation of Two Models of Cladogenesis”. en. In: *SYSTEMATIC BIOLOGY* vol. 51, p. 7.
- Barthelemy, E. et al. (2021). “Imprints of past habitat area reduction on extant taxonomic, functional and phylogenetic composition”. In: *Frontiers in Ecology and Evolution* vol. 9. ISBN: 2296-701X Publisher: Frontiers, p. 332.
- Bartish, I. V. et al. (2016). “Different habitats within a region contain evolutionary heritage from different epochs depending on the abiotic environment”. In: *Global Ecology and Biogeography*, n/a–n/a.
- Beaumont, M. A. (2010). “Approximate Bayesian Computation in Evolution and Ecology”. In: *Annual Review of Ecology and Systematics* vol. 41, pp. 379–406.
- Blomberg, P. and Garland, T. (2002). “Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods”. In: *Journal of Evolutionary Biology* vol. 15, pp. 899–910.
- Blum, M. G. and François, O. (2006). “Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance”. In: *Systematic Biology* vol. 55, no. 4. ISBN: 1076-836X Publisher: Society of Systematic Zoology, pp. 685–691.
- Bortolussi, N. et al. (2006). “apTreeshape: statistical analysis of phylogenetic tree shape”. en. In: *Bioinformatics* vol. 22, no. 3, pp. 363–364.
- Cadotte, M. et al. (2010). “Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history”. In: *Ecology Letters* vol. 13, pp. 96–105.
- Clavel, J., Escarguel, G., and Merceron, G. (2015). “mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data”. In: *Methods in Ecology and Evolution* vol. 6, no. 11. ISBN: 2041-210X Publisher: Wiley Online Library, pp. 1311–1319.
- Cornell, H. V. and Harrison, S. P. (2014). “What Are Species Pools and When Are They Important?” In: *Annual Review of Ecology, Evolution, and Systematics* vol. 45, pp. 45–67.
- Costa-Pereira, R. et al. (2019). “Individual niche trajectories drive fitness variation”. In: *Functional Ecology* vol. 33, no. 9. ISBN: 0269-8463 Publisher: Wiley Online Library, pp. 1734–1745.
- Denelle, P., Violle, C., and Munoz, F. (2019). “Distinguishing the signatures of local environmental filtering and regional trait range limits in the study of trait–environment relationships”. In: *Oikos*.
- Enquist, B. et al. (2015). “Scaling from traits to ecosystems: Developing a general Trait Driver Theory via integrating trait-based and metabolic scaling theories”. In: *Advances in Ecological Research* vol. 52.
- Felsenstein, J. (1985). “Phylogenies and the Comparative Method”. In: *Am Nat* vol. 125, pp. 1–15.
- Folk, R. A., Siniscalchi, C. M., and Soltis, D. E. (2020). “Angiosperms at the edge: Extremity, diversity, and phylogeny”. In: *Plant, Cell & Environment* vol. n/a, no. n/a.

- Gross, N., Bagousse-Pinguet, Y. L., et al. (2017). “Functional trait diversity maximizes ecosystem multifunctionality”. In: *Nature ecology & evolution* vol. 1, no. 5. ISBN: 2397-334X Publisher: Nature Publishing Group, pp. 1–9.
- Gross, N., Le Bagousse-Pinguet, Y., et al. (2021). “Unveiling ecological assembly rules from commonalities in trait distributions”. In: *Ecology Letters* vol. 24, no. 8. ISBN: 1461-023X Publisher: Wiley Online Library, pp. 1668–1680.
- Heard, S. B. (1996). “Patterns in phylogenetic tree balance with variable and evolving speciation rates”. en. In: *Evolution* vol. 50, no. 6, pp. 2141–2148.
- Lessard, J.-P. et al. (2012). “Strong influence of regional species pools on continent-wide structuring of local communities”. In: *Proceedings of the Royal Society of London B: Biological Sciences* vol. 279, pp. 266–274.
- Lortie, C. J. et al. (2004). “Rethinking plant community theory”. In: *Oikos* vol. 107, no. 2, pp. 433–438.
- Maliet, O., Gascuel, F., and Lambert, A. (2018). “Ranked tree shapes, nonrandom extinctions, and the loss of phylogenetic diversity”. In: *Systematic Biology* vol. 67, no. 6. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 1025–1040.
- Mittelbach, G. G. and McGill, B. J. (2019). *Community ecology*. Oxford University Press.
- Morin, P. J. (2009). *Community ecology*. John Wiley & Sons.
- Morlon, H. (2014). “Phylogenetic approaches for studying diversification”. en. In: *Ecology Letters* vol. 17, no. 4. Ed. by Mooers, A., pp. 508–525.
- Nee, S. and May, R. M. (1997). “Extinction and the loss of evolutionary history”. In: *Science* vol. 278, no. 5338. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science, pp. 692–694.
- Paradis, E., Claude, J., and Strimmer, K. (2004). “APE: Analyses of Phylogenetics and Evolution in R language”. In: *Bioinformatics* vol. 20, pp. 289–290.
- Pearse, W. D. et al. (2019). “The interaction of phylogeny and community structure: Linking the community composition and trait evolution of clades”. In: *Global Ecology and Biogeography* vol. 28, no. 10, pp. 1499–1511.
- R_Development_Core_Team (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ricklefs, R. (2008). “Disintegration of the ecological community: American Society of Naturalists Sewall Wright award winner address”. In: *The American Naturalist* vol. 172, no. 6, pp. 741–750.
- Spasojevic, M. J. et al. (2018). “Integrating species traits into species pools”. In: *Ecology* vol. 99, no. 6, pp. 1265–1276.
- Swenson, N. G. (2019). *Phylogenetic Ecology: A History, Critique, and Remodeling*. en. University of Chicago Press.
- Tucker, C. M. et al. (2018). “On the relationship between phylogenetic diversity and trait diversity”. en. In: *Ecology* vol. 99, no. 6, pp. 1473–1479.
- Vellend, M. (2016). *The Theory of Ecological Communities*. Princeton University Press.
- Webb, C. O. et al. (2002). “Phylogenies and Community Ecology”. In: *Annual Review of Ecology and Systematics* vol. 33, pp. 475–505.
- Zurell, D. et al. (2010). “The virtual ecologist approach: simulating data and observers”. en. In: *Oikos* vol. 119, no. 4, pp. 622–635.

Article 7

Diversification and divergence of rainforest woody plants in South India and Madagascar relate to geomorphological history

Elizabeth Barthélemy¹, Riya Mamitiana Olivier Ranaivosoa², Nofy Khristiane Zoelinirina², François Munoz¹

published in *Journal of Biogeography*.

Abstract

Aim Indian and Malagasy tropical rainforests share a common history that dates back to the fragmentation of Gondwana. We partitioned phylogenetic diversity across historical periods to uncover how such continental displacement led to the divergence of lineages, through vicariance and/or independent migration from other regions, influencing extant phylogenetic composition.

Location Madagascar and South India (Western Ghats).

Taxon Woody vascular plants (tracheophytes).

Methods Using a time-calibrated megaphylogeny, we partitioned the phylogenetic diversity of Indian and Malagasy tropical rainforest assemblages, across evolutionary periods defined based on milestone events in these regions of geological history. We decomposed the relative proportion of lineages with extant representatives in each area as well as in both areas, to quantify the common and separate biogeographical legacies in the two regions. We identified an excess or deficit of shared and unique lineages for each period and region.

Results The proportion of shared lineages between Madagascar and India progressively decreased following the successive breakoff and isolation of Madagascar from the Indian-Malagasy plate. Lineages unique to India are more numerous than expected at random in more recent time periods, whereas, lineages unique to Madagascar were significantly more present in older time periods.

Main conclusions The flora of Madagascar harbours a greater number of ancient lineages than does the Western Ghats, consistent with the higher number of endemic families and genera of trees in Madagascar. More recent diversification in India could be related to the collision of India with Eurasia, which enabled subsequent biological interchanges.

¹LIPhy, Université Grenoble Alpes, Grenoble, France

²Office National pour l'Environnement, Antananarivo, Madagascar

1 Introduction

Indian and Malagasy tropical rainforests share a common history that dates back to the breakup of Gondwana circa 162 million years ago (Figure 1). Between 135 and 100 million years ago, India and Madagascar were a single island located c. 30°S (Yoder and Nowak 2006). Some 20 million years later, India separated from Madagascar. During its northward journey, India experienced an equatorial climate that favored tropical rainforests (Prasad et al. 2009 ; Shukla and Mehrotra 2014). The meeting of India and the Laurasian Plate c. 50 million years ago began the Himalayan orogeny and caused an great drought of India (citeBriggs2003). This great drought confined Indian rainforests to smaller geographic areas, such as the Western Ghats on the western slopes of India (Gunnell 1997) and some Himalayan foothills in the north (Karanth 2003). During this time, the geological plate of Madagascar became attached to the African plate and caused the island of Madagascar to move slightly northward, between 10°S and 25°S. Prior to timeframes earlier than 40 million years ago (mya), Madagascar was likely protected from moisture coming from the Proto-Indian Ocean by the presence of India to its northeast and thus had a very dry climate. Only when India had moved further north did Madagascar have a wetter climate, resulting in the expansion of Miocene tropical rainforest, 50 to 40 mya (Yoder and Nowak 2006). The monsoon regime began in India 8 mya, coincident with the rise of the Tibetan plateau (Yoder and Nowak 2006 ; Yoshida, Rajesh, and Santosh 1999).

The composition of extant plant assemblages is the outcome of contemporary ecological dynamics. In particular, human activities have shaped extant ecosystems as, e.g., paleoecological, ecological and molecular data support that anthropogenic grassland recently replaced forest within the last millennium in Madagascar (Joseph et al. 2021). Extant forest composition still keeps tracks of the way taxa have evolved and diversified in a given region over the long term (Hubbell 2001 ; Ricklefs

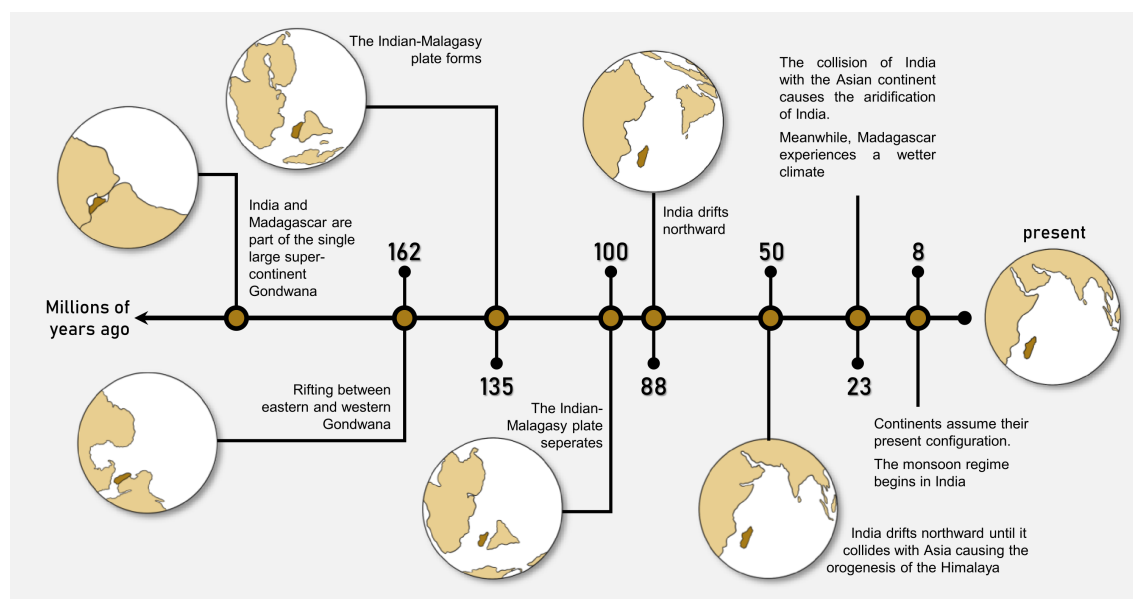


Figure 7.1: Summary of the fragmentation of Gondwana that led to Madagascar's gradual geographic isolation from India and associated biogeographic events. These events defined the 8 time periods used to compute the α LPP and sLPP. Redrawn from Scotese (2000).

2008). For instance, an analysis of the relative frequency of lineages of different epochs in plant assemblages currently occurring in different habitats has shown that over- or under-representation of lineages in an epoch could reflect over- or under-diversification of taxa at this epoch (Bartish et al. 2016). There are different ways of decomposing and analyzing the phylogenetic diversity across periods within assemblages (Pavoine, Love, and M. Bonsall 2009 ; Pavoine, Baguette, and M. B. Bonsall 2010 ; Skeels 2019). The separation of the Indo-Malagasy plate from the rest of Gondwana would have favored the isolation and divergence of India and Malagasy tree lineages, through vicariance (Toussaint, Fikáček, and Short 2016), and independent input of taxa from other regions (Warren et al. 2010). Dipterocarpaceae - a major tree family that now dominates forests in India and South-East Asia - originated in the Malagasy-Indian plate and subsequently dispersed into Asia after India collided with the plate (“out-of-India” hypothesis, Ducouso et al. 2004 ; Dutta et al. 2011), while a vicariant sister family, Sarcolaenaceae, meanwhile differentiated and diversified in Madagascar. Despite the continental shifts that drove the physical separation and isolation of India and Madagascar, fauna capable of crossing small marine barriers may have maintained some connectivity between the two regions (Briggs 2003) and with adjacent areas. Studies based on genetic data have highlighted the recurrent role of long-distance dispersal on the diversification and distribution of organisms (Cook and Crisp 2005 ; De Queiroz 2005). Furthermore, other lineages, such as the Memecylon genus in Melastomataceae with divergence times that are inconsistent with geographic separation events (Amarasinghe et al. 2021 ; Renner 2004), emphasize the need to consider transoceanic migration events in elucidating the biogeographic history of these regions.

Here, we partitioned the phylogenetic diversity of tropical rainforest trees across evolutionary periods, based on milestone events in Malagasy and Indian geologic history, so as to elucidate how this history has influenced the divergence and accumulation of lineages in the two regions. Specifically, we devised a method to analyze lineages that are unique or shared between Madagascar or India for a given evolutionary period. We expected there to be more shared lineages in periods when Madagascar and India were united, with the progressive separation and increasing distance between India and Madagascar resulting in decreasing numbers of shared lineages over time. We also expected there to be more unique lineages in Madagascar or India at periods when local diversification, vicariance and immigration prevailed.

2 Material and Methods

2.1 Indian and Malagasy tropical rainforest data

We used data of forest composition from Madagascar based on a national network of 753 plots covering the diversity of forest types along major environmental gradients. The inventories were conducted in 1995-1996 (Collectif 1996) selecting only plots occurring in wet evergreen forests (303 plots). Plots were triangular with an 80m side, in which six circular sampling units were considered. Trees over 15cm DBH were counted, while smaller trees (between 3cm and 14.9cm DBH) were counted in smaller nested circles. We used a set of 96 1ha forest plots in Karnataka, central Western Ghats, South India, in which all trees above 10cm DBH were identified and counted

7. Diversification and divergence of rainforest woody plants in South India and Madagascar relate to geomorphological history

between 1996 and 1997 (Ramesh et al. 2010). A map detailing the location of the plots is provided in Supplementary Material 7.S1.

Based on these forest inventories we used the time-calibrated megaphylogeny of Zanne et al. 2014, updated by Qian and Jin 2016, built from gene sequences of land plants to derive the phylogenetic relationships between species in both regions using the R package V.Phylomaker (Jin and Qian 2019). Any taxa present in India or Madagascar, but missing from the phylogeny, were added to the next genus- or family-level basal node.

We calculated metrics of taxonomic composition, i.e., the number of species in India and in Madagascar and the percentage of sampled uniqueness of species (number of species found only in one region relative to all sampled species), genera and families in each region.

2.2 Phylogenetic diversity per period

We defined 8 historical periods based on biogeographic events relating to the gradual dislocation of Gondwana and subsequent climate regime shifts in India and Madagascar (Yoder and Nowak 2006), that is 162, 135, 100, 88, 50, 23 and 8 mya (Figure 1). We took care to define periods with decreasing resolution in the past (i.e., longer ancient periods) so as to include sufficient lineages in each period to allow robust statistical analysis. Within each period, the summed branch lengths with extant descendants in both Indian and Malagasy communities were divided by the total branch lengths over each given period (see Figure 7.2). This accounted for the shared lineage history per period, further referred to as Shared Lineages Per Period (sLPP). Consequently, sLPP was bounded between 0 and 1 (where 0 means none of the branches in a given period have descendants from both India and Madagascar, and 1 denotes all branches have descendants from both India and Madagascar). In other words, a smaller sLPP indicates a period with greater divergence between India and Madagascar.

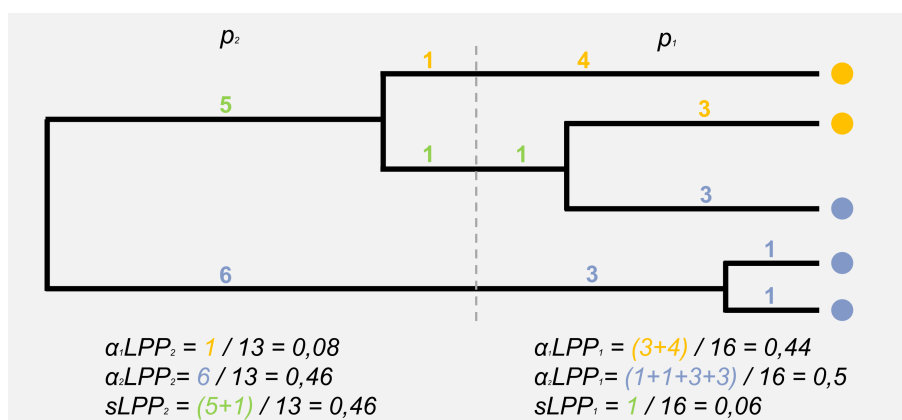


Figure 7.2: Computation of the Lineages Per Period (LPP) metric of phylogenetic diversity through time. This example shows 5 hypothetical species distributed in two regions (blue or yellow) along with their phylogenetic relations. The sLPP is computed for each period (p1 or p2 – separated by a dashed vertical line) by dividing the summed branch lengths upstream from nodes with extant descendants in both regions specific to a given period (branch lengths labelled in green) by the total branch lengths in a given period. The α LPP is computed by dividing the summed branch lengths descending to species found only in one region over a specific period (branch lengths labeled in either yellow or blue) by the total branch lengths in a given period.

We also used branch lengths with extant descendants unique to Indian or Malagasy communities to derive the α LPP in India and Madagascar, respectively (so that α LPP_{India} + α LPP_{Madagascar} + sLPP = 1).

A detail of the code in R language used to compute the sLPP and α LPP for the above-mentioned time periods using the phylogeny and associated community data is provided in the Supplementary Materials.

2.3 Null model-based testing

We designed a null model to quantify, as an effect size (ES), the extent to which observed sLPP values differed from their expected distributions if the phylogenetic distribution of global species was random across the two regions. This phylogenetic randomization was obtained by reshuffling species positions among the tips of the phylogenetic tree. Tree topology and branch lengths were unchanged by the randomization process. The resulting null model thus constrained the diversity of species per region by leaving the species by site matrix unchanged.

For each period, we compared the observed sLPP to the distribution generated by 999 iterations of these null models. Effect sizes (ES) were computed based on the probability for the observed sLPP to be greater than expected under the null model and re-scaled to vary from -1 to 1. If the ES was close to zero, then the observed sLPP values were close to the median of the null distribution. More negative values implied that India and Madagascar share fewer lineages over a given period than expected by chance, and vice versa.

We also compared observed α LPP values for India and Madagascar to their distributions generated by 999 iterations of these null models and derived the associated ES. In this case, positive ES values implied that a given period included more unique lineages than expected by chance, and thus greater local diversification or input from external regions at this period.

2.4 Contribution of taxa to phylogenetic diversity per period

We quantified the contribution of families and genera to the sLPP and α LPP in India and Madagascar for each time period using the distance between the observed values and those computed while removing species belonging to said families and genera, respectively. The larger the distance the more a given family or genus contributed to the metric of phylogenetic divergence per period. We performed Principal Component Analysis of the family and genera contributions per period and for sLPP and α LPP, to identify taxa that contributed the most to these metrics.

All analyses were performed using the R (3.6.3) statistical platform (R_Development_Core_Team 2010).

3 Results

Table 7.S1 summarizes the taxonomic composition of plots in India and Madagascar, and is provided in the supplementary materials.

The proportion of shared lineages between extant Malagasy and Indian rainforest plots progressively decreased from the dates of separation of the Indian-Malagasy plate (100 mya) (Section 3a). The associated negative ES values (Section 3b) showed significant divergence after 100 mya - i.e. Madagascar and India shared fewer lineages than expected under the null expectation. Furthermore, our results showed a drop in the proportion of shared lineages (sLPP) following the separation of the Indian-Malagasy plate from the rest of Gondwana (at -135 Myr).

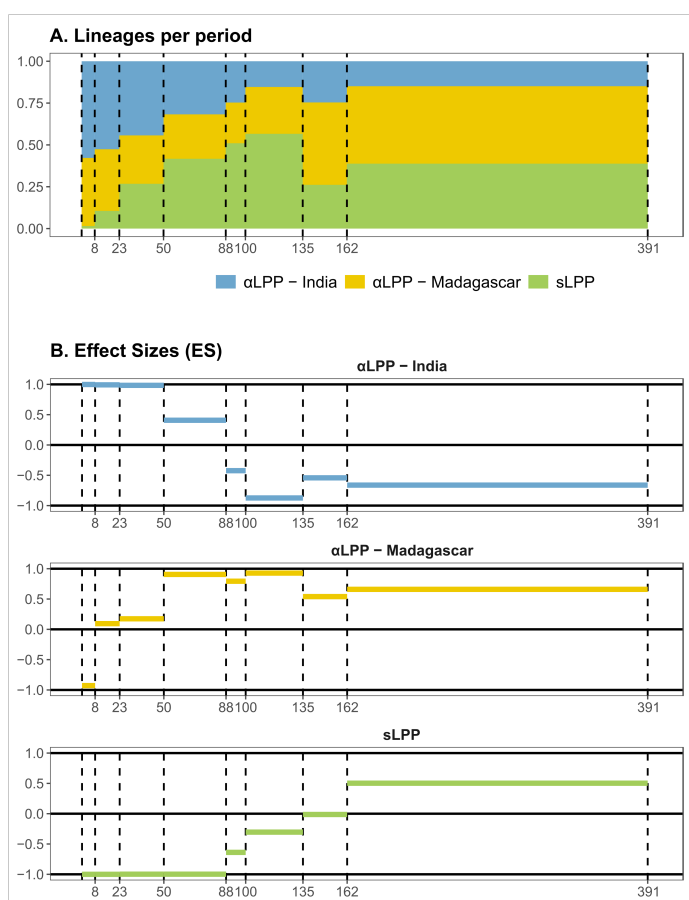


Figure 7.3: Observed shared Lineage Per Period (sLPP) and unique Lineage Per Period values (α LPP) for each time period defined in Figure 1 (A) and associated effect sizes (B) in Indian and Malagasy rainforests' woody plants. sLPP values are shown in green, Indian α LPP in blue and Malagasy α LPP in yellow for each time period (dashed vertical lines). sLPP and α LPP values are bound between 0 and 1 with smaller sLPP values indicating a period of higher divergence between India and Madagascar. The effect sizes (ES) associated with the sLPP, Indian α LPP or Malagasy α LPP are shown in green, blue or yellow, respectively for each time period (dashed vertical lines). If the ES is close to zero (horizontal line), then the observed values are close to the median of the null distribution. ES values are bounded between -1 and 1 (horizontal lines) with negative values suggesting a period with more lineage divergence (fewer shared lineages between Indian and Malagasy communities) than expected by chance and vice versa.

Negative ES associated with α LPP-India at earlier periods dropped between 100 and 135 million years ago then increased and became positive following the separation of India from the Indian-Malagasy plate and its subsequent journey north (between 88 mya until present). Thus, lineages unique to India were more numerous than expected at random in more recent time periods. Conversely, positive ES values associated with α LPP-Madagascar started to decrease after the breakup of the Indian-Malagasy plate, specifically during India's northward drift (88 mya) and became negative after the Miocene (23 mya). ES values closer to 1 between 50 and 135 mya suggested that lineages unique to Madagascar were more extensive in these older time periods.

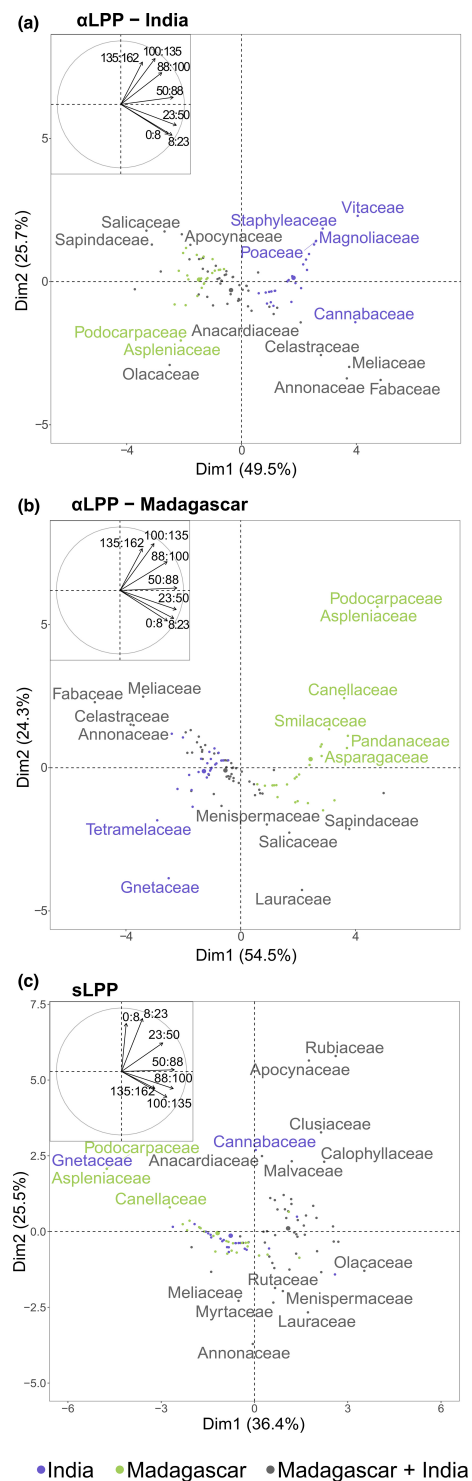


Figure 7.4: Principal component analysis of family contributions to α LPP in India (A), Madagascar (B) and to sLPP (C). Correlation circles in the top left corner of each panel show the contributions of the different time periods to the two first PCA axes. Families in green were sampled only in Madagascar and those in blue only in India while those in grey were found in both regions.

The ES was highest after the collision of India with Eurasian plate c. 50 mya.

The 2 gymnosperm and 1 pteridophyte families (Podocarpaceae, Gnetaceae and Aspleniaceae respectively) contributed exclusively to the sLPP and α LPP values in periods older than 135 mya. The appearance of angiosperms 135 mya coincides with families from these groups showing greater contributions to sLPP and α LPP values in subsequent time periods (Figure 7.4). Celastraceae, Meliaceae, Fabaceae, Annonaceae and Cannabaceae contributed substantially to the Indian α LPP values in geological periods after the collision of India with Asia (50 mya). Families, such as Staphyleaceae and Vitaceae, that were not found in Madagascar plots, contributed to lineages unique to India in earlier time periods (Figure 7.4a). Meanwhile, Canellaceae an older basal Angiosperm family strongly contributed to lineages unique to Madagascar (α LPP-Madagascar) at ancient periods, whilst the endemic family Asteropeiaceae contributed strongly to the Malagasy α LPP between 50 and 88 mya (after Madagascar's separation from India) (Figure 7.4b, Figure 7.S2). Lauraceae, Salicaceae, Sapindaceae and Menispermaceae contributed to the α LPP-Madagascar in more recent periods (after 50 Myr). Shared lineages that contributed the most to the observed patterns were the basal Angiosperms families, such as Lauraceae and Annonaceae. Shared lineage diversity was mostly driven in recent time periods by Rubiaceae and Apocynaceae families (Figure 7.4c, Figure 7.S2).

4 Discussion

Using a dated phylogeny and a network of rainforest plots in the two regions, we found different patterns of phylogenetic diversity in India and Madagascar following their separation. Overall, our results showed that the phylogenetic divergence between Malagasy and Indian rainforest plots progressively increased following Madagascar's isolation from the Indian-Malagasy plate. Furthermore, unique extant Malagasy lineages dated from older periods while extant unique lineages in India were more recent. This

7. Diversification and divergence of rainforest woody plants in South India and Madagascar relate to geomorphological history

‘museum’ effect (Bartish et al. 2016) in Malagasy rainforest phylogeographic history was arguably driven by families of Gondwanan origin. For example, Canellaceae yielded a number of extant taxa in Madagascar, notably the endemic genus *Cinnamosma*, which could not persist in or disperse to India after the separation from Madagascar. This effect could also have been caused by endemic families like the Asteropeiaceae, close to Theaceae, which could have differentiated by vicariance in Madagascar after its separation from the Indian-Malagasy plate 88 mya.

Our results show opposite trends for the diversification of the Indian flora following its separation from Madagascar, as the over-representation of more recent lineages in India rather suggests some “cradle” effect (Bartish et al. 2016). Thus, taxa in the Malagasy flora could diversify earlier than those in the flora in the Western Ghats, which is consistent with the higher number of endemic families (Asteropeiaceae, Barbeuiaceae, Physenaceae, Sarcolaenaceae and Sphaerosepalaceae) and genera of trees and liana in Madagascar, while there are no endemic families of spermatophytes in Western Ghats rainforests. More recent diversification in India follows the collision with the Eurasian plate and suggests substantial immigration from Africa and Asia, with possible recent diversification in Western Ghats. This is consistent with a high level of endemism in Western Ghats at mostly infrageneric level. Colonization of African elements into the Indian subcontinent during the Plio-Pleistocene, especially of Fabaceae and Meliaceae taxa, has been reported based on fossil data (Shukla, Mehrotra, and Guleria 2013). Meliaceae are of western Gondwanan origin, and frequent dispersal shaped their present-day distribution (Muellner et al. 2006). Celastraceae form a spatially restricted group of taxa in South India that were able to drift with India after separation with Madagascar (Bacon et al. 2016).

Among the lineages that contribute the most to the shared LPP between India and Madagascar, some that evolved more recently support the hypothesis of long-distance dispersal, as has already been suggested for Clusiaceae (Ruhfel et al. 2016). Some shared recent genera could have diversified subsequently, producing many endemic species in each area (e.g., *Memecylon*, *Diospyros*). Some old families contributed to both shared lineages (e.g., Annonaceae and Lauraceae) and to lineages unique to India (Annonaceae) and unique to Madagascar (Lauraceae), suggesting both the museum and the cradle effects. Both India and Madagascar were able to retain groups inherited from primary Gondwanan forests (e.g., *Beilschmiedia* and *Cryptocarya* genera in Lauraceae, *Polyathia* complex in Annonaceae), and our results support the possible colonization of Annonaceae in Asia from India (Doyle and Le Thomas 1997). On the other hand, long-distance dispersal might have played a role in establishing extant patterns of Annonaceae (Richardson et al. 2004), as suggested by the recent discovery of a disjunct representative of the African *Friesodielsia* genus in India (Page and Surveswaran 2014).

We compared the composition of mature rainforest plots and considered these as a representative, conserved and unmanaged sample of Indian and Malagasy tropical rainforest biodiversity. However, substantial anthropogenic impacts are likely to have blurred the long-term biogeographical signals in phylodiversity. We selected forest plots situated in more natural ecosystems to limit an influence of recent inputs in managed and planted forests, yet the plot-based sampling could also undersample native woody biodiversity in the two study areas. Ongoing initiatives to provide checklists of rainforest biodiversity with accurate information on the biogeographical status (e.g., TreeGhats in

South India, (Ravel et al. 2017) should enable the scope of future research to be extended.

In this study we devised a new method that allowed us to partition the phylogenetic diversity of tropical rainforests not among the nodes of a phylogeny (Pavoine, Love, and M. Bonsall 2009) but across defined evolutionary periods based on milestone events in the geologic history of one or more regions, so as to uncover major changes in diversification and migration pathways. Future perspectives could make use of this approach to decompose trait diversity (Pavoine, Baguette, and M. B. Bonsall 2010) across geological periods to examine trait evolution in relation to past environmental changes.

Data availability. The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s. Malagasy forest census data used in this study were obtained from the Inventaire Ecologique Forestier National (<https://inventaire-forestier.ign.fr/>), and the complete list of sampled taxa is provided here in Supplementary Material. Indian forest census data on which the analyses of this study were based are available at <https://doi.org/10.6084/m9.figshare.c.3303531.v1>

Author contributions. All authors contributed to design the study. EB performed the analysis and wrote the first draft of the manuscript. All authors contributed revisions to the manuscript and gave final approval for publication.

Acknowledgements. No collecting permits were required for this research. We thank Pr. B.R. Ramesh for stimulating discussions on the biogeography of Western Ghats and Madagascar. We warmly thank Ghislain Vieilledent and Jean-Roger Rakotoarijaona for having digitized and made available the data of forest inventory in Madagascar.

References

- Amarasinghe, P. et al. (2021). “Evolution and biogeography of Memecylon”. In: *American Journal of Botany* vol. 108, no. 4. ISBN: 0002-9122 Publisher: Wiley Online Library, pp. 628–646.
- Bacon, C. D. et al. (2016). “Biogeography of the Malagasy Celastraceae: Multiple independent origins followed by widespread dispersal of genera from Madagascar”. In: *Molecular Phylogenetics and Evolution* vol. 94, Part A, pp. 365–382.
- Bartish, I. V. et al. (2016). “Different habitats within a region contain evolutionary heritage from different epochs depending on the abiotic environment”. In: *Global Ecology and Biogeography*, n/a–n/a.
- Briggs, J. C. (2003). “The biogeographic and tectonic history of India”. In: *Journal of Biogeography* vol. 30, pp. 381–388.
- Collectif (1996). *Inventaire Ecologique Forestier National*. Tech. rep. Direction des Eaux et Forêts DFS Deutsche Forstservice GmbH Entreprise d’Etudes de Développement Rural »Mamokatra« Foiben-Taosarintanin’I Madagasikara.
- Cook, L. G. and Crisp, M. D. (2005). “Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography”. In: *Journal of biogeography* vol. 32, no. 5. ISBN: 0305-0270 Publisher: Wiley Online Library, pp. 741–754.

7. Diversification and divergence of rainforest woody plants in South India and Madagascar relate to geomorphological history

- De Queiroz, A. (2005). “The resurrection of oceanic dispersal in historical biogeography”. In: *Trends in ecology & evolution* vol. 20, no. 2. ISBN: 0169-5347 Publisher: Elsevier, pp. 68–73.
- Doyle, J. A. and Le Thomas, A. (1997). “Phylogeny and geographic history of Annonaceae”. In: *Géographie physique et Quaternaire* vol. 51, pp. 353–361.
- Ducouso, M. et al. (2004). “The last common ancestor of Sarcolaenaceae and Asian dipterocarp trees was ectomycorrhizal before the India-Madagascar separation, about 88 million years ago”. en. In: *Molecular Ecology* vol. 13, no. 1, pp. 231–236.
- Dutta, S. et al. (2011). “Eocene out-of-India dispersal of Asian dipterocarps”. In: *Review of Palaeobotany and Palynology* vol. 166, pp. 63–68.
- Gunnell, Y. (1997). “Relief and climate in South Asia: the influence of the Western Ghats on the current climate pattern of peninsular India”. In: *International journal of climatology: a journal of the Royal Meteorological Society* vol. 17, no. 11. ISBN: 0899-8418 Publisher: Wiley Online Library, pp. 1169–1182.
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton and Oxford: Princeton University Press.
- Jin, Y. and Qian, H. (2019). “V. PhyloMaker: an R package that can generate very large phylogenies for vascular plants”. In: *Ecography* vol. 42, no. 8. ISBN: 0906-7590 Publisher: Wiley Online Library, pp. 1353–1359.
- Karant, K. P. (2003). “Evolution of disjunct distributions among Wet-zone species of the Indian subcontinent: Testing various hypotheses using”. In: *Current Science* vol. 85.
- Muellner, A. N. et al. (2006). “The mahogany family “out-of-Africa”: Divergence time estimation, global biogeographic patterns inferred from plastid rbcL DNA sequences, extant, and fossil distribution of diversity”. In: *Molecular Phylogenetics and Evolution* vol. 40, pp. 236–250.
- Page, N. V. and Surveswaran, S. (2014). “Friesodielsia sahyadrica (Annonaceae), a peculiar new species from the Western Ghats, India”. In: *Phytotaxa* vol. 158, pp. 275–282.
- Pavoine, S., Baguette, M., and Bonsall, M. B. (2010). “Decomposition of trait diversity among the nodes of a phylogenetic tree”. en. In: *Ecological Monographs* vol. 80, no. 3, p. 23.
- Pavoine, S., Love, M., and Bonsall, M. (2009). “Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically-structured species assemblages: application to rockfish (genus: Sebastes) in the Southern California Bight”. In: *Ecology Letters* vol. 12, pp. 898–908.
- Prasad, V. et al. (2009). “Evidence of Late Palaeocene-Early Eocene equatorial rain forest refugia in southern Western Ghats, India”. In: *Journal of Biosciences* vol. 34, pp. 777–797.
- Qian, H. and Jin, Y. (2016). “An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure”. In: *Journal of Plant Ecology* vol. 9, no. 2. ISBN: 1752-9921 Publisher: Oxford University Press, pp. 233–239.
- R_Development_Core_Team (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raevel, V. et al. (2017). “Taxonomic and ecological database of trees of Western Ghats-TreeGhatsData (Version 1.1)[Data set]”. In: *Zenodo*.
- Ramesh, B. R. et al. (2010). “Forest stand structure and composition in 96 sites along environmental gradients in the central Western Ghats of India: Ecological Archives E091-216”. In: *Ecology* vol. 91, no. 10. ISBN: 1939-9170 Publisher: Wiley Online Library, pp. 3118–3118.

- Renner, S. S. (2004). “Multiple Miocene Melastomataceae dispersal between Madagascar, Africa and India”. en. In: p. 10.
- Richardson, J. E. et al. (2004). “Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 359, pp. 1495–1508.
- Ricklefs, R. (2008). “Disintegration of the ecological community: American Society of Naturalists Sewall Wright award winner address”. In: *The American Naturalist* vol. 172, no. 6, pp. 741–750.
- Ruhfel, B. R. et al. (2016). “Dispersal largely explains the Gondwanan distribution of the ancient tropical clusioid plant clade”. In: *American Journal of Botany* vol. 103, no. 6. ISBN: 0002-9122 Publisher: Wiley Online Library, pp. 1117–1128.
- Shukla, A. and Mehrotra, R. C. (2014). “Paleoequatorial rain forest of western India during the EECO: evidence from *Uvaria* L. fossil and its geological distribution pattern”. In: *Historical Biology* vol. 26, no. 6. ISBN: 0891-2963 Publisher: Taylor & Francis, pp. 693–698.
- Shukla, A., Mehrotra, R. C., and Guleria, J. S. (2013). “African elements (fossil woods) from the upper Cenozoic sediments of western India and their palaeoecological and phytogeographical significance”. In: *Alcheringa: An Australasian Journal of Palaeontology* vol. 37, no. 1. ISBN: 0311-5518 Publisher: Taylor & Francis, pp. 1–18.
- Skeels, A. (2019). “Lineages through space and time plots: Visualising spatial and temporal changes in diversity”. In: *Frontiers of Biogeography* vol. 11, no. 2.
- Toussaint, E. F. A., Fikáček, M., and Short, A. E. Z. (2016). “India–Madagascar vicariance explains cascade beetle biogeography”. In: *Biological Journal of the Linnean Society* vol. 118, pp. 982–991.
- Warren, B. H. et al. (2010). “Why does the biota of the Madagascar region have such a strong Asiatic flavour?” In: *Cladistics* vol. 26, pp. 526–538.
- Yoder, A. D. and Nowak, M. D. (2006). “Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell”. In: *Annu. Rev. Ecol. Evol. Syst.* vol. 37, pp. 405–431.
- Yoshida, M., Rajesh, H. M., and Santosh, M. (1999). “Juxtaposition of India and Madagascar: a perspective”. In: *Gondwana Research* vol. 2, no. 3. ISBN: 1342-937X Publisher: Elsevier, pp. 449–462.
- Zanne, A. E. et al. (2014). “Three keys to the radiation of angiosperms into freezing environments”. In: *Nature* vol. 506, no. 7486, pp. 89–92.

7. Diversification and divergence of rainforest woody plants in South India and Madagascar relate to geomorphological history

Appendix 7.A Supplementary Figures

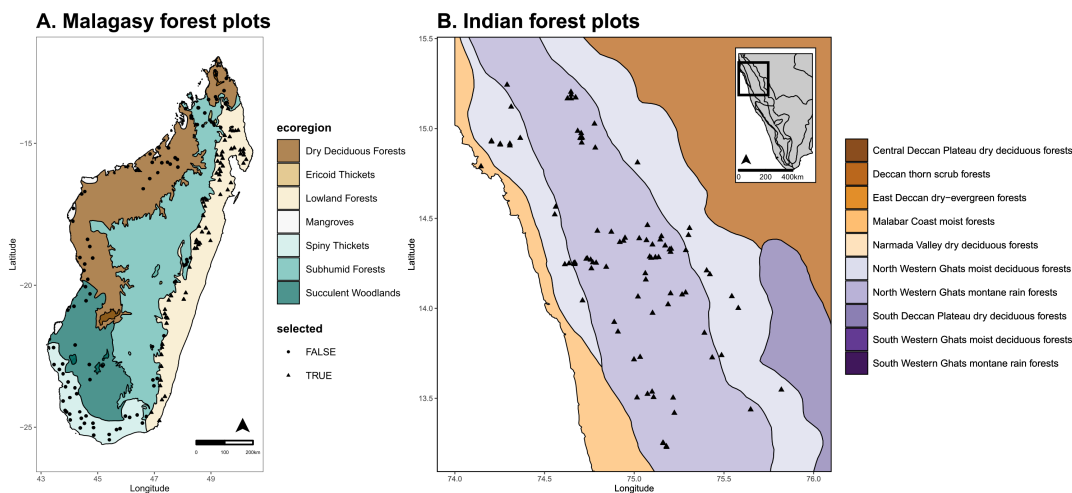
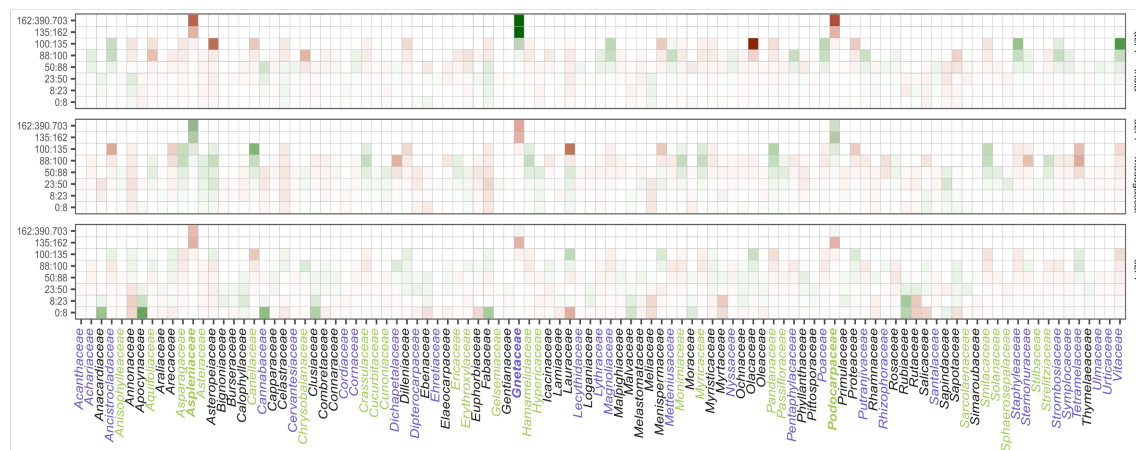


Figure 7.S1: Distribution of forest plots available in Madagascar (A) and in Western Ghats, India (B). Plots in Madagascar cover all kinds of forests (circles), among which we will select plots occurring in rainforest context (triangles). Projection: WGS84 - World Geodetic System 1984.



Appendix 7.B Supplementary Tables

Table 7.S1: Description of taxonomic and phylogenetic tree diversity in Indian and Malagasy rainforest plots

	India	Madagascar	Shared
Species richness	363	277	3
Species uniqueness	43%	56.5%	0.5%
Genus uniqueness	53.8%	33.6%	5.6%
Family uniqueness	28.8%	26%	45.2%
PD	15120.55	12620.80	
MPD	240.67	246.86	
MNTD	47.29	58.14	

Appendix 7.C R Scripts

```

1 BetaPhyloDiv <- function(tree, spmatrix, period){
2
3   getLeaves <- function(tree, node, index = F){
4     leaves <- c()
5     while(length(node)>0){
6       children <- tree$edge[which(tree$edge[,1] %in% node),2]
7       if(length(children) == 0) if(index) return(node) else
8       return(tree$tip.label[node])
9       node <- children[which(children %in% tree$edge[,1])]
10      leaves <- c(leaves, children[which(!children %in% tree$edge[,1])])
11
12    }
13    if(index) return(leaves) else return(tree$tip.label[leaves])
14  }
15
16  # get the ages of the nodes
17  node_ages <- tree.age(tree)
18  node_ages$elements <- as.character(node_ages$elements)
19  node_ages$elements[which(node_ages$elements == "")] <-
20  rownames(node_ages[which(node_ages$elements == ""),])
21
22  root <- node_ages[which(node_ages$ages == max(node_ages$ages)),]
23  if(any(period > root$ages)) period <- period[- which(period > root$ages)]
24  period <- c(0, period, root$ages)
25  per_begin <- period[-length(period)]
26  per_end <- period[-1]
27  per_names <- paste0(per_begin, ":", per_end)
28  periods.lengths <- per_end - per_begin
29  names(periods.lengths) <- per_names
30
31  for(i in length(per_end):1){
32    node_ages[which(node_ages[,1] <= per_end[i]),"period"] <- per_names[i]
33    node_ages[which(node_ages[,1] <= per_end[i]),"period_begin"] <- per_begin[i]
34    node_ages[which(node_ages[,1] <= per_end[i]),"period_end"] <- per_end[i]
35  }
36
37  # get the branch lengths between pairs of nodes
38  dist_mat <- dist.nodes(tree)
39
40  # initialize vectors for return
41  tot.br.len.per <- rep(0, length(per_names))
42  names(tot.br.len.per) <- per_names
43  comm.br.len.per <- rep(0, length(per_names))
44  names(comm.br.len.per) <- per_names

```

7. Diversification and divergence of rainforest woody plants in South India and Madagascar relate to geomorphological history

```
45 first.comm.br.len.per <- rep(0, length(per_names))
46 names(first.comm.br.len.per) <- per_names
47 second.comm.br.len.per <- rep(0, length(per_names))
48 names(second.comm.br.len.per) <- per_names
49
50 # loop over nodes and assign branch lengths to correct period
51 for(node in 1:nrow(node_ages)){
52   first.comm = F
53   second.com = F
54   common = F
55   br.to.node <- which(tree$edge[,2] == node)
56   if(length(br.to.node)!=0){
57     parent <- tree$edge[br.to.node,1]
58     node.period = node_ages$period[node]
59     parent.period = node_ages$period[parent]
60     br.length = dist_mat[node, parent]
61
62     leaves = getLeaves(tree, node)
63
64     first.comm = all(spmatrix[1,leaves] > spmatrix[2,leaves])
65     second.comm = all(spmatrix[1,leaves] < spmatrix[2,leaves])
66     common = !any(rowSums(spmatrix[leaves])==0)
67
68     if(node.period == parent.period){
69       tot.br.len.per[node.period] <- tot.br.len.per[node.period] + br.length
70       if(common) comm.br.len.per[node.period] <- comm.br.len.per[node.period]
71       + br.length
72       if(first.comm) first.comm.br.len.per[node.period] <-
73       first.comm.br.len.per[node.period] + br.length
74       if(second.comm) second.comm.br.len.per[node.period]
75       <- second.comm.br.len.per[node.period] + br.length
76
77     }else{
78
79       traversed.periods <- per_names[seq(which(per_names == node.period),
80       which(per_names == parent.period))]
81       traversed.periods <- traversed.periods[-c(1,length(traversed.periods))]
82
83       if(length(traversed.periods) == 0){
84         trav.br.length = 0}else{
85         trav.br.length <- periods.lengths[traversed.periods]
86         tot.br.len.per[traversed.periods] <- tot.br.len.per[traversed.periods] + trav.br.length
87         if(common) comm.br.len.per[traversed.periods] <- comm.br.len.per[traversed.periods] + trav.br.length
88         if(first.comm) first.comm.br.len.per[traversed.periods] <- first.comm.br.len.per[traversed.periods]
89         + trav.br.length
90         if(second.comm) second.comm.br.len.per[traversed.periods]
91         <- second.comm.br.len.per[traversed.periods] + trav.br.length
92       }
93
94       par.br.length <- node_ages$ages[parent] - node_ages$period_begin[parent]
95       tot.br.len.per[parent.period] <- tot.br.len.per[parent.period] + par.br.length
96       tot.br.len.per[node.period] <- tot.br.len.per[node.period] + (br.length - (sum(trav.br.length) +
97       par.br.length))
98       if(common){
99         comm.br.len.per[parent.period] <- comm.br.len.per[parent.period] + par.br.length
100        comm.br.len.per[node.period] <- comm.br.len.per[node.period] + (br.length - (sum(trav.br.length) +
101        par.br.length))
102
103      }
104      if(first.comm){
105        first.comm.br.len.per[parent.period] <- first.comm.br.len.per[parent.period] + par.br.length
106        first.comm.br.len.per[node.period] <- first.comm.br.len.per[node.period] + (br.length -
107        (sum(trav.br.length) + par.br.length))
108      }
```

```
109     }
110     if(second.comm){
111         second.comm.br.len.per[parent.period] <- second.comm.br.len.per[parent.period] + par.br.length
112         second.comm.br.len.per[node.period] <- second.comm.br.len.per[node.period] + (br.length -
113             (sum(trav.br.length) + par.br.length))
114     }
115     }
116 }
117 }
118 }
119 }
120
121 Alpha = rbind(first.comm.br.len.per,
122               second.comm.br.len.per)
123 rownames(Alpha) <- rownames(spmatrix)
124
125 Ratio = rbind(Alpha[1,] / tot.br.len.per,
126              Alpha[2,] / tot.br.len.per,
127              comm.br.len.per/tot.br.len.per)
128 rownames(Ratio) <- c(rownames(spmatrix), "Shared")
129
130 ret <- list(Alpha= Alpha,
131            SharedBrLen = comm.br.len.per,
132            TotalBrLen = tot.br.len.per,
133            Ratio = Ratio)
134 return(ret)
135
136 }
```


Discussion

Discussion

Identifier parmi une multitude de processus, ceux qui contribuent réellement aux patrons de biodiversité observés dans les milieux naturels et à quelles échelles de temps est un enjeu de recherche majeur en écologie. Les dynamiques individuelles dans un environnement fini et hétérogène (perspective écologique) affectent les dynamiques de spéciation et d'extinction sur le long terme (perspective évolutive), en fonction des contraintes exercées par des changements environnementaux. Le premier chapitre de cette thèse développe une nouvelle approche de modélisation éco-évolutive, basée sur la simulation des dynamiques des individus par coalescence, pour prédire les trajectoires évolutives des espèces dans un contexte de fluctuations environnementales. L'utilisation d'une approche par simulation explicite des processus, dans le deuxième chapitre de la thèse, permet de caractériser à la fois l'influence de la dispersion limitée et de la forme d'un filtre environnemental qui limite l'établissement d'adventices dans des parcelles agricoles en fonction de la culture en place et du contexte paysager. Pour finir, le dernier chapitre explore certaines des signatures laissées par des événements biogéographiques anciens tels que les refuges et la séparation de masses continentales. Par exemple, comment les abondances relatives reflètent l'influence durable des refuges plus récents sur la dynamique démographique, tandis que la composition phylogénétique reflète l'influence des changements d'habitat plus anciens. Ou encore, comment des patrons de diversité fonctionnelle peuvent témoigner de la présence de lignées anciennes qui se sont très peu diversifiées et qui persistent dans les assemblages actuels. Et enfin, comment la fragmentation progressive de deux régions peut se lire dans la proportion de lignées partagée dans la phylogénie au cours de différentes périodes historiques.

1 Intégration des processus éco-évolutifs à de multiples échelles

1.1 *ecophylo*: héritier de la théorie neutre?

1.1.1 Couplage des processus écologiques et évolutifs

Vingt ans après le livre exposant la théorie neutre (Hubbell 2001), les modèles associant les dynamiques individuelles dans les communautés locales et leurs dynamiques de spéciation à large échelle, sont rares (Morlon 2014). La démarche particulière de modélisation de Hubbell qui prend comme point de départ la démographie des individus pour prédire l'évolution des espèces en fait un des rares modèles individus-centrés qui peut être utilisé en macroévolution. Cette démarche est fondamentalement basée sur une envie de décrire les processus à l'échelle locale pour prédire les propriétés et les processus à large échelle. Mais a-t-on besoin de décrire les processus microscopiques d'un système pour en comprendre les phénomènes macroscopiques? Le parti pris de

ce travail de thèse est que oui. Les individus sont les constituants élémentaires des communautés, et donc l'échelle individuelle est pertinente pour comprendre la diversité des espèces, quelle que soit l'échelle d'observation. Prenons l'exemple de la sélection: c'est la différence de performance individuelle qui détermine, pour un environnement donné, la persistance et donc l'abondance locale ou régionale d'espèces avec des propriétés biologiques plus ou moins adaptées. La sélection intervient donc au niveau individuel, qu'elle opère à l'échelle de quelques générations dans le cas d'un filtre environnemental, ou sur le long terme dans le cas de la sélection naturelle. Tout effort de modélisation implique forcément un certain niveau d'abstraction. De ce fait, il peut être valide - selon nos hypothèses - d'ignorer certains processus opérant à des échelles d'organisation inférieures ou supérieures à celle qui nous intéresse. A-t-on vraiment besoin de décrire les mécanismes génétiques sous-jacents pour comprendre que les individus d'une population varient naturellement? De même, est-il nécessaire de décrire les mécanismes qui structurent la diversité des espèces d'un pool régional, quand bien même ces mécanismes n'agissent pas aux mêmes échelles de temps que ceux régissant des communautés? Peut-être que non. Cela-dit, le couplage des échelles par la mise en relation des processus interdépendants est un objectif de recherche intéressant en soi, et constitue un aspect central pour l'étude des dynamiques éco-évolutives. L'étude des dynamiques éco-évolutives demande forcément un effort d'intégration des processus multi-échelles, car les changements environnementaux passés peuvent avoir modifié les dynamiques des populations et de migration des individus, qui à leur tour influencent les trajectoires évolutives des espèces (Pelletier, Garant, and Hendry 2009). C'est la nature transitoire et lente des dynamiques éco-évolutives pour de nombreux groupes d'espèces, couplée au caractère contingent des événements climatiques et géologiques qui rend le rôle de l'histoire particulièrement important. Les modèles éco-évolutifs nécessitent donc d'intégrer d'une part la biogéographie historique des espèces, et d'autre part la dynamique des individus formant des populations dans des environnements fluctuants. En négligeant les dynamiques de populations, les approches phylogéographiques traditionnelles (par exemple, Pigot et al. 2010 ; Tello and Stevens 2012) ne tiennent pas compte de comment la spéciation, la migration et la dérive conduisent à la variation des abondances des espèces (comme envisagé dans la théorie neutre) sous l'influence des variations historiques de la configuration de l'environnement.

1.1.2 L'environnement non stationnaire

De nouvelles approches qui font cet effort d'intégration des dynamiques individuelles commencent à émerger (Overcast, Emerson, and Hickerson 2019 ; Overcast, Ruffley, et al. 2021), mais peu qui puissent rendre compte de l'influence à long terme des changements historiques sur les patrons de biodiversité actuels, et encore moins qui puissent facilement être utilisées pour faire de l'inférence (Prates et al. 2016). L'influence de la variabilité temporelle de l'environnement a été prise en compte dans les modèles phylogéographiques, qui prédisent la diversité génétique en fonction de la taille des populations (Avice 2009), mais rarement dans les modèles biogéographiques multi-espèces de la dynamique de la biodiversité. L'algorithme de simulation *ecophylo*, présenté dans l'Article 1, permet de représenter explicitement la manière dont les mécanismes écologiques élémentaires tels que la dérive écologique et la migration (Vellend 2010) dans un environnement fini et fluctuant, façonnent les trajectoires évolutives et aboutissent aux patrons actuels de diversité phylogénétique

et taxonomique. Celui-ci permet d'étendre le cadre de modélisation coalescent (Kingman 1982 ; Hudson 1983) à des assemblages d'espèces multiples (Manceau, Lambert, and Morlon 2015). Il peut donc être utilisé dans le contexte de la phylogéographie comparative (Avice 2009) pour étudier l'influence des événements biogéographiques passés sur plusieurs taxons co-occurrents, comme il est fait exemple dans l'Article 5. De plus, ce modèle permet de relâcher l'hypothèse de stationnarité de la taille des métacommunautés de la théorie neutre (Hubbell 2001), et peut donc rendre compte de la vicariance et des histoires démographiques distinctes de métacommunautés dont la taille fluctue dans le temps. Cette caractéristique est un avantage essentiel d'*ecophylo* pour modéliser l'influence des changements environnementaux passés sur la dynamique biogéographique, et les patrons de biodiversité.

1.1.3 La spéciation prolongée

Un autre avantage de l'approche de modélisation développée dans cette thèse consiste à modéliser la spéciation comme un processus graduel. Outre le fait que cela répond à des attentes plus réalistes - la spéciation étant un processus inévitablement long - cela permet d'aboutir à de meilleures prédictions en termes de patrons de diversité, comme mentionné dans Rosindell et al. (2010) et démontré dans l'Article 1. La spéciation est modélisée en deux étapes. Une première étape consiste à répartir des événements de mutation le long des branches de la généalogie obtenue par coalescence et contrainte par un environnement fluctuant. Les descendants d'une branche présentant au moins une mutation définissent un clade considéré comme génétiquement distinct. En d'autres termes, si suffisamment de générations séparent un individu de son ancêtre il y a des chances pour que celui-ci ait accumulé assez de mutations pour qu'il soit différent. Cela-dit le modèle ne représente pas explicitement les mécanismes génétiques qui aboutissent à cette variation. Dans un deuxième temps, les branches de la généalogie sont fusionnées pour former des clades monophylétiques d'espèces (Manceau, Lambert, and Morlon 2015). Pour rendre compte du temps dit de fixation des variations génétiques, seules les lignées monophylétiques présentant des génotypes distincts et plus anciennes que τ générations sont considérées comme une espèce distincte (Manceau, Lambert, and Morlon 2015) dans le modèle.

Ce lien entre la généalogie des individus et la phylogénie des espèces est ce qui rend le modèle fondamentalement éco-évolutif. En effet, il permet de rendre compte de l'effet de l'écologie des individus dans un environnement variable sur les trajectoires évolutives des espèces. Supposons que la taille d'une population augmente, par exemple dans le cas où de nouvelles opportunités écologiques se présentent. La probabilité de coalescence dans une population de grande taille étant plus faible, les branches dans la généalogie d'individus seront plus longues. Des branches plus longues impliquent qu'il y a de plus grande chance d'introduire des individus variants, et donc de former de nouvelles espèces. Il s'agit donc d'un cas de spéciation sympatrique. Prenons un autre exemple. Dans le cas d'une population fragmentée, la recherche d'un ancêtre commun par le processus de coalescence va dépendre de l'ancienneté de la séparation géographique. En l'absence de migration, la date minimale de coalescence de deux individus issus de sous-populations différentes dépendra de la date de séparation ou, auquel cas, du taux de migration qui relie ces populations. Toujours est-il que les temps de coalescence seront rallongés par l'apparition d'une barrière, et

donc de nouvelles espèces apparaîtront sous l'effet de cette séparation. Il s'agit donc d'un cas de spéciation allopatrique. Ainsi, bien que le modèle éco-évolutif ne décrive pas explicitement les mécanismes génétiques exacts, il est toutefois possible de représenter des cas de spéciations classiques.

1.2 Perspectives de modélisation futures

1.2.1 Modélisation des dynamiques écologiques par coalescence

Les limites de la neutralité Dans la continuité du modèle éco-évolutif de Hubbell, *ecophylo* suppose la neutralité. C'est-à-dire que les espèces sont soumises de manière équivalente aux fluctuations de la taille de l'assemblage qu'elles forment, sans tenir compte de la manière dont des différences de performance individuelles peuvent influencer leur probabilité de persister dans un environnement donné et/ou de migrer. L'hypothèse de neutralité est fondamentalement peu réaliste, cependant le modèle peut être utilisé comme une base de référence neutre pour étudier comment les mécanismes élémentaires de mutation, de dérive et de migration (Vellend 2010) dans un environnement fluctuant influencent les trajectoires évolutives, et ainsi façonnent les patrons de diversité. Pour accroître la complexité du modèle, et donc son niveau de réalisme, il faudrait tenir compte de la façon dont les processus déterministes tels que les filtres environnementaux, les interactions biotiques et la sélection naturelle affectent la diversification et la dynamique des assemblages au fil du temps (Ewing and Hermisson 2010 ; Ovaskainen et al. 2011 ; Karhunen et al. 2013 ; Manceau, Lambert, and Morlon 2016). En effet, des interactions répétées et fortement liées aux performances des individus peuvent avoir des répercussions sur leur évolution à long terme, pouvant engendrer de la coévolution dans le cas d'une interaction bénéfique, ou au contraire favoriser l'extinction d'une espèce moins compétitive (Dayan and Simberloff 2005). Les processus déterministes impactent directement les taux d'extinction et/ou de migration des individus, et donc influencent forcément leurs généalogies. D'ailleurs, il existe des modèles éco-évolutifs récents qui permettent de prendre en compte les filtres environnementaux (Rangel et al. 2018 ; Garwood, Spencer, and Sutton 2019), mais avec une approche « forward-in-time » qui simule tous ces mécanismes élémentaires pour tous les individus chronologiquement, et donc avec un coût important en temps de calcul. Il est fait mention dans Rangel et al. 2018 que ces approches sont plus adaptées pour simuler et explorer certains scénarios théoriques (i.e problème direct), mais ne sont pas conçues pour l'inférence à partir de données réelles (i.e problème inverse). Le cadre de la modélisation par coalescence a certes ses limites, mais il permet au moins de limiter le temps de calcul pour espérer faire de l'inférence basée sur des simulations. Prendre en compte les processus déterministes, tels que l'effet des interactions biotiques et des filtres environnementaux, dans la modélisation par coalescence constitue pour cela un enjeu de recherche important, et dont des modèles comme *ecophylo* pourront bénéficier dans le futur.

Vers un modèle éco-évolutif spatialement explicite? Il existe également des modèles coalescents spatialement explicites (*pycoalescence/rcoalescence* voir Thompson, Chisholm, and Rosindell 2020) qui permettent de représenter les conséquences de la dérive et de la migration d'individus dans un

maillage spatial. L'habitat spatialement explicité dans le modèle peut éventuellement changer en introduisant des matrices différentes au cours du temps dans la simulation, et de fait, représenter l'effet de fluctuations environnementales spatialement explicites, là où *ecophylo* demeure spatialement implicite. On pourrait utiliser ces nouvelles approches dans notre modèle, pour obtenir des généalogies résultant de fluctuations environnementales spatialement explicitées par une série de matrices représentant un habitat changeant. Cette généalogie pourrait par la suite tout à fait être transformée en phylogénie grâce au modèle de spéciation d'*ecophylo*. Cependant, afin d'utiliser de tels modèles pour l'inférence bayésienne de paramètres, il est nécessaire d'exécuter des simulations en tirant les paramètres dans des distributions a priori. La difficulté dans le cas spatialement explicite vient de la paramétrisation de ces distributions régissant la série de matrices, censée représenter le scénario biogéographique souhaité. Cela dit, spatialement expliciter les changements de configuration de l'environnement dans *ecophylo* n'est peut-être pas nécessaire, car le modèle, bien que spatialement implicite, permet d'introduire de la structuration de population de façon à représenter des cas classiques en biogéographie ; telle que la spéciation géographique, sans avoir besoin d'être spatialement explicite.

1.2.2 Modélisation des dynamiques de spéciation

Limites du modèle de spéciation L'hypothèse de neutralité du modèle éco-évolutif s'étend aux modalités de spéciation. La probabilité d'apparition de nouvelles mutations est identique pour tous les individus, et ne dépend in fine que de la longueur des branches de leur généalogie. Un modèle d'expansion démographique - comme illustré dans l'Article 5 - peut cependant permettre de représenter une radiation évolutive : c'est-à-dire la diversification de nombreuses espèces à partir d'un seul ancêtre. En effet, on peut envisager de représenter la prolifération d'opportunités écologiques - longtemps considérée comme une condition préalable importante, voire nécessaire, à la radiation évolutive (Simpson 1953 ; Schluter 2000) par une augmentation de la taille de la communauté. Par exemple, un épisode de radiation évolutive peut se produire lorsqu'une espèce colonise une zone dans laquelle des ressources sont disponibles, et qu'elle a la capacité évolutive de se diversifier pour en tirer parti. L'expansion démographique qui suit l'expansion d'un habitat s'accompagne alors d'une augmentation de la diversification, car il y a moins d'extinction due à la dérive - plus forte dans des communautés de petite taille. Néanmoins, tous les individus demeurent égaux dans ce nouvel habitat, alors qu'on se doute bien que certaines espèces s'en sortiront mieux si elles possèdent des traits qui leur confèrent un avantage dans ce nouvel environnement. Une radiation évolutive peut se produire sans la colonisation d'un nouveau milieu lorsque, par exemple, l'espèce est déjà présente dans une zone, mais qu'un événement d'extinction élimine une espèce dominante sur le plan écologique (Stroud and Losos 2016), auquel cas les dynamiques de niches interviendront. Si de nouveaux filtres environnementaux apparaissent au cours du temps, il est probable qu'ils modifient la trajectoire évolutive des espèces, en favorisant la survie et donc la sélection de certaines lignées ; processus dont *ecophylo* ne tient pas compte pour le moment. Cela-dit, on pourrait concevoir la sélection naturelle comme l'application répétée d'un filtre environnemental qui s'exerce génération après génération. Dans ce cas, la prise en compte de la sélection naturelle dans un modèle, qui part de la dynamique des individus pour prédire l'évolution des espèces, passe par l'intégration des processus de niche à l'échelle individuelle, et donc dans le cas d'*ecophylo*

par l'architecture de la généalogie. Encore une fois, cette logique relève de la volonté assumée de décrire le comportement microscopique des communautés pour en comprendre les propriétés macroscopiques.

Il n'est cependant pas nécessaire de modéliser les dynamiques individuelles pour représenter le rôle des filtres dans les dynamiques adaptatives. Rendre compte de l'évolution des propriétés des espèces est possible en utilisant directement la phylogénie ; c'est d'ailleurs ce qui a été fait dans l'Article 5. Celui-ci fait usage, comme dans beaucoup d'études en phylogénie comparative, de mouvements browniens (BM) qui considèrent la phylogénie comme un canevas, permettant de modéliser l'évolution des traits le long de ses branches. Le mouvement brownien suppose que la variance du trait s'accumule comme une fonction linéaire du temps, et que le taux d'évolution est stochastiquement constant sur toutes les branches (Ricklefs 2006). Ce type de modélisation est utile pour représenter l'évolution neutre d'un caractère, i.e l'absence de sélection. Mais il est également possible de dépendre les conséquences sur l'évolution d'un trait d'une sélection qui s'accroît, ou au contraire qui se relâche au cours du temps, par l'utilisation de modèles Late-Burst

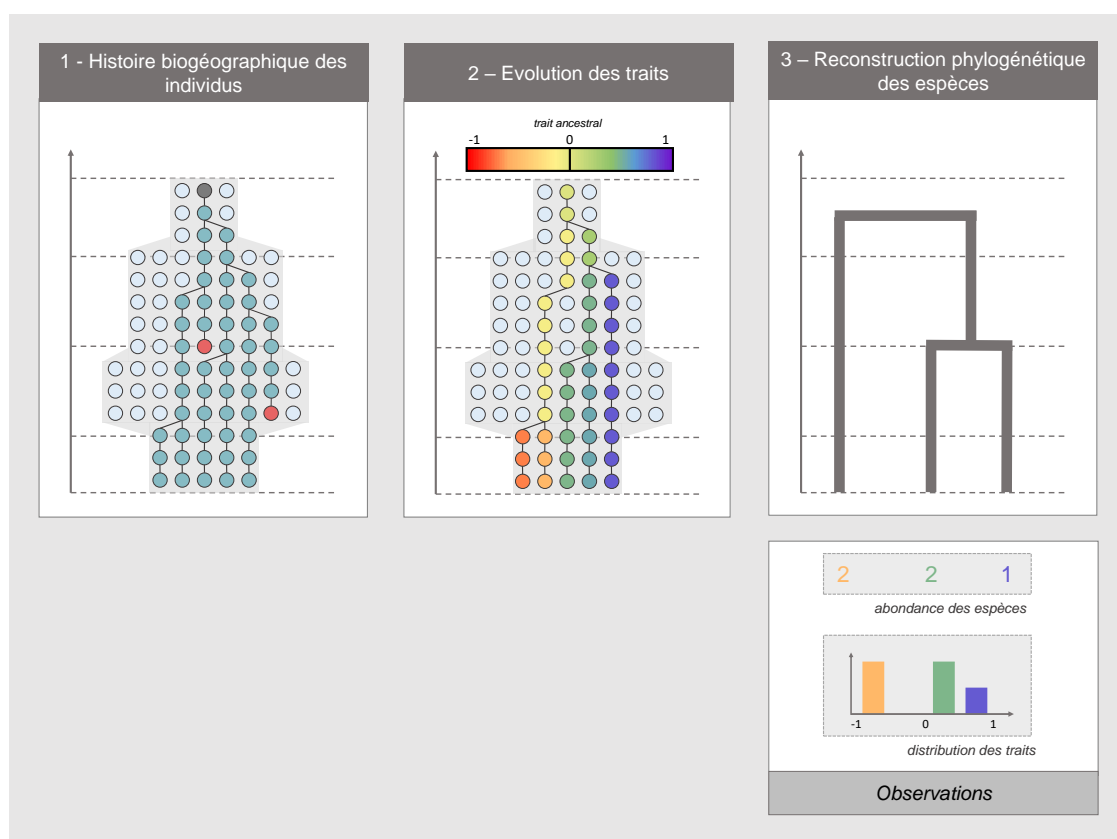


Figure D.1: Proposition d'un modèle de spéciation par différenciation fonctionnelle qui pourrait être implémenté dans *ecophylo*. L'ascendance commune des individus échantillonnés est reconstruite en utilisant la théorie du coalescent sans supposer à quelle espèce ils appartiennent jusqu'à ce qu'un ancêtre commun soit trouvé, en fonction de fluctuations environnementales (zones grises). Notons que les lignées aboutissant à des extinctions de lignées locales (cercles rouges) sont représentées sur le schéma mais ne sont pas explicitement représentées dans le processus de coalescence (ce qui permet de gagner du temps de calcul). Puis les traits de ces individus sont simulés, par exemple par un mouvement brownien. Basé sur la distance entre les traits des individus et en respectant la monophylie les branches de la généalogie sont fusionnées pour obtenir les relations phylogénétiques entre espèces. Il est ainsi possible d'obtenir des prédictions sur (i) l'abondance des espèces, (ii) leurs relations phylogénétiques et (iii) la distribution de leurs traits en fonction de changements de l'environnement dans le passé

(LB) ou Earl-Burst (EB), respectivement. Dans une vision classique de la radiation adaptative, un clade bénéficiant d'une opportunité écologique subit une spéciation et une évolution des traits précoce et rapide, suivie d'une réduction des taux de diversification lorsque l'espace écologique se sature (Simpson 1953). L'utilisation de mouvements browniens contraste toutefois avec la notion communément admise selon laquelle de nombreux traits biologiques évoluent à des rythmes différents le long des différentes branches de l'arbre du vivant. C'est pourquoi il peut être intéressant d'étudier comment des variations dans la vitesse et la direction de changements phénotypiques dans les phylogénies impactent les patrons de diversité fonctionnelles (Smaers, Mongle, and Kandler 2016 ; Blomberg, Rathnayake, and Moreau 2020).

Spéciation par différenciation fonctionnelle (SFD) Dans une volonté de représenter les conséquences de l'évolution des propriétés des espèces sur le processus de spéciation, on pourrait apporter les modifications suivantes au modèle de spéciation dans *ecophylo*. En s'inspirant de l'utilisation actuelle de mouvements browniens le long des branches de la phylogénie pour représenter l'évolution d'un trait pour différentes espèces - on pourrait envisager d'utiliser ces mêmes familles de modèles pour représenter la micro-évolution de caractères le long des branches de la généalogie pour différents individus (Figure D.1). Il serait donc possible, en théorie, de regrouper les individus apparentés et semblables ainsi obtenus en groupes monophylétiques, et de baser ainsi le processus de spéciation sur la fusion des branches de la généalogie reliant des individus apparentés et fonctionnellement proches (Higgs and Derrida 1992). Je vois deux avantages à cette méthode. La première est que l'utilisation de différentes familles de modèles d'évolution de trait permettrait de représenter les conséquences d'une sélection directionnelle et/ou dont les contraintes se relâchent, ou à l'inverse s'intensifient avec le temps, sur les phylogénies. La deuxième est qu'en conservant les valeurs de traits simulées pour chaque individu, avant la fusion des branches qui permet d'obtenir la phylogénie, il devient possible d'obtenir la composition fonctionnelle intra et inter spécifique des communautés simulées.

2 Le lien entre processus et patrons

2.1 Quelles approches en écologie et biogéographie ?

2.1.1 L'approche inductive

Il y a une dualité entre la réalité des patrons qu'on observe et les modèles utilisés pour se représenter les processus sous-jacents. Historiquement, l'écologie a souvent eu recours à une approche inductive pour appréhender ce lien entre modèle et observations ; le but étant la recherche de lois universelles et robustes à partir de relations entre variables observées. Par exemple on pourra observer une relation positive entre la richesse et la surface d'échantillonnage, ou bien observer une plus grande diversité d'espèces à l'équateur qu'aux pôles. Derrière cette généralisation d'observations de patrons, il n'y a pas de modèle mécaniste, mais plutôt une description de la réalité et une recherche de lois empiriques qui correspondent à des généralités observables. Selon certains philosophes comme David Hume et Francis Bacon (Malherbe 2001 ; Bacon, Vert, and Le Doeuff 1986), c'est l'approche qui serait la plus ancrée dans la réalité car basée sur l'identification des propriétés de l'objet d'étude,

et non sur l'imaginaire du chercheur.

2.1.2 L'approche déductive

L'approche inductive s'oppose à la démarche inverse, dite déductive, qui part de la représentation abstraite d'un système pour en déduire des propriétés. Par exemple, on peut construire un modèle mathématique avec un certain choix de variables et de paramètres, et utiliser ce modèle pour prédire les variations de grandeurs écologiques. Dans ce cas, on va déduire à partir de l'intuition ce que devrait être la réalité, et donc fonder notre rapport à la réalité par un raisonnement abstrait. L'approche déductive s'oppose à l'empirisme dans le sens où le lien entre processus et patrons est basé sur la représentation abstraite des processus. Ces deux approches constituent deux manières différentes d'appréhender le lien entre la réalité et les modèles, notamment pour examiner la causalité.

2.1.3 L'approche hypothético-déductive

L'approche la plus consensuelle développée en écologie moderne est un cas particulier de l'approche déductive, qu'on appelle hypothético-déductive. Cette démarche ne consiste plus à la formalisation d'un seul modèle associé à une prédiction, mais l'énonciation d'une séquence d'hypothèses alternatives pour lesquelles on formule un attendu. Cet attendu doit être confronté à la réalité par des protocoles expérimentaux et/ou d'échantillonnage, afin d'évaluer si celui-ci est conforme à la réalité. La logique de l'approche repose non plus sur une hypothèse unique, mais sur une succession d'hypothèses de plus en plus complexes. La première hypothèse, la plus simple, est appelée hypothèse nulle et correspond aux attendus en l'absence du processus d'intérêt. À l'inverse, l'hypothèse alternative postule l'implication du ou des processus que l'on cherche à mettre en évidence. Ainsi, plutôt que la validation d'une hypothèse, l'approche hypothético-déductive est conçue de manière à invalider une hypothèse - en l'occurrence l'hypothèse nulle - afin d'aller dans le sens de l'hypothèse alternative. Et donc la théorie sur laquelle est basé le modèle est soumise à l'expérience par un test d'hypothèse, jusqu'à sa falsification: tant qu'il n'a pas été démontré qu'une hypothèse est fautive, celle-ci peut être vraie. Un modèle simple est ainsi mis à l'épreuve pour voir jusqu'où il doit être complexifié par hypothèses alternatives successives, afin de s'approcher de la réalité.

2.2 Simulation pour l'inférence: une nouvelle approche?

2.2.1 Modélisation explicite et mécaniste des processus

Bien que non biaisée par l'imaginaire du chercheur, l'approche inductive est, à mon sens, limitante car elle ne propose pas d'explications des mécanismes élémentaires qui sont à l'origine de nos observations. Si on se donne comme objectif la recherche de relations de cause à effets entre des processus biologiques et les patrons de diversité observés, alors il est nécessaire de faire appel à des modèles ; ce qui sous-tend l'implication directe du modélisateur. Les méthodes corrélatives, bien

qu'utiles pour identifier des tendances dans les propriétés des communautés soumises à différents environnements, relèvent néanmoins d'une logique largement inductive et ne proposent aucune explication des processus à l'œuvre. De plus, ce type d'approche est particulièrement inefficace si plusieurs processus agissent simultanément et potentiellement en contradiction (Chave, Muller-Landau, and Levin 2002 ; Mayfield and Levine 2010). Parallèlement, les modèles nuls sont souvent utilisés en écologie des communautés, s'inscrivant de fait dans une logique hypothético-déductive. Cependant l'absence de processus d'intérêt (bien souvent des processus de niche) est représentée par la randomisation des propriétés des espèces (Bernard-Verdier et al. 2012) ou de leurs abondances (Gotelli 2000). Ces modèles n'explicitent donc à aucun moment la nature et la forme des processus de niche qu'ils cherchent à mettre en évidence, et ne reconnaissent pas non plus l'influence des dynamiques stochastiques neutres (Gotelli and McGill 2006 ; Bausman 2018). A l'inverse, la formalisation explicite du problème direct permet de représenter formellement des mécanismes élémentaires d'intérêt et leurs interactions (Gaucherel, Bérard, and Munoz 2011). Grâce aux simulations il est possible de reproduire le comportement des entités élémentaires d'un système et d'en observer les conséquences sur la structuration de ses propriétés macroscopiques. Non seulement il est possible de considérer comment s'influencent des processus opérants à différentes échelles spatiales et temporelles, mais également d'examiner comment des processus contraires structurent conjointement la biodiversité. Dans le cas des adventices de culture, nous mettons en évidence comment la diversité de ces communautés est influencée à la fois par les méthodes de gestion agricoles, qui contraignent l'installation d'espèces mal-adaptées, et en même temps par la stochasticité démographique et la migration.

2.2.2 L'intérêt du problème direct

L'apport des expériences virtuelles Comment savoir quels processus écologiques et, pire encore, évolutifs, opèrent dans les milieux naturels déjà en place? Outre la difficulté d'échelle des dispositifs expérimentaux, se pose également le problème de la contingence liée aux événements rares, couplée à des dynamiques transitoires. Cela dit, l'observation de la relation de cause à effet entre un processus et les patrons en résultant peut s'appréhender par des modèles qui simulent les mécanismes d'intérêt et qui relèvent du problème direct. Il y a deux avantages qui motivent la considération du problème direct pour la résolution du problème inverse. Premièrement, il est beaucoup plus simple de considérer la question du lien entre processus et patron sous l'angle du problème direct plutôt que du problème inverse. Mais surtout, si l'on espère conclure sur les paramètres qui influencent les patrons de diversité en milieux naturels, alors nous devons savoir comment ces patrons changent lorsque les paramètres changent. De cette façon, les effets confondants peuvent être identifiés en examinant virtuellement comment des patrons similaires émergent à partir de différents mécanismes. Cela permet d'identifier les signatures propres aux processus que l'on souhaite découvrir dans des données réelles. Ensuite, la comparaison des patrons de diversité observées, à ceux qui ont explicitement été simulés, permet de déduire lesquels des modèles mécanistes et de leurs paramètres sont susceptibles d'avoir influencé nos observations. Ce travail de thèse fait usage d'expériences virtuelles à plusieurs reprises. Principalement, l'Article 5 et l'Article 6 font usage de simulations pour explorer comment des statistiques de diversité spécifiques, phylogénétiques et fonctionnelles dépendent de paramètres éco-évolutifs les ayant générés. Ces études montrent dans quelle mesure des

processus d'intérêts - l'augmentation de la dérive dans des refuges de petite taille et des variations de vitesse d'évolution, respectivement - façonnent différentes facettes de la diversité. L'analyse comparative des statistiques choisies, au moyen de simulations intensives, pour différents modèles, sur de larges plages de valeurs de paramètres, peut aider à caractériser la façon dont les patrons de diversité sont affectées par les processus écologiques et/ou évolutifs et la façon dont nous choisissons de les simuler (Zurell et al. 2010). Une fois que nous savons si un ensemble de mesures de diversité porte ou non la signature distinctive d'un processus d'intérêt, celles-ci peuvent être utilisées pour résoudre le problème inverse.

L'importance des validations croisées Cependant, si les simulations ne diffèrent pas significativement entre les modèles, soit parce qu'elles ne portent pas la signature des processus simulés, soit parce que les distributions de paramètres a priori choisies aboutissent à des simulations trop similaires, la sélection d'un modèle et l'estimation de ses paramètres sera imprécise, au mieux, ou dans le pire des cas complètement aléatoire. Il est possible de vérifier si l'on se trouve dans de tels cas en générant des ensembles de données artificielles à partir des valeurs de paramètres simulées, puis d'évaluer ultérieurement la façon dont une méthode ABC donnée attribue les simulations au bon modèle (sélection de modèle), ou évalue les vraies valeurs de paramètres (estimation de paramètres) (Sunnåker et al. 2013). Outre la présentation du modèle éco-évolutif, l'Article 1 établit qu'il est possible de reconnaître correctement les simulations issues de différents modèles biogéographiques générés par *ecophylo*, mais aussi que les valeurs des paramètres de ces modèles peuvent être estimées avec précision à l'aide des méthodes ABC par des protocoles de validation croisées. Ces résultats justifient donc l'utilisation de ce modèle pour réaliser des inférences sur des données réelles observées, et donc de résoudre le problème inverse.

2.2.3 La résolution du problème inverse

La conditionnalité La difficulté est qu'une expérience virtuelle ne peut pas montrer le résultat de tous les modèles imaginables pour chaque combinaison de paramètres. Le choix des mécanismes et de la façon de les simuler incombe au modélisateur. La comparaison d'une observation à des simulations ne révélera jamais que la meilleure des options simulées, et non pas si les simulations et donc les choix des modèles sont pertinents. L'architecture de l'expérience virtuelle détermine donc explicitement les limites de validité et d'interprétation des résultats (Hickerson 2014). Cette question de la conditionnalité n'est pas exclusive aux approches basées sur la simulation. Je dirais même qu'elle est inhérente à toute démarche déductive qui, effectivement, fait intervenir l'imaginaire du modélisateur. Dans le cas où l'on définit un pool régional d'espèces, les hypothèses sur l'assemblage des communautés locales sont formulées - et de fait validées ou non - systématiquement par rapport à la définition a priori de ce pool. Dans le modèle d'assemblage des communautés d'adventices de culture (Article 3), le pool régional est défini comme l'ensemble des espèces d'adventice présentes dans la zone d'étude. Celui-ci est donc déjà basé sur la liste d'espèces, dont les préférences écologiques incluent les milieux ouverts et cultivés. Ainsi le filtre environnemental dont on estime les paramètres correspond à celui qui contraint l'installation des espèces dans les différents champs, et dont on a considéré le type de culture comme un bon proxy des pratiques agricoles en place

(et donc des contraintes abiotiques associées). À l'inverse, l'inférence des mêmes paramètres à partir d'un pool d'espèces limité à la composition des communautés adjacentes aurait mis l'accent sur le rôle des interactions biotiques à une échelle plus fine (Perronne et al. 2017), ou aurait pu atténuer l'importance du filtre environnemental lié au type de pratique si nous avions considéré ce facteur dans la définition du pool d'espèces. Par conséquent, lorsque nous tirons des conclusions sur l'assemblage des communautés, celles-ci ne peuvent être que conditionnelles au pool d'espèces que nous avons nous-mêmes, en tant que modélisateur, conçu.

Calcul Bayésien Approché et biais de confirmation La conditionnalité peut poser particulièrement problème avec des méthodes par Calcul Bayésien Approché (ABC), car celles-ci sont particulièrement susceptibles au biais de confirmation. L'ABC établit les valeurs de paramètres d'un modèle les plus plausibles comme étant celles ayant généré les simulations les plus proches des observations. Ainsi, ces méthodes ne donnent jamais tort au modélisateur: elles ne font que lui révéler la moins pire des options simulées. Par construction des méthodes bayésiennes, on introduit volontairement des a priori dans le modèle, alors que baser les a priori sur des connaissances indépendantes peut induire des conclusions trompeuses, soit en ignorant des résultats possibles qui peuvent correspondre aux données, soit en introduisant des informations erronées dans le modèle (Betini, Avgar, and Fryxell 2017). Par exemple, dans une logique purement déductive où le modélisateur ne considère qu'une seule option (qu'il peut lui-même avoir biaisée en introduisant des a priori erronés), alors celui-ci risque de confirmer son propre attendu fallacieux. En réalité, rien ne lui garantit que l'adéquation de son modèle aux données est parfaite, ni même bonne, mais simplement moins pire que les autres options simulées. Il est donc utile de savoir pourquoi un modèle est faux. La "non-adéquation" du modèle aux données peut nous indiquer quels aspects sont mal prédits, et nous aider ainsi à définir de manière itérative des modèles alternatifs plus pertinents. Parce que le design de l'expérience virtuelle conditionne l'interprétation des résultats, il est essentiel de s'affranchir, autant que possible, des biais cognitifs qui limitent notre capacité à considérer et tester des hypothèses alternatives. Le risque existe de concentrer notre attention sur un ensemble étroit de théories préconçues (Elliott and Brook 2007 ; Betini, Avgar, and Fryxell 2017) que les méthodes ABC ont le grand potentiel de confirmer aveuglément. Il est donc important d'envisager des modèles concurrents, qui nous permettent de concevoir des hypothèses multiples, de tester leur validité par comparaison avec les patrons observés, et enfin de les réévaluer pour affiner et améliorer successivement les choix de modélisation (Platt 1964 ; Betini, Avgar, and Fryxell 2017 ; Yanco et al. 2020). Cette approche itérative proposée dans l'Article 2 vise justement à éviter ce type de biais de confirmation en se rapprochant le plus possible d'une démarche hypothético-déductive, mais dans un cadre bayésien.

Limites de l'approche hypothético-déductive en biogéographie historique Cela dit, l'imaginaire du modélisateur peut également être une limite à la formulation d'hypothèses alternatives, puisque nous ne pourrions jamais simuler que ce que nous sommes capables de concevoir. Et quand bien même, il faut faire des choix parmi la multitude d'hypothèses alternatives à notre disposition, car la multiplication de modèles alternatifs peut diminuer la puissance statistique des tests visant à les discriminer (Burnham 1998). Par conséquent, l'inférence de changements environnementaux passés avec *ecophylo*, restera en grande partie déductive. Postulons l'absence de changements passés, dans ce cas l'hypothèse nulle est équivalente à celle du modèle de la théorie neutre, à

un modèle de spéciation près. Cependant, sans aucune information a priori, il y a une infinité d'hypothèses alternatives à lui opposer. Doit-on considérer des changements environnementaux s'étant hypothétiquement produits au cours de deux périodes, trois périodes, quatre périodes ou plus ? Quelles dates doit-on retenir? Doit-on considérer que la migration a, elle aussi, changé au cours du temps? Etc. Prendre en compte explicitement l'histoire, augmente considérablement la complexité du modèle en termes d'hypothèses alternatives, et donc de paramètres.

2.3 Le cas des dynamiques éco-évolutives: quels patrons de diversité pour l'inférence?

Est-ce seulement possible d'inférer des processus datant de millions d'années à partir d'un aperçu de la diversité? La difficulté vient du fait qu'il y a extrêmement peu de données fossiles, et que l'inférence de processus très anciens se base donc uniquement sur des données actuelles. L'expérience virtuelle permet de répondre à cette question. L'Article 5 examine comment l'augmentation d'habitats favorables, suite à la colonisation depuis un refuge, influence la diversité taxonomique, phylogénétique et fonctionnelle, en simulant explicitement l'expansion démographique des assemblages, l'émergence de nouvelles espèces, leurs relations phylogénétiques, et les valeurs des traits associés. Cette expérience virtuelle permet de mettre en évidence que différentes facettes de la diversité n'ont pas la même temporalité de réponse à des changements anciens.

2.3.1 Statistiques unidimensionnelles

La richesse spécifique et le nombre d'espèces rares reflètent l'influence de changements environnementaux passés lorsque ceux-ci sont suffisamment récents. Pour des changements trop anciens, le signal sur la diversité taxonomique est perdu, et correspond aux communautés les plus riches et avec le maximum d'espèces rares. En effet, une réduction d'habitat dans un refuge intensifie le phénomène de dérive, et donc entraîne une perte d'espèces - en particulier d'espèces rares. Cela dit, même dans le cas de changements environnementaux très anciens, les patrons actuels de diversité phylogénétique (PD, MPD, etc.) portent encore la signature de la perte par dérive de lignées passées. Ainsi, il conviendra d'utiliser l'information phylogénétique pour l'étude et l'inférence de processus éco-évolutifs anciens, alors que l'information taxonomique peut s'avérer suffisante pour des changements plus récents. A noter ici, que l'ancienneté d'un événement est relative au taux de mutation. Les espèces rares en particulier, sont un bon marqueur des fluctuations démographiques et donc peuvent être particulièrement utiles pour l'inférence. La composition d'une communauté en termes des propriétés biologiques, quant à elle, résulte de l'interaction entre la vitesse d'évolution de ces propriétés, et la distribution des événements de diversification dans la phylogénie (Tucker et al. 2018 ; Pearse et al. 2019). L'asymétrie phylogénétique peut être liée à l'irrégularité des événements de diversification dans les modèles évolutifs (Kirpatrick and Slatkin 1993 ; Blum and François 2006) ; soit lorsque ces événements sont distribués irrégulièrement entre la racine de l'arbre et les noeuds terminaux, soit lorsqu'ils se concentrent sur quelques lignées. L'utilisation d'une mesure de l'asymétrie racine-feuille (Lewitus and Morlon 2016) dans l'Article 5 a permis de mettre en

évidence l'augmentation de la variance des traits, comme une conséquence de l'effet combiné de l'intensification récente des événements de diversification, et de l'accélération de l'évolution des traits. Tandis que l'Article 6 montre que dans le cas où la majeure partie de l'évolution des traits se produit dans les branches profondes de l'arbre, la distribution des traits traduit l'apparition d'individus rares, qui témoignent de la présence d'anciennes lignées qui se sont très peu diversifiées depuis, et qui persistent dans les assemblages actuels. Cet effet peut s'apparenter à l'effet musée décrit dans Bartish et al. 2016, appliqué à la diversité fonctionnelle.

2.3.2 Statistiques multidimensionnelles

Spectre de fréquence d'espèces Le spectre de fréquence des espèces (SFS) est une extension multi-espèces des spectres de fréquence des allèles utilisés en génétique des populations (Fisher 1930). L'intérêt particulier du SFS est qu'il ne constitue pas une mesure agrégée de la diversité spécifique et fournit donc des informations plus complètes sur la variation de l'abondance des espèces dans un assemblage. Dans l'Article 5, celui-ci s'est avéré particulièrement utile pour détecter l'influence des expansions démographiques plus récentes, car il permet de détecter la signature spécifique portée par des classes d'abondance particulières, notamment celle des espèces rares.

Statistiques de diversité par période De même, il est possible de décomposer l'information phylogénétique afin de reconnaître l'influence d'événements historiques sur la diversification des lignées. L'Article 7 présente une nouvelle méthode ayant permis de partitionner la diversité phylogénétique des forêts tropicales humides, non pas entre les nœuds d'une phylogénie (Pavoine, Love, and Bonsall 2009), mais à travers des périodes évolutives définies, basées sur des événements marquants de l'histoire géologique d'une ou plusieurs régions. Cette nouvelle méthode permet d'examiner des changements majeurs dans la diversification et les voies de migration.

3 Conclusion

La prise en compte des processus éco-évolutifs nécessite d'intégrer des processus qui opèrent à des échelles spatiales et temporelles contrastées. Ce travail de thèse propose de nouvelles méthodes, qui intègrent la dynamique des individus dans une perspective écologique, et les dynamiques de diversification des espèces dans une perspective évolutive ; le tout dans un environnement fluctuant soumis aux aléas de l'histoire sur un long terme. Cependant, la volonté de prendre explicitement en compte la complexité des processus éco-évolutifs à de multiples échelles a pour inévitable conséquence la complexification des modèles que l'on utilise pour représenter ce type de processus. Sans nul doute que la prise en compte de mécanismes tels que la sélection, ou les interactions biotiques, alourdirait considérablement la complexité des modèles éco-évolutifs tels que développés dans ce travail. La complexité de tels modèles pose un problème pour l'inférence à partir de données réelles. Envisager le rôle de l'histoire mène à une infinité d'hypothèses alternatives, et donc de paramètres de modèles. Sans méthodes d'inférence robustes pour des modèles complexes, l'étude des dynamiques éco-évolutives se limitera à la considération du problème direct, jamais à la résolution du problème inverse.

Références

- Avice, J. C. (2009). “Phylogeography: retrospect and prospect”. In: *Journal of Biogeography* vol. 36, pp. 3–15.
- Bacon, F., Vert, F., and Le Doeuff, M. (1986). *Valerius Terminus of the Interpretation of Nature*. Méridiens Klincksieck.
- Bartish, I. V. et al. (2016). “Different habitats within a region contain evolutionary heritage from different epochs depending on the abiotic environment”. In: *Global Ecology and Biogeography*, n/a–n/a.
- Bausman, W. (2018). “Modeling: Neutral, Null, and Baseline”. In.
- Bernard-Verdier, M. et al. (2012). “Community assembly along a soil depth gradient: contrasting patterns of plant trait convergence and divergence in a Mediterranean rangeland”. In: *Journal of Ecology* vol. 100, pp. 1422–1433.
- Betini, G. S., Avgar, T., and Fryxell, J. M. (2017). “Why are we not evaluating multiple competing hypotheses in ecology and evolution?” In: *Royal Society Open Science* vol. 4.
- Blomberg, S. P., Rathnayake, S. I., and Moreau, C. M. (2020). “Beyond Brownian motion and the Ornstein-Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters”. In: *The American Naturalist* vol. 195, no. 2. ISBN: 0003-0147 Publisher: The University of Chicago Press Chicago, IL, pp. 145–165.
- Blum, M. G. and François, O. (2006). “Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance”. In: *Systematic Biology* vol. 55, no. 4. ISBN: 1076-836X Publisher: Society of Systematic Zoology, pp. 685–691.
- Burnham, K. P. (1998). “Model selection and multimodel inference”. In: *A practical information-theoretic approach*. Publisher: Springer.
- Chave, J., Muller-Landau, H. C., and Levin, S. A. (2002). “Comparing classical community models: theoretical consequences for patterns of diversity”. In: *The American Naturalist* vol. 159, no. 1, pp. 1–23.
- Dayan, T. and Simberloff, D. (2005). “Ecological and community-wide character displacement: the next generation”. In: *Ecology Letters* vol. 8, pp. 875–894.
- Elliott, L. P. and Brook, B. W. (2007). “Revisiting Chamberlin: Multiple Working Hypotheses for the 21st Century”. en. In: *BioScience* vol. 57, no. 7, pp. 608–614.
- Ewing, G. and Hermisson, J. (2010). “MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus”. In: *Bioinformatics* vol. 26, pp. 2064–2065.
- Fisher, R. A. (1930). “The Distribution of Gene Ratios for Rare Mutations”. In: *Proceedings of the Royal Society of Edinburgh* vol. 50, pp. 205–220.
- Garwood, R. J., Spencer, A. R., and Sutton, M. D. (2019). “RE voSim: Organism-level simulation of macro and microevolution”. In: *Palaeontology* vol. 62, no. 3. ISBN: 0031-0239 Publisher: Wiley Online Library, pp. 339–355.
- Gauchere, C., Bérard, S., and Munoz, F. (2011). “Equation or Algorithm: Differences and Choosing Between Them”. In: *Acta Biotheoretica* vol. 59, pp. 67–79.

- Gotelli, N. J. (2000). “Null model analysis of species co-occurrence patterns”. In: *Ecology* vol. 81, pp. 2606–2621.
- Gotelli, N. J. and McGill, B. J. (2006). “Null Versus Neutral Models: What’s The Difference?” en. In: *Ecography* vol. 29, no. 5, pp. 793–800.
- Hickerson, M. J. (2014). “All models are wrong”. In: *Molecular Ecology* vol. 23, pp. 2887–2889.
- Higgs, P. G. and Derrida, B. (1992). “Genetic-Distance and Species Formation in Evolving Populations”. In: *Journal of Molecular Evolution* vol. 35, pp. 454–465.
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton and Oxford: Princeton University Press.
- Hudson, R. R. (1983). “Properties of a neutral allele model with intragenic recombination”. In: *Theoretical population biology* vol. 23, no. 2. ISBN: 0040-5809 Publisher: Elsevier, pp. 183–201.
- Karhunen, M. et al. (2013). “DRIFTSEL: an R package for detecting signals of natural selection in quantitative traits”. In: *Molecular Ecology Resources* vol. 13, no. 4, pp. 746–754.
- Kingman, J. F. C. (1982). “On the genealogy of large populations”. In: *Journal of Applied Probability*, pp. 27–43.
- Kirxpatrick, M. and Slatkin, M. (1993). “Searching for evolutionary patterns in the shape of a phylogenetic tree”. In: *Evolution* vol. 47, no. 4. ISBN: 0014-3820 Publisher: Wiley Online Library, pp. 1171–1181.
- Lewitus, E. and Morlon, H. (2016). “Characterizing and comparing phylogenies from their Laplacian spectrum”. In: *Systematic biology* vol. 65, no. 3. ISBN: 1063-5157 Publisher: Oxford University Press, pp. 495–507.
- Malherbe, M. (2001). *La philosophie empiriste de David Hume*. Vrin.
- Manceau, M., Lambert, A., and Morlon, H. (2015). “Phylogenies support out-of-equilibrium models of biodiversity”. In: *Ecology Letters* vol. 18, pp. 347–356.
- Manceau, M., Lambert, A., and Morlon, H. (2016). “A Unifying Comparative Phylogenetic Framework Including Traits Coevolving Across Interacting Lineages”. In: *Systematic Biology* vol. 66, no. 4, pp. 551–568.
- Mayfield, M. M. and Levine, J. M. (2010). “Opposing effects of competitive exclusion on the phylogenetic structure of communities”. In: *Ecology Letters* vol. 13, pp. 1085–1093.
- Morlon, H. (2014). “Phylogenetic approaches for studying diversification”. en. In: *Ecology Letters* vol. 17, no. 4. Ed. by Mooers, A., pp. 508–525.
- Ovaskainen, O. et al. (2011). “A New Method to Uncover Signatures of Divergent and Stabilizing Selection in Quantitative Traits”. en. In: *Genetics* vol. 189, no. 2, pp. 621–632.
- Overcast, I., Emerson, B. C., and Hickerson, M. J. (2019). “An integrated model of population genetics and community ecology”. en. In: *Journal of Biogeography* vol. 46, no. 4, pp. 816–829.
- Overcast, I., Ruffley, M., et al. (2021). “A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities”. In: *Molecular ecology resources* vol. 21, no. 8. ISBN: 1755-098X Publisher: Wiley Online Library, pp. 2782–2800.
- Pavoine, S., Love, M., and Bonsall, M. (2009). “Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically-structured species assemblages: application to rockfish (genus: *Sebastes*) in the Southern California Bight”. In: *Ecology Letters* vol. 12, pp. 898–908.

- Pearse, W. D. et al. (2019). “The interaction of phylogeny and community structure: Linking the community composition and trait evolution of clades”. In: *Global Ecology and Biogeography* vol. 28, no. 10, pp. 1499–1511.
- Pelletier, F., Garant, D., and Hendry, A. P. (2009). “Eco-evolutionary dynamics”. In: Issue: 1523 Pages: 1483-1489 Publication Title: Philosophical Transactions of the Royal Society B: Biological Sciences.
- Perronne, R. et al. (2017). “How to design trait-based analyses of community assembly mechanisms: insights and guidelines from a literature review”. In: *Perspectives in Plant Ecology, Evolution and Systematics* vol. 25, pp. 29–44.
- Pigot, A. L. et al. (2010). “The shape and temporal dynamics of phylogenetic trees arising from geographic speciation”. In: *Systematic biology* vol. 59, no. 6. ISBN: 1076-836X Publisher: Oxford University Press, pp. 660–673.
- Platt, J. R. (1964). “Strong inference”. In: *science* vol. 146, no. 3642. ISBN: 0036-8075 Publisher: JSTOR, pp. 347–353.
- Prates, I. et al. (2016). “Inferring responses to climate dynamics from historical demography in neotropical forest lizards”. In: *Proceedings of the National Academy of Sciences* vol. 113, no. 29, p. 7978.
- Rangel, T. F. et al. (2018). “Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves”. en. In: *Science* vol. 361, no. 6399, eaar5452.
- Ricklefs, R. (2006). “Time, species, and the generation of trait variance in clades”. In: *Systematic Biology* vol. 55, no. 1. ISBN: 1076-836X Publisher: Society of Systematic Zoology, pp. 151–159.
- Rosindell, J. et al. (2010). “Protracted speciation revitalizes the neutral theory of biodiversity”. In: *Ecology Letters* vol. 13, pp. 716–727.
- Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.
- Simpson, G. G. (1953). *The Major Features of Evolution*. Columbia University Press.
- Smaers, J. B., Mongle, C. S., and Kandler, A. (2016). “A multiple variance Brownian motion framework for estimating variable rates and inferring ancestral states”. In: *Biological Journal of the Linnean Society* vol. 118, no. 1. ISBN: 0024-4066 Publisher: Oxford University Press, pp. 78–94.
- Stroud, J. T. and Losos, J. B. (2016). “Ecological Opportunity and Adaptive Radiation”. en. In: *Annual Review of Ecology, Evolution, and Systematics* vol. 47, no. 1, pp. 507–532.
- Sunnåker, M. et al. (2013). “Approximate Bayesian Computation”. In: *PLOS Computational Biology* vol. 9, no. 1, e1002803.
- Tello, J. S. and Stevens, R. D. (2012). “Can stochastic geographical evolution re-create macroecological richness–environment correlations?” In: *Global Ecology and Biogeography* vol. 21, no. 2. ISBN: 1466-822X Publisher: Wiley Online Library, pp. 212–223.
- Thompson, S. E., Chisholm, R. A., and Rosindell, J. (2020). “pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity”. In: *Methods in Ecology and Evolution* vol. 11, no. 10. ISBN: 2041-210X Publisher: Wiley Online Library, pp. 1237–1246.
- Tucker, C. M. et al. (2018). “On the relationship between phylogenetic diversity and trait diversity”. en. In: *Ecology* vol. 99, no. 6, pp. 1473–1479.
- Vellend, M. (2010). “Conceptual synthesis in community ecology”. In: *The Quarterly review of biology* vol. 85, no. 2, pp. 183–206.

- Yanco, S. W. et al. (2020). “A modern method of multiple working hypotheses to improve inference in ecology”. en. In: *Royal Society Open Science* vol. 7, no. 6, p. 200231.
- Zurell, D. et al. (2010). “The virtual ecologist approach: simulating data and observers”. en. In: *Oikos* vol. 119, no. 4, pp. 622–635.

Annexes

Encouraging undergraduate engagement in large lecture groups through Think-Pair-Share and peer-learning

Elizabeth Barthélemy¹

¹LIPhy, Université Grenoble Alpes, Grenoble, France

Active learner participation is often associated with beneficial learning outcomes for students at many levels. A challenge for educators is thus to encourage in class participation in lectures with large groups of undergraduates. This study investigates the perceived benefit of two types of in class activities meant to promote students' participation and engagement toward regular coursework uptake by undergraduates.

Undergraduates enrolled in the international biology program at Université Grenoble Alpes (UGA) were proposed two different activities aimed at promoting in class engagement during lectures. The first aimed at stimulating regular coursework uptake by students, by asking them to compare notes and answer questions based on Think-Pair-Share (TPS) methods. The second was implemented in response to online lecture formats brought upon by the COVID-19 pandemic, to promote interactions between students and with the lecturer. Students were invited to prepare a small tidbit of information to interrupt the lecture through peer-learning.

At the end of the semester these same students were invited to complete a survey canvassing their general interest and the perceived learning benefit they found to these activities. These were repeated over 2 to 3 consecutive years of teaching and cohorts of students between January 2020 and April 2022.

Introduction

Active learner participation and engagement tend to favor students' understanding, learning, enjoyment and appreciation of what they have learned. From there a major challenge for educators is to encourage their students to actively engage in the learning process [1].

Lectures in large groups have been criticized as these may fail to engage students as these favor a passive learning experience [2], [3]. However, because of their

low cost and high acceptability by both students and faculty, lectures in large groups are likely to remain a popular element of higher education. University students often consider lectures both an essential component of their journey through higher education and highly beneficial to their interactions with their peers and their lecturers [1]. Hence, a challenge for lecturers is to promote active learning by centering students in the learning process during lectures through innovative methods [4], [5]. Figure 1 classifies approaches of in-class participation centering students depending on the type of target knowledge [6].

One of the objectives in teaching my first undergraduate lecture was to stimulate regular coursework uptake by students. Typically, this can be obtained by having the lecturer inquire about a previous lecture or about a reading, for instance at the beginning of a lecture. However, this approach while useful for inviting learners to recall information does not usually fuel higher level thinking and rarely promote students to focus [6]. In the case of Frank Lyman's Think-Pair-Share (TPS) strategy in which questions are considered by pairs of students comparing notes and sharing ideas about the answer, offers a more structured form of discussion and thus increase the likelihood of productive student engagement with the content of the lecture (Figure 1).

Later, online lecture formats brought upon by the COVID-19 pandemic, made it even more crucial – *and challenging* – to engage students in lectures. Students report “negative classmates”, “personal worries” and “non-conductive physical settings” among the factors most deterrent to in class participation [7] – which we

Glossary

Peer-teaching: a broad collection of educational activities whereby students learn from and with each other in both formal and informal ways'

Think-Pair-Share (TPS): a collaborative learning strategy where students work together to solve a problem or answer a question about an assigned reading

Corresponding author: Barthélemy, E. (elizabeth.barthelemy@univ-grenoble-alpes.fr)

Available online 1 May 2022

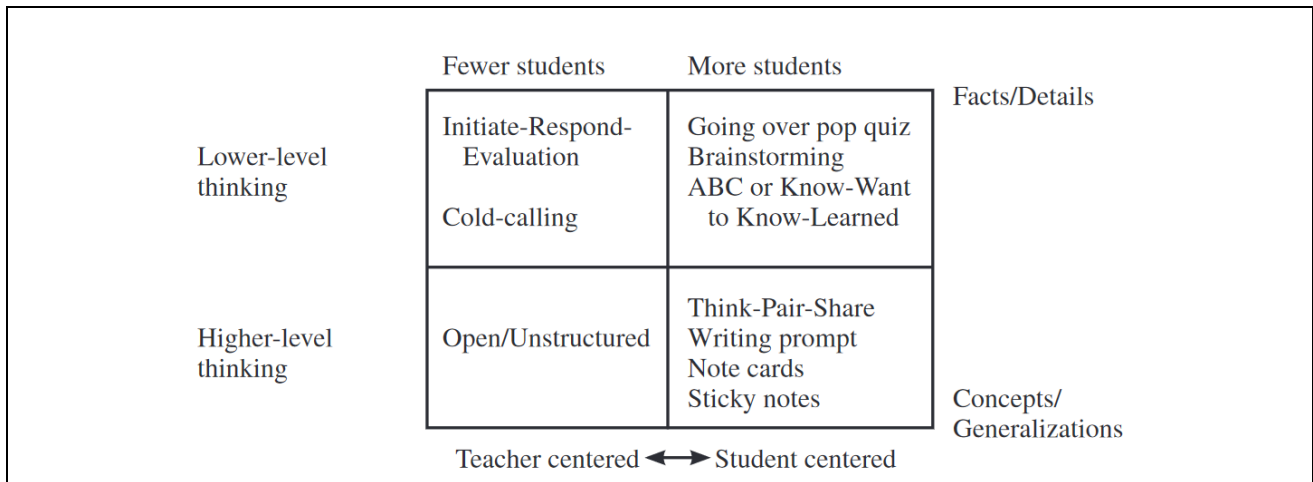


Figure 1. Approaches to classroom participation [6]

can assume affected learners during the pandemic. From there, an additional objective was to promote positive student interactions with the lecturer but also between students, in spite of a context which discouraged students from directly interacting with their peers. Bearing this in mind, peer-teaching, by having learners directly involved in teaching part of the course content seemed like a promising way to promote student engagement, in an online context.

From here I describe my attempts at promoting student engagement, both offline and online, using approaches based on TPS and peer-teaching. Later I investigate the perceived benefit of these methods by students through surveys.

Methods

Setting and Data Collection

The study was conducted at the Université Grenoble Alpes, on a cohort of first year undergraduate students enrolled in the international biology program in 2020, 2021 and 2022. I was in charge of delivering a 9 lecture course in introductory plant biology. The lectures were taught in English as a part of the international bachelor program held at the university. In 2020, lectures were attended by 60 students from both the biology and biochemistry programs, on campus from January 2020 to March 2020 (n=60). The final lecture was cancelled due to the COVID-19 pandemic. The following year, all 9 lectures had to be delivered online with the regular participation of 45 students on Zoom again from the biology and biochemistry programs, from January 2021 to April 2021 (n = 45). In 2022, on site courses resumed yet, only the biology students were required to partake

in the plant biology lectures, amounting to 21 students (n = 21).

Regular learning of lecture content through TPL

Educators often expect university students to be autonomous in their regular studying of lecture content. Most of the time we only provide limited opportunities for students to verify whether their studying is efficient (most often at the time of midterms and final exams). An initial objective during my first experience of teaching a large first year undergraduate group was to ease student's transition from secondary to higher education. During my introduction to plant biology lectures, I therefore invited students to take a moment to both individually and collectively study the material from the previous lecture, at the beginning of every class. First, students were asked to silently read their notes from the previous lecture for 5 minutes, and complete them by comparing notes with their neighbors if need be. At this point I encouraged students to call me over if they had noted something they could no longer make sense of and helped them make corrections. Then, I would display a list of roughly 10 questions pertaining to the essential

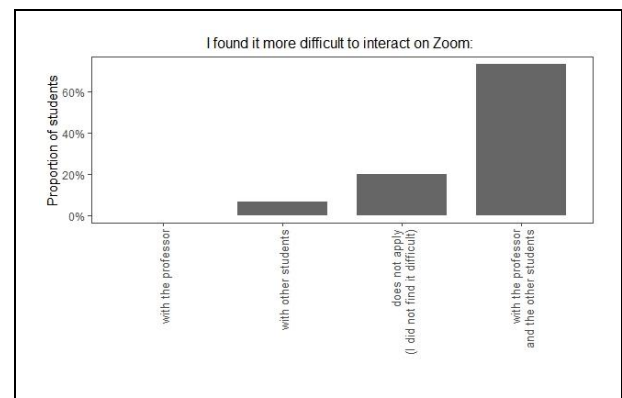


Figure 2. Proportion of students with difficulties interacting during lectures on Zoom in 2021 (n = 45)

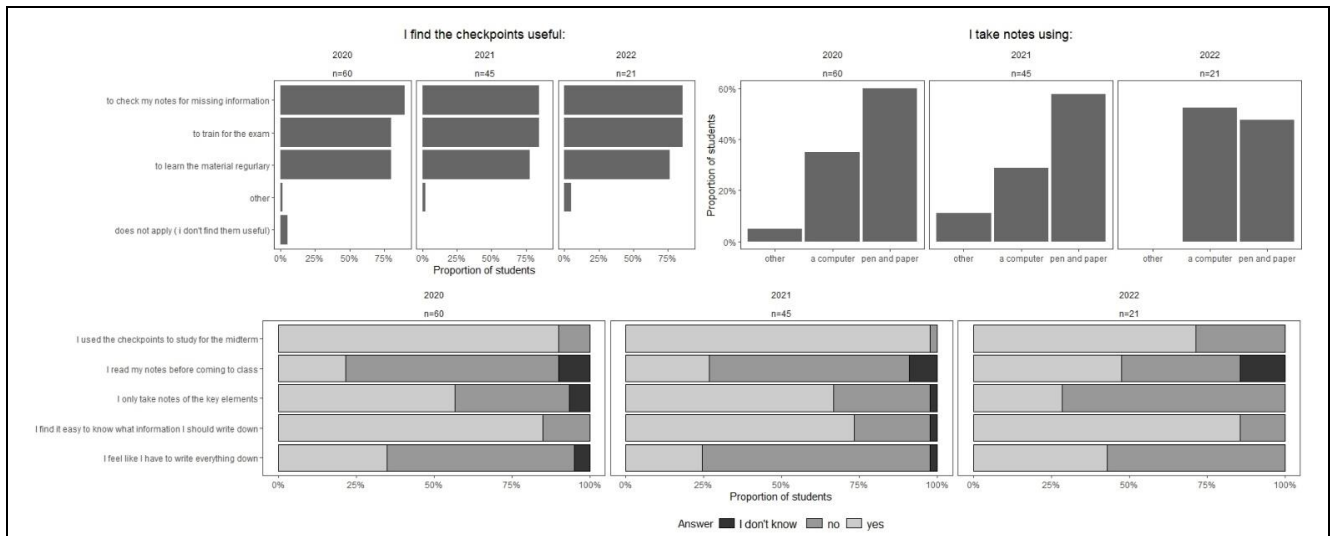


Figure 3. Students' self-reported note-taking habits and assessment of their appreciation of the checkpoints in 2020, 2021 and 2022. In 2021, lectures and checkpoints were dispensed online via Zoom.

points of the previous lecture (see Figure S1). I define essential or key elements of the lecture as fundamental information or overarching concepts that students need to assimilate in order to understand more complex notions. Finally, students had 5 to 10 minutes to answer the questions collectively, by contributing either individually or in pairs to the answers which I would use to synthesize the previous lectures. This activity which I further refer to as “checkpoints” was undertaken at the begging of every lecture, in 2020 and 2022.

In 2021, the entirety of the course was taught online. Thus small changes were made to the checkpoints to accommodate an online format. The first phase in which students were asked to read their notes were no longer conducted in pairs but individually – although it must be reported that some students were roommates and thus could partake in this activity together. Secondly, questions were no longer open-ended but multiple choice so as to allow direct participation on Zoom.

Offline and online peer-teaching

Brought by the COVID-19 pandemic, online classes via Zoom considerably decreased student opportunities to interact with lecturers and/or each other in a large group lecture context. In 2021, over 60% of undergraduates taking the plant biology course reported difficulties in interacting with both lecturer and other students (Figure 2). As a lecturer, I find myself relying on visual cues from students in the audience to gauge whether to spend more time on a given information. These cues disappeared, as students vanished behind blank computer screens. When I disclosed these difficulties with my students, they

reported that their limited access to a stable internet connection hindered them turning on their camera (for those who even possessed one). As for opportunities for them to interact with each other, these were also very limited by the fact that most of them were locked-down alone (self-reported information gathered by after-class discussions).

In response to these difficulties, I asked students to research any information they found interesting about one of the plants from the lecture either individually or in pairs. The instructions went as follows:

“Prepare a 1 min presentation about an assigned topic. You will present your findings during class. Feel free to include fun facts or anything you found interesting: the objective is to provide fun conversation starters about random facts about plants, so you and your friends can sound smart at a dinner party”

Students had access to a table with a list of plant related topics to choose from. Then, during lectures I invited them to interrupt me (and eventually turn on their cameras) to present what they had found, which was open to questions from the rest of the class.

In 2022, lectures were on campus. Yet I decided to continue the “plant fun facts” as we had called them. In this case, students interrupted the lecture but were invited to stand up to share what they had found with the rest of the lecture hall.

Survey instrument

A survey was designed to inquire about students note-taking habits and to assess their appreciation of the checkpoints (if and how they had found them useful). Students were asked to take the survey after the midterm but before the final exam, in 2020, 2021 and 2022. As of 2021, this survey was completed to include questions about student appreciation of the “plant fun facts”. The complete survey is provided in the Supplementary Materials (S2 and S3).

Results

Checkpoints and note taking

Most undergraduate students reported not reading their notes before coming to class in 2020 and 2021, while half of students did so in 2022. A majority of students declared no difficulties in identifying the information that they should note down. Results varied between years but mostly less than half of students felt the need to write everything down (Figure 3).

Overall, few undergraduate students in plant biology did not find the checkpoints useful. The most cited use they found to the checkpoints was to “check their notes for missing information”, followed closely by “studying for the exam”. Over half of students did in fact use these to study for the midterm (Figure 3). Amongst the “other” reasons for enjoying the checkpoints, students cited “identifying what information was important”.

Plant fun facts

A majority of students enjoyed researching and listening to others present their fun facts and none of them did not find them useful (Figure 4). However, whether they enjoyed them more coming from lecturer or other students greatly varied. More students reported learning more by doing their own research in 2021 – when classes were being held online rather than on campus.

The most reported use to the fun facts were to increase their “general culture” and to “illustrate the course”, both in 2021 and 2022. Yet, more students found they the fun facts “changed the pace of the lecture” in 2022 when these were conducted on campus rather than online. Few students found that the fun facts helped them interact with other students yet more did in 2021 – when classes were being held online rather than on campus (Figure 4).

Discussion

Using approaches based on TPS and peer-teaching have attempted to promote student engagement in a large lecture group both during online and offline classes. My

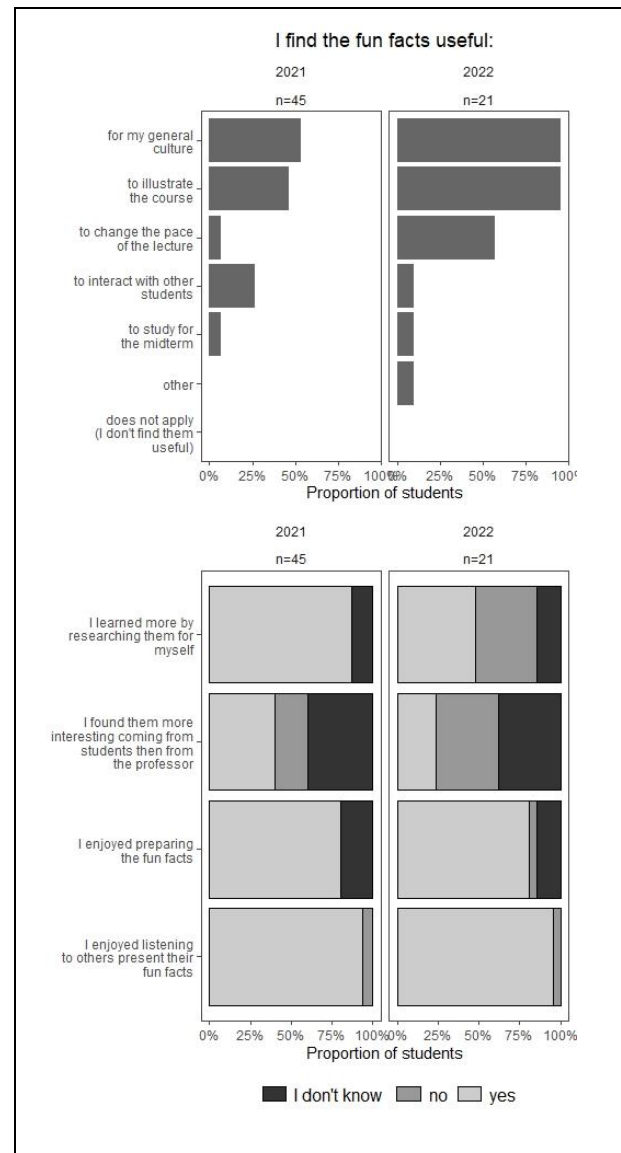


Figure 4. Students' self-reported assessment of their appreciation of the plant fun facts in 2021 and 2022. In 2021, lectures and fun facts were dispensed online via Zoom.

results show that overall students enjoyed these approaches. Checkpoints showed to be an effective method to engage students in learning the coursework more regularly and as a tool for students to evaluate how they assimilate information whether through their note taking or in preparation for an examination. There was little difference in the perceived appreciation of the checkpoints between in person and online contexts despite the latter removing the possibility to compare notes and discussing the questions with neighbors. Thus, I am tempted to believe that the perceived benefit by

undergraduates was more of a personal nature rather than something to benefit the group.

Fun facts, while not having shown to improve student interactions, at least were enjoyable to most students. These also seem to have different perceived benefits whether lectures were held online or in person. In the case of online lectures, more respondents found that fun facts benefited their interactions with their peers, and how well they learned the information. Yet in they found that the pace of the lecture was changed by the interruptions provided by the fun facts in an offline context.

This may be due to the change in the setting of the lecture. Online, students may have been more craving for interactions, while in person the fun facts did not offer any added benefit. However, in 2022 the cutback in the number of students taking the course (from roughly 50 to 20), removed the large lecture context and thus may have naturally promoted student interactions.

Acknowledgements

I wish to thank Aline Cisse for our many constructive conversations, her remarkable insight into teaching and for her invaluable friendship.

About the author

Elizabeth Barthélemy's interests center around biogeography and eco-evolutionary modelling, more specifically on the influence of past environmental fluctuations on the current diversity of tropical rainforest trees. This work represents a component of her PhD work at Université Grenoble Alpes as part of the label RES.

References

- [1] V. Gysbers, J. Johnston, D. Hancock, et G. Denyer, « Why do students still bother coming to lectures, when everything is available online? », *Int. J. Innov. Sci. Math. Educ.*, vol. 19, n° 2, 2011.
- [2] G. Brown et M. Manogue, « AMEE Medical Education Guide No. 22: Refreshing lecturing: a guide for lecturers », *Med. Teach.*, vol. 23, n° 3, p. 231- 244, 2001.
- [3] R. M. Harden et J. M. Laidlaw, *Essential skills for a medical teacher: an introduction to teaching and learning in medicine*. Elsevier Health Sciences, 2020.
- [4] J. D. Walker, S. H. Cotner, P. M. Baepler, et M. D. Decker, « A delicate balance: integrating active learning into a large lecture course », *CBE—Life Sci. Educ.*, vol. 7, n° 4, p. 361- 367, 2008.
- [5] D. Allen et K. Tanner, « Infusing active learning into the large-enrollment biology class: seven strategies, from the simple to complex », *Cell Biol. Educ.*, vol. 4, n° 4, p. 262- 268, 2005.
- [6] R. C. Jones, « The " why" of class participation: A question worth asking », *Coll. Teach.*, vol. 56, n° 1, p. 59- 63, 2008.
- [7] S. M. Mustapha, N. S. N. A. Rahman, et M. Md.Yunus, « Factors influencing classroom participation: a case study of Malaysian undergraduate students », *Procedia - Soc. Behav. Sci.*, vol. 9, p. 1079- 1084, 2010, doi: 10.1016/j.sbspro.2010.12.289.

Supplementary materials

Checkpoint #3

You have 10mins to look
over your notes from the
last lecture

Checkpoint #3

- 1- What is positive phototropism? Negative geotropism?
- 2- Name 2 examples of adaptive variations of aerial stems and give their functions.
- 3- What is secondary growth? Which group of angiosperms is mainly concerned?
- 4- What are the meristems responsible for secondary growth? Where are they located?
- 5- What tissues arise from each of these meristems? What are their functions?
- 6- Which tissue is primarily responsible for the stem's secondary growth?
- 7- What are lenticels? How are they formed?
- 8- Why is it possible to tell a tree's age by counting the rings in its trunk?
- 9- What is the duramen? What is it made from? What is its function?
- 10- What is the difference between monocot and dicot secondary growth?

Figure S1. Example of checkpoint questions

In the context of my PhD I am participating in a training curriculum to improve my teaching skills and would very much appreciate your feedback about the plant bio course (especially the regular checkpoints and the plant fun facts). Your answers are **anonymous** and for my benefit alone.

ABOUT THE CHECKPOINTS

I take my notes using:

A good old pen and paper

A computer / tablet

Other: _____

	YES*	NO*
I read my notes before coming to class	<input type="checkbox"/>	<input type="checkbox"/>
I find it easy to know what information to write down	<input type="checkbox"/>	<input type="checkbox"/>
I only take notes of the key elements	<input type="checkbox"/>	<input type="checkbox"/>
I feel like I should write everything down	<input type="checkbox"/>	<input type="checkbox"/>
I used the checkpoints for study for the midterm	<input type="checkbox"/>	<input type="checkbox"/>

**No answers = I don't know*

I found the checkpoints useful:

To check my notes for missing information

To learn the material regularly

To train for the exam

Does not apply (I don't find them useful)

Other reason(s): _____

Figure S2. Survey designed to assess students' appreciation of the checkpoints

ABOUT THE FUN FACTS

	YES*	NO*
I enjoyed preparing the fun facts	<input type="checkbox"/>	<input type="checkbox"/>
I enjoyed listening to other present their fun facts	<input type="checkbox"/>	<input type="checkbox"/>
I found them more interesting coming from students then from the professor	<input type="checkbox"/>	<input type="checkbox"/>
I learned more by researching them for myself	<input type="checkbox"/>	<input type="checkbox"/>

I found the fun facts useful:

<input type="checkbox"/> To study for the exam	<input type="checkbox"/> To interact with other students
<input type="checkbox"/> To illustrate the course	<input type="checkbox"/> Does not apply (I don't find them useful)
<input type="checkbox"/> For my general culture	<input type="checkbox"/> Other reasons:
<input type="checkbox"/> To change the pace of the lecture

OTHER

If you want to give me any more feedback, I would very much enjoy it! Let me know if you have any ideas to improve next year's students learning experience (I can't do anything about the content of the course but I can help make it more fun and easy to learn!) Even if its negative feedback, we are all here to learn, and so am I so as long as you keep it civil I won't be offended :-)

Figure S3: Survey designed to asses students' appreciation of the plant funfacts

RESUME

L'occurrence locale et la fréquence régionale des espèces dépendent de processus écologiques fondamentaux tels que les processus de niche, la migration, et la stochasticité démographique, mais aussi de la diversification des espèces et des changements de leurs caractères au cours du temps. De plus, la dynamique d'assemblage dans un environnement fini et hétérogène (perspective écologique) affecte la dynamique de spéciation et d'extinction sur le long terme (perspective évolutive), en fonction des changements environnementaux. Cette combinaison complexe de processus opérant à différentes échelles spatiales et temporelles implique d'intégrer les processus écologiques, biogéographiques et évolutifs dans de nouveaux modèles éco-évolutifs, pour une meilleure compréhension de la dynamique de la biodiversité. La thèse propose une nouvelle approche de modélisation de telles dynamiques, basées sur la théorie du coalescent, pour explicitement représenter, sous la forme d'une généalogie, l'effet d'un environnement fluctuant sur la démographie des individus sur le long terme. Le couplage entre démographie des individus et diversification des espèces est permis par un modèle de spéciation prolongée, et permet d'obtenir des prédictions sur différentes facettes de la biodiversité actuelle en fonction de changements anciens. L'utilisation de simulations à partir de ce modèle permet d'évaluer la signature sur le long terme de processus historiques sur la biodiversité actuelle. Dans une perspective où les communautés d'espèces ne sont pas à l'équilibre avec les conditions environnementales actuelles, mais restent influencées par des événements historiques anciens, du fait de dynamiques internes lentes, l'émergence de dynamiques transitoires permet de détecter l'influence de tels événements. De même, l'utilisation d'une approche par simulation explicite des processus permet d'inférer par Calcul Bayésien Approché (ABC), la contribution relative de processus combinés, tels que les processus de niche et les processus neutres, à partir de l'observation des patrons de diversité. Ainsi, la thèse propose des formalisations originales de la dynamique écologique et éco-évolutive de la biodiversité, et clarifie comment les processus fondamentaux peuvent être représentés et inférés au moyen de modèles dynamiques.

ABSTRACT

Both the local occurrence and regional frequencies of species depend on basic ecological processes such as niche dynamics, migration, and demographic stochasticity, as well as on the diversification of species and on changes in their traits over time. In addition, assembly dynamics in a finite and heterogeneous environment (ecological perspective) affect speciation and extinction dynamics in the long-term (evolutionary perspective), depending on environmental fluctuations. This complex combination of processes operating at different spatial and temporal scales implies that the ecological, biogeographical and evolutionary processes be integrated in new eco-evolutionary models, to better understand biodiversity dynamics. This thesis proposes a new approach for modeling such dynamics, based on coalescent theory, which explicitly represents, in the form of a genealogy, the effect of a fluctuating environment on the demography of individuals over time. The coupling between the demography of individuals and the diversification of species is achieved by a model of protracted speciation, and yields predictions regarding different components of current diversity patterns depending on past changes. The use of simulations from this model can be used to assess the long-term signature of historical events on current biodiversity. In a perspective where communities of species are not in equilibrium with current environmental conditions, but continue to be impacted by ancient historical events, as a result of slow internal dynamics, the emergence of transient dynamics will reveal the influence of such events. Similarly, the use of an explicit process-based simulation approach can be applied to infer by Approximate Bayesian Computation (ABC), the relative contribution of compound processes, such as niche-based and neutral processes, from the observation of diversity patterns. Thus, the thesis proposes new ways to formalize ecological and eco-evolutionary biodiversity dynamics, and clarifies how fundamental processes can be depicted and inferred using dynamic models.

