



HAL
open science

En dehors de la norme : déviation de l'optimalité écologique et originalité fonctionnelle

Matthias Grenié

► **To cite this version:**

Matthias Grenié. En dehors de la norme : déviation de l'optimalité écologique et originalité fonctionnelle. Biodiversité et Ecologie. Université Montpellier, 2020. Français. NNT : 2020MONTG027 . tel-03157339

HAL Id: tel-03157339

<https://theses.hal.science/tel-03157339v1>

Submitted on 1 Jul 2021

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 DE MONTPELLIER

En Écologie, Évolution, Ressources Génétiques, Paléobiologie

École Doctorale GAIA (n°584)

Unité de recherche UMR 5175 – Centre d'Écologie Fonctionnelle et Évolutive

En dehors de la norme :
déviation de l'optimalité écologique et originalité fonctionnelle

Présentée par Matthias GRENIÉ

Le 28 septembre 2020

Sous la direction de François MUNOZ
et Cyrille VIOLLE

Devant le jury composé de

Sébastien BROSSE, Professeur Université Paul Sabatier Toulouse III, UMR 5174 EDB, Toulouse

Stéphane DRAY, Directeur de Recherche CNRS, UMR 5558 LBBE, Lyon

Emmanuelle PORCHER, Professeure Muséum National d'Histoire Naturelle, UMR 7204 CESCO, Paris

David MOUILLOT, Professeur Université de Montpellier, UMR Marbec, Montpellier

François MUNOZ, Professeur Université de Grenoble Alpes, UMR 5553 LECA Grenoble

Cyrille VIOLLE, Directeur de Recherche CNRS, UMR 5178 CEFE, Montpellier

Rapporteur

Rapporteur

Examinatrice

Examineur

Directeur

Co-Directeur



UNIVERSITÉ
DE MONTPELLIER

Sommaire

	Page
REMERCIEMENTS	I
CONTRIBUTIONS SCIENTIFIQUES	4
1 Publications	4
2 Paquets R	5
INTRODUCTION	6
1 Des règles d'assemblages à la théorie des filtres	7
2 L'approche fonctionnelle des communautés	9
3 Une seule valeur optimale de trait : le paradigme de l'optimum fonctionnel	11
4 Des espèces qui échappent à l'optimalité fonctionnelle : les espèces rares	12
5 Une approche fonctionnelle de la rareté : la rareté fonctionnelle	14
6 Espèces originales et conservation de la biodiversité	15
7 Modéliser les communautés pour mieux les comprendre	17
8 Synthèse	19
CHAPITRE I : MODÉLISER LES COMMUNAUTÉS POUR IDENTIFIER LES PROCESSUS QUI POURRAIENT EXPLIQUER UNE DÉVIATION À L'OPTIMALITÉ	20
1 L'interaction de processus neutres et non-neutres peut causer une déviation de l'optimalité	22
2 La contribution des traits à différents processus peut causer des déviations de l'optimum fonctionnel	33

CHAPITRE 2 :	DES PROCESSUS À L'ÉCHELLE DE LA COMMUNAUTÉ PEUVENT FAIRE DÉVIER DE L'OPTIMALITÉ	70
1	Les espèces ne saturent pas forcément leurs habitats ce qui peut causer une déviation à l'optimalité	72
2	Dimensionnalité des indices de diversité fonctionnelle	80
CHAPITRE 3 :	LES ESPÈCES FONCTIONNELLEMENT ORIGINALES DÉVIENT DE L'OPTIMUM ENVIRONNEMENTAL	108
1	Un outil pour calculer la rareté fonctionnelle	110
2	L'originalité fonctionnelle peut se calculer à de multiples échelles spatiales et biologiques	118
3	Déterminants environnementaux de l'originalité fonctionnelle globale et relations avec les autres raretés	121
CHAPITRE 4 :	RARETÉS & ORIGINALITÉS ÉCOLOGIQUES DANS LA CONSERVATION	153
1	Points chauds de rareté fonctionnelle et prise en compte de l'originalité fonctionnelle dans la conservation	155
2	Menaces sur les raretés écologiques dans l'Anthropocène	167
DISCUSSION		186
1	Dévier de l'optimalité fonctionnelle	186
1.1	Des processus spécifiques	186
1.2	Apports des approches de modélisation	187
2	L'originalité fonctionnelle	190
2.1	Rareté fonctionnelle ou originalité fonctionnelle?	190
2.2	Quels traits utiliser pour calculer l'originalité fonctionnelle?	192
2.3	Aux sources de l'originalité	194
2.4	Combien de traits pour calculer l'originalité fonctionnelle?	199
2.5	Originalité, échelle spatiale, échelle biologique, et échelle taxonomique	200
2.6	L'originalité fonctionnelle le long de l'arbre du vivant	202
2.7	Faut-il conserver spécifiquement les espèces rares fonctionnellement?	206
3	Perspectives	208
3.1	Besoins en données et descriptions écologiques	208
3.2	De l'importance des outils en écologie	210

3.3	L'originalité fonctionnelle : un nouveau concept ou un indice de plus?	213
ANNEXE A :	ANNEXES À L'ARTICLE 1.1	215
ANNEXE B :	ANNEXES À L'ARTICLE 4.1	242
ANNEXE C :	ANNEXES À L'ARTICLE 4.2	261
ANNEXE D :	BARNAGAUD ET. AL (2019) GLOBAL ECOLOGY AND BIOGEOGRAPHY	285
BIBLIOGRAPHIE		300
RÉSUMÉS		321

Remerciements

« Si j'ai vu plus loin, c'est en montant sur les épaules de géants. » Isaac Newton, 1675.

Cette citation est souvent utilisée pour illustrer le caractère cumulatif de la science. Cependant, en plus d'être cumulatif, le travail de recherche est fondamentalement collectif. Même si cette thèse ne présente qu'un seul nom en couverture, je souligne ici les contributions des autres personnes qui ont rendu possible sa réalisation.

Tout d'abord je remercie les membres de mon jury de thèse qui ont accepté d'évaluer mon travail : Sébastien Brosse, Stéphane Dray, Emmanuelle Porcher, et David Mouillot.

Je remercie également les membres de mon comité de thèse avec qui j'ai beaucoup échangé : Gaël Grenouillet, Ana Rodrigues, Vincent Devictor, et Sébastien Villéger.

Bien évidemment, je remercie chaleureusement le « duo de choc » formé par mes deux directeurs de thèse, auprès de qui j'ai pu apprendre énormément et avec qui j'ai pris beaucoup de plaisir à travailler.

Depuis cinq ans maintenant, Cyrille, nous travaillons ensemble et je l'espère pour encore au moins cinq ans ! Grande ta porte est restée ouverte toutes ces années, je te remercie pour toute l'écoute et le temps que tu m'as consacré, afin de toujours me pousser plus loin dans mon travail. Si nous ne sommes pas d'accord sur tous les sujets, nos désaccords ont même parfois été l'occasion de m'ouvrir à davantage. Immanquablement, je souhaite aussi te remercier pour l'environnement scientifique dont tu m'as fait bénéficier, à la fois sur la grande liberté que tu m'as laissé exercer – tout en t'assurant que je ne m'égarais pas en chemin – ainsi que sur les nombreuses possibilités de collaborations, de conférences et de participations à des ateliers que tu m'as données.

François, merci pour ces années à m'aider à avancer sur le chemin de la recherche et à parfaire la rigueur de mes raisonnements. La bienveillance dont tu as toujours fait preuve à mon égard m'a aussi permis d'apprendre à mieux développer mon propos. Évidemment, je mentionnerais ton humour à toute épreuve qui a toujours égayé nos réunions tri-, quadri-, ou multi-latérales. Un souvenir tout particulier pour cette soirée dans un restaurant de Moulis qui reste l'un des meilleurs moments de ma thèse. Rasséréiné, je l'ai été de nombreuses fois par ton impressionnante connaissance botanique qui nous a à la fois permis de valider certaines analyses mais aussi de profiter de joyeuses balades en ta compagnie.

Alors encore merci à tous les deux, qui faites une excellente paire de directeurs.

Je remercie le Centre de Synthèse et d'Analyse sur la Biodiversité. D'abord, l'ensemble des personnes qui y travaillent : Nicolas Mouquet, Maud Calmet, Nicolas Casajus et Marie-Claire Danner. Ensuite, je remercie l'ensemble des membres du groupe *Functional Rarity in Ecology and Evolution* grâce à qui une partie de ces travaux de thèse a pu être menée.

Un grand merci aux collaborateur-riche-s qui n'ont pas été cité-e-s : merci Jean-Yves pour nos conversations toujours passionnantes et nos échanges scientifiques ; thank you Caroline for your insights and for the journal club ; merci David pour tes remarques toujours pertinentes ; merci Pierre Gaüzère pour nos discussions sur la rareté fonctionnelle ; merci Nicolas L. pour nos discussions autour de nos questions partagées.

Merci aux chercheur-euse-s m'ayant permis d'avoir une expérience d'enseignement complémentaire à la recherche : Élodie Brunel-Piccini, Éric, et Sylvain C.

Merci aux stagiaires que j'ai pu encadrer : Nathan Mazet et Charlotte Guérineau.

Étant donné l'importance de l'environnement local, je souhaite remercier ici les membres de l'équipe ECOPAR. Merci aux membres permanents : merci Catherine pour ta gestion de l'équipe, merci Éric pour nos discussions animées scientifiques ou non, merci Florence pour ta grande sensibilité, merci Elena pour ta sympathie, merci Florian pour ton regard acéré et ces trajets ariégeois, merci Grégoire pour nos chouettes échanges, et merci à François V. tout dernier permanent arrivé à ECOPAR.

Dans l'équipe, je remercie en particulier toutes les personnes avec qui j'ai pu partager mon bureau : les membres de Turbopar, Marianne G., Étienne, Laura, et Pierre; Caroline et Mar, merci pour ces chouettes discussions multiculturelles; merci aux membres de Viollopar/Surledépar; merci aux derniers arrivé-e-s, Stefania pour ton ouverture, ton rire communicatif et ta chaleur, Léo pour tes questionnements scientifico-politiques toujours pertinents, ainsi que Marie-Charlotte et Aurélien pour l'ambiance potache, bon courage à vous pour les années qui arrivent!

Merci aux à tous les non-permanents d'ECOPAR : merci à Pauline et Iris B. pour avoir été les premières à m'accueillir; merci à Marianne d'avoir été l'une de mes premières complices; merci Jules pour les innombrables discussions philosophico-botaniques après 18h; merci à Julie pour nos discussions et ta bonne humeur; merci Yamei pour avoir apporté un peu de Chine à ECOPAR; merci à Lucie d'en avoir si gros sous la chapelle; merci Helena pour la découverte des desmides.

Je remercie mes comparses de thèse avec qui nous avons vécu cette expérience unique et transformatrice.

Tout d'abord, un grand merci à Pierre D., dont j'ai partagé le bureau et les directeurs pendant quatre ans. Cette thèse n'aurait pu voir le jour sans toi ou alors dans une forme bien différente. J'ai eu de la chance de grandir à tes côtés et de me nourrir de nos questions mutuelles aussi bien d'un point de vue fondamental que d'un point de vue technique. Merci d'avoir été là pendant la thèse et je sais que nous continuerons à échanger et travailler au-delà. À bientôt Spolète;)

Floriane, tu étais (snif) la deuxième personne la plus ancienne de ce bureau. Merci pour ton organisation sans faille pour animer nos vies au sein et en dehors du labo! Beaucoup de courage à toi qui es la dernière des mousquetaires à finaliser sa thèse, Marcatrude! Kevin, merci pour toute ton énergie qui a permis de se concentrer sur les bonnes choses dans les moments difficiles. Merci Apruncule pour tes intérêts sur le travail du bois et les sports originaux (et vive le Spikeball!).

Germain, le plus discret et parfait des co-bureaux. Merci Dagoulf pour ton enthousiasme constant qui ne t'a fait douter de ta thèse que dans les deux dernières semaines (!) avant ta soutenance.

Lauren, propagule d'ECOPAR établie à peine plus loin chez nos collègues de BIOFLUX, même si nous n'avons pas partagé de bureau, merci d'avoir été là toutes ses années et à bientôt.

En plus de l'environnement local de l'équipe, une thèse se mène dans le contexte plus global d'un laboratoire. Merci à Jordane pour son accueil au secrétariat toujours chaleureux; merci à Marguerite, Marjorie, Mélanie, et Natacha du service gestion pour les différentes missions; merci à Florence; merci à Cyril et Marie-Claude de la plateforme SIE, pour nos nombreuses discussions techniques et politiques; merci à Anne d'animer la bibliothèque, précieuse ressource de notre laboratoire; merci aux personnes du service informatique en particulier à Aïcha, Damien, et Christian à qui j'ai adressé bien trop de tickets; merci aussi à Enrique pour son émulation à propos de la programmation scientifique.

Je remercie l'ensemble des chercheur-e-s du département d'Écologie Fonctionnelle pour les interactions scientifiques. Merci à Stephan d'avoir animé le département.

Cette thèse aura aussi été l'occasion d'expérimenter le rôle de représentant des non-permanent au Conseil de Laboratoire. Merci à mes camarades élu-e-s : Jonathan, Marion J., Rémi, Pierre G. et Victor. Merci à tous

les autres membres des conseils du laboratoire. Un merci particulier à Marie-Laure pour avoir mené, avec l'aide notamment de Véronique, toutes ces réunions.

Merci aux personnes ayant soutenu ou participé au groupe des utilisateur·rice·s de R au laboratoire qui j'espère me survivra au laboratoire.

Merci aussi à mes complices d'organisation du séminaire *Models in Ecology and Evolution* édition 2016 : Marianne B., Maxime, Valentin, et Hugo.

Merci aux différents comités d'animations du laboratoire qui se sont succédés et ont permis un mélange et des rencontres entre personnes du laboratoire. Un merci tout particulier à Ana pour son implication sans faille dans l'animation scientifique et humaine de la vie du laboratoire. Merci à toutes les personnes qui se sont impliquées et s'impliquent dans l'organisation et la vie collective du laboratoire.

Merci aux nombreux·ses ami·es du laboratoire. Merci au groupe des déjeuners du mercredi : Nina, Sarah, Françoise, Marion C., et Eva. Merci à Manon, Anne-Sophie, Marion J., Juliane, Maude, Nafissa, Thomas, Marianne B., Tatiana, Pierre G., Adrien, Victor, Hugo, Stéphane, Nathalie (et Leïa) et Rémi. Merci aussi à celles et ceux en dehors du laboratoire : Robin, Alexis, Sylvain L., Évanne, Benjamin et Mathilde. Et merci aux ami·e·s plus ancien·ne·s : Valentine, Emma, Flora, Ronan, Fanny, Laure.

Plus largement, je remercie les personnes qui croient en des projets qui profitent à tou·te·s. Alors merci aux fondateurs de R, Ross Ihaka et Robert Gentleman pour avoir créé un outil ouvert et accessible. Merci à Hadley Wickham pour la création de ggplot2, qui m'a permis de me plonger dans la visualisation de données. Merci à Greg Wilson d'avoir co-fondé *The Carpentries*, qui a permis à de nombreuses personnes de pouvoir agir sur le monde *via* l'apprentissage des « bonnes » pratiques de programmation. Merci aux utopistes de rOpenSci : Karthik Ram, Carl Boettiger, Scott Chamberlain, Stefanie Butland, et Maëlle Salmon qui, par leur implication dans un projet résolument ouvert et collaboratif, démontrent tous les jours la force d'une communauté. Merci à Hugo, comparse des side-projects ouverts qui nous en apprennent plus en rendant service à d'autres.

Enfin, je remercie les personnes qui m'ont permis ce parcours. En particulier, merci à Nelly Kieffer pour m'avoir ouvert la première fois les portes d'un laboratoire; merci à Éric Baptiste et Philippe Lopez de m'avoir confié mon premier sujet de recherche; merci à Jean-François Goût pour le début de mon expérience internationale; merci à Ophélie Ronce pour la découverte de l'écosystème montpelliérain; et merci à Bruno Hérault pour la découverte de l'écologie fonctionnelle, ce qui m'a ensuite permis de contacter un certain Cyrille Violle.

Évidemment, merci à ma famille pour leur soutien indéfectible. Merci à Iris d'être à mes côtés.

Contributions scientifiques

I PUBLICATIONS

Chapitre 1

1. Munoz, F., **Grenié, M.**, Denelle, P., Taudière, A., Laroche, F., Tucker, C., Violle, C., 2018. ecolottery : Simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution* 9, 693–703. <https://doi.org/10.1111/2041-210X.12918>
2. Denelle*, P., **Grenié***, M., Violle, C., Munoz, F., Tucker, C. How hierarchical competition and limiting similarity affect trait-abundance relationships in population-dynamic models. *in prep* for *Journal of Ecology*. * : ces auteurs ont contribué également au travail

Chapitre 2

1. **Grenié, M.**, Violle, C., Munoz, F., 2020. Is prediction of species richness from stacked species distribution models biased by habitat saturation? *Ecological Indicators* 111, 105970. <https://doi.org/10.1016/j.ecolind.2019.105970>
1. **Grenié, M.**, Munoz, F., Violle, C. When more is less : adding more traits dilutes the Functional Diversity-Environment Relationship. *in prep* for *Ecological Indicators*.

Chapitre 3

1. **Grenié, M.**, Denelle, P., Tucker, C.M., Munoz, F., Violle, C., 2017. funrar : An R package to characterize functional rarity. *Diversity and Distributions* 23, 1365–1371. <https://doi.org/10.1111/ddi.12629>
2. Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N.J.B., Cadotte, M.W., Livingstone, S.W., **Grenié, M.**, Mouillot, D., 2017. A Common Toolbox to Understand, Monitor or Manage Rarity? A Response to Carmona et al. *Trends in Ecology & Evolution* 32, 891–893. <https://doi.org/10.1016/j.tree.2017.10.001>
3. **Grenié, M.**, Munoz, F., FREE Consortium, Violle, C. Distribution and environmental determinants of plant ecological originality the New World. *in prep* for *Global Ecology and Biogeography*.

Chapitre 4

1. **Grenié, M.**, Mouillot, D., Villéger, S., Denelle, P., Tucker, C.M., Munoz, F., Violle, C., 2018. Functional rarity of coral reef fishes at the global scale : Hotspots and challenges for conservation. *Biological Conservation* 226, 288–299. <https://doi.org/10.1016/j.biocon.2018.08.011>
2. Loiseau*, N., Mouquet*, N., Casajus, N., **Grenié, M.**, Gueguen, M., Maitner, B., Mouillot, D., Ostling, A., Renaud, J., Tucker, C., Velez, L., Thuiller, W., Violle C. Worldwide ecological rarity of mammals and birds. *in revision* in *Nature Ecology and Evolution*. * : ces auteurs ont contribué également au travail

Annexes

1. Barnagaud, J.-Y., Mazet, N., Munoz, F., **Grenié, M.**, Denelle, P., Sobral, M., Kissling, W.D., Şekercioğlu, Ç.H., Violle, C., 2019. Functional biogeography of dietary strategies in birds. *Global Ecology and Biogeography* 28, 1004–1017. <https://doi.org/10.1111/geb.12910>

Non-inclus dans la thèse

1. Joffard, N., Massol, F., **Grenié, M.**, Montgelard, C., Schatz, B., 2019. Effect of pollination strategy, phylogeny and distribution on pollination niches of Euro-Mediterranean orchids. *Journal of Ecology* 107, 478–490. <https://doi.org/10.1111/1365-2745.13013>

2. PAQUETS R

1. `ecolottery` : Munoz et al. 2018. `ecolottery` : simulating and assessing neutral and niche-based community dynamics in R
2. `funrar` : **Grenié M.**, Denelle P., Caroline M. Tucker, François Munoz, Cyrille Violle (2017). *funrar : An R package to characterize functional rarity*. R package version 1.2.1, <URL :<https://cran.r-project.org/package=funrar>>.
3. `rromeo` : **Grenié M.**, Gruson H (2020). *rromeo : An R Client for SHERPA/RoMEO API*. R package version 0.1.1, <URL :<https://CRAN.R-project.org/package=rromeo>>.
4. `rtaxref` : **Grenié M.**, Gruson H (2020). *rtaxref : an R client to the French Taxonomical Reference*. R package version 0.1.

Introduction

L'extraordinaire diversité des formes du vivant passionne les écologues depuis la création de leur discipline (Von Humboldt, 1877). Une des ambitions de l'écologie est d'identifier les mécanismes qui permettent d'expliquer le maintien de cette diversité. Au sein de l'écologie, différentes sous-disciplines mobilisent et analysent la diversité à différentes échelles spatio-temporelles et à différents niveaux biologiques (Relyea, Ricklefs, & Lempérière, 2019). L'écologie des communautés s'intéresse par exemple à expliquer la diversité des espèces qui coexistent au sein d'un environnement relativement homogène, les communautés écologiques (Southwood, 1987). Au travers de différentes théories, l'écologie des communautés définit des principes qui permettent de comprendre et prédire la composition des communautés écologiques.

Historiquement, les principes qui sont supposés régir les communautés écologiques ont été appelés « règles d'assemblages » (Diamond, 1975). Les écologues des communautés conceptualisent ces règles comme si les communautés observées étaient des puzzles dont il suffirait de suivre l'ensemble des règles pour identifier les patrons, c'est-à-dire les états observés. Ces règles, fondamentalement phénoménologiques qui décrivent les processus écologiques qui expliquent la composition des communautés. L'objectif premier des études en écologie des communautés était de déterminer si les communautés observées n'étaient que des sous-échantillons aléatoires de l'ensemble des espèces présentes (Diamond, 1975; Connor & Simberloff, 1979; Gotelli & Graves, 1996). Si les communautés observées étaient différentes, en nombre d'espèces notamment, de celles obtenues par un tirage aléatoire, alors cela signifiait que la communauté était structurée par une « règle » d'assemblage. Beaucoup de travaux ont été consacrés au raffinement des méthodes statistiques pour générer des communautés aléatoires en conservant les propriétés des communautés observées comme le nombre total d'espèces, l'abondance totale par espèce ou l'abondance totale par site (Gotelli & Graves, 1996). Certaines espèces occupent certains environnements de manière préférentielle. Pour s'assurer que cette observation n'est pas juste liée à la répartition des espèces dans des communautés de manières aléatoires, on peut utiliser des communautés « nulles ». Pour ce faire, on permute les espèces se trouvant dans les communautés observées de manière aléatoire, en conservant le nombre total d'espèces par site et le nombre de sites où une espèce est présente. Cela conserve la structure de « répartition » des espèces mais perturbe, s'il existe, le lien entre espèces et environnement. Si l'on observe une différence entre ces communautés nulles et les communautés observées, on en déduit que certaines espèces occupent des niches écologiques particulières, c'est-à-dire qu'elles se maintiennent et se reproduisent dans un ensemble restreint de conditions

abiotiques et biotiques (Hutchinson, 1957). Comme deux espèces ne peuvent pas coexister en occupant strictement la même niche selon le principe d'exclusion compétitive (Gause, 1934), cela contribue aussi à structurer les communautés de manière non-aléatoire.

1 DES RÈGLES D'ASSEMBLAGES À LA THÉORIE DES FILTRES

Les règles d'assemblages se conçoivent comme un ensemble de processus plus ou moins indépendants qui peuvent être d'intensité variable en fonction des contextes plutôt que comme une série de règles rigides qui expliquent les patrons observés (Kraft *et al.*, 2015). On visualise le plus souvent cette théorie sous la forme de processus successifs qui agissent sur l'ensemble des espèces présentes, du niveau régional au niveau local (logique *top-down*, voir Figure 1). Selon cette vision, les espèces observées aujourd'hui sont issues de l'ensemble des espèces qui pourraient atteindre la communauté locale par dispersion (Lortie *et al.*, 2004).

Tout d'abord, pour qu'on observe des espèces dans la communauté locale, elles doivent l'atteindre physiquement, y disperser. La diversité d'espèces dans la communauté locale dépend donc de leur capacité à atteindre un environnement donné dans l'espace, leur capacité de dispersion. La dispersion limitée (ou limite à la dispersion) décrit le processus selon lequel, du fait de différences de capacités de dispersion entre espèces, tous les habitats de celles-ci, même favorables, ne seront pas également occupés (Levins (1969), Figure 1). Par exemple, une plante dispersant à longue distance, pourra facilement atteindre tous ses habitats favorables, là où une plante dispersant à moins longue distance ne les atteindra pas forcément. Toutes choses étant égales par ailleurs, on s'attend donc qu'une espèce qui disperse à longue distance soit présente dans davantage de communautés qu'une autre qui ne disperse qu'à courte distance. De manière similaire on s'attend à observer une plus grande diversité d'espèces dans une communauté où davantage d'espèces peuvent disperser qu'une communauté atteignable par moins d'espèces (Hanski & Hanski, 1999).

L'environnement abiotique affecte l'implantation et la survie des espèces à un endroit donné, en sélectionnant celles qui peuvent se maintenir dans la communauté locale (Kraft *et al.*, 2015). Certaines espèces mal adaptées à l'environnement local ne peuvent pas s'y implanter malgré l'arrivée de propagules. Les interactions avec les autres espèces peuvent aussi empêcher le maintien de ces espèces, en sélectionnant à leur tour les autres espèces qui peuvent se maintenir dans la communauté, il s'agit du filtre biotique. La communauté observée résulte de l'influence combinée de ces différents filtres (Figure 1).

La série des filtres qui s'appliquent pour passer du bassin d'espèces régional à la communauté locale agit à différentes échelles spatiales (Lortie *et al.*, 2004). Tout d'abord, la dispersion a lieu à une échelle spatiale relativement large, comme on peut l'identifier à travers les barrières biogéographiques à la dispersion comme les chaînes de montagne ou les océans. Puis, à une échelle un peu plus petite, le filtre environnemental sélectionne les espèces qui peuvent se maintenir à travers le climat moyen d'une région. Enfin, les espèces qui ont pu atteindre la communauté locale et survivre dans le climat moyen subissent la compétition les unes des autres à l'échelle locale. La succession des filtres agit donc *a priori* le long d'une succession d'échelles

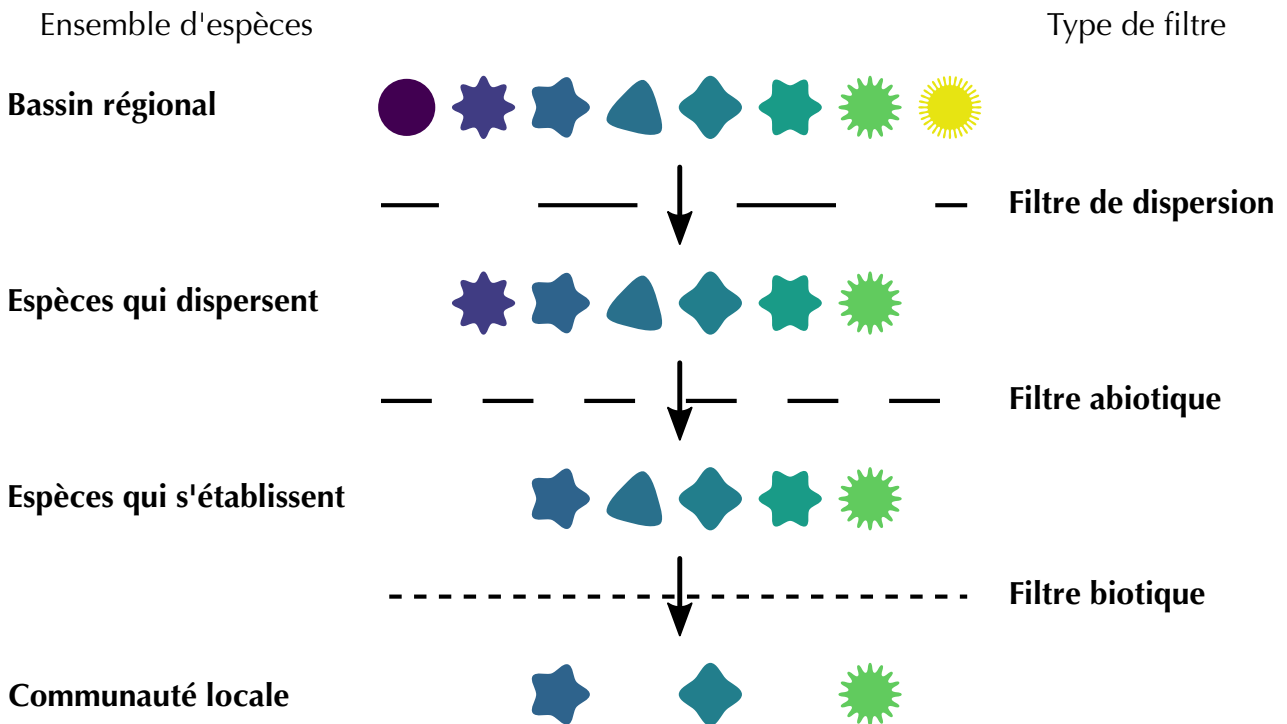


FIGURE 1 – Représentation schématique de la théorie des filtres. Différents processus ont lieu successivement et déterminent la composition de la communauté observée à l'échelle locale par rapport au bassin régional. Les espèces doivent d'abord disperser dans la communauté locale, c'est-à-dire qu'elles passent le filtre de dispersion. Puis le filtre environnemental sélectionne les espèces capables de se maintenir. Enfin les interactions entre espèces sélectionnent à leur tour les espèces localement *via* le filtre biotique. Adapté de Lortie et collab. (2004).

spatiales emboîtées.

Bien que les filtres soient souvent conçus comme indépendants et agissant successivement de plus en plus de travaux soulignent leurs interactions et rétro-actions à même échelle spatiale ou même entre échelles spatiales (Lortie *et al.* (2004); Mittelbach & Schemske (2015), voir Figure 2). Par exemple, un filtre environnemental très local peut être modifié par la présence d'une espèce « nurse » qui modifie les conditions micro-climatiques et facilite la présence d'autres espèces qui ne pourraient pas se maintenir autrement (Danet *et al.*, 2018). Il peut donc y avoir rétro-action entre filtre biotique et abiotique, car la plante « nurse » modifie les conditions abiotiques pour les autres espèces. En outre, ces processus peuvent être liés à différentes échelles spatiales. Par exemple à large échelle on peut considérer que les déterminismes climatiques dominent les communautés écologiques (Swenson *et al.*, 2012), alors qu'à plus petite échelle on suppose une importance plus grande des interactions biotiques (Kneitel & Chase, 2004), et qu'à une échelle encore plus fine, ce sont individus qui modifient les conditions micro-climatiques autour d'eux et il devient difficile de séparer le filtre abiotique du filtre biotique (Blonder *et al.*, 2018).

La stochasticité démographique des espèces, bien que longtemps négligée, est aujourd'hui montrée comme

étant un processus structurant de nombreuses communautés écologiques. La stochasticité démographique (ou dérive écologique ou processus neutres) correspond au fait que les processus démographiques (naissance, mort, et reproduction) sont stochastiques par nature, ce qui influe sur la composition attendue des communautés en fonction des tailles de celles-ci. Par exemple, dans une communauté composée d'individus d'espèces différentes mais partageant les mêmes paramètres démographiques, les processus neutres seront les seuls déterminants de la composition de la communauté. On peut alors dériver le nombre d'espèces attendu à l'équilibre en fonction de l'abondance relative initiale de chacune des espèces et des paramètres démographiques de la communauté (Hubbell, 2001). Ainsi, une communauté avec un faible nombre d'individus subira davantage l'influence des processus neutres qu'une communauté contenant des espèces issues de la même source de migrants mais avec un plus grand nombre d'individus. Toutes choses étant égales par ailleurs, on s'attend à observer une plus grande diversité d'espèces dans la communauté contenant davantage d'individus que dans celle en contenant moins. Si les processus neutres ont longtemps été opposés aux processus déterministes (comme le filtre environnemental), en partie du fait de l'histoire de l'étude de la stochasticité des communautés écologiques comme citée plus haut, la vision contemporaine de l'écologie des communautés les intègre pleinement (Leibold & McPeck, 2006). En effet, les processus neutres affectent toutes les communautés écologiques et c'est leur importance relative qui dépend des paramètres de celles-ci.

Cette vision synthétique de l'écologie des communautés *via* la théorie des filtres permet une vision large et flexible applicable à de nombreux écosystèmes. L'intensité des différents processus varie en fonction du contexte et permet d'expliquer la diversité des communautés observées. Une grande part du travail de l'écologie des communautés est alors d'identifier les processus dominants à l'œuvre dans ses communautés d'intérêt. Une fois ces processus identifiés, il devient possible de faire des prédictions quantitatives de certains paramètres de la communauté comme par exemple le nombre total d'espèces attendues ou bien la distribution des abondances relatives. Seulement, ces prédictions sont souvent agrégées à l'échelle de la communauté et ne permettent pas d'identifier les caractéristiques des espèces qui pourraient la composer.

2 L'APPROCHE FONCTIONNELLE DES COMMUNAUTÉS

Afin de mieux comprendre les causes des processus d'assemblages des communautés, l'approche fonctionnelle de l'écologie des communautés privilégie l'étude des traits phénotypiques comme *proxy* de la performance locale des espèces (McGill *et al.*, 2006). Ces traits peuvent être des traits morphologiques, phénologiques ou comportementaux mesurables à l'échelle de l'individu (Violle *et al.*, 2007). L'idée est d'étudier les traits et leurs distributions plutôt que l'identité des espèces pour comprendre l'effet des processus d'assemblages (McGill *et al.*, 2006). On choisit d'étudier les traits directement liés à la valeur sélective des espèces (*fitness*) ou à ses composantes : survie, reproduction, et croissance. La performance locale des espèces est donc la capacité des espèces à survivre, se reproduire et croître dans un environnement donné.

De nombreux travaux ont souligné l'intérêt de l'approche basée sur les traits pour caractériser les processus

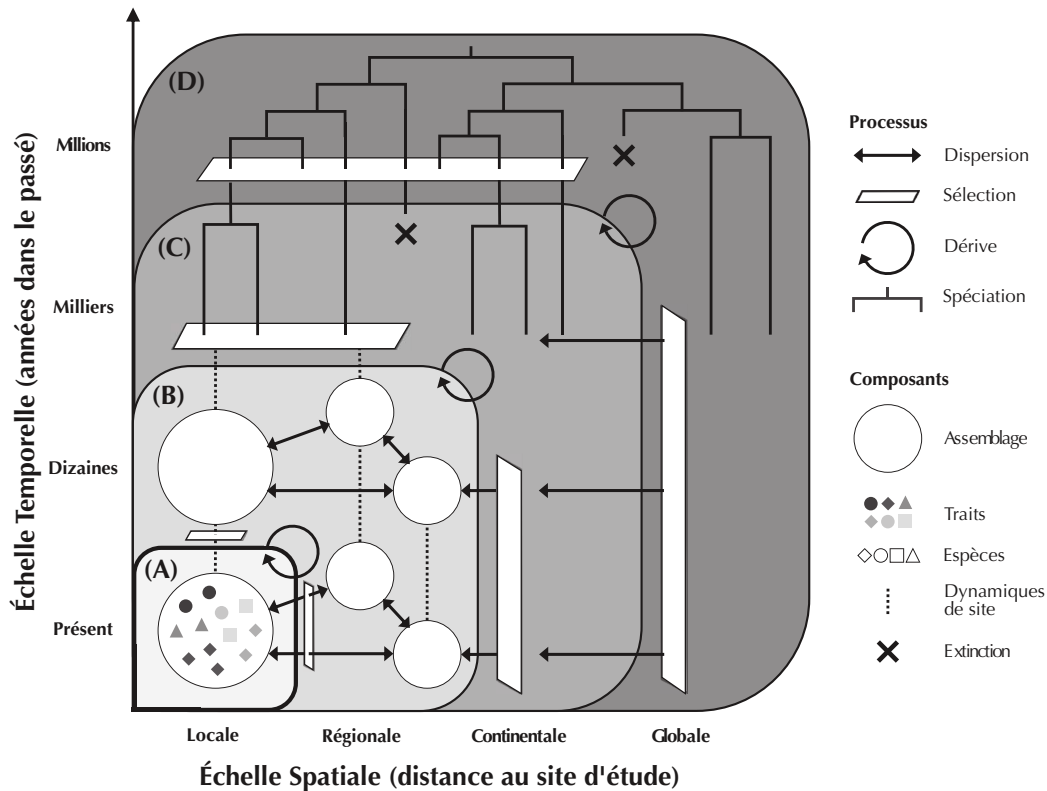


FIGURE 2 – Schéma synthétique de la vision contemporaine de l'écologie des communautés. Traduction de Rapacciolo & Blois (2019).

qui structurent les communautés (Enquist *et al.*, 2015). Par exemple on peut exprimer le filtre environnemental directement comme fonction des traits phénotypiques. Le plus couramment, on suppose un filtre environnemental une relation Gaussienne entre valeur du trait phénotypique d'une espèce et performance de celle-ci dans cet environnement. Cette fonction suggère qu'il existe une valeur optimale d'un trait qui permet une performance maximale dans cet environnement. De manière similaire les processus de compétition entre espèces et la dispersion des espèces peuvent s'exprimer en fonction des traits qui leur sont associés. Par exemple, les espèces de plantes les plus grandes ont tendance à disperser plus loin que les espèces petites (Thomson *et al.*, 2011); en outre, elles dispersent d'autant plus loin que leur masse de graine est petite.

En utilisant des traits bien choisis, il est possible de construire des paysages qui décrivent des valeurs de traits optimales dans des environnements précis, et ainsi comprendre les causes de la performance locale des espèces (Laughlin & Messier, 2015). Les liens observés entre valeurs de traits fonctionnels et environnement témoignent de l'adaptation des espèces à leur environnement. Cependant, la performance des espèces n'est pas forcément liée à un seul trait mais peut-être le fruit de la contribution de plusieurs traits (Pistón Caballero *et al.*, 2019; Laughlin *et al.*, n.d.). En outre, la performance locale est souvent entendue à l'échelle individuelle, mais le succès d'une espèce peut être mesuré à d'autres échelles biologiques : individuelle, po-

pulationnelle ou spécifique (Salguero-Gómez *et al.*, 2018).

3 UNE SEULE VALEUR OPTIMALE DE TRAIT : LE PARADIGME DE L'OPTIMUM FONCTIONNEL

L'écologie fonctionnelle des communautés quantifie les processus d'assemblages des communautés à la lumière des distributions de traits dans celles-ci (McGill *et al.*, 2006). De nombreuses études expérimentales ont montré que les différences de traits fonctionnels expliquaient les mécanismes d'exclusion des espèces (examinés dans Weiher & Keddy, 1995). De ces études émerge le postulat suivant : il existe un optimum de trait local qui maximise la performance locale dans un environnement donné (Shipley, Vile, & Garnier, 2006; Enquist *et al.*, 2015). Les espèces ayant les valeurs de traits les plus proches de cette valeur optimale seront donc plus abondantes que les espèces plus éloignées de celle-ci (Figure 3). En considérant l'influence du filtre environnemental qui induit une valeur optimale de trait, on peut prédire les abondances des espèces dans la communauté locale en fonction de leurs valeurs de trait (Lavorel & Garnier, 2002; Shipley, Vile, & Garnier, 2006; Shipley, 2010).

La plus grande part des études en écologie fonctionnelle utilise ce postulat pour associer des valeurs de trait optimales le long de gradients environnementaux. Étant donné le fait que les espèces les plus proches de l'optimum seraient plus abondantes, la valeur de trait la plus abondante dans la communauté devrait représenter fidèlement la valeur optimale. Un estimateur de la valeur optimale de trait est donc la moyenne des valeurs de traits pondérées par les abondances des espèces de la communauté, c'est le *Community-Weighted Mean* (CWM, pointillés noirs sur la Figure 3) (Garnier *et al.*, 2004). Comme les traits ont été choisis pour représenter la performance locale des espèces, le CWM doit varier le long de gradients environnementaux, car une valeur de trait n'est optimale que dans un environnement donné. On peut alors associer les variations de CWM aux variations de l'environnement (Cornwell & Ackerly, 2009; Fournier *et al.*, 2015). Ainsi les espèces les plus abondantes devraient correspondre plus fortement au filtre environnemental que les espèces plus rares.

Si l'on suit le postulat de l'optimalité, les espèces les plus abondantes possèdent donc les traits les plus proches de la valeur de trait optimale. Mais du fait de la multi-dimensionnalité des traits fonctionnels il est possible qu'il existe non pas une seule valeur de trait optimal, mais une combinaison de traits fonctionnels qui maximise la performance locale le long du gradient environnemental. L'environnement étant lui aussi multi-dimensionnel, avec des traits différents variants selon des gradients distincts, il peut donc exister plusieurs combinaisons de traits qui permettent des performances locales équivalentes (Marks & Lechowicz, 2006). Enfin, si certains traits peuvent répondre à l'environnement selon le filtre environnemental, d'autres, moins liés à la valeur sélective des espèces, peuvent avoir une variance plus grande le long de l'environnement. Ainsi, il n'existe pas forcément un seul optimum de trait, dans la mesure où il est nécessaire de tenir compte de la contribution de traits différents aux différents processus d'assemblages des communautés pour prédire correctement les traits des espèces les plus abondantes.

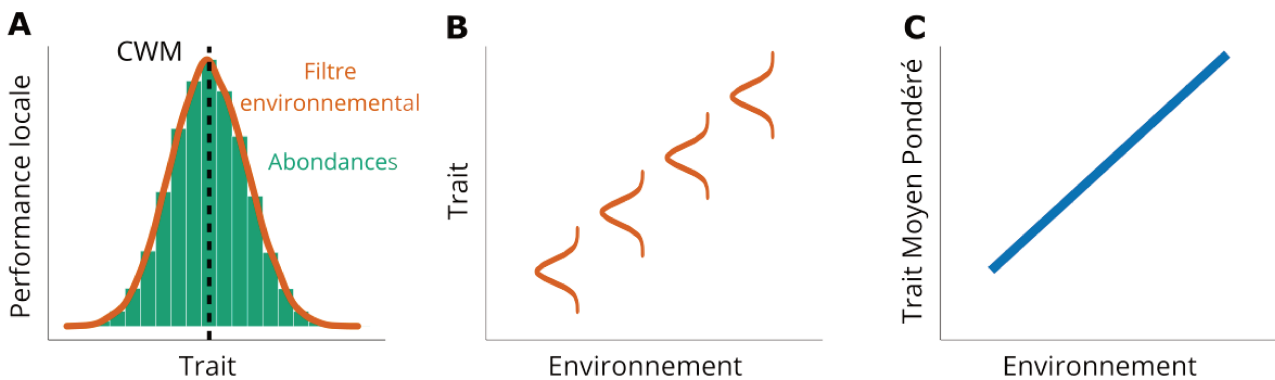


FIGURE 3 – Représentation du paradigme de l'optimalité fonctionnelle. (A) Relation entre trait et performance des espèces, Filtre environnemental Gaussien en orange, distribution des abondances dans la communauté locale en vert, le trait moyen pondéré (CWM) par les abondances est figuré en pointillé noir. (B) Si toutes les communautés suivent l'optimalité fonctionnelle le long d'un gradient environnementale on obtient des distributions de traits légèrement décalées. (C) En calculant le CWM de chaque communauté le long d'un gradient environnemental, on obtient une relation linéaire.

Certains processus d'assemblages des communautés ne correspondent pas au postulat de l'optimalité. Les processus neutres, par exemple, affectent toutes les espèces quels que soient leur valeur de trait (Hubbell, 2001). Si au départ considérée comme provocatrice et biologiquement insensée, car considérant toutes les espèces comme équivalentes alors même qu'il existe des variations évidentes entre espèces (Díaz *et al.*, 2016), la théorie neutre a fait des prédictions valides notamment dans les écosystèmes les plus riches (forêts tropicales, récifs coralliens). Aujourd'hui, les processus neutres sont pleinement intégrés dans le schéma théorique de l'écologie des communautés en plus des processus de niches dont découle l'optimalité (Leibold & McPeck, 2006; Adler, HilleRisLambers, & Levine, 2007; Haegeman & Loreau, 2011). Ainsi, pour aller plus loin que l'optimalité fonctionnelle, il faut tenir compte de l'influence des multiples processus d'assemblages.

4 DES ESPÈCES QUI ÉCHAPPENT À L'OPTIMALITÉ FONCTIONNELLE : LES ESPÈCES RARES

Comme nous l'avons évoqué dans les parties précédentes, l'optimalité fonctionnelle s'applique moins aux espèces rares qu'aux espèces abondantes. Les espèces rares peuvent donc être aussi la source d'une déviation par rapport à l'optimalité attendue. Cette partie va donc tâcher de mieux définir ce qu'est une espèce rare et d'expliquer les processus qui pourraient permettre le maintien des espèces rares.

Avec les premières extinctions observées dues aux humains est venue la conscience que les espèces les plus rares courraient un risque d'extinction plus grand que les espèces communes (Diamond, 1989). L'Union Internationale pour la Conservation de la Nature (UICN) utilise ainsi le critère de faible abondance locale pour classer les espèces dans la liste rouge des espèces les plus menacées (Vié *et al.*, 2009; IUCN, 2017). Seulement la rareté des espèces est longtemps restée un concept vague et mal défini (Gaston, 1997). Le cadre conceptuel proposé par Rabinowitz (1981) a permis de mieux définir ce qu'était la rareté d'une es-

Table 1. Les sept formes de raretés telles que définies par Rabinowitz (1981).

Aire de répartition		Large		Petite	
Gamme d'habitats (spécificité d'habitat)		Large	Fine	Large	Fine
Abondance Locale					
Grande	Espèce commune	Espèce spécifique	Espèce restreinte	Espèce restreinte et spécifique	
Petite	Espèce peu abondante	Espèce peu abondante et spécifique	Espèce restreinte et peu abondante	Espèce restreinte, spécifique, et peu abondante	

èce selon trois composantes (Table 1). Premièrement, une espèce peut être rare parce qu'elle est faiblement abondante localement. Deuxièmement, une espèce peut être rare parce qu'elle n'est présente que dans un habitat spécifique (elle a une forte spécificité d'habitat). Enfin, une espèce peut être rare à une échelle plus large parce que son aire de répartition est petite. En combinant ces différentes composantes, Rabinowitz (1981) définit sept façons d'être rare et une façon d'être commune. Une espèce pourra être rare par sa faible abondance locale, tout en n'ayant pas d'habitat spécifique et une aire de répartition assez large. Une espèce pourra aussi être rare sur les trois composantes en même temps, en ayant une faible taille de population, une forte spécificité d'habitat et une aire de répartition restreinte. La rareté n'est ici pas définie comme une seule dimension absolue, afin d'utiliser les catégories de rareté de la manière la plus pertinente en fonction du contexte. La grille proposée par Rabinowitz permet d'identifier les espèces rares sur l'ensemble des axes et qui seraient donc les plus menacées d'extinction. Cela permet de se focaliser sur les principaux déterminants de la probabilité d'extinction des espèces afin de mettre en place des politiques de conservation adéquates (Espeland & Emam, 2011). Étant donné que la majorité des espèces sont rares, les théories générales de l'écologie pour être largement applicables, doivent donc pouvoir prédire correctement les abondances des espèces rares. L'étude des distributions des abondances s'est par exemple intéressée très tôt à la distribution des espèces faiblement abondantes (Preston, 1948, 1962). Comme nous l'avons rappelé dans les parties précédentes, la théorie neutre s'est particulièrement intéressée aux espèces rares afin de mieux prédire la distribution de leurs abondances (Hubbell, 2001; Rosindell, Hubbell, & Etienne, 2011). Dans le cadre de l'optimalité, les espèces rares devraient avoir tendance à présenter des traits différents de l'optimum théorique attendu.

Table 2. Les douze formes de rareté fonctionnelle. Adapté de Violle et collab. (2017) à partir de Rabinowitz (1981).

		Fréquence de l'espèce				
		Régionalement étroite		Régionalement étendue		
		Localement rare	Localement abondante	Localement rare	Localement abondante	
Traits de l'espèce	Régional. uniques	Localement originaux	Traits originaux à toutes les échelles	Trait originaux liés à la présence d'espèces spécialistes	Traits originaux liés à des espèces présentes largement	Traits originaux liés à quelques espèces communes
		Localement communs	Impossible	Impossible	Impossible	Impossible
		Localement originaux	Traits originaux liés à des espèces rares qui ne co-occurrent pas	Traits spécialisés liés à de nombreuses espèces	Traits originaux localement liés à la présence d'espèces ne co-occurrent pas à large répartition mais localement rare	Traits originaux localement lié à la présence de nombreuses espèces communes ne co-occurrent pas
	Régional. communs	Localement communs	Traits originaux liés à de nombreuses espèces rares	Traits spécialisés liés à de nombreuses espèces	Traits originaux localement liés à la présence d'espèces à large répartition mais localement rare	Traits communs à toutes les échelles

5 UNE APPROCHE FONCTIONNELLE DE LA RARETÉ : LA RARETÉ FONCTIONNELLE

D'après l'hypothèse du *mass-ratio*, les espèces dominantes localement sont celles qui ont le plus fort effet sur le fonctionnement des écosystèmes (Grime, 1998). Pour comprendre les écosystèmes il faut donc étudier les espèces abondantes. Seulement, les espèces rares peuvent aussi présenter des contributions uniques à ce fonctionnement (Mouillot *et al.*, 2013; Jain *et al.*, 2014; Leitão *et al.*, 2016). Dans le postulat de l'optimalité fonctionnelle, l'écologie fonctionnelle des communautés s'intéresse à la prédiction de la valeur de trait optimal de la communauté, négligeant souvent les variations des stratégies autour de l'optimum supposé (Rolhauser & Pucheta, 2017). Or, ces espèces pourraient présenter des traits originaux, et donc des contributions originales aux écosystèmes. L'objectif de cette partie est de présenter le cadre d'étude de l'originalité fonctionnelle. L'originalité fonctionnelle des espèces est le fait que des espèces possèdent des traits peu communs, différents des autres espèces considérées. Un cadre conceptuel pour l'originalité, proche du cadre proposé pour la rareté par Rabinowitz (1981), a récemment été proposé par Violle et collab. (2017). L'idée est de séparer la dimension de rareté liée à l'abondance et l'aire de répartition des espèces à la dimension d'originalité fonctionnelle locale et régionale. On peut définir ces deux dimensions à l'échelle locale et à l'échelle régionale ce qui fait qu'on obtient un total de seize combinaisons possibles (Table 2), dont douze

façons d'être rare car quatre d'entre elles sont impossibles (une espèce ne peut pas être localement commune tout en étant régionalement unique). Par exemple, une espèce peut être localement originale, tout en étant régionalement commune, en étant localement abondante, et avec une petite aire de répartition. Les différentes dimensions ne sont *a priori* pas corrélées.

Ce cadre conceptuel propose d'examiner la rareté fonctionnelle selon différentes dimensions. Pour quantifier ces dimensions, Violle et collab. (2017) ont proposé d'utiliser des grandeurs continues, afin d'éviter une catégorisation binaire hâtive entre espèces rares et espèces communes qui pourrait limiter la portée des analyses (Gaston, 1997). Toutefois, il peut être intéressant de catégoriser la rareté fonctionnelle, notamment dans le cadre de la biologie de la conservation. Par exemple, en fixant des seuils à partir desquels on définit qu'une espèce est originale par rapport aux autres, critères que l'on peut cumuler sur différents traits et à différentes échelles, à l'image des critères utilisés par l'UICN pour l'évaluation des espèces sur la liste rouge des espèces menacées d'extinction.

À partir de ce cadre, on peut se demander s'il existe des processus particuliers qui favorisent le maintien des espèces originales à large échelle. Il existe une variabilité des combinaisons de traits existantes entre espèces à l'échelle globale (Díaz *et al.*, 2016), avec des stratégies communes qui représenteraient des optimums globaux de valeurs sélectives. Pourtant, si ces stratégies étaient globalement optimales, comment expliquer le maintien des stratégies originales qui devraient être contre-sélectionnées. Une hypothèse est que ces espèces possèdent des traits originaux comme réponse à des adaptations à des milieux particuliers. En effet les milieux extrêmes engendrent un fort filtre environnemental qui réduit la diversité fonctionnelle locale en ne sélectionnant que les espèces tolérantes au stress (Cornwell & Ackerly, 2009; Lhotsky *et al.*, 2016; Zanne *et al.*, 2018). Les stratégies de tolérances ou d'évitement nécessaires en milieux extrêmes sont en outre coûteuses métaboliquement (Parsons, 1990). Dans un environnement peu extrême, les espèces utilisant ces stratégies devraient donc être moins compétitives que les espèces qui ne survivent en milieu extrême. Il devrait donc y avoir davantage d'espèces tolérantes au stress (et donc originales globalement) en milieu extrême qu'en milieu moins extrême.

6 ESPÈCES ORIGINALES ET CONSERVATION DE LA BIODIVERSITÉ

Les espèces originales ont une contribution originale au fonctionnement des écosystèmes, et même parfois plus importante que certaines espèces communes (Mouillot *et al.*, 2013; Jain *et al.*, 2014; Leitão *et al.*, 2016). De ce fait, si les espèces originales sont davantage menacées d'extinction que les espèces communes, alors le fonctionnement des écosystèmes pourrait être perturbé. Il faut donc évaluer les menaces qui portent sur les espèces originales. En outre, si les espèces originales sont faiblement abondantes, elles peuvent être doublement menacées, à la fois du fait de leurs traits originaux et du fait de leur faible abondance (Gaston, 1998; Hughes *et al.*, 2014; Giam & Olden, 2017). Il est donc fondamental d'étudier les relations entre l'originalité et les différentes autres facettes de la rareté.

En biologie de la conservation, une approche consiste à désigner les espèces qu'il faut prioritairement cibler par des politiques de conservation (Arponen, 2012). Or, il n'existe pas de critère absolu pour prioriser les espèces, c'est ce que l'on appelle l'« agonie du choix » (Vane-Wright, Humphries, & Williams, 1991). Certains critères peuvent être esthétiques, symboliques ou utilitaires, ces critères intègrent souvent des dimensions socio-économiques (Arponen, 2012). Des critères biologiques sont aussi utilisés, par exemple en priorisant les espèces les plus menacées d'extinction, ou les plus « importantes » pour maintenir une fonction biologique (Arponen, 2012). Il a par exemple été suggéré de prioriser les espèces étant rares selon de multiples facettes dans le cadre de la rareté de Rabinowitz (1981). D'autres ont suggéré de prioriser les espèces les plus originales phylogénétiquement afin de préserver le « potentiel évolutif » des communautés (Vane-Wright, Humphries, & Williams, 1991; Mace, Gittleman, & Purvis, 2003; Forest *et al.*, 2007; Winter, Devictor, & Schweiger, 2013; Tucker *et al.*, 2019). De manière similaire, on pourrait suggérer de prioriser les espèces originales, d'une part pour leur contribution non-remplaçable au fonctionnement des écosystèmes, d'autre part pour préserver une diversité maximale de traits.

Les relations phylogénétiques entre espèces reflètent leur histoire évolutive. En particulier, pour des traits évoluant lentement les distances phylogénétiques entre espèces représentent directement les distances fonctionnelles entre espèces (Ackerly, 2009). Des espèces séparées par une longue histoire évolutive devraient voir leurs valeurs de traits diverger. Ainsi, les espèces étant séparées d'une plus grande distance évolutive des autres espèces, c'est-à-dire les plus originales phylogénétiquement, devraient aussi avoir les traits les plus éloignés des autres espèces, et donc les plus originaux fonctionnellement. Si c'est le cas, alors les schémas de priorisation qui proposent de prioriser les espèces en utilisant l'originalité phylogénétique prennent implicitement en compte l'originalité fonctionnelle (*evolutionary distinctiveness*; voir Isaac *et al.*, 2007; Redding, DeWolff, & Mooers, 2010; Jetz *et al.*, 2014; mais voir Tucker *et al.*, 2019). Si ça n'est pas le cas, alors il peut devenir urgent d'intégrer l'originalité fonctionnelle dans les schémas de priorisation.

La conservation peut aussi passer par la protection d'habitats particulièrement riches ou uniques. Il existe des écosystèmes considérés comme des « points chauds » de biodiversité à l'échelle mondiale car abritant une part très importante de la biodiversité compte tenu de leur aire géographique (Prendergast *et al.*, 1993; Myers *et al.*, 2000; Marchese, 2015). Même si ces points chauds ont d'abord été identifiés à partir du simple nombre d'espèces, d'autres études ont identifié des points chauds de diversité phylogénétique et de diversité fonctionnelle, plus ou moins congruents avec les points chauds identifiés par la richesse spécifique seule (Rodrigues & Gaston, 2002; Stuart-Smith *et al.*, 2013; Mazel *et al.*, 2014). La description de ces points chauds de la biodiversité permet d'identifier de grandes zones aux échelles biogéographiques qui nécessitent notre attention afin de permettre leur conservation. Existe-t-il des points chauds d'originalité fonctionnelle à large échelle? Dans quelles mesures sont-ils différents des autres points chauds identifiés avec d'autres métriques de diversité?

La planification systématique de la conservation (*systematic conservation planning*) travaille au contraire à plus petite échelle et de manière plus opérationnelle, afin de proposer des habitats à conserver prioritairement (Margules & Pressey, 2000). On sélectionne les habitats qui sont les plus représentatifs d'une région

ou qui maximise la diversité représentée par l'ensemble des habitats sélectionnés. Ainsi, on construit un réseau de zones protégées (ou à protéger) qui atteint une certaine représentativité de la biodiversité. Les algorithmes utilisés dans ces approches peuvent tenir compte de nombreuses contraintes afin d'atteindre des objectifs de conservation à moindre coût (McDonald, 2009). Il a d'ailleurs été récemment montré que pour un coût minime, en ajoutant peu d'aires protégées par rapport à celles actuellement protégées, il était possible d'obtenir des gains substantiels dans la diversité phylogénétique et fonctionnelle représentée par les aires protégées mondiales (Pollock, Thuiller, & Jetz, 2017). En ajoutant de contraintes de représentativité des espèces originales fonctionnellement dans ces approches, il est probable qu'elles donnent des résultats distincts de ceux obtenus en cherchant à maximiser la représentation des différentes diversités.

En complément des approches de priorisation, l'évaluation des menaces pesant sur les espèces d'intérêt permet d'envisager des actions de conservation à plus long terme. Une fois les menaces identifiées, il devient possible de proposer des actions de conservation ciblées qui auraient le plus d'impact sur les espèces d'intérêts. Les menaces peuvent être associées à des traits particuliers. Il a par exemple été montré des interactions entre traits fonctionnels des abeilles et sensibilités à des perturbations multiples (Williams *et al.*, 2010). Les espèces fonctionnellement originales, du fait de leur écologie particulière, pourraient être davantage sensibles aux perturbations de toutes origines.

7 MODÉLISER LES COMMUNAUTÉS POUR MIEUX LES COMPRENDRE

Les études de terrain sont capitales dans la description des communautés observées, mais elles ne permettent pas forcément de décomposer des gradients environnementaux multiples influençant les communautés. Elles ne permettent pas non plus de tester de manière exhaustive l'influence des différents processus indépendamment les uns des autres. Enfin leur coût élevé et l'important temps de travail qui leur est nécessaire rend aussi l'étude à de multiples échelles spatiales difficiles. Les études expérimentales, menées par exemple en serre pour les végétaux, permettent de répondre partiellement à ces contraintes en testant différentes contraintes environnementales ou en associant différentes espèces. Même si les études expérimentales permettent de mieux contrôler les processus, les contraintes pratiques empêchent de les utiliser pour mener des expériences le long de grands gradients environnementaux multiples à large échelle spatiale.

La modélisation des communautés est complémentaire des approches de terrain et expérimentale car elle permet de tenir compte de multiples processus de manière indépendante. En effet, nous avons souligné dans les parties précédentes la diversité des processus qui affectent à la fois la distribution des abondances locales et la distribution des valeurs de traits fonctionnels ainsi que leurs liens. Nous avons aussi insisté sur le fait que ces différents processus pouvaient se dérouler à différentes échelles spatiales, et qu'il est donc nécessaire de prendre en compte ces différentes échelles pour mieux comprendre l'effet des processus d'assemblages. La modélisation des communautés permet tester de manière indépendante l'influence des différents processus d'assemblages.

Les modèles de simulation des communautés peuvent permettre d'établir des ensembles de cas pour pouvoir ensuite inférer les processus d'assemblages en comparant les simulations aux communautés observées. Ces approches se développent largement en écologie des communautés (Münkemüller & Gallien, 2015 ; van der Plas *et al.*, 2015 ; Botta-Dukát & Czúcz, 2015 ; Cabral, Valente, & Hartig, 2016). Pour autant, pour pouvoir utiliser ces modèles pour l'inférence de processus, il peut être nécessaire de simuler de nombreuses communautés ce qui peut s'avérer coûteux en temps de calcul et nécessite donc des algorithmes de simulations et d'échantillonnage efficaces (Gimenez *et al.*, 2014). Il est donc nécessaire de développer un modèle de simulation de communauté, couplant les processus neutres et non-neutres, qui soit efficace en temps de calcul.

Afin de comprendre l'influence de traits multiples dans les différents processus, on peut formaliser ces différents processus en fonction de la valeur des traits des espèces. Comme je l'ai expliqué plus haut, on peut étendre la définition du filtre environnemental qui décrit la performance locale d'une espèce en fonction de la valeur d'un de ses traits à plusieurs traits, en utilisant par exemple une distribution Gaussienne multivariée plutôt qu'une distribution Gaussienne unidimensionnelle. De manière similaire, les filtres abiotiques comme les processus de compétition entre espèces peuvent s'écrire explicitement en fonction des valeurs de traits des espèces considérées. La limite à la similarité prédit que deux espèces aux niches trop similaires, c'est-à-dire avec des traits proches, ne peuvent pas coexister car elles sont plus fortement en compétition que des espèces plus éloignées (MacArthur & Levins, 1967). On pourra donc écrire une fonction de compétition liée à la limite à la similarité pour une espèce qui diminuera au fur et à mesure que le trait fonctionnel de l'espèce d'intérêt sera différent des traits des autres espèces présentes localement. La compétition hiérarchique prédit que la valeur de certains traits est d'autant plus avantageuse qu'elle est forte (ou faible) (Goldberg, 1990). On peut exprimer cela en utilisant une fonction asymétrique : si l'espèce possède un trait proche de la valeur la plus élevée (ou la plus basse) dans la communauté alors elle subira peu la compétition hiérarchique. Ainsi un modèle explicitant la contribution de différents traits dans les différents processus d'assemblage des communautés pourrait permettre de mieux comprendre les causes de la déviation des communautés de l'optimalité attendue.

8 SYNTHÈSE

En mobilisant des outils de l'écologie fonctionnelle, de l'écologie des communautés et de la biogéographie, cette thèse est composée de deux grandes parties : une première partie centrée sur la déviation des communautés à l'optimalité attendue sous l'hypothèse d'un filtre environnemental uni-modal, puis une deuxième partie à propos des espèces qui dévient de cet optimum de trait, les espèces originales.

Pour déterminer les processus qui peuvent favoriser une déviation de l'optimalité, l'apport de démarche de modélisation des communautés est crucial afin d'étudier l'influence relative des différents processus d'assemblages des communautés (Chapitre 1). À l'aide ces modèles il devient possible de mettre en évidence des processus qui dévient de l'optimalité fonctionnelle, comme la non-saturation de l'habitat par certaines espèces, ou la variation neutre de certains traits en fonction de l'environnement (Chapitre 2). Puis, un cadre quantitatif pour estimer l'originalité fonctionnelle, le pendant de l'optimalité fonctionnelle, est présenté avant que ne soit testée l'hypothèse selon laquelle l'originalité fonctionnelle globale est liée à une adaptation à des environnements extrêmes (Chapitre 3). Enfin, étant donné que les espèces originales fonctionnellement devraient avoir une importance particulière pour le fonctionnement des écosystèmes, il est important d'évaluer dans quelle mesure ces espèces sont d'ores et déjà prises en compte dans les schémas globaux de conservation (Chapitre 4).

1

Modéliser les communautés pour identifier les processus qui pourraient expliquer une déviation à l'optimalité

De multiples processus peuvent structurer les communautés. Pour mieux comprendre ces processus, nous avons souligné en introduction l'intérêt de la démarche de modélisation. Les modèles permettent en effet de réaliser des expériences virtuelles en isolant les processus indépendamment. Il est aussi possible de les combiner afin d'étudier leur influence conjointe. L'objectif de ce chapitre est de présenter deux modèles complémentaires dans leurs approches pour mieux identifier les processus qui pourraient expliquer que des communautés dévient du paradigme de l'optimalité fonctionnelle.

Les processus neutres, c'est-à-dire n'étant pas liés aux traits d'espèces, peuvent contribuer à structurer les communautés. Par exemple, la dispersion limitée peut empêcher des espèces les plus adaptées de s'établir dans certaines communautés. De manière similaire, les phénomènes source-puits peuvent maintenir des espèces dans des communautés où elles ne sont pas adaptées. Ce chapitre présente un modèle, *ecolottery*, qui permet de tenir compte des influences relatives de la neutralité et des processus de niche, à partir d'un ensemble régional d'espèce. Ce modèle peut en outre être couplé à des méthodes d'inférences Bayésiennes pour estimer l'importance relatives des processus neutres et non-neutres dans des communautés observées. Cet article a fait l'objet d'une publication dans *Methods in Ecology and Evolution*, accompagné d'un paquet R pour utiliser le modèle.

Certains traits peuvent être impliqués dans différents processus en même temps. Par exemple, la hauteur des plantes peut à la fois être filtrée par l'environnement et contribuer à la compétition hiérarchique entre espèces (Le Bagousse-Pinguet *et al.*, 2017). Malgré l'influence d'un filtre environnemental uni-modal sur un trait pour une espèce à l'optimum fonctionnel, si le trait est impliqué dans d'autres processus, ce trait sera affecté et l'on n'observera pas forcément une optimalité fonctionnelle. La deuxième partie de ce chapitre présente un modèle, *fdcoexist*, qui exprime explicitement les différents processus d'assemblages des communautés en fonction des valeurs de traits espèces et de leurs contributions aux différents processus. L'article associé est en préparation pour *Journal of Ecology*.

I L'INTERACTION DE PROCESSUS NEUTRES ET NON-NEUTRES PEUT CAUSER UNE DÉVIATION DE L'OPTIMALITÉ

Received: 2 January 2017 | Accepted: 21 July 2017

DOI: 10.1111/2041-210X.12918

RESEARCH ARTICLE



ecolottery: Simulating and assessing community assembly with environmental filtering and neutral dynamics in R

François Munoz¹ | Matthias Grenié² | Pierre Denelle² | Adrien Taudière² | Fabien Laroche^{2,3} | Caroline Tucker^{2,4} | Cyrille Violle²

¹University Grenoble-Alpes, LECA, Grenoble Cedex 9, France

²CEFE UMR 5175, CNRS-Université de Montpellier-Université Paul-Valéry Montpellier-EPHE, Montpellier Cedex 5, France

³Irstea, UR EFNO, Centre de Nogent-sur-Vernisson, Nogent-sur-Vernisson, France

⁴Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Correspondence

François Munoz
Email: francois.munoz@cirad.fr

Funding information

European Research Council, Grant/Award Number: ERC-StG-2014-639706-CONSTRAINTS; Marie Curie IIF, Grant/Award Number: H2020-MSCA-IF-2014-657951; French Ecole Normale Supérieure

Handling Editor: Ryan Chisholm

Abstract

1. We introduce the R package *ecolottery* dedicated to quick and efficient simulation of communities undergoing local neutral dynamics with environmentally filtered immigration from a reference species pool (spatially implicit model). The package includes an Approximate Bayesian Computation (ABC) tool to estimate the parameters of these processes. We present the rationale of the approach and show examples of simulations and ABC analysis.
2. The species in the reference pool differ in their abundances and trait values. Environmental filtering weights the probability of immigration success depending on trait values, while the descendants of established immigrants undergo neutral stochastic drift. The reference pool can be defined in a flexible way as representing, e.g. the composition of a broad biogeographical region, or available dispersers around local communities. The package provides a process-based alternative to the use of randomization-based null models.
3. The package proposes a coalescent-based simulation algorithm that presents significant advantages over alternative algorithms. It does not require simulating community dynamics from an initial state forward in time but does still allow measurement of the influence of environmental filtering. Because of its high calculation speed, this approach allows simulating many communities within a reasonable amount of time.
4. Diverse patterns of taxonomic, functional and phylogenetic compositions can be generated. The package can be used to explore the outcome of ecological and evolutionary processes playing at local and regional scales, and to estimate the parameters of these processes based on observed patterns.

KEYWORDS

Approximate Bayesian Computation, coalescent, community assembly, dispersal limitation, environmental filtering, species pool, stochastic and deterministic processes

1 | INTRODUCTION

A major aim of community ecology is to identify the influence of ecological processes on species dynamics and coexistence from descriptive features of realized biodiversity patterns (McGill, Enquist, Weiher,

& Westoby, 2006; Weiher & Keddy, 1995). Heuristic assembly rules have been proposed to relate basic ecological drivers to typical features (Weiher & Keddy, 1995), but the interplay of multiple drivers at multiple scales requires more mechanistic and comprehensive understanding of their combined influence (Denny & Benedetti-Cecchi,

2012). Experiments are useful to address the influence of a given driver while controlling for others. However, practical constraints prevent devising experiments at the large spatial and temporal scales at which ecological, biogeographical and evolutionary dynamics are entangled. In this context, a simulation-based approach represents a sound alternative to assess the extent to which a given pattern can be related to a given process (Peck, 2004; Zurell et al., 2010).

Simulating ecological communities according to different scenarios can help to establish a benchmark against which to infer the signatures of assembly processes from resulting patterns of taxonomic, functional or phylogenetic diversity (Botta-Dukát & Czúcz, 2016; Pearse et al., 2015). In silico experiments are developing quickly in ecology and allow researchers to address the possible outcome of (meta)community models (Gravel, Canham, Beaudet, & Messier, 2006; Kembel, 2009; May, Giladi, Ristow, Ziv, & Jeltsch, 2013). With increasing computational power, it becomes possible to simulate many communities across broad ranges of parameter values. Approximate Bayesian Computation (ABC; Csillery, Blum, Gaggiotti, & François, 2010) and Monte Carlo Markov Chain (MCMC) methods allow exploring the outcome of complex models given prior distributions of their parameters, and then determining posterior distributions of the parameters consistent with observed patterns (Hartig, Calabrese, Reineking, Wiegand, & Huth, 2011).

Although simulation-based inference of parameter values from observed patterns of biodiversity is a major opportunity for ecological studies (Gimenez et al., 2014), complex algorithms for community simulation still require intensive and time-consuming computation. Here we introduce a novel simulation toolkit for community and functional ecologists who wish to investigate the signatures of ecological processes in community composition. Each community is assembled by combining immigration from outside the community with stochastic variations in the abundance of the descendants of successful immigrants (ecological drift, Vellend, 2016). The immigrants are drawn from a reference external pool that can represent either a large-scale biogeographic or habitat context (Lessard, Belmaker, Myers, Chase, & Rahbek, 2012), or dispersers available nearby the community (Lessard et al., 2016). Environmental filtering during immigration determines varying success of immigrants depending on the way their ecological attributes allow establishment and survival in a local environmental context (Jabot, Etienne & Chave, 2008; Keddy, 1992). This spatially-implicit framework follows a hierarchical conception of community dynamics embedded in a regional context (Lortie et al., 2004; Ricklefs, 2008), and provides a process-based alternative to randomization-based null models (Gotelli & Graves, 1996; Munoz & Huneman, 2016). It is intended to be quicker and more efficient than alternative algorithms, and is thus suited to intensive simulation schemes. We provide a complementary ABC tool to assess parameters of neutral dynamics and environmental filtering from the composition of real communities. The tool allows addressing whether we can unambiguously relate a given pattern (either taxonomic, functional or phylogenetic) to some hypothesized drivers.

We propose a coalescent-based algorithm of community simulation, which rebuilds the shared ancestry of coexisting individuals in

a community. Whether individuals share (or do not share) ancestors backward in time depends on immigration and local stochastic drift (Etienne & Olf, 2004). The genealogy of individuals is also called the *coalescent*. While coalescent-based modelling is popular in population genetics (Wakeley, 2004), it is seldom used in ecology (Etienne & Olf, 2004; Munoz, Couteron, Ramesh, & Etienne, 2007; Munoz, Couteron, & Ramesh, 2008; Munoz, Ramesh, & Couteron, 2014; Rosindell, Wong, & Etienne, 2008). We expose here the logic and advantages of the approach to examine community assembly. It is suited to simulation of neutral communities (Etienne & Olf, 2004; Hubbell, 2001; Munoz et al., 2007) but also allows deviations from neutrality based on niche-dependent immigration biases (Jabot, 2010; Janzen, Haegeman, & Etienne, 2015; Liu & Zhou, 2011; Munoz et al., 2014). Therefore, it allows simulation of ecological communities over a continuum of relative importance of neutral dynamics and environmental filtering.

This sampling scheme was previously implemented in ecology in MATLAB and PARI-GP softwares (Etienne & Olf, 2004; Munoz et al., 2007, 2014), or in C language (Rosindell et al., 2008), and the lack of easy-to-use tools may have limited the application for community simulation. Thus we here introduce the method in R language with the *ecolottery* package. The package also includes an alternative forward-in-time option of community simulation, which generalizes the algorithm of neutral dynamics formerly proposed in the *untb* package (Hankin, 2007) to the context of non-neutral dynamics. Although slower than the coalescent-based algorithm, the forward-in-time option allows exploring more diverse options of niche-based dynamics, such as limiting similarity related to competitive interactions (Abrams, 1983).

2 | GENEALOGY OF A LOCAL COMMUNITY: THE FRAMEWORK

The coalescent-based simulation rebuilds the genealogy of individuals in a community by sampling their immigrant ancestors from a reference pool of species while conditioning establishment and survival of their descendants to environmental filtering and neutral drift dynamics (Figure 1). In an initial neutral version of the model akin to Wright-Fisher/Moran models of population genetics, Etienne and Olf (2004) investigated the influence of a single immigration parameter, I . This parameter represents the role of successful dispersal, establishment of immigrants, and survival of descendants in determining the composition of the simulated community. Subsequent use of “immigration” will encompass all these components. Etienne and Olf (2004) further considered neutral evolutionary dynamics determining the species abundance distribution in a regional reference pool (or metacommunity) (Munoz et al., 2007, 2008). This approach was later generalized for any user-defined reference pool (Munoz et al., 2008).

In a version with environment filtering, the probability of successful immigration and subsequent survival of descendants also depends on the correspondence between species niche

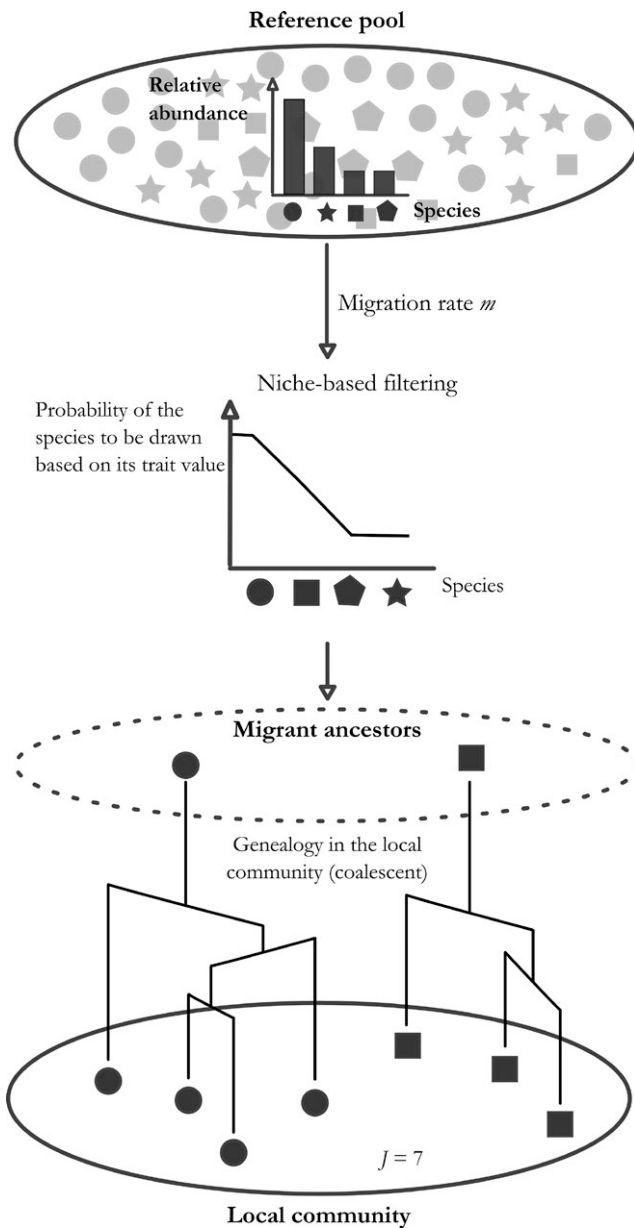


FIGURE 1 Analyzing assembly dynamics through the genealogy of individuals in a local community. The reference pool (top) includes individuals with varying ecological niche requirements (different symbols) and varying regional abundances (numbers of each symbol type). The individuals found in a local community (bottom) are the descendants of ancestors that immigrated and established in the past. The coalescent tree represents the genealogy of local individuals back to these migrant ancestors. A pool of ancestors (dashed ellipses) is drawn from the reference pool either randomly (neutral case), or depending on their ability to disperse, establish and survive in the local environment (niche-based filtering)

preferences (or some related functional trait values) and local environmental conditions (Figure 1) (Jabot et al., 2008; Munoz et al., 2014). The probability of drawing an ancestor immigrant of the genealogy from the pool is then weighted according to how well its functional trait values allow successful establishment (symbols in Figure 1). The relative species weights capture fitness differences

due to variation in functional traits among immigrants (Munoz et al., 2014).

The framework thereby integrates basic dispersal and environmental filtering of immigrants from a reference pool to a local community (Lortie et al., 2004), and stochastic drift dynamics of the descendants. As such, it represents a synthetic approach to jointly investigate environmental filtering and neutral dynamics in community assembly (Gravel et al., 2006; Leibold & McPeck, 2006; Vellend, 2016). An important aspect of the approach is that ecological and evolutionary dynamics can be decoupled in the community and in a regional pool of immigrants, respectively (Munoz et al., 2014). Community dynamics are assumed to be fast enough to neglect speciation events and trait evolution at the local scale; conversely, evolutionary and biogeographical dynamics (e.g. extinction and speciation events) can occur and determine biodiversity patterns at the larger spatial and temporal scales represented by a reference pool. The immigration process is here analogous to a mainland-island model in island biogeography (Hubbell, 2001; MacArthur & Wilson, 1967), where the regional pool undergoes biogeographical and evolutionary dynamics, while the community islands follow migration-drift dynamics with possible environmental filtering.

The reference pool can thus represent a biogeographic context at large spatial scale, but it can also be restricted to species occupying a specific habitat (Kraft, Valencia, & Ackerly, 2008), or to a spatially delimited neighbourhood surrounding the community (Lessard et al., 2016). In the latter case, the coalescent-based framework represents combined influence of species dispersal from a neighbourhood, environmental filtering and drift dynamics in the community. Alternative definitions of the pool can thus be used to address specific hypotheses on the respective roles of biogeography, environmental filtering, and neutral dynamics.

3 | SIMULATION ALGORITHM IN COALESC

coalesc is the core function in *ecolottery* performing coalescent-based simulation of ecological communities. Let us consider a given pool of species s with relative abundances f_s , and a community of size J . The migration rate is denoted as m . When an individual dies in the community, it can be replaced by the offspring of one of the $J - 1$ remaining individuals with probability $1 - m$, or by a migrant with probability m . The immigration parameter $I = m(J - 1)/(1 - m)$ represents the number of immigrants competing with local offspring to replace a dead individual in the community at each mortality event (Etienne & Olff, 2004).

The simulation algorithm determines which individuals in the sample share the same immigrant ancestor. The simulation begins by drawing a first individual from the reference pool (Figure 2): it is the ancestor of the first individual of the community. The taxonomic identity of this first individual is the identity of its ancestor. The second individual drawn from the community can be (i) either a descendant of the same ancestor as the first individual (in which case the second and the first individuals are conspecific), or (ii) a

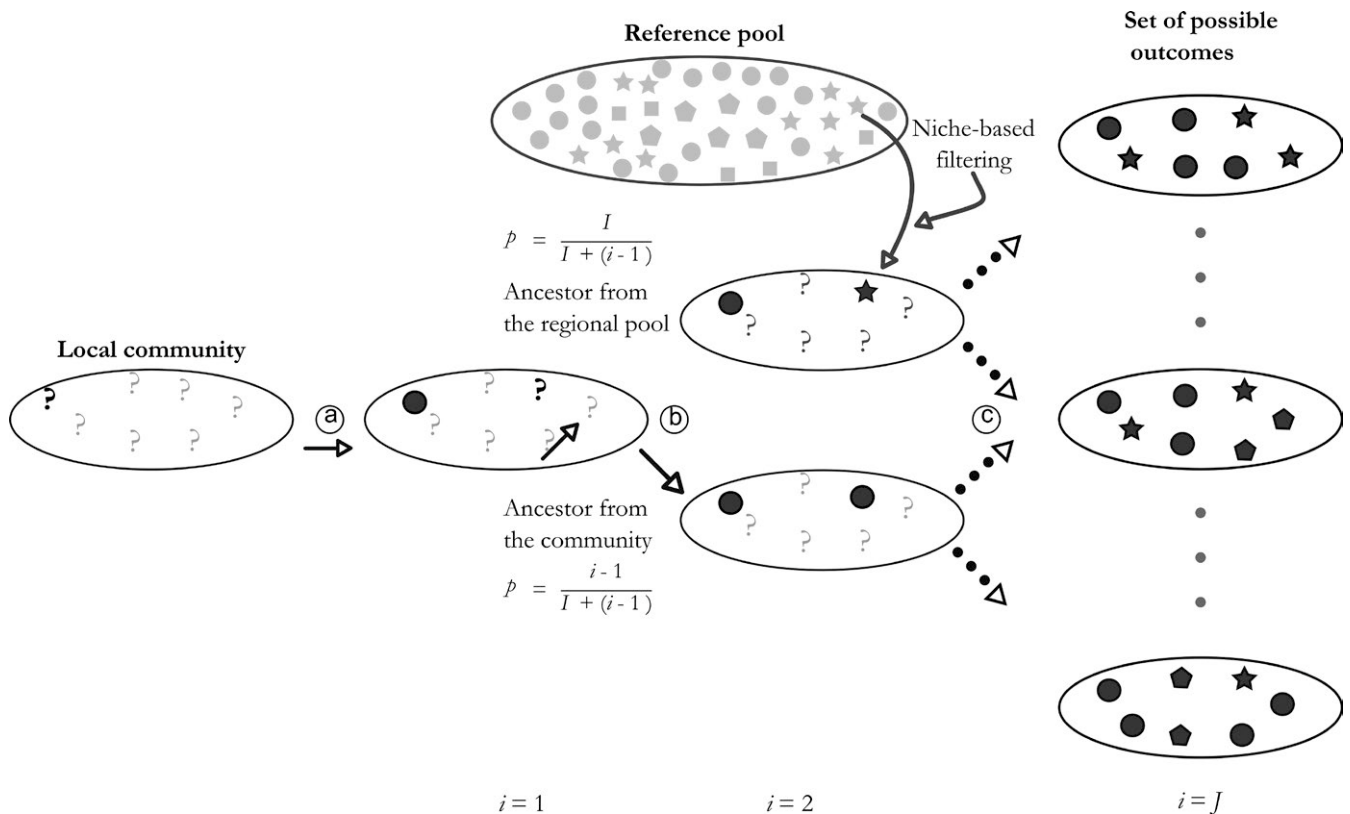


FIGURE 2 Coalescent-based sampling of a simulated community. Individuals with different trait values have different symbols. The top large ellipse represents the reference pool; the smaller bottom ellipses represent successive stages of coalescent-based sampling to generate a simulated community. In (a), the first individual sampled in the community is the descendant of an immigrant ancestor drawn from the reference pool. In (b), the second sampled individual is either the descendant of the same ancestor, or of another ancestor also drawn from the reference pool. Subsequent draws of individuals are performed with the same two choices of sharing an ancestor with previously sampled individuals, or being the descendant of a new immigrant ancestor, until sampling of J individuals in (c). Sampling of immigrants from the reference pool can depend on niche-based filtering, i.e., the probability of immigration success is weighted by the fit of species niche requirements to local environment

descendant of another immigrant ancestor. The probabilities of (i) and (ii) depend on the immigration parameter, I . If the descendants of immigrant ancestors have the same probability of survival until present, the probability of (i) is $1/(I + 1)$ and the probability of (ii) is $I/(I + 1)$. A binomial lottery is used to calculate whether the second individual descends from a new immigrant ancestor. If the ancestor is a new immigrant, it can be drawn from the reference pool and the identity of the second individual can be assigned accordingly. Note that the new immigrant can be, but need not to be, from a different taxon from the first immigrant. The third individual drawn from the community can be the descendant of one of the ancestors of the two first individuals, with probability $2/(I + 2)$, or the descendant of a new immigrant ancestor, with probability $I/(I + 2)$. In the former case, we randomly select one of the two first individuals as being conspecific of the third individual. Otherwise, we select a new immigrant ancestor of the pool as before. This process is repeated until reaching the desired number of individuals, J . Thus, individual i can be the descendant of one of the ancestors of the $(i - 1)$ first individuals, with probability $(i - 1)/(I + i - 1)$, or of a new immigrant ancestor, with probability $I/(I + i - 1)$.

With environmental filtering, the probabilities of sampling immigrant ancestors are affected by the match of their niche preferences with local environmental conditions. Let us consider that an immigrant bears a trait value t , which influences immigration success in the community depending on a filtering function $filt(t)$. The same sampling procedure as above is applicable, except that the immigrants are drawn from the pool with relative probability $filt(t)$. The higher this probability, the better adapted the immigrant to the local environment, and the greater chance it has to successfully establish in the community (Jabot, 2010; Munoz et al., 2014). Figure 3 illustrates different kinds of environmental filtering depending on the definition of $filt(t)$.

The function *coalesc* implements this basic simulation scheme. In the example below, a community of size $J = 400$ is simulated. $m = 1$ means that any dead individual in the community is replaced by an immigrant from the pool, while $m < 1$ means that both immigrants and local offspring can replace a dead individual. The user must also provide information on the composition of the reference pool, by setting either a θ value, to simulate a pool of size J_{pool} with log-series abundance distribution, or explicit species abundances in the pool for *pool*.

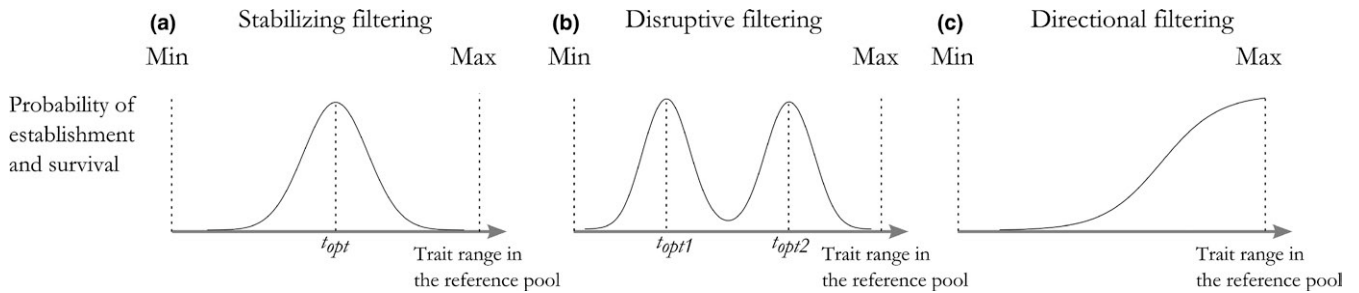


FIGURE 3 Several types of environmental filtering. The probability that an individual immigrates and establishes in a community can depend on its trait value(s) when environmental filtering occurs. Different kinds of filtering are illustrated: (a) a stabilizing environmental filtering around an optimal trait value called t_{opt} ; (b) a disruptive environmental filtering generating two modes around t_{opt1} and t_{opt2} ; (c) a directional environmental filtering with increasing probability of success along a gradient of trait values. Other types of niche-based filtering can be defined using the argument *filt*

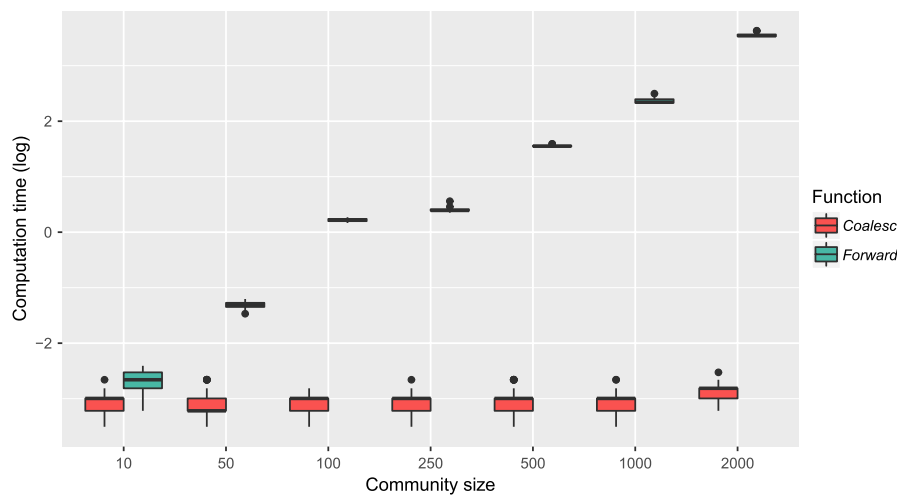


FIGURE 4 Comparison of computation time of alternative coalescent-based (red) and forward-in-time (blue) algorithms in *ecolottery* package. The simulation time in seconds is log-transformed. It is given for a single simulated community related to a given reference pool with log-series distribution of abundances and uniform distribution of trait values between 0 and 1 across species. In forward-in-time simulation, one individual dies at each time step and is replaced by either local offspring or by an immigrant. The migration rate is fixed to $m = 0.5$ and the immigrants undergo Gaussian environmental filtering around $t_{opt} = 0.5$, with standard deviation $\sigma = 0.1$. We simulated varying community size (abscissa). Simulation time quickly increased with community size in forward-in-time simulation, while it remained very low in coalescent-based simulation. See Appendix S2 for more details and the complete code

The *pool* object contains at least three columns; an individual index, a species index and the values of one or multiple traits. If trait values are not provided in *pool*, the user can provide them separately, using the *traits* object. In this case, it is assumed that *traits* includes species trait values, without intraspecific variation. If no trait information is provided, species trait values are drawn from a uniform distribution between 0 and 1.

```
res <- coalesc(J = 400, m = 0.5, theta = 50,
             filt = NULL, pool = NULL, traits = NULL)
summary(res)

##      Length Class      Mode
## com  3      data.frame list
## pool 3      data.frame list
```

The resulting object *res* has two components: community composition in *com* and reference pool composition in *pool*. A tutorial in Appendix S1 shows how to simulate neutral dynamics and environmental filtering in diverse situations with *coalesc*, and how to characterize the taxonomic, functional and phylogenetic compositions of simulated communities.

We also provide in *ecolottery* a sister function *forward* to simulate communities using a classical forward-in-time approach. From a given initial community composition, *forward* simulates at each time step a number of mortality events and replacement through local reproduction and immigration from an external pool. The *forward* function thus simulates communities with neutral dynamics and environmental filtering following the spatially-implicit framework of Figure 1. It can also condition the mortality events to trait

differences among the individuals present at each time step, and thus simulate limiting similarity (Abrams, 1983). It extends the scope of the *untb* function of Hankin (2007) to the context of non-neutral dynamics.

In Appendix S2, we compare the computation times of *forward* and *coalesc* for comparable parameterization of immigration and environmental filtering, and for varying community sizes.

Figure 4 shows a comparison of computation times between the two approaches. While calculation time of *coalesc* remains low and increases slowly with community size, calculation time of *forward* increases dramatically and is almost 3,000 times greater than *coalesc* for a community with 2,000 individuals. In addition, checking the convergence to stationarity of *forward* can be challenging, while there is no such issue with *coalesc*, since it simulates community composition at equilibrium (see Appendix S2).

4 | INFERENCE OF PARAMETERS USING APPROXIMATE BAYESIAN COMPUTATION

A long-standing objective of community ecology is to infer the ecological processes yielding observed patterns in community composition. A challenging issue is to disentangle the respective influences of neutral and non-neutral dynamics from their combined signatures in local communities. Because it is quick and efficient, *coalesc* can be used to simulate a large number of communities over a broad range of parameters of environmental filtering and neutral dynamics. By performing Approximate Bayesian Computation (ABC, Csillery et al., 2010), we can then compare the structure of an observed community to the simulated communities and determine the plausible parameter values corresponding to the observed community structure. We provide the *coalesc_abc* function in *ecolottery* to perform such ABC inference in a flexible way. Parallel computing is proposed to improve calculation speed in multi-core computers.

In the following example, a reference community is generated with known parameter values of immigration and environmental filtering. Stabilizing environmental filtering along a Gaussian distribution provides an establishment advantage to immigrants displaying trait values close to an optimum t_{opt} ; this advantage decreases from this optimum value according to the standard deviation, σ , of the Gaussian curve (see Figure 3).

The ABC analysis is based on simulating a large number of communities using *coalesc*, for varying values of migration rate m , of filtering parameters t_{opt} and σ , and for a user-defined reference pool. Summary statistics (calculated by the function *f.sumstats*) are used to characterize resulting patterns of taxonomic and functional composition in communities. Posterior distributions of parameter values for a given community are then derived according to a comparison of the summary statistics from this community with the summary statistics from simulated communities.

```
# Trait-dependent filtering function
filt_gaussian <- function(t, params)
  exp(-(t-params[1])^2/(2*params[2]^2))

# Definition of parameters of environmental filtering
and their range
params <- data.frame(rbind(c(0, 1), c(0.05, 1)))
row.names(params) <- c("topt", "sigmaopt")

# Number of values to sample in prior distributions
nb.samp <- 1000000 # Should be large

# Basic summary statistics
f.sumstats <- function(com) array(dimnames=
  list(c("cwm", "cwg", "cws", "cwk", "S", "Es")),
  c(mean(com[,3]), var(com[,3]),
  e1071::skewness(com[,3]),
  e1071::kurtosis(com[,3]),
  vegan::specnumber(table(com[,2])),
  vegan::diversity(table(com[,2]))))

# A reference community is here simulated
(know parameters)
comm <- coalesc(J = 400, m = 0.5, theta = 50,
  filt = function(x) filt_gaussian(x, c(0.2, 0.1)))

# ABC estimation of the parameters based on summary
statistics of the observed community
# The function makes vary the migration rate, m,
and the parameters of environmental
# filtering defined in params
res <- coalesc_abc(comm$com, comm$pool,
  f.sumstats = f.sumstats,
  filt.abc = filt_gaussian,
  params = params, nb.samp = nb.samp,
  parallel = T, tol = 1,
  pkg = c("e1071", "vegan"),
  method = "neuralnet")
plot(res$abc, param=res$par)

The posterior distributions of the parameters can be compared to
the expected values used to simulate the reference community. In addition,
cross-validation can be performed to evaluate the accuracy of
estimation for a broad range of parameter values, e.g. using the cv4abc
function of package abc.

pos.na <- which(rowSums(is.na(res$ss))!=0)
res$cv <- cv4abc(param = res$par[-pos.na,],
  sumstat = res$ss[-pos.na,], nval = 1000,
  tols = c(0.01, 0.1, 1), method = "neuralnet")
```

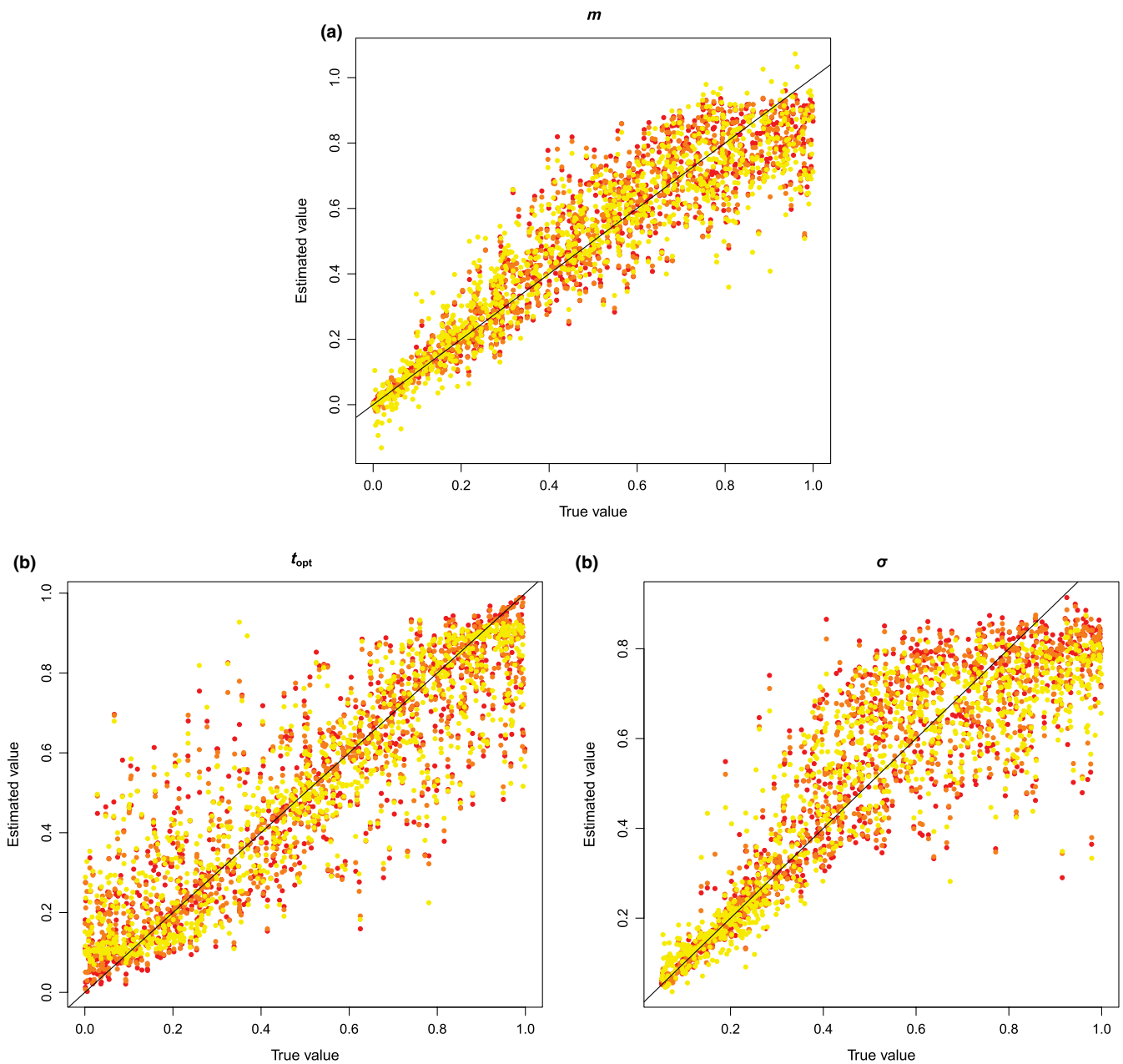



FIGURE 5 Cross-validation analysis of simulated communities with limited migration and stabilizing environmental filtering, using the function *plot.cv4abc* function of package *abc*. Neutral (migration rate m in a) and non-neutral (Gaussian stabilizing filtering with mean t_{opt} in b and standard deviation σ in c) parameters are estimated using *coalesc_abc*, based on the four first moments of the local trait composition, species richness and Shannon taxonomic diversity. 1,000 validation simulations are considered, expected parameter values are on abscissa and estimated values are on ordinates. Cross-validation is performed for three tolerance levels with increasing value from red to yellow (0.01, 0.1 and 1)

Figure 5 shows the result of cross-validation analysis. All the parameters are estimated reasonably accurately. m and t_{opt} are well estimated over the whole range of values between 0 and 1. σ is less well estimated when it becomes large, i.e. in this case a wide environmental filter selects all trait values almost equivalently and may not be discernible from pure neutral dynamics.

In Appendix S3, we also provide an example of ABC analysis to estimate neutral parameters in the Barro Colorado Island rainforest of Panama, with step-by-step explanation of the calculations.

5 | DISCUSSION

We propose a coalescent-based framework for quick and efficient simulation of ecological communities assembled through migration and environmental filtering from a regional species pool, and neutral stochastic drift (Vellend, 2016). The *coalesc_abc* function combines the simplicity and speed of coalescent-based simulation with powerful estimation of parameters using Approximate Bayesian Computation (ABC, Csillery et al., 2010). In this regard, it should be useful for a

broad audience of ecologists interested in exploring the linkage of ecological processes to biodiversity patterns and the role of processes in producing observed community composition.

Compared to other available simulation tools, a coalescent-based approach does not simulate the dynamics forward in time, but determines the shared ancestry of individuals sampled in a community at equilibrium. This offers several appealing properties for community simulation: first, we do not need to simulate the dynamics of all individuals that were present through time, but only to determine the lineages that have provided descendants in the observed community. Therefore, the coalescent-based approach is expected to be quicker by several orders of magnitude compared to a forward-in-time approach, as illustrated in Figure 4. Second, the sampling algorithm directly produces a community at a stationary state, so that there is no difficulty related to the choice of the number of generations to simulate or checks for convergence (see Appendix S2). Despite the significant gain of computation time, ABC analysis can still be computationally intensive when increasing the number of parameters and thus the number of simulations required. Therefore, reasonable complexity must be considered in these models.

The approach is in line with the current conceptual synthesis combining neutral and niche-based processes in the study of community dynamics (Munoz & Huneman, 2016; Vellend, 2016). The reference pool considered in simulations can result from neutral speciation-extinction dynamics (default case, Hubbell, 2001), be defined following specific scenarios of trait evolution (e.g. Appendix S1.6), or be based on some predefined flora or fauna information. We can then address how ecological processes driving local community assembly, and biogeographical and evolutionary processes underlying the composition of the reference pool, jointly influence patterns of taxonomic, functional and phylogenetic diversity. We can simulate one or several communities related to the same reference pool, and address patterns of *beta* diversity among communities depending on environmental filtering and neutral dynamics (Munoz et al., 2008, e.g. Appendix S3).

The spatial structure of communities can influence metacommunity dynamics, and this influence cannot be addressed based on a spatially implicit framework (Economato & Keitt, 2007). Recent advances in the modelling of spatially explicit coalescent should allow extending *coalesc* to model community assembly in spatially explicit networks (Kelleher, Etheridge, & Barton, 2014). Comparing the outcome of dynamics through the spatially implicit scheme of *ecolottery* to alternative spatially-explicit schemes will help better understand the emergent properties of spatial biodiversity dynamics. Currently it is not clear how reliably a spatially implicit approach can represent the outcome of spatially explicit dynamics and so how local to regional biodiversity dynamics can scale (Morozov & Poggiale, 2012; Munoz, Beeravolu, Pélissier, & Couteron, 2013).

Although adopting a spatially implicit approach, some recent works have proposed defining "process-based species pools" restricted to potential immigrants of a given habitat (Kraft et al., 2008), or to immigrants available from the neighbourhood of communities (Lessard et al., 2016). These alternative pools are considered in randomization schemes to define more or less restrictive null models (Gotelli & Graves,

1996), in order to address the influence of specific processes on simulated patterns. *coalesc* can also be used with alternate species pools and alternate parameterizations of environmental filtering and neutral dynamics, to address and test the relative influence of hypothesized drivers. It differs from randomization approaches that shuffle species trait values, site environmental characteristics, or community composition, but do not consider actual parameters of the ecological processes.

Addressing the influence of environmental filtering on species abundances and coexistence is a basic objective of community and functional ecologists (Cornwell & Ackerly, 2009; McGill et al., 2006; Shipley, Vile, & Garnier, 2006). The functions *coalesc* and *forward* allow custom definition of environmental filtering, through the input function *filt*. The filtering function can be designed to represent, for instance, stabilizing, directional, or disruptive filtering (Figure 3 and Appendix S1.4), in analogy with the basic types of selective pressures in evolutionary biology (Shipley, 2013). The filtering process represents a weighted lottery drawing immigrants from the reference pool (Lortie et al., 2004), where assigned weights depend on the properties of species (here based on functional traits defined in the *pool* or *traits* input arguments) and local environment (in the example of stabilizing filtering used in ABC estimation above, through the definition of t_{opt} and σ) and so can be visualized as the fit of each species' niche to local environmental conditions. The parameters m and *filt* of the weighted lottery then represent the respective roles of neutral dispersal limitation and trait-dependent environmental filtering during immigration and establishment.

The definition of environmental filtering is flexible and can integrate the influence of many biological traits (Cadotte & Tucker, 2017; Kraft et al., 2015). Therefore, we must not consider our definition of "trait" in the examples as a specific attribute, but as a more integrative measure of species fitness. Likewise, relating this species' "trait" to a local environment optimum t_{opt} in the example of environmental filtering represents a synthetic response of species to a set of possibly numerous environmental factors. The user can decompose the basic functional dimensions representing different components of fitness by addressing, for instance, the influence of dispersal traits versus the influence of traits related to local reproduction and competition. In the multi-trait example of Appendix S1.5, we show how to integrate the combined influence of filtering on several traits. The user may consider contrasted filtering operating on different traits, the outcome of this filtering on phylogenetic composition when niche conservatism differs among traits, etc.

Another major objective in community ecology is to address the extent to which limiting similarity and niche differentiation determine species coexistence (Abrams, 1983). It is possible to use *coalesc* to model disruptive filtering, i.e. only groups of species with contrasted traits can coexist in the community (see Appendix S1.4, in which we define such a filtering function). This would represent differentiation among ecological guilds, which may reflect, for instance, a vertical stratification of plants in vegetation, and where each guild represents a group of functionally similar species (Vergnon, Dulvy, & Freckleton, 2009). Another specific case of competitive interactions concerns the priority effect (Fukami, 2015). In this case, the order of species

arrival determines community composition, as first arrived species have a competitive advantage and prevent subsequent establishment of other species. Because the logic of priority effect is basically forward-in-time, it represents a case where coalescent-based approach cannot be applied. For this reason, we provide the sister function *forward* to simulate communities forward in time: with input argument *limit.sim = T*, community assembly is conditioned to the trait dissimilarity of coexisting species at each time step (Appendix S1.7). A forward-in-time simulation approach is also preferable to address the influence of density dependence, i.e. when species fitness varies over time depending on species abundance variations. Nevertheless, coalescent-based approaches can be designed to address the influence of fitness differences in local community dynamics as, e.g. by applying a rejection algorithm on the output of the coalescent process (Donnelly, Nordborg, & Joyce, 2001).

The user can analyse a number of metrics of taxonomic, functional and phylogenetic diversity in simulated communities (Munoz et al., 2007, 2008, 2014) to test the influence of entangled environmental filtering and neutral dynamics. In the examples of Appendix S1.4, we calculate average trait values (classically known as Community Weighted Mean, CWM, Garnier et al., 2004) in local communities undergoing environmental filtering along an environmental gradient. We also illustrate in Appendix S1.6 how niche conservatism and environmental filtering can produce phylogenetic clustering in the community (Mouquet et al., 2012). However, these approaches often consider *ad hoc* hypotheses about the influence of niche-based processes on these metrics, as, e.g. environmental filtering should decrease the range and variance of local trait variation, while shift of CWM should occur along environmental gradients (Cornwell & Ackerly, 2009). With the interplay of multiple processes driving local composition, how trait-based metrics reflect underlying drivers is expected to be less straightforward. Specifically, we can disentangle the influence of these drivers by performing cross-validation of parameter estimation in ABC estimation with *coalesc_abc*. As shown in the basic example of ABC estimation with neutral dynamics and Gaussian environmental filtering (Figure 5), cross-validation indicates that we can disentangle the influence of these drivers from summary statistics representing functional composition (the four first moments, CWM, CWV, CWS, CWK; Enquist et al., 2015) and taxonomic composition (here richness and Shannon diversity).

Another crucial issue in community and functional ecology is how intraspecific variation of ecological properties influences community dynamics (Bolnick et al., 2011; Violle et al., 2012). The niche breadth and overlap of coexisting species are basic components of competition theory (MacArthur & Levins, 1967). Coalescent-based simulation of community assembly in *ecolottery* represents individual dynamics based on individual properties, and is therefore suited to address the role of intraspecific trait variation among immigrants, when assuming perfect heritability and no mutation during community dynamics. In Appendix S1.3, an example of community simulation with intraspecific variation is provided. Changes in inter- and intra-specific trait variations are expected between regional and local scales depending on community assembly processes and their comparison allows inferring the influence of these processes (Violle et al., 2012). As the

descendants of immigrants display the same trait values than their ancestor in this approach, we can assess the diversity of phenotypes related to distinct genotypes, which is also appropriate to address the relationship between genetic and specific diversity in communities (Laroche, Jarne, David, Lamy, & Massol, 2015; Vellend & Geber, 2005). Further refinements should also allow addressing the influence of phenotypic plasticity.

6 | CONCLUSION

Even though the coalescent-based simulation scheme is a simplified approach to community assembly, it allows straightforward and quick simulation of community composition in diverse situations involving environmental filtering and neutral dynamics. It is suited for intensive simulation schemes, and thus allows ABC estimation of the parameters of assembly dynamics based on observed community composition. Thanks to great flexibility in parameterization of simulations and in the choice of summary statistics for ABC analyses, *ecolottery* will allow in-depth investigation of how the nature and quality of functional and phylogenetic information (e.g. types of phylogenetic and functional distances, structural versus ultrametric type of tree, resolution of the tree, continuous and categorical traits), and how the scale and the composition of reference pools (nature and uncertainty of regional abundance data, inclusion or not of rare species) can affect correct inference of underlying ecological processes. Therefore, *ecolottery* offers options to test theoretical hypotheses on the role of entangled processes, as well as a benchmark for estimating uncertainty in hypothesis testing depending on methodological choices (Perronne, Munoz, Borgy, Reboud, & Gaba, 2017).

ACKNOWLEDGEMENTS

P.D. and C.V. were supported by the European Research Council (ERC) Starting Grant Project “Ecophysiological and biophysical constraints on domestication of crop plants” (Grant ERC-StG-2014-639706-CONSTRAINTS). C.T. acknowledges the European Commission for the Marie Curie IIF (H2020-MSCA-IF-2014-657951). M.G. received PhD funding from the French Ecole Normale Supérieure.

AUTHORS' CONTRIBUTIONS

F.M. conceived the study and built the basic architecture of the *ecolottery* package. M.G. and P.D. did substantial work on adding functionalities, testing and cleaning code. All the authors contributed substantially to setting up the framework and to writing the manuscript.

DATA ACCESSIBILITY

The version 1.0.0 of *ecolottery* is available on CRAN, <https://cran.r-project.org/web/packages/ecolottery/index.html>. The development version is available on Github (<https://github.com/frmunoz/>)

ecolottery/tree/master/pkg), and can be installed by using the following function from package *devtools*,

```
devtools::install_github("frmunoz/ecolottery/pkg")
```

ORCID

François Munoz  <http://orcid.org/0000-0001-8776-4705>

Cyrille Violle  <http://orcid.org/0000-0002-2471-9226>

REFERENCES

- Abrams, P. (1983). The theory of limiting similarity. *Annual Review of Ecology and Systematics*, 14, 359–376.
- Bolnick, D. I., Amarasekare, P., Araujo, M. S., Burger, R., Levine, J. M., Novak, M., ... Vasseur, D. A. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology and Evolution*, 26, 183–192.
- Botta-Dukát, Z., & Czúcz, B. (2016). Testing the ability of functional diversity indices to detect trait convergence and divergence using individual-based simulation. *Methods in Ecology and Evolution*, 7, 114–126.
- Cadotte, M. W., & Tucker, C. M. (2017). Should environmental filtering be abandoned? *Trends in Ecology & Evolution*, 32, 429–437.
- Cornwell, W. K., & Ackerly, D. D. (2009). Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, 79, 109–126.
- Csillery, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate bayesian computation (abc) in practice. *Trends in Ecology & Evolution*, 25, 410–418.
- Denny, M., & Benedetti-Cecchi, L. (2012). Scaling up in ecology: Mechanistic approaches. *Annual Review of Ecology, Evolution, and Systematics*, 43, 1–22.
- Donnelly, P., Nordborg, M., & Joyce, P. (2001). Likelihoods and simulation methods for a class of nonneutral population genetics models. *Genetics*, 159, 853–867.
- Economu, E. P., & Keitt, T. H. (2007). Species diversity in neutral metacommunities: A network approach. *Ecology Letters*, 11, 52–62.
- Enquist, B., Norberg, J., Bonsor, S., Violle, C., Webb, C., Henderson, A., ... Savage, V. (2015). Scaling from traits to ecosystems: Developing a general trait driver theory via integrating trait-based and metabolic scaling theories. *Advances in Ecological Research*, 52, 249–318.
- Etienne, R. S., & Olff, H. (2004). A novel genealogical approach to neutral biodiversity theory. *Ecology Letters*, 7, 170–175.
- Fukami, T. (2015). Historical contingency in community assembly: Integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, 46, 1–23.
- Garnier, E., Cortez, J., Billes, G., Navas, M. L., Roumet, C., Debussche, M., ... Toussaint, J. P. (2004). Plant functional markers capture ecosystem properties during secondary succession. *Ecology*, 85, 2630–2637.
- Gimenez, O., Buckland, S. T., Morgan, B. J. T., Bez, N., Bertrand, S., Choquet, R., ... de Rexstad, E. (2014). Statistical ecology comes of age. *Biology Letters*, 10, 20140698.
- Gotelli, N. J. & Graves, G. R. (1996). *Null models in ecology*. Washington, DC: Smithsonian Institution.
- Gravel, D., Canham, C. D., Beaudet, M., & Messier, C. (2006). Reconciling niche and neutrality: The continuum hypothesis. *Ecology Letters*, 9, 399–409.
- Hankin, R. (2007). Introducing *untb*, an *r* package for simulating ecological drift under the unified neutral theory of biodiversity. *Journal of Statistical Software*, 22, 1–15.
- Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. (2011). Statistical inference for stochastic simulation models – Theory and application. *Ecology Letters*, 14, 816–827.
- Hubbell, S. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton, NJ and Oxford, UK: Princeton University Press.
- Jabot, F. (2010). A stochastic dispersal-limited trait-based model of community dynamics. *Journal of Theoretical Biology*, 262, 650–661.
- Jabot, F., Etienne, R. S., & Chave, J. (2008). Reconciling neutral community models and environmental filtering: Theory and an empirical test. *Oikos*, 117, 1308–1320.
- Janzen, T., Haegeman, B., & Etienne, R. S. (2015). A sampling formula for ecological communities with multiple dispersal syndromes. *Journal of Theoretical Biology*, 374, 94–106.
- Keddy, P. A. (1992). Assembly and response rules – 2 goals for predictive community ecology. *Journal of Vegetation Science*, 3, 157–164.
- Kelleher, J., Etheridge, A. M., & Barton, N. H. (2014). Coalescent simulation in continuous space: Algorithms for large neighbourhood size. *Theoretical Population Biology*, 95, 13–23.
- Kembel, S. W. (2009). Disentangling niche and neutral influences on community assembly: Assessing the performance of community phylogenetic structure tests. *Ecology Letters*, 12, 949–960.
- Kraft, N. J. B., Adler, P. B., Godoy, O., James, E., Fuller, S., & Levine, J. M. (2015). Community assembly, coexistence, and the environmental filtering metaphor. *Functional Ecology*, 29, 592–599.
- Kraft, N. J. B., Valencia, R., & Ackerly, D. D. (2008). Functional traits and niche-based tree community assembly in an Amazonian forest. *Science*, 322, 580–582.
- Larocque, F., Jarne, P., David, P., Lamy, T., & Massol, F. (2015). A neutral theory for interpreting correlations between species and genetic diversity in communities. *The American Naturalist*, 185, 59–69.
- Leibold, M. A., & McPeck, M. A. (2006). Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, 87, 1399–1410.
- Lessard, J.-P., Belmaker, J., Myers, J. A., Chase, J. M., & Rahbek, C. (2012). Inferring local ecological processes amid species pool influences. *Trends in Ecology & Evolution*, 27, 600–607.
- Lessard, J.-P., Weinstein, B. G., Borregaard, M. K., Marske, K. A., Martin, D. R., McGuire, J. A., ... Graham, C. H. (2016). Process-based species pools reveal the hidden signature of biotic interactions amid the influence of temperature filtering. *The American Naturalist*, 187, 75–88.
- Liu, J., & Zhou, S. (2011). Asymmetry in species regional dispersal ability and the neutral theory. *PLoS ONE*, 6, e24128.
- Lortie, C. J., Brooker, R. W., Choler, P., Kikvidze, Z., Michalet, R., Pugnaire, F. I., & Callaway, R. M. (2004). Rethinking plant community theory. *Oikos*, 107, 433–438.
- MacArthur, R., & Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*, 101, 377–385.
- MacArthur, R. & Wilson, E. (1967). *The theory of island biogeography*. Princeton, NJ: Princeton University Press.
- May, F., Giladi, I., Ristow, M., Ziv, Y., & Jeltsch, F. (2013). Metacommunity, mainland-island system or island communities? Assessing the regional dynamics of plant communities in a fragmented landscape. *Ecography*, 36, 842–853.
- McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, 21, 178–185.
- Morozov, A., & Poggiale, J.-C. (2012). From spatially explicit ecological models to mean-field dynamics: The state of the art and perspectives. *Ecological Complexity*, 10, 1–11.
- Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Bersier, L.-F., Chave, J., ... Thuiller, W. (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, 87, 769–785.
- Munoz, F., & Huneman, P. (2016). From the neutral theory to a comprehensive and multiscale theory of ecological equivalence. *The Quarterly Review of Biology*, 91, 321–342.
- Munoz, F., Beeravolu, C. R., Péliissier, R., & Coutron, P. (2013). Do spatially-implicit estimates of neutral migration comply with seed dispersal data in tropical forests? *PLoS ONE*, 8, e72497.

- Munoz, F., Couteron, P., & Ramesh, B. (2008). Beta-diversity in spatially implicit neutral models: A new way to assess species migration. *The American Naturalist*, *172*, 116–127.
- Munoz, F., Couteron, P., Ramesh, B., & Etienne, R. S. (2007). Estimating parameters of neutral communities: From one single large to several small samples. *Ecology*, *88*, 2482–2488.
- Munoz, F., Ramesh, B. R., & Couteron, P. (2014). How do habitat filtering and niche conservatism affect community composition at different taxonomic resolutions. *Ecology*, *95*, 2179–2191.
- Pearse, W. D., Cadotte, M., Cavender-Bares, J., Ives, A. R., Tucker, C., Walker, S., & Helmus, M. R. (2015). Pez: Phylogenetics for the environmental sciences. *Bioinformatics*, *31*, 2888–2890.
- Peck, S. L. (2004). Simulation as experiment: A philosophical reassessment for biological modeling. *Trends in Ecology & Evolution*, *19*, 530–534.
- Perronne, R., Munoz, F., Borgy, B., Reboud, X., & Gaba, S. (2017). How to design trait-based analyses of community assembly mechanisms: Insights and guidelines from a literature review. *Perspectives in Plant Ecology, Evolution and Systematics*, *25*, 29–44.
- Ricklefs, R. E. (2008). Disintegration of the ecological community. *The American Naturalist*, *172*, 741–750.
- Rosindell, J., Wong, Y., & Etienne, R. S. (2008). A coalescence approach to spatial neutral ecology. *Ecological Informatics*, *3*, 259–271.
- Shipley, B. (2013). *From plant traits to vegetation structure: Chance and selection in the assembly of ecological communities*. Cambridge, UK: Cambridge University Press.
- Shipley, B., Vile, D., & Garnier, E. (2006). From plant traits to plant communities: A statistical mechanistic approach to biodiversity. *Science*, *314*, 812–814.
- Vellend, M. (2016). *The theory of ecological communities (mpb-57)*. Princeton, NJ: Princeton University Press.
- Vellend, M., & Geber, M. A. (2005). Connections between species diversity and genetic diversity. *Ecology Letters*, *8*, 767–781.
- Vergnon, R., Dulvy, N. K., & Freckleton, R. P. (2009). Niches versus neutrality: Uncovering the drivers of diversity in a species-rich community. *Ecology Letters*, *12*, 1079–1090.
- Violle, C., Enquist, B. J., McGill, B. J., Jiang, L., Albert, C. H., Hulshof, C., ... Messier, J. (2012). The return of the variance: Intraspecific variability in community ecology. *Trends in Ecology & Evolution*, *27*, 244–252.
- Wakeley, J. (2004). *Coalescent theory – An introduction*. Greenwood Village, CO: Roberts & Company Publishers.
- Weiher, E., & Keddy, P. A. (1995). Assembly rules, null models, and trait dispersion – New questions front old patterns. *Oikos*, *74*, 159–164.
- Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Munkemuller, T., ... Schroder, B. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, *119*, 622–635.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Munoz F, Grenié M, Denelle P, et al. *ecolottery*: Simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods Ecol Evol*. 2018;9:693–703. <https://doi.org/10.1111/2041-210X.12918>

2 LA CONTRIBUTION DES TRAITS À DIFFÉRENTS PROCESSUS PEUT CAUSER DES DÉVIATIONS DE L'OPTIMUM FONCTIONNEL

How hierarchical competition and limiting similarity affect trait-abundance relationships in population-dynamic models

Pierre DENELLE^{1*}, Matthias GRENIÉ^{1*}, Cyrille VIOLLE¹, François MUNOZ², Caroline TUCKER³

¹ CEFE UMR 5175, CNRS - Univ Montpellier - CNRS – EPHE – IRD - Univ Paul-Valéry Montpellier 3 - 1919 route de Mende, F-34293 Montpellier, CEDEX 5, France

² University Grenoble-Alpes, LiPhy, 38041 Grenoble Cedex 9, France

³ Department of Biology, University of North Carolina at Chapel Hill, Coker Hall, CB #3280 120 South Road, Chapel Hill, NC 27599-3280, United States of America

* These authors contributed equally to this work.

Running title: Competition and species mismatches

Keywords: coexistence, environmental filtering, functional traits, hierarchical competition, limiting similarity, population dynamic

Abstract

Functional ecology stresses that the local performance of species depend on their functional trait values. Environmental filtering assumes that abiotic conditions determine an optimal trait value, such as species with trait value closer to the optimum should perform better. Trait-environment relationships should enable ecologists to detect this process. However, competitive interactions between species also influence local trait distributions and can lead to mismatches between the abiotic and the observed optima. Limiting similarity should be influential by entailing more even, still symmetric trait distributions, while hierarchical competition can favor an optimal strategy in terms of competitive ability.

A synthetic view on how these entangled processes can ultimately affect trait-environment relationships and the inferences from them is still lacking. It is also not clear whether using the abundances of species or their relative growth rates allows better inference of optimal trait values due to environmental filtering. Finally, whether a functional trait contribute to environmental filtering or to a competition process can entail different trait-environment relationships.

We devised a model integrating population dynamics, trait-based performance, environmental filtering, limiting similarity and hierarchical competition within a metacommunity. With this model, we analyzed mismatches between theoretical abiotic optima of species and the observed sites where species performed the best. We evaluated i) the impacts of competitive processes on species performance; ii) how the accuracy of inference of trait-environment relationships depends on performance metrics and biotic processes; and iii) the impact of multiple traits on inference of trait-environment relationships.

We found that hierarchical competition can entail strong mismatches between the environmental optimal sites of species and the one in which they perform best. While being strong at the species

level, these mismatches were averaged at the community-level, leading to trait-environment relationships close to the theoretical expectation. Due to its intrinsic symmetry, limiting similarity led to smaller mismatches. Both abundances and growth rates of species led to a similar accuracy of the trait-environment relationships. Higher trait contribution to the environmental filtering led to steeper trait-environment relationship.

Our model bridging population biology and trait-based ecology highlights the importance of accounting for assembly mechanisms when investigating trait-environment relationships. It questions the use of these relationships in functional biogeography as a predictive tool to track community response to environmental changes. Rather, we suggest that trait-environment relationships can be used as a diagnostic tool to reveal the role of the trait under scrutiny in the regulation of abiotic and biotic filtering.

Introduction

It has long been recognized that the diversity and composition of ecological communities change in response to underlying environmental gradients. Some phenotypic traits capture physiological tolerances or life history strategies (Violle et al., 2007). Measuring such trait values have therefore proven a useful tool for understanding the impacts of abiotic environment on the identity and abundances of species in a site (Mcgill, Enquist, Weiher, & Westoby, 2006; Bill Shipley, 2010). Furthermore, measuring trait values should be sufficient to predict performance (growth, survival, fecundity) in a given environment (Laughlin, Joshi, van Bodegom, Bastow, & Fulé, 2012; Lavorel & Garnier, 2002; Shipley, Vile, & Garnier, 2006; Weiher, Clarke, & Keddy, 1998). Multiple studies have reported strong relationships between specific trait values and performance under different environments ('trait-environment relationships') for a variety of organisms (Amatangelo, Johnson, Rogers, & Waller, 2014; Borgy et al., 2017; Pease, González-Díaz, Rodiles-Hernández, & Winemiller, 2012; Reich et al., 1999; Wright et al., 2005).

Knowledge of trait-environment relationships allows predicting the spatiotemporal distribution of species from data on relevant traits and site environmental conditions (Carboni et al., 2018; Verheijen et al., 2015; C. Violle, Reich, Pacala, Enquist, & Kattge, 2014; Wüest, Münkemüller, Lavergne, Pollock, & Thuiller, 2018). Specifically, in case of environmental filtering (Kraft et al., 2015), the abundance of a particular trait value is expected to reflect the performance of species bearing this value in a specific environment (Cingolani, Cabido, Gurvich, Renison, & Díaz, 2007; Garnier et al., 2004; Garnier, Navas, & Grigulis, 2016). Such relationship between performance and trait value is typically assumed unimodal, implying that a single optimal trait t_{opt} exists. The mean trait value in a community, weighted by the frequency of trait values, is commonly considered a proxy of the optimal trait value (community-weighted mean, CWM;

Borgy et al., 2017; Cornwell & Ackerly, 2009) (Fig. 1a). In addition to the mean trait value, environmental filtering is also expected to influence other trait moments. For a same t_{opt} , the environmental filtering strength can vary. As a response, the variance in trait values should decline as environmental filtering becomes stronger, since the cost of bearing trait values away from the optimum increase (Enquist et al., 2015) (Fig. 1a).

Despite the popularity of using the CWM to infer optimal trait values at a site (Enquist et al., 2015; Shipley, 2010; Warton, Shipley, & Hastie, 2015), there is still uncertainty regarding how well trait-environment relationships inferred from observation data reflect the true (eco-physiological) relationships (Blonder et al., 2017; Denelle, Violle, & Munoz, 2019; Forrestel et al., 2017; Garnier et al., 2016; Muscarella & Uriarte, 2016; Reich & Oleksyn, 2004; Bill Shipley et al., 2016; Yang, Cao, & Swenson, 2018). In principle, if environmental filtering is the dominant mechanism determining community composition, there should be a perfect linkage between a species' performance and their trait value, relative to the local optimum (Bohner & Diez, 2020; Kraft et al., 2015). It is still not straightforward to demonstrate this assumption, given the difficulty in obtaining experimental data over large environmental gradients or for long-lived species, and the likely presence of dispersal limitation and multiple biotic processes, including competition, predation, and mutualisms, in observational data (Cadotte & Tucker, 2017). These biotic processes can impact the distribution of traits in the community (Givnish, 1982; Grime, 2006; Mayfield & Levine, 2010; Swenson & Enquist, 2009), but their influence on trait-environment relationships is poorly understood, particularly when traits contribute to both biotic and abiotic processes.

Traits can contribute to competition in several ways, most commonly described in terms of 'limiting similarity' (MacArthur & Levins, 1967) or 'hierarchical competition' (Calcagno, Mouquet, Jarne, & David, 2006). Limiting similarity is usually thought as a force promoting the

coexistence of species by minimizing functional overlap and competition (MacArthur & Levins, 1967) (Fig. 1b). If a trait contributes to both environmental fitting and limiting similarity, environmental filtering should result in a final trait distribution in a community having a lower peak since similar competitors will reduce growth rates and allow species that are less affected by competition and slightly less adapted to the environment to survive. Hierarchical competition between species can act as a directional filtering towards trait values giving a direct competitive advantage (Fig. 1c) (Givnish, 1982; Herben & Goldberg, 2014; Keddy & Shipley, 1989; Loranger, Munoz, Shipley, & Violle, 2018). If a trait related to the environmental filtering is also associated with the outcome of hierarchical competition, the trait distribution in a site may be skewed towards the most competitive phenotypes, away from the true t_{opt} . Scenarios in which traits contribute to multiple processes, or in which processes are affected by several traits, entail even greater complexity in the linkage of trait values and actual performance. Deriving trait ~ environment relationship from competition-related traits should therefore lead to different patterns than with traits only related to the environmental filtering. Usually interpreted from the observation of local functional patterns (Spasojevic & Suding, 2012), we here evaluate how the contribution of one or more trait(s) to each of the three coexistence mechanisms affect the resulting trait ~ environment relationships.

Competition can produce mismatches between the sites where species perform best and the sites in which they should perform best according their trait value and the environmental filtering (Bohner & Diez, 2020). To estimate species' performance, studies of traits and environmental gradients most often rely on species abundances. However, observed abundances measured at a single given time may, because of temporal fluctuations or dispersal, not relate to the actual abiotic optimal trait t_{opt} . Weighting the local trait distribution by species' abundances may therefore lead

to incorrect inference of the actual t_{opt} . Demographic rates may better capture a species' performance in a site, for example, if instantaneous growth rate captures early success after colonization when competition processes are low (Aikens & Roach, 2014; Laughlin, Strahan, Adler, & Moore, 2018; Salguero-Gómez, Violle, Gimenez, & Childs, 2018). However, while acknowledging for the overall performance of species (Salguero-Gómez et al., 2018), growth rates also cover a multiplicity of processes, and can therefore be imperfect proxies (Aikens & Roach, 2014), or only provide short-term estimates of performance, especially in long-lived species. Identifying the circumstances in which different performance measures accurately depict trait-environment relationships remain to be determined.

Little information, regarding appropriate model or data, is available to guide expectations about how biotic interactions or choice of performance measure might affect inference on trait-environment interactions. One approach is to explore this issue by generating community data in which known abiotic conditions and biotic processes are manipulated and then measuring the resulting trait-environment relationships. Virtual ecology is, in that sense, extremely useful as it provides an approach that is carefully controlled, suitable for complex treatments and extensive replications (Zurell et al., 2010). Using a model integrating population dynamics, trait-based performance, and species interactions within a metacommunity, we evaluated a) impacts of competition processes on species mismatches between observed and theoretical optimal performance; b) how competition and the use of abundances or growth rates affect the accuracy trait-environment relationships at the community level; and c) when multiple traits are involved, the impact of trait contribution to environmental fitting on inference of trait-environment relationships.

Methods

Model Design. We developed a meta-community, discrete-tile model representing temporal changes in traits, species abundances, and distributions across sites arranged along an environmental gradient. Within sites, species dynamics are rules by a trait-based version of the classic Beverton-Holt competition model (Beverton & Holt, 1957; Leslie & Gower, 1960). We modelled and varied the strength of three ecological processes, specifically environmental filtering, hierarchical competition, and limiting similarity.

The model represents the temporal dynamics of a population, based on its intrinsic growth rate and the biotic interactions occurring in the site, as follows:

$$\text{Eqn (1)} \quad N_{t+1,i,x} = \frac{R_{i,x} \times N_{t,i,x}}{1 + A \times \alpha_i + B \times N_{t,i,x}}$$

with $N_{t+1,i,x}$ the abundance of species i in site x at time $t+1$, $R_{i,x}$ the growth rate of species i in site x , A is a constant tuning the global intensity of interspecific competition and α_i is the competition coefficient for species i . We also quantify the strength of intraspecific competition with parameter B . Intrinsic growth rates ($R_{i,x}$) are determined by how well a species is adapted to the local environment, which is a function of both the environment and the trait(s) values. This function takes the form of a Gaussian curve, where the growth rate depends on how close a species' trait is to the optimum trait for that environment (Munoz et al., 2018; Cyrille Violle et al., 2007):

$$\text{Eqn (2)} \quad R_{i,x} = k \times \exp\left(-\frac{(t_i - t_x)^2}{2l_x^2}\right)$$

with t_i the trait value of species i , t_x the optimal trait value in the environment of site x and l_x^2 the strength of the environmental filter in site x (e.g. the standard deviation of the Gaussian curve).

This function defines the proximal, physiologically determined relationship between traits,

environment and population growth. The environmental growth rate ($R_{i,x}$) is scaled by a constant k that sets the maximum possible rate.

Limiting similarity

Limiting similarity is incorporated in coefficients α , which are made dependent on functional distances between species,

$$\text{Eqn (3)} \quad \alpha_i = \sum_{j=1, j \neq i}^S N_{t,j,x} [\max(\delta_{ij}') - \delta_{ij}']$$

with δ_{ij}' the square of Euclidean distance between the traits of species i and species j :

$$\text{Eqn (4)} \quad \delta_{ij}' = (t_i - t_j)^2$$

To dissociate intraspecific and interspecific competition, the focal species does not contribute to calculation of α_i . In this way, the denominator of equation (1) increases when focal species i and its competitors j are functionally similar and as j are more abundant.

Hierarchical competition

In case of competitive hierarchy, the numerator of equation (1) is modified to include both the environmentally determined intrinsic growth rate and a hierarchical competition component,

$$\text{Eqn (5)} \quad R_{i,x} = R_{i,x,env} + R_{i,x,hierarch}$$

$R_{i,x,env}$ is the growth term due to environmental filtering as defined in equation (2) and $R_{i,x,hierarch}$ is the growth term related to hierarchical competition (Givnish, 1982; Grime, 2006). For instance, it can occur when smaller species are outcompeted for light and other resources by taller species.

In this logic, the term related to hierarchical competition is written as,

$$\text{Eqn (6)} \quad R_{i,x,hierarch} = -H \sum_{j=1, t_j \geq t_i}^S (t_j - t_i) N_{t,j,x}$$

With H a scalar for hierarchical competition, $N_{t,j,x}$ the abundance of species j at time t in site x , t_j and t_i the traits related to hierarchical competition for species j and i . Since we consider only species taller than the focal species, the difference t_j minus t_i is always positive. Hierarchical competition can then only lead to a decrease in growth rates as compared to the maximum possible.

Dispersal

Equation (1) represents local assembly dynamics in a single site. Individuals can also disperse and establish in communities. In our set of simulations, we defined $n = 25$ communities. At the end of each time step, and after the calculation of species growth rates, a proportion d of the individuals in a given site is evenly redistributed among all other communities:

$$\text{Eqn (7)} \quad D_{t,i,x} = -dN_{t,i,x} + d \frac{\sum_{k=1}^{25} N_{t,i,k}}{25}$$

With only one trait involved in the three coexistence processes, the full equation is:

$$\text{Eqn (8)} \quad N_{t+1,i,x} = \frac{k \times \exp\left(-\frac{(t_i - t_x)^2}{2l_x^2}\right) - H \sum_{j=1, t_j \geq t_i}^S N_{t,j,x} (t_j - t_i)}{1 + A \sum_{j=1, j \neq i}^S N_{t,j,x} [\max(\delta_{ij}') - \delta_{ij}'] + B \times N_{t,i,x}} N_{t,i,x} + D_{t,i,x}$$

Contributions of multiple traits

Traits can contribute to community dynamics through weighted contribution to one (or all) of hierarchical competition, limiting similarity and environmental filtering (i.e., w_h, w_c, w_g), e.g.:

$$\text{Eqn (9)} \quad R_{i,x} = k \times \exp\left(-\frac{(\sum_{g=1}^T w_g t_{i,g} - t_{x,g})^2}{2l_x^2}\right),$$

T being the total number of traits, w_g the weight of the focal trait in growth such that $t_{i,g}$ is the g^{th} trait of species i and $t_{x,g}$ the optimum value for the g^{th} trait in site x . Trait dissimilarity becomes:

$$\text{Eqn (10)} \quad \delta_{ij}' = \sum_{c=1}^T w_c (t_{i,c} - t_{j,c})^2,$$

With T the number of traits, w_c the weight of the considered trait in limiting similarity such that $t_{i,c}$ is the c^{th} trait of species i and $t_{j,c}$ the c^{th} trait of species j .

Hierarchical competition becomes:

$$\text{Eqn (11)} \quad R_{i,x,hierarch} = -H \sum_{h=1}^T w_h \sum_{j=1, t_j \geq t_i}^S N_{t,j,x} (t_{j,h} - t_{i,h}),$$

T being the number of traits, w_h the weight of the focal trait in hierarchical competition such that $t_{i,h}$ is the h^{th} trait of species i and $t_{j,h}$ is the h^{th} trait of species j .

The equation of community dynamics incorporating the varying contributions of functional traits is then:

$$\text{Eqn (12)} \quad N_{t+1,i,x} = \frac{k \times \exp\left(-\frac{(\sum_{g=1}^T w_g t_{i,g} - t_{x,g})^2}{2l_x^2}\right) - H \sum_{h=1}^T w_h \sum_{j=1, t_j \geq t_i}^S N_{t,j,x} (t_{j,h} - t_{i,h})}{1 + A \sum_{j=1, j \neq i}^S N_{t,j,x} [\max(\delta_{ij}') - \delta_{ij}'] + B \times N_{t,i,x}} N_{t,i,x} + D_{t,i,x}$$

Carrying capacity

Since the early discovery of the logistic equation describing population dynamics (Lotka, 1926; Verhulst, 1838; Volterra, 1927), a crucial parameter has been put forward to model the limitations each species encounter. The carrying capacity K models the density-dependence affecting a population of a given species and in a given site and implies a linear decline of per capita growth rate. While density-dependence is already at play with inter and intraspecific competitions, carrying capacity enables to have comparable number of individuals across range of parameters.

We added the parameter K to our global equation such as:

$$\text{Eqn (13)} \quad N_{t+1,i,x} = \frac{R_{i,x} + R_{i,x,hierarch}}{1 + A \times \alpha_i + B \times N_{t,i,x}} N_{t,i,x} \times \frac{(K - N_{t,i,x})}{K} + D_{t,i,x}$$

Importantly, in our multi-species coexistence model, the parameter K allows controlling for the overall number of individuals in a given site and make simulations with distinct sets of parameters comparable with each other.

Range of parameter values and initial conditions

We simulated 50 species across 25 sites; all species had initial populations of 50 individuals in all sites. A species went extinct when its number of individuals in a given site became inferior to 2, since we assume it cannot reproduce. Species were assigned continuous trait values drawn from a random, uniform distribution between 1 and 25. The optimal trait value t_{opt} across sites were matched with site number, such that t_{opt} equals 1 for site 1, continuing linearly to 25 in site 25. Our discrete-time simulations extended to 50 generations, which was sufficient to reach equilibrium for all sets of parameter values. For a given set of trait values, the model is deterministic with a single equilibrium.

We examined species coexistence across a broad range of parameter values (Table 1). We first considered scenarios in which only a single trait contributes to the processes. In a second step, two traits with varying contributions to the three coexistence processes were at play.

Simulations

1. Species and community mismatches. We generated three scenarios in which we included environmental filtering, environmental filtering and moderate limiting similarity, or environmental filtering and moderate hierarchical competition. To determine these moderate values of competition, we first looked at a broad range of limiting similarity and hierarchical competition parameter values. Then, we identified the parameter values for which species richness was high enough and where some mismatches arose (Fig. S1). At the species level, equation (2) determines the site (and associated environment) in which each species should reach its maximal performance: the true optimal site for a given trait value. However, due to the influence of competitive processes,

the sites where species reach their maximal performance could differ from those that are theoretically most suitable. For each species, we measured the observed mismatch in proportion to the environmental gradient. This proportional difference is calculated as the difference between the observed site in which a species has the highest performance and the true optimal site, divided by the true optimal site. We repeated this procedure for each of the performance measures obtained from the simulated data – observed abundance, and instantaneous, average, and maximum observed growth rates. Using the same data, we then generated the per-site community weighted mean trait value: for each of the performance metrics, we calculated the community-weighted trait values for each site. We calculated metrics growth rate using the temporal data generated by simulations. The average growth rate was taken across all 50 generations; the maximum growth rate was that maximum rate observed over all generations; and the instantaneous growth rate which relies on an estimation of the increase of species abundances. Specifically, we fit a linear model between log-transformed abundances and time, and then define the instantaneous growth rate as the exponential of the slope coefficient.

2. *Quantifying the contribution of competition to inaccuracies in inference of trait-environment relationships.* We performed 90 simulations across a wide range of H and A parameter combinations (Table 1) and calculated the observed trait-environment relationships using each of the four performance measures as described above. We fit linear regressions of the form (Observed community-weighted mean trait value – t_{opt}) $\sim A + H$.

We also modelled the overall variance in abundance relative to competition:

3. *Contribution to the environmental growth.* We also designed simulations with two traits involved, with varying contributions to the environmental growth component. Across the two traits, the contribution to one coexistence process sums to 100%. The first trait contributed to each process at either 0, 50 or 100%, while the second trait had either 100, 50 or 0% of contribution. For a given set of parameters, we computed CWM and CWV values for the first trait in each community, and compared it depending on the way the trait contributed to population growth.

Results

Species-level impacts of competition on performance

We asked whether competition led to differences in performance among species, and specifically, whether this led to mismatches between the sites in which a species had highest performance compared to the expected optimal site (determined by Eqn 2). The deviation or mismatch between the observed and expected sites varied in strength and direction (Fig. 2), as compared to the expectation (where $x=0$). Further, we wanted to understand how the result depended on the performance metric. Notably, when competition occurred via limiting similarity, most species had small or no mismatch, although the species with extreme trait values ($\sim <10$ or >40) tended to be found in sites with slightly closer to the edge of the gradient than expected. All performance metrics provided similar conclusions, with instantaneous growth rate providing the least accurate estimates for species at the edge of the environmental gradient. Pairwise correlations among the other metrics were above 50%, suggesting that all captured the same trait-performance relationship (Fig. 2a-inset). When competition was driven by trait hierarchies and when the same trait played on environmental filtering and hierarchical competition, we found that species mismatches were

common and much larger (leading to mismatches which were up to ~20% of the length of the environmental gradient (Fig. 2b)). These mismatches were asymmetrical – species with low trait values were observed in sites with significantly higher t_{opt} than expected, while species with moderate and very high trait values were observed in sites with significantly lower t_{opt} . Apart from few species with low traits and t_{opt} , all species had negative mismatches regardless of performance measure. Compared to abundance, the growth-rate measures of performance performed worse and lead to larger mismatches over at least part of the environmental gradient. For this form of competition, the performance metrics were not closely correlated, and abundance and average growth rate tends to be only weakly correlated with the maximum, or instantaneous rates of growth (Fig. 2b-inset).

Community-level impacts of competition on trait-environment relationships

We further addressed how mismatches at the species level could alter inference of trait-environment relationship among sites. Using the same data as in Fig 2, Fig 3 shows the relationship between t_{opt} values (grey line) and estimated values estimated using of the four potential measures of performance, averaged at community-level, plotted for comparison. Even when simulated communities experience only environmental filtering (Fig 3a), sites in extreme environments (edge of the environmental gradient) tend to deviate from the expectation as a result of truncated trait range and the limited numbers of species with positive growth rates in those sites. Limiting similarity (Fig. 3b) shows similar results between different performance metrics, and has lowered edge effects relative to Fig. 3c. When hierarchical competition is present, most measures of performance overestimate t_{opt} , but maximum growth rate appears to most closely estimate the true t_{opt} .

Contribution of competition type and strength to site-level mismatches

We performed linear regressions of observed community-weighted mean trait values as a function of competitive parameters (standardized A and H parameter values). A separate model was fit for each of the four performance measures, and the coefficient and t-value for a variable reported. The strength of the A parameter (limiting similarity) had similar impacts on the CWM trait value for all performance measures, suggesting that no measure performs better or worse for inferring trait environment-relationships when limiting similarity is present. The effect of the H parameter (hierarchical competition) on the CWM was much stronger: hierarchical competition was associated with positive coefficient values, that is, sites have higher average trait values than expected under the sole influence of environmental filtering. We also measured an abundance-weighted measure of trait variance for each site and modelled it as a function of the A and H parameters. Both limiting similarity and hierarchical competition have strong negative impacts on the trait variation in a site (See Fig 1).

Trait contribution and CWM ~ Environment relationship

With two functional traits involved in the model and a varying contribution to the environmental growth, we observed that the higher the contribution of the first trait to growth, the closer the CWM of this first trait from the theoretical expectation under the sole influence of the environmental filtering (Fig. 4). The community-weighted variance (CWV) decreased along with the contribution of the first functional trait to the environmental growth (Fig. 4). When the contribution of one trait to the hierarchical competition increased, the CWM values of this trait were higher through the whole environmental gradient (Figure S2). Increasing this contribution

also decreased CWV. The contribution to limiting similarity had no effect on CWM and CWV ~ Environment relationships (Fig. S3).

Discussion

Many processes operate simultaneously and interact in complex ways in community assembly, making the elementary mechanisms difficult to infer from observed patterns (Spasojevic & Suding, 2012). Mathematical simulations lend themselves to improving inference about such processes by allowing carefully controlled, highly replicated manipulations to test relevant hypotheses. The use of trait-environment relationships has become well-established in functional ecology, but the extent to which mismatches between the most frequent trait values at a site and that site's optimal trait value arise is an open question (Muscarella & Uriarte, 2016). Using simulations, we found that mismatches arise frequently, from a variety of causes. Importantly, we found that the mismatches seen at the species-scale can be strongly affected by competition but the inference of trait-environment relationships across sites can still be accurate.

Inaccuracies in estimating true trait-environment relationships were an expected outcome of the simulation's structure: the environmental gradient was linear and traits were constrained to values between 1 and 25. As a result, even when only environmental filtering occurred, the limited subset of traits for which positive growth was possible in those extreme environments creates asymmetrical trait distributions, which has been coined as the Trait Gradient Boundary Effect (Denelle et al., 2019). This asymmetry leads to higher mean trait values for communities in edge sites with low t_{opt} values and lower mean trait values than expected for high value t_{opt} sites. For this reason, even under environmental filtering only, the observed trait-environment relationship diverged from the true values at the edges (Fig. 3A). This effect is apparent for all scenarios. Where

environmental gradients with extreme conditions at one or more end occur, similar asymmetric effects on trait distributions may arise.

Limiting similarity was predicted to select species with suitable traits for environmental conditions, but which are sufficiently functionally distant so as to minimize competition. In fact, limiting similarity tended to result in only small mismatches in species' observed distribution and abundances relative to their expected optimal site (Fig. 2). Thus, trait-environment relationships were relatively consistent in the presence of limiting similarity, regardless of the performance metric used. In fact, the inference about the true trait-environment relationships with limiting similarity was better as compared to when only environmental filtering was present (Fig. 3A vs. 3B). The set of species suitable for a given environment already have reduced trait variation and so the relative functional distances and strength of limiting similarity in a site are fairly similar for all species. The net effect is that for all species performance is reduced and population sizes remain below carrying capacity, where low density, environmental-driven growth occurs. This strengthens the relationship between traits and each measure of performance. Accordingly, the strength of limiting similarity is only a moderate predictor of mismatches (Table 2).

Hierarchical competition had much stronger effects on how accurately observed data reflected true trait-environment relationships, and also had strong effects on species-level mismatches. Highly competitive species, provided their traits are within reasonable distance of the local t_{opt} , tended to dominate sites when hierarchical competition is at play. One key difference compared to limiting similarity is that hierarchical competition leads to asymmetry in species' competitive abilities, and if high values of a trait are most competitive, we expect the observed trait-environment

relationships to be skewed towards these higher trait values (as in Fig 1, Fig 3B). Thus, estimates of trait-environment relationships were typically higher than the true relationship when hierarchical competition is occurring (Fig 3) making it difficult to correctly infer the true t_{opt} directly (Bohner & Diez, 2020). The site-level results are in accordance with the observed species-level mismatches. Just as sites under these conditions tend to have higher trait values than expected, species with highly competitive traits (large values) tend to be present in sites with lower t_{opt} values and those with the least competitive trait values (small) tend to shift to sites with lower t_{opt} values. These species with higher trait values than optimal with result in higher mean traits at the site level. One exception to this pattern is at the very lowest trait values, in which species are more successful in sites with higher trait values than their optimum. These species are reasonably environmentally suited for neighboring sites and escape the strongest competitors, which are not environmentally suited for distant edge sites. Interestingly, despite our predictions, both forms of competition strongly decrease variation in the traits present in a site. As with environmental filtering, increasing strength of any process leads to a reduction in the suitable traits.

Increasingly, authors have suggested that demographic rates may be more informative than measuring abundances for these questions (Salguero-Gómez et al., 2018). For example, the maximum species growth rate may be better estimation of the local environmental optimum. Our model illustrated that each measure of performance varies in terms of when they provide the most accurate estimates. As seen in the pairwise correlations reported in Fig 2, the type of competition, the trait of a species, and its position along the environmental gradient all affect which measure performs best. In fact, abundances performed similar to measures of growth rates, despite being

common concerns that abundances can confound performance related to environmental conditions with biotic processes and dispersal (Fox, 2012).

There are other important processes potentially affecting coexistence and local functional patterns that we did not consider in our model. We did not study the explicit role of spatial dispersion, even though a dispersal component linking every community to the other ones was included. Dispersal limitation (Mouquet & Loreau, 2003) and mass-effects (Leibold et al., 2004; Pulliam, 2000) could indeed prevent the best adapted species from being present in a site and therefore generate mismatches affecting coexistence and the CWM values. Similarly, the model we used neglects important sources of demographic stochasticity (Hubbell, 2001). No external species pool was explicitly modeled and every species was present at the same initial abundance, preventing the appearance of any priority effect known as potential factor acting on trait distributions (Fukami, 2015).

Conclusions

Studies establishing trait-environment relationship at the community level rely on the assumption that we can use observational data to better characterize trait-environment relationships (e.g. Borgy et al., 2017). We proposed an original model explicitly representing of how functional traits contribute to environmental filtering, limiting similarity and hierarchical competition. Such models are needed in ecology (Herben & Goldberg, 2014) and can give insights about the generalization of trait-environment relationships (Forrestel et al., 2017). Our results highlight that even in the absence of competition, non-linearities can arise in the relationship between observed trait values and environmental conditions; where competition occurs, it appears that the type of

competition strongly affects the reliability of such inference (Table 2). Surprisingly, the choice of performance metric was relatively unimportant for the accuracy of a trait-environment relationship, and different performance metrics were accurate under different conditions. These results suggest that further work is needed to estimate the impacts of competition on trait distributions in field systems, and to validate field-measured estimates of trait-environment relationships with experimental and greenhouse data. Questions about how well observational data allows us to describe trait-environment relationships are difficult to address, but important if we are to be confident about their usage for predictive models.

Acknowledgements

This study was supported by the European Research Council (ERC) Starting Grant Project ‘Ecophysiological and biophysical constraints on domestication in crop plants’ (Grant ERC-StG-2014-639706-CONSTRAINTS and by the French Foundation for Research on Biodiversity (FRB; www.fondationbiodiversite.fr) in the context of the CESAB project ‘Causes and consequences of functional rarity from local to global scales’ (FREE). Computations were made on the Montpellier Biodiversity Bioinformatics (MBB) cluster computing platform of the Labex CeMEB “Centre Méditerranéen de l’Environnement et de la Biodiversité” (ANR-10-LABX-0004).

Authors’ contributions

PD, MG and CT designed the population-dynamic model, and performed the analyses and produced the results. PD and MG created the R package structure. All authors contributed critically to all drafts of the manuscript.

Data availability

R code to generate the simulated data is available on Github and can be downloaded as a R package.

References

- Aikens, M. L., & Roach, D. A. (2014). Population dynamics in central and edge populations of a narrowly endemic plant. *Ecology*, *95*(7), 1850–1860.
- Amatangelo, K. L., Johnson, S. E., Rogers, D. A., & Waller, D. M. (2014). Trait–environment relationships remain strong despite 50 years of trait compositional change in temperate forests. *Ecology*, *95*(7), 1780–1791.
- Beverton, R. J. H., & Holt, S. J. (1957). On the Dynamics of Exploited Fish Populations. Fisheries Investment Series 2. Vol. 19. *UK Ministry of Agriculture and Fisheries: London*.
- Blonder, B., Moulton, D. E., Blois, J., Enquist, B. J., Graae, B. J., Macias-Fauria, M., ... Sandel, B. (2017). Predictability in community dynamics. *Ecology Letters*, *20*(3), 293–306.
- Bohner, T., & Diez, J. (2020). Extensive mismatches between species distributions and performance and their relationship to functional traits. *Ecology Letters*, *23*(1), 33–44.
- Borgy, B., Violle, C., Choler, P., Denelle, P., Munoz, F., Kattge, J., ... Bahn, M. (2017). Plant community structure and nitrogen inputs modulate the climate signal on leaf traits. *Global Ecology and Biogeography*, *26*(10), 1138–1152.
- Cadotte, M. W., & Tucker, C. M. (2017). Should environmental filtering be abandoned? *Trends in Ecology and Evolution*, *32*(6), 429–437. doi: 10.1016/j.tree.2017.03.004
- Calcagno, V., Mouquet, N., Jarne, P., & David, P. (2006). Coexistence in a metacommunity: the competition–colonization trade-off is not dead. *Ecology Letters*, *9*(8), 897–907.
- Carboni, M., Calderon-Sanou, I., Pollock, L., Violle, C., Consortium, D., & Thuiller, W. (2018). Functional traits modulate the response of alien plants along abiotic and biotic gradients. *Global Ecology and Biogeography*, *27*(10), 1173–1185.
- Cingolani, A. M., Cabido, M., Gurvich, D. E., Renison, D., & Díaz, S. (2007). Filtering processes in the assembly of plant communities: Are species presence and abundance driven by the same traits? *Journal of Vegetation Science*, *18*(6), 911–920.
- Cornwell, W. K., & Ackerly, D. D. (2009). Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, *79*(1), 109–126.

- Denelle, P., Violle, C., & Munoz, F. (2019). Distinguishing the signatures of local environmental filtering and regional trait range limits in the study of trait–environment relationships. *Oikos*.
- Enquist, B. J., Norberg, J., Bonser, S. P., Violle, C., Webb, C. T., Henderson, A., ... Savage, V. M. (2015). Chapter Nine-Scaling from Traits to Ecosystems: Developing a General Trait Driver Theory via Integrating Trait-Based and Metabolic Scaling Theories. *Advances in Ecological Research*, *52*, 249–318.
- Forrestel, E. J., Donoghue, M. J., Edwards, E. J., Jetz, W., du Toit, J. C., & Smith, M. D. (2017). Different clades and traits yield similar grassland functional responses. *Proceedings of the National Academy of Sciences*, *114*(4), 705–710.
- Fox, J. W. (2012). When should we expect microbial phenotypic traits to predict microbial abundances? *Frontiers in Microbiology*, *3*, 268.
- Fukami, T. (2015). Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 1–23. doi: 10.1146/annurev-ecolsys-110411-160340
- Garnier, E., Cortez, J., Billès, G., Navas, M.-L., Roumet, C., Debussche, M., ... others. (2004). Plant functional markers capture ecosystem properties during secondary succession. *Ecology*, *85*(9), 2630–2637.
- Garnier, E., Navas, M.-L., & Grigulis, K. (2016). *Plant functional diversity: organism traits, community structure, and ecosystem properties*.
- Givnish, T. J. (1982). On the adaptive significance of leaf height in forest herbs. *The American Naturalist*, *120*(3), 353–381.
- Grime, J. P. (2006). Trait convergence and trait divergence in herbaceous plant communities: mechanisms and consequences. *Journal of Vegetation Science*, *17*(2), 255–260.
- Herben, T., & Goldberg, D. E. (2014). Community assembly by limiting similarity vs. competitive hierarchies: testing the consequences of dispersion of individual traits. *Journal of Ecology*, *102*(1), 156–166.
- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography* Princeton University Press Princeton.
- Keddy, P. A., & Shipley, B. (1989). Competitive hierarchies in herbaceous plant communities. *Oikos*, 234–241.
- Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., & Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, *29*(5), 592–599. doi: 10.1111/1365-2435.12345
- Laughlin, D. C., Joshi, C., van Bodegom, P. M., Bastow, Z. A., & Fulé, P. Z. (2012). A predictive model of community assembly that incorporates intraspecific trait variation. *Ecology Letters*, *15*(11), 1291–1299.

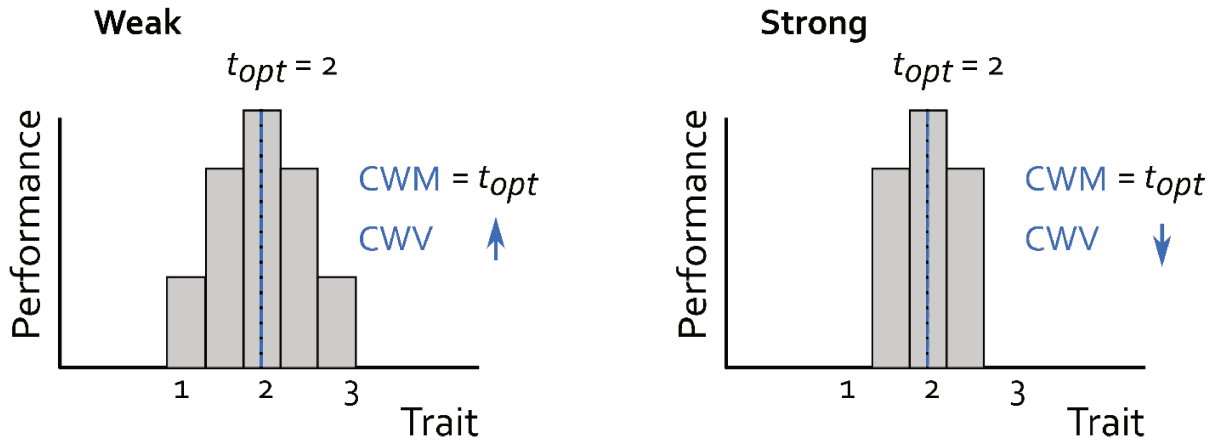
- Laughlin, D. C., Strahan, R. T., Adler, P. B., & Moore, M. M. (2018). Survival rates indicate that correlations between community-weighted mean traits and environments can be unreliable estimates of the adaptive value of traits. *Ecology Letters*, *21*(3), 411–421.
- Lavorel, S., & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, *16*(5), 545–556.
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., ... Gonzalez, A. (2004). The metacommunity concept: a framework for multi-scale community ecology: The metacommunity concept. *Ecology Letters*, *7*(7), 601–613. doi: 10.1111/j.1461-0248.2004.00608.x
- Leslie, P. H., & Gower, J. C. (1960). The properties of a stochastic model for the predator-prey type of interaction between two species. *Biometrika*, *47*(3/4), 219–234.
- Loranger, J., Munoz, F., Shipley, B., & Violle, C. (2018). What makes trait–abundance relationships when both environmental filtering and stochastic neutral dynamics are at play? *Oikos*, *127*(12), 1735–1745.
- Lotka, A. J. (1926). Elements of physical biology. *Science Progress in the Twentieth Century (1919-1933)*, *21*(82), 341–343.
- MacArthur, R., & Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*, *101*(921), 377–385.
- Mayfield, M. M., & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities: Phylogeny and coexistence. *Ecology Letters*, *13*(9), 1085–1093. doi: 10.1111/j.1461-0248.2010.01509.x
- Mcgill, B., Enquist, B., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, *21*(4), 178–185. doi: 10.1016/j.tree.2006.02.002
- Mouquet, N., & Loreau, M. (2003). Community patterns in source-sink metacommunities. *The American Naturalist*, *162*(5), 544–557.
- Munoz, F., Grenié, M., Denelle, P., Taudière, A., Laroche, F., Tucker, C., & Violle, C. (2018). ecolottery: Simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution*, *9*(3), 693–703.
- Muscarella, R., & Uriarte, M. (2016). Do community-weighted mean functional traits reflect optimal strategies? *Proceedings of the Royal Society B: Biological Sciences*, *283*(1827), 20152434. doi: 10.1098/rspb.2015.2434
- Pease, A. A., GONZÁLEZ-DÍAZ, A. A., RODILES-HERNÁNDEZ, R., & Winemiller, K. O. (2012). Functional diversity and trait–environment relationships of stream fish assemblages in a large tropical catchment. *Freshwater Biology*, *57*(5), 1060–1075.
- Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology Letters*, *3*(4), 349–361.

- Reich, P. B., Ellsworth, D. S., Walters, M. B., Vose, J. M., Gresham, C., Volin, J. C., & Bowman, W. D. (1999). Generality of leaf trait relationships: a test across six biomes. *Ecology*, *80*(6), 1955–1969.
- Reich, P. B., & Oleksyn, J. (2004). Global patterns of plant leaf N and P in relation to temperature and latitude. *Proceedings of the National Academy of Sciences*, *101*(30), 11001–11006.
- Salguero-Gómez, R., Violle, C., Gimenez, O., & Childs, D. (2018). Delivering the promises of trait-based approaches to the needs of demographic approaches, and vice versa. *Functional Ecology*, *32*(6), 1424–1435.
- Shipley, B., Vile, D., & Garnier, E. (2006). From Plant Traits to Plant Communities: A Statistical Mechanistic Approach to Biodiversity. *Science*, *314*(5800), 812–814. doi: 10.1126/science.1131344
- Shipley, Bill. (2010). *From plant traits to vegetation structure: chance and selection in the assembly of ecological communities*.
- Shipley, Bill, De Bello, F., Cornelissen, J. H. C., Laliberté, E., Laughlin, D. C., & Reich, P. B. (2016). Reinforcing loose foundation stones in trait-based plant ecology. *Oecologia*, *180*(4), 923–931.
- Spasojevic, M. J., & Suding, K. N. (2012). Inferring community assembly mechanisms from functional diversity patterns: the importance of multiple assembly processes. *Journal of Ecology*, *100*(3), 652–661.
- Swenson, N. G., & Enquist, B. J. (2009). Opposing assembly mechanisms in a Neotropical dry forest: implications for phylogenetic and functional community ecology. *Ecology*, *90*(8), 2161–2170. doi: 10.1890/08-1025.1
- Verheijen, L. M., Aerts, R., Brovkin, V., Cavender-Bares, J., Cornelissen, J. H., Kattge, J., & Van Bodegom, P. M. (2015). Inclusion of ecologically based trait variation in plant functional types reduces the projected land carbon sink in an earth system model. *Global Change Biology*, *21*(8), 3074–3086.
- Verhulst, P.-F. (1838). Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.*, *10*, 113–126.
- Violle, C., Reich, P. B., Pacala, S. W., Enquist, B. J., & Kattge, J. (2014). The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences*, *111*(38), 13690–13696. doi: 10.1073/pnas.1415442111
- Violle, Cyrille, Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, *116*(5), 882–892. doi: 10.1111/j.2007.0030-1299.15559.x
- Volterra, V. (1927). *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi*. C. Ferrari.
- Warton, D. I., Shipley, B., & Hastie, T. (2015). CATS regression—a model-based approach to studying trait-based community assembly. *Methods in Ecology and Evolution*, *6*(4), 389–398.
- Weihner, E., Clarke, G. P., & Keddy, P. A. (1998). Community assembly rules, morphological dispersion, and the coexistence of plant species. *Oikos*, 309–322.

- Wright, I. J., Reich, P. B., Cornelissen, J. H., Falster, D. S., Garnier, E., Hikosaka, K., ... Osada, N. (2005). Assessing the generality of global leaf trait relationships. *New Phytologist*, *166*(2), 485–496.
- Wüest, R. O., Münkemüller, T., Lavergne, S., Pollock, L. J., & Thuiller, W. (2018). Integrating correlation between traits improves spatial predictions of plant functional composition. *Oikos*, *127*(3), 472–481.
- Yang, J., Cao, M., & Swenson, N. G. (2018). Why functional traits do not predict tree demographic rates. *Trends in Ecology & Evolution*, *33*(5), 326–336.
- Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., ... Schröder, B. (2010). The virtual ecologist approach: simulating data and observers. *Oikos*, *119*(4), 622–635.

Figures and Tables.

a) Trait distribution with environmental filtering



b) Trait distribution with competition

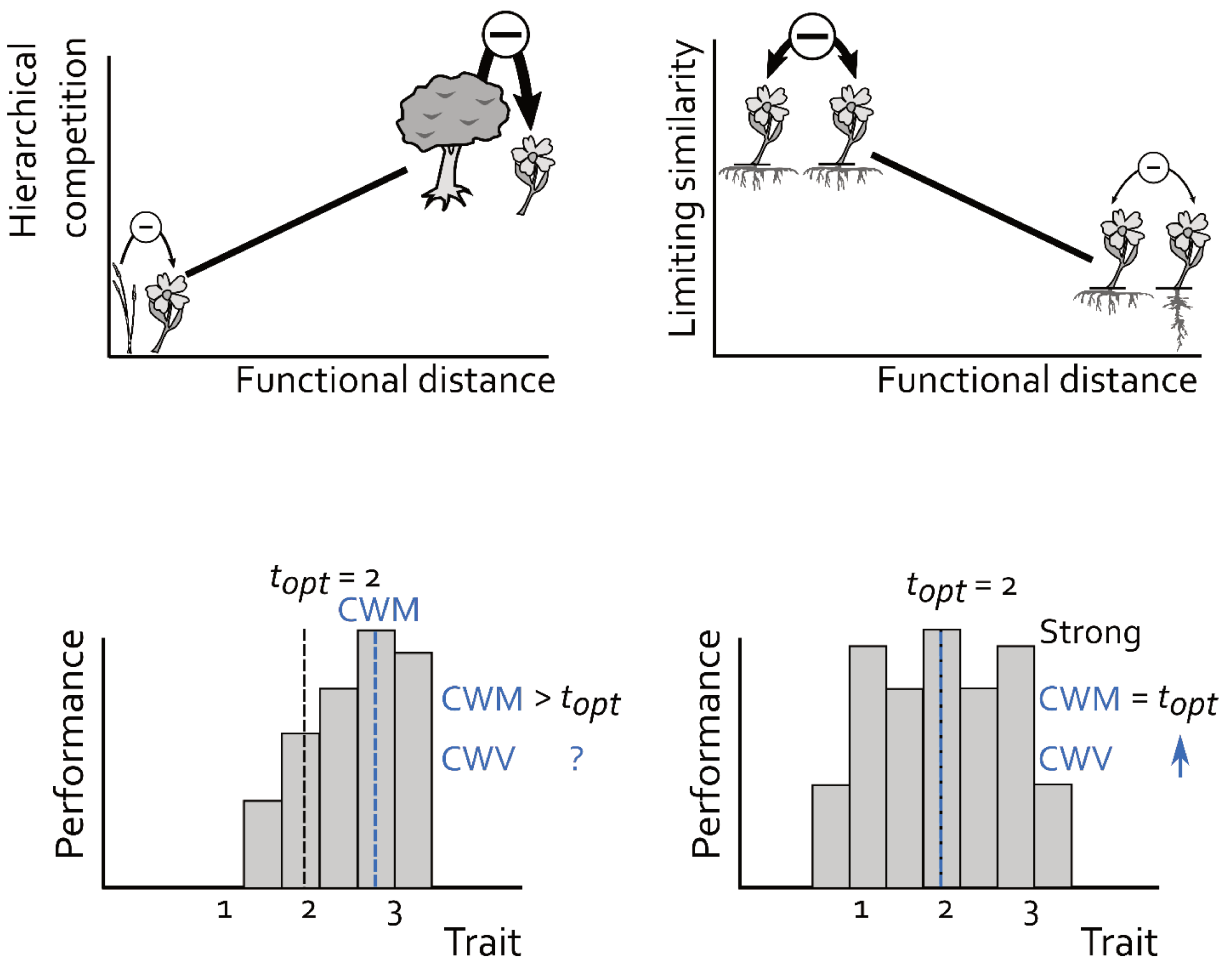


Figure 1. Predicted impacts of environmental filtering, hierarchical competition, and limiting similarity on the distribution of traits present in a particular environment. The true (i.e. eco-physiological) relationship between traits and a given environment is represented as a Gaussian function, with a local optimal trait value t_{opt} , providing the highest performance. **A)** Under environmental filtering, the community-weighted mean trait value (CWM) is expected to be equal to t_{opt} . The variation in traits present is predicted to decline as the strength of environmental filtering increases. **B)** The impacts of competition may depend on the form of competitive interactions: hierarchical competition favours species with specific trait values (e.g. those species with greater biomasses) and so may skew the observed CWM by allowing highly competitive but less environmentally suited species to survive in a site. Limiting similarity – in which functionally similar species compete most – may reduce the performance of similar, locally adapted species and allow greater local coexistence.

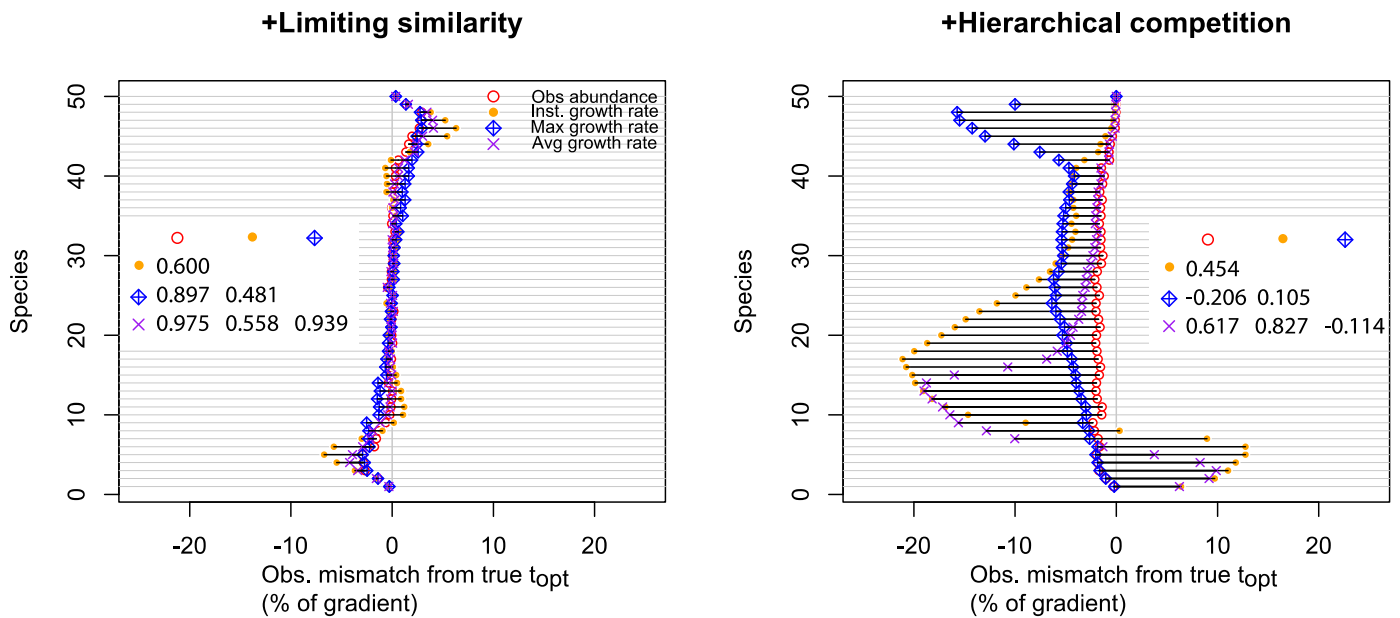
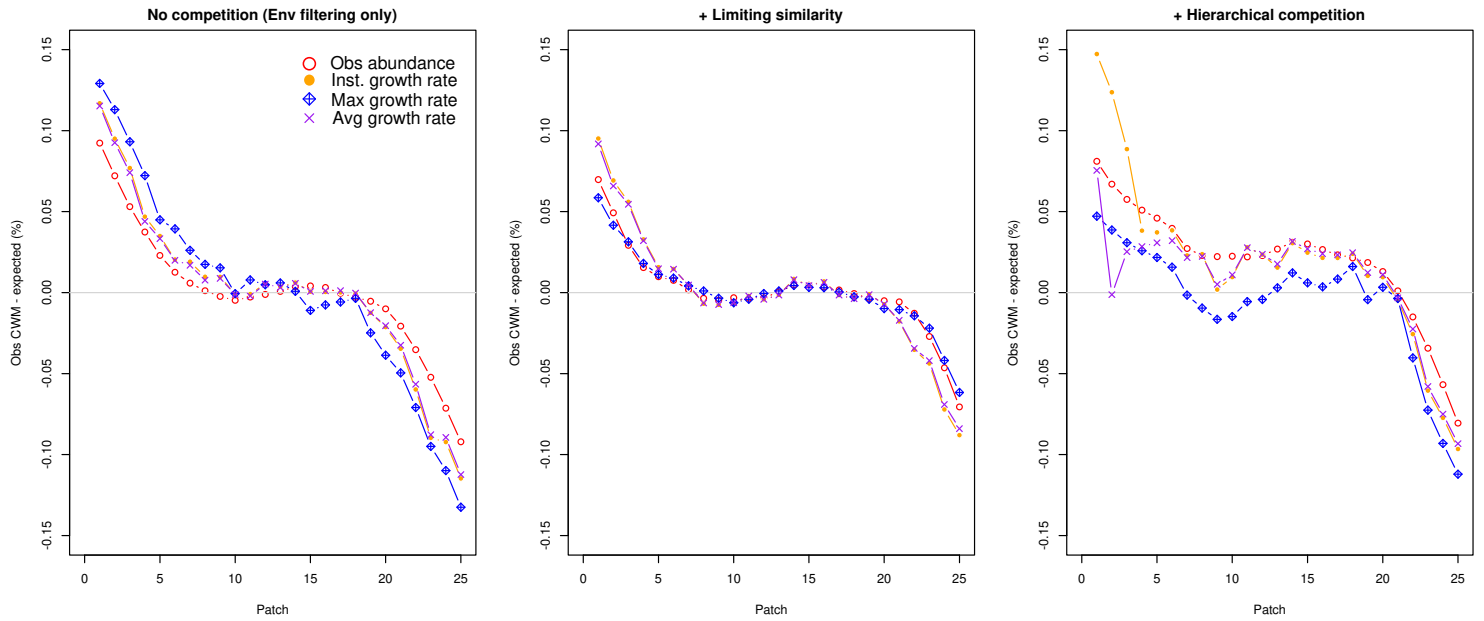
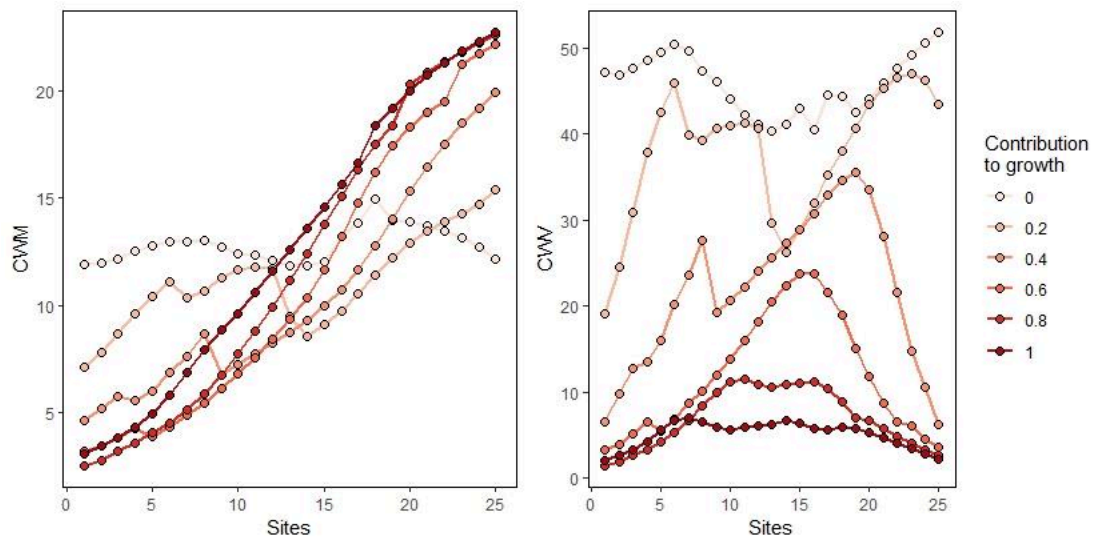


Figure 2. Mismatches between the site in which species' observed performance is highest and the true optimal site for their trait, expressed as a percentage of the gradient length. The true relationship between performance and a trait is determined by Eqn. 2 (leading to the expectation that trait 1 has the highest performance in site 1, etc.) and so if species are successful in the environment they are best suited to there will be a mismatch of 0 (grey line). However, species may have the highest performance in environmental conditions that differ from the one to which they are theoretically best suited. Mismatches show species with highest performance in environments with value below (negative values) or above (positive values) their optimum. Panels show simulated scenarios with competition occurring via limiting similarity or via hierarchical competition. Symbols represent four different measures of species performance – the observed abundance, and the instantaneous, maximum observed, and average growth rates. Insets show the Spearman pairwise correlations between each performance measure.

Figure 3. Deviation between the observed CMW trait values along the gradient and the true optimal trait

value (t_{opt}) as determined by the theoretical function (zero line, grey) used to generate the simulated communities. CWM values were generated using four potential measures of performance – abundance (red), instantaneous growth rate (orange), maximum growth rate (blue), and average growth rate (purple). Communities experienced either no competition, competition resulting from limiting similarity, or hierarchical competition. Note that simulated communities at the extremes of environmental conditions show the effect of truncation, and divergences are expected.



1
 2 **Figure 4.** Community-Weighted Mean (CWM) and Variance (CWV) for a scenario in which two
 3 independent traits contribute to environmental growth, but a single trait used to measure CWM values.
 4 The relative contribute of the two traits ranges from 0% and 100%: the contribute of trait 1 is indicated
 5 by the color gradient.
 6

7 **Table 1.** Parameter values used in the simulations. Values in square brackets represent those
 8 used in Scenario 2 simulations.
 9

Parameter	Description	Range/value
k	Growth rate scalar	2
A	Limiting similarity scalar	1e-3, [1e-8; 0.01]
B	Intra-specific competition scalar	1e-7
H	Hierarchical competition scalar	1e-3, [1e-5; 0.05]
d	Dispersal	0.05
K	Carrying capacity	100
I	Standard deviation of the Gaussian environmental function	5
	Replicates	100

10
 11

12 **Table 2.** A) Regression analysis of the community weighted trait value in a site as a function of the
 13 strength of limiting similarity (*A*) and hierarchical competition (*H*). Each performance measure was
 14 used to measure the community-weighted value and modelled as the response variable. (columns). B)
 15 The total variation in trait abundance modelled as a function of *A* and *H*. Columns show the
 16 standardized coefficients and the t value.

17

18 A)

	Abundance		Maximum GR		Instantaneous GR		Average GR	
	Coef	t value	Coef	t value	Coef	t value	Coef	t value
Intercept	-1.90	1.6	1.46	1.2	-1.5	1.1	-1.2	0.96
A	-6.90	2.2	2.9	0.889	-5.8	1.6	-5.1	1.43
H	0.87	18.2	0.96	20.74	0.93	1.726	0.94	18.38

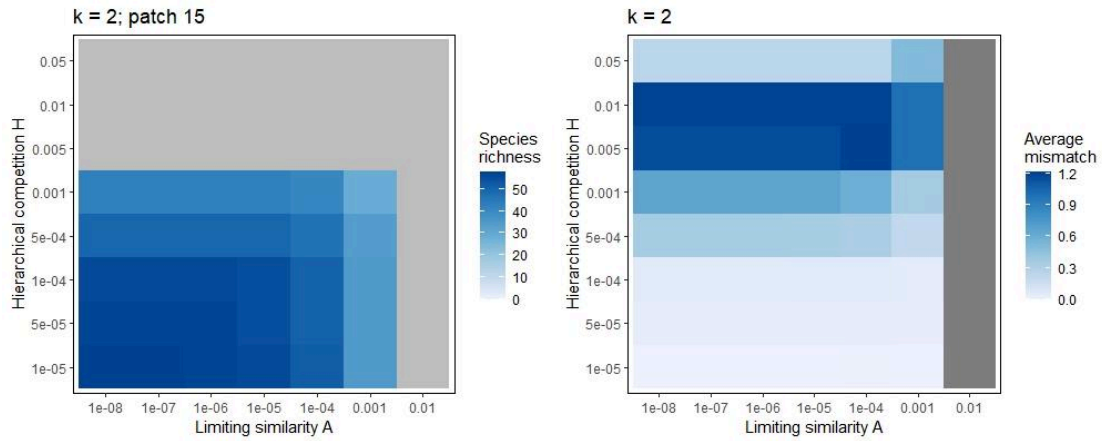
19

20 B)

Community weighted variation (Abundance)		
Intercept	-50.0	23.9
A	-146.8	-25.89
H	-337.5	-22.7

21

22

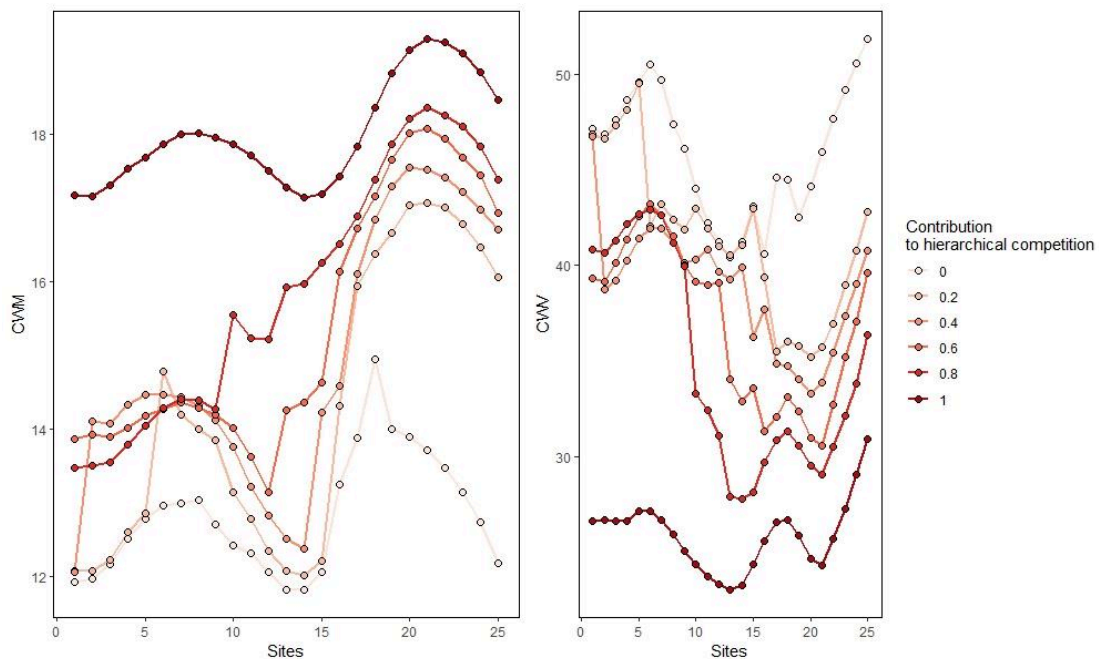
23 **Supplementary**

24

25 **Figure S1.** Landscapes of species richness and average mismatch between species optimum
 26 and their maximal abundances.

27 In both landscapes, the intensity of the limiting similarity A is increasing along the x-axis and the
 28 intensity of the hierarchical competition H is increasing along the y-axis. Each tile corresponds to a
 29 simulation ran with the corresponding values of both competitive processes. In the left panel, the color
 30 gradient corresponds to the species richness in the site 15 when the equilibrium is reached. Grey cells
 31 correspond to situations with no species. In the right panel, the color gradient corresponds to the average
 32 mismatch for all the species between the site where they have the highest theoretical environmental
 33 growth rate and the site where they reach their maximal abundances.

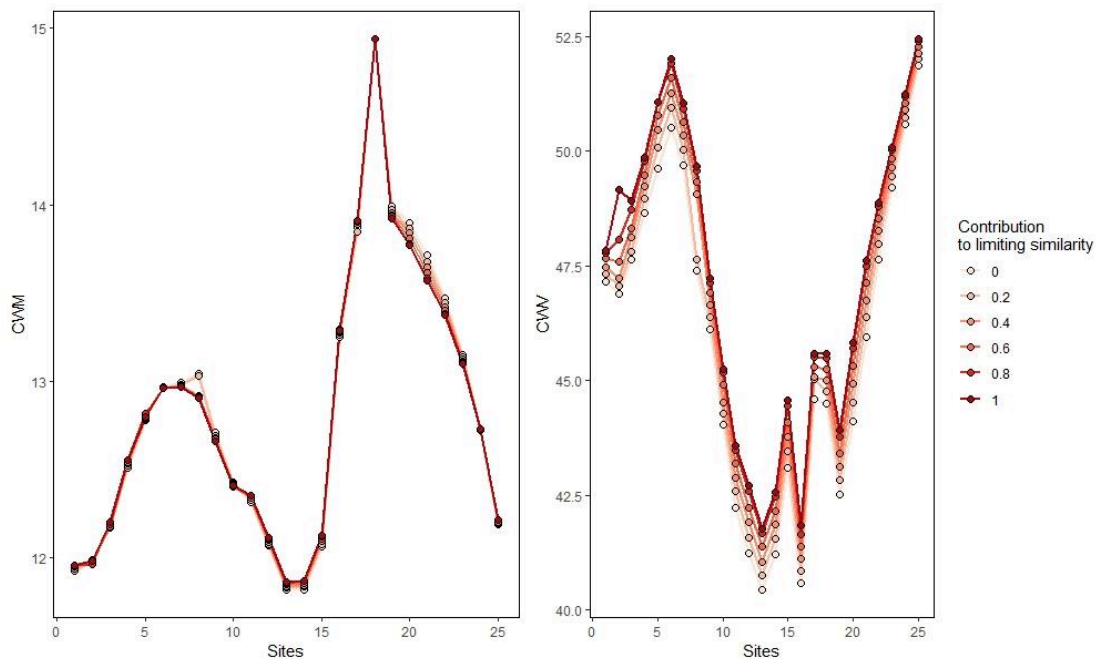
34



35
 36
 37
 38
 39
 40
 41
 42
 43

Figure S2. Community-Weighted Mean (CWM) and Variance (CWV) for varying contributions to the hierarchical competition term.

In these simulations, each species has two traits and both moments only integrate one of the two traits. The contribution of this focal trait to the growth varies between 0% and 100%, as figured by the color gradient. For each site and contribution, a Community-Weighted Mean, CWM, and Variance, CWV, is calculated across the range of parameters of Table 1. The mean of the CWMs and CWVs for a given contribution to growth is represented.



44

45 **Figure S3.** Community-Weighted Mean (CWM) and Variance (CWV) for varying
 46 contributions to the limiting similarity term.

47 In these simulations, each species has two traits and both moments only integrate one of the two traits.
 48 The contribution of this focal trait to the growth varies between 0% and 100%, as figured by the color
 49 gradient. For each site and contribution, a Community-Weighted Mean, CWM, and Variance, CWV, is
 50 calculated across the range of parameters of Table 1. The mean of the CWMs and CWVs for a given
 51 contribution to growth is represented.

52

2

Des processus à l'échelle de la communauté
peuvent faire dévier de l'optimalité

Plusieurs mécanismes peuvent empêcher l'optimalité fonctionnelle de s'établir dans les communautés. En effet, dans le chapitre précédent nous avons montré que les processus neutres pouvaient contribuer à modifier les abondances des espèces sans tenir compte de leurs traits. Ce chapitre présente des processus distincts du filtre environnemental, qui peuvent générer des attendus différents de composition des communautés.

Dans le chapitre précédent, nous avons évoqué les dynamiques source-puits qui peuvent permettre à des espèces d'être présentes dans des environnements où elles ne sont pas adaptées. De ce fait, ces espèces sont présentes dans des habitats non-favorables pour elles. D'autres phénomènes comme la dispersion limitée peuvent empêcher certaines espèces d'atteindre des habitats favorables. On appelle saturation d'habitat le fait pour une espèce d'être présente dans son habitat favorable. Certains processus peuvent donc conduire une espèce à "sur-saturer" son habitat, en étant présente à davantage d'endroits que sous l'hypothèse d'une stricte correspondance à son environnement. D'autres processus, peuvent conduire une espèce à ne pas saturer son habitat, en étant absente de certains habitats favorables. Or, cette saturation d'habitat variable peut avoir des effets sur les modèles de distribution d'espèce qui supposent que les espèces saturent leurs habitats. Ces effets, sont en particulier visibles quand on cumule les prédictions de plusieurs modèles individuels pour prédire la richesse spécifique. La première partie de ce chapitre, présente une étude utilisant des espèces virtuelles qui vise à évaluer l'influence de la saturation d'habitat sur les prédictions de richesse spécifique qui utilisent des modèles de distribution d'espèces empilés. Cette partie a fait l'objet d'une publication dans *Ecological Indicators*.

Lorsque qu'une communauté ne subit que le filtre environnemental et que l'on considère un trait unique, la valeur de ce trait pondérée par les abondances nous donne l'optimum fonctionnel de la communauté. Seulement, il est possible de considérer davantage de traits, en particulier en l'absence d'information préalable sur les traits les plus pertinents à utiliser le long de gradients environnementaux. Il est alors possible de relier l'environnement à des valeurs de diversité fonctionnelle et d'observer si le long d'un gradient environnemental, celle-ci varie selon l'attendu de l'optimalité fonctionnelle. Mais si les différents traits utilisés pour calculer la diversité fonctionnelle ne sont pas filtrés par l'environnement, alors il est possible que l'on ne détecte pas la tendance liée au filtre environnemental. La deuxième partie présente un article qui vise à évaluer le nombre de traits qu'il convient d'utiliser avant d'être exposé à cette dilution de signal. L'article est en préparation pour *Ecological Indicators*.

I LES ESPÈCES NE SATURENT PAS FORCÉMENT LEURS HABITATS CE QUI PEUT CAUSER UNE DÉVIATION À L'OPTIMALITÉ

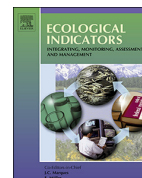
Ecological Indicators 111 (2020) 105970



ELSEVIER

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

Is prediction of species richness from stacked species distribution models biased by habitat saturation?

Matthias Grenié^{a,*}, Cyrille Violle^a, François Munoz^b^a CEFE, Univ. Montpellier, CNRS, EPHE, IRD, Univ. Paul Valéry Montpellier 3, Montpellier, France^b University Grenoble-Alpes, LECA, 2233 Rue de la Piscine, 38041 Grenoble Cedex 9, France

ARTICLE INFO

Keywords:

Habitat saturation
 Species richness
 Stacked species distribution models
 Predicted presence probabilities
 Threshold-based presence prediction

ABSTRACT

Several studies have proposed to predict Species Richness (SR) by combining the predictions of independent Species Distributions Models (SDMs) (the predict first-assemble later strategy). Alternative methods propose to combine outputs from SDMs differently, by either summing predicted presence probabilities at each location, or summing binary presence predictions after thresholding the probabilities. Species can occupy various proportions of their suitable habitats (i.e. have various levels of habitat saturation), which can cause discrepancy when predicting their presences through SDMs. Furthermore, these discrepancies can be increased when combining the predictions of individual SDMs to predict SR. In this article, we performed simulations of species distributions with varying habitat saturation (i.e., the amount of suitable habitat occupied by a species), and we compared observed richness with that predicted by the alternative approaches. We found that probability-based richness is not biased by the level of habitat saturation, while threshold-based richness over-predicts richness at low habitat saturation and under-predicts it as high habitat saturation. Probability-based richness should thus be used in priority when predicting species richness locally. Nonetheless, threshold-based richness represents species richness constrained by environmental filtering only and thus is a useful indicator of potential species richness when species fully saturate their habitats. Thus the systematic comparison of probability-based and threshold-based richness predictions can reveal the importance of habitat saturation and can thus help identify community assembly mechanisms at play.

1. Introduction

Species Richness is an Essential Biodiversity Variable (EBV) (Pereira et al., 2013), which should be assessed, monitored and compared across space, time, and ecological contexts. Different models have been proposed for richness prediction in diverse ecological contexts and at large spatial scale (Dodson, 1992; Graham and Hijmans, 2006; O'Brien, 1998), with the perspective of identifying biodiversity hotspots (Mazel et al., 2014; Myers et al., 2000), targeting effective management practices (Chown et al., 2003), quantifying biodiversity changes (Newbold et al., 2015) and predicting ecosystem functioning (Cardinale et al., 2012).

Several methods can be used to predict richness depending on which ecological processes are at play. For example, Macro-Ecological Models (MEMs) directly predict richness at any location as a function of local environmental variables. These models consider the influence of environmental filtering and energy limits on richness (Hurlbert and Stegen, 2014). Because site-species data are first aggregated to estimate

richness and then used to predict the variation with the environment, these approaches are called 'assemble first, predict later' (Ferrier and Guisan, 2006). Conversely, more and more global and local biodiversity databases include species occurrences instead of local assemblage composition (GBIF, 2019; Sullivan et al., 2009; Tedesco et al., 2017). An alternative approach has been to first model occurrences, independently for each species, at any location using environmental variables through species distribution models (SDMs) (Guisan and Thuiller, 2005; Guisan and Zimmermann, 2000), then to deduce potential local richness by combining (=stacking) the predictions of individual SDMs (Calabrese et al., 2014; D'Amen et al., 2015b; Gavish et al., 2017; Scherrer et al., 2018; Schmitt et al., 2017), which is known as the 'predict first, assemble later' approach (Ferrier and Guisan, 2006). When stacking SDMs, each SDM predicts occurrences for species independently using environmental variables (Guisan and Zimmermann, 2000). Then, predictions of SDMs for different species are summed to predict richness at assemblage-level. Stacked-SDMs (S-SDMs) predict observed richness as well as or better than macro-

* Corresponding author.

E-mail addresses: matthias.grenie@ens-lyon.fr (M. Grenié), cyrille.violle@cefe.cnrs.fr (C. Violle), fmunoz@univ-grenoble-alpes.fr (F. Munoz).

ecological models (Dubuis et al., 2011; Guisan and Rahbek, 2011), but there is still no consensus on the stacking method to be used so as to reliably predict richness with S-SDMs (Scherrer et al., 2018).

Two main methods exist to stack SDMs (Dubuis et al., 2011; Pineda and Lobo, 2009; Scherrer et al., 2018). Some authors suggested using thresholds to convert probabilities to binary predictions (presence and absence) (Jiménez-Valverde and Lobo, 2007; Liu et al., 2005). These binary predictions are then summed to predict richness at local scale (hereafter threshold-based richness). One of the main arguments for conversion of probabilities provided by SDMs to binary predictions is that most of practical applications need binary maps (Jiménez-Valverde and Lobo, 2007). A caveat of binary predictions is that they translate continuous responses of species along environmental gradients into binary responses, which imply more abrupt shifts from presence to absence between suitable and unsuitable conditions (Meynard and Kaplan, 2012). When predicted probabilities are under the threshold, the model only predicts absences, while it only predicts presences when predicted probabilities are above it. Close to the threshold value, a small change in predicted presence probability can change the binary prediction from absence to presence. Meynard and Kaplan (2012) showed that presence predictions using thresholds fit observed presences only when species has a threshold-like response, while error increases when a species response is more gradual. The more species considered that have a gradual response along the environment, the greater the error when predicting richness. SDMs also directly provide continuous presence probabilities as outputs (Guisan and Thuiller, 2005), and threshold conversion to binary predictions adds a step compared to the direct sum of individual model predictions. Summing the probabilities of individual species model provides the mathematical expectation of the number of species locally present, assuming that species occurrences are independent (Calabrese et al., 2014; Violle et al., 2011), hereafter called probability-based richness.

A basic implicit assumption of SDMs is that only environmental conditions determine species occurrence, depending on a species fundamental niche (Guisan and Zimmermann, 2000). Additional processes should affect the realized occupancy patterns, such as dispersal limitation, competitive exclusion, local extinction dynamics (Pulliam, 2000). SDM predictions and thus richness predictions are likely to be biased by neglecting the contribution of processes shaping realized species distributions beyond their fundamental niche requirements (Václavík and Meentemeyer, 2012), thereby affecting SDM predictions and thus richness predictions. For instance, due to source-sink dynamics, some species can occupy less suitable sites, and thus be distributed outside the suitable habitat delimited based on presence probabilities predicted by SDMs. In addition, a species that is less often present across its suitable habitat would have a lower predicted presence probability than a species that is present in all its suitable habitats, even though the predicted binary distribution of an SDM would be the same. We define habitat saturation of a species as a parameter that affects species occurrence probability based on environmental suitability. Here saturation is a species-level property and not an upper bound for richness in assemblages as proposed by Mateo et al. (2017). When species display low habitat saturation, their realized presence probabilities decrease, so that the predicted summed probability gets lower. On the contrary, the threshold-based presence prediction is not affected, by habitat saturation. Indeed, even if the determined species threshold changes with habitat saturation, the prediction will still be binary (presence or absence) (Meynard and Kaplan, 2012), thus we expect to observe increasing difference between threshold-based and observed richness with lower (or higher) habitat saturation.

Predicted presence probabilities partly reflect the ability of species to saturate their niche. Therefore, we expect probability-based richness to best predict actual richness. While we expect threshold-based richness to over-predict actual richness. Threshold-based richness rather represents a pool of species able to occur in given environmental conditions. To test these expectations we simulated virtual species with

varying saturation and niche requirements (Hirzel et al., 2001; Meynard et al., 2019). We performed S-SDMs to predict richness given environmental conditions using both threshold- and probability-based richness and compared how the predictions were affected by habitat saturation. Probability-based richness followed observed richness whatever the habitat saturation, while threshold-based richness only matched observed richness when habitat saturation was 100%. Threshold-based richness only considered the environmental requirements of species, and could thus be used as the prediction of potential richness based solely on local environmental conditions. Potential richness could then be compared with other richness predictions that incorporate other ecological processes.

2. Material and methods

2.1. Species assemblage simulations

Individual species simulation. We simulated a linear environmental gradient of 2000 values, from 1 to 2000. We then used the virtual species package version 1.4-2 (Leroy et al., 2016) to define 100 species independently, with quadratic environmental response $s_{i,k} = a \times Env_k^2 + b \times Env_k$, with $s_{i,k}$ the environmental suitability of species i in assemblage k and Env_k the environmental variable. a was drawn from a uniform distribution between -20 and -0.01 . b was chosen as $b = -m * 2 * a$ where m was drawn from a uniform distribution between 1 and 2000 and represents the environment of maximum suitability. The suitability was then scaled between 0 and 1 by subtracting its minimum and dividing by the difference of its maximum and minimum. We used the function `generateRandomSp()` in `virtualespecies` to get suitability probabilities for each species and each environmental value (see Fig. 1 left column).

Habitat saturation and predicted assemblages. We simulated species presences along the environmental gradient by performing binomial draws based on the presence probabilities. The presences probabilities $p_{i,k} = s_{i,k} \times \beta$ depend on (i) the suitability probabilities defined above, $s_{i,k}$ for species i and assemblage k , reflecting fundamental niche requirements, and (ii) an additional habitat saturation coefficient β representing the ability of species to occupy their suitable habitat (realized niche). When saturation is below 100%, the species tend to be less often present in suitable sites than species at 100% saturation (e.g., due to dispersal limitation or extinction). Species can also reach a saturation over 100% when they are present in less suitable conditions than according to their fundamental niche (e.g., through source-sink dynamics). We simulated 8 values of β : 10%, 40%, 70%, 100%, 120%, 150% and 170%. If the weighted probability of presence was greater than one, we reduced it to a maximum of one. We thus simulated each species assemblage k for each value β .

2.2. Individual and stacked species distribution models

We performed Species Distribution Models (SDM) based on simulated species presences.

Modeling and Predicting Presences. We modeled the presence of each species using two predictors: the environmental value and the square of this value (see Fig. 1 middle column) in Generalized Linear Models (GLM) of the binomial family:

$$\text{logit}(p_{i,k}) = \beta_0 Env_k + \beta_1 Env_k^2, \quad (1)$$

with $p_{i,k}$ the presence of species i in assemblage k and Env_k its associated environmental variable. We thus estimated in each assemblage the probability of finding each species. For each species we determined the best threshold to get binary predictions by maximizing the True Skill Statistic (TSS) (Allouche et al., 2006). The TSS balances the proportion of presences correctly predicted and the proportion of absences correctly predicted.

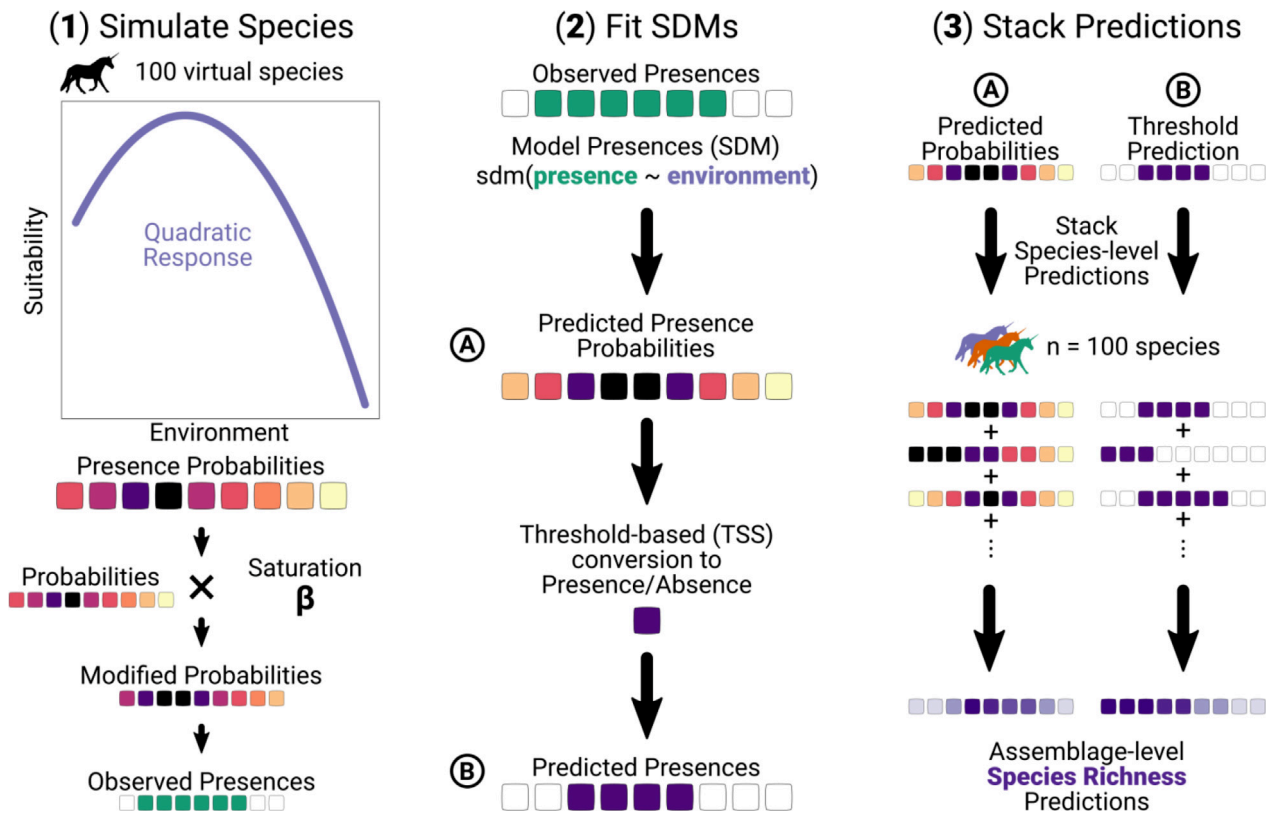


Fig. 1. Full Simulation routine. (1) We first simulated 100 virtual species with quadratic environmental suitability curves with randomly sampled coefficients. We multiplied each predicted presence probability by the habitat saturation level then use these probabilities to draw realized presences (see Material and Methods for details). Then, using the modified probabilities we drew presences in each assemblage following a binomial distribution. (2) We analyzed the realized presences with a binomial Generalized Linear Model (GLM), independently for each species, which provided predicted presence probability of each species in each assemblage (A). We defined a threshold based on True Skill Statistic (see Material and Methods for details, Allouche et al., 2006). This gave the second set of predictions: (B) binary predictions. (3) Finally, we summed individual predictions in each assemblage for all the species to get two richness predictions.

Predicting Species Richness. We stacked SDM predictions in each assemblage to get a prediction of richness, with two approaches. We first summed the predicted presence probability for each species (probability-based richness, prediction (A) in Fig. 1):

$$pred_{rich,prob,k} = \sum_{i=1}^S p_i(k), \quad (2)$$

with $pred_{rich,prob,k}$ the probability-based predicted richness in assemblage k , S the total number of species in the species pool, and $p_i(k)$ the predicted presence probability of species i in assemblage k . Using these probabilities we determine a species-specific threshold t_i using the True Skill Statistic (Allouche et al., 2006) that defines a binary function $1_i(k)$ to predict the presence of the species in each assemblage:

$$1_i(k) = \begin{cases} 1 & \text{if } p_i(k) \geq t_i \\ 0 & \text{if } p_i(k) < t_i \end{cases}, \quad (3)$$

with $p_i(k)$ the predicted presence probability of species i in assemblage k and t_i the species i threshold defined using TSS. We then compared the sum of predicted presence probabilities $pred_{rich,prob,k}$ to the sum of predicted presences with species-specific threshold (threshold-based richness, prediction (B) in Fig. 1):

$$pred_{rich,thresh,k} = \sum_{i=1}^S 1_i(k), \quad (4)$$

with $pred_{rich,thresh,k}$ the threshold-based predicted richness in assemblage k , S the total number of species in the species pool and $1_i(k)$ the indicator function defined as above.

We examined how predicted richness fitted observed richness across

the whole environmental gradient, for different levels of habitat saturation. We quantified the deviation with Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N_k} \sum_{k=1}^{N_k} (pred_{rich,k} - obs_{rich,k})^2}, \quad (5)$$

with $pred_{rich,k}$ the predicted richness of a given method in assemblage k , $obs_{rich,k}$, the observed richness in this assemblage, and N_k the total number of assemblages. We defined Bias and Variance components:

$$Bias = \frac{1}{N_k} \sum_{k=1}^{N_k} (pred_{rich,k} - obs_{rich,k}) \quad (6)$$

$$Variance = \frac{1}{N_k} \sum_{k=1}^{N_k} (pred_{rich,k} - \overline{pred_{rich}})^2 \quad (7)$$

with $\overline{pred_{rich}}$ the average predicted richness of a given method across all assemblages.

All analyses and SDMs were performed using R version 3.5.2 (R Core Team, 2019). A version of the code used in this article is archived on Zenodo (<https://doi.org/10.5281/zenodo.3552836>).

3. Results

Binary predictions (solid segments above and below the plot) showed few differences whatever habitat saturation (Fig. 2). There were the same from environment 1–273, then between environment 467 and 1514, and for environments greater than 1720. In total binary predictions were the same whatever habitat saturation for over 80% of

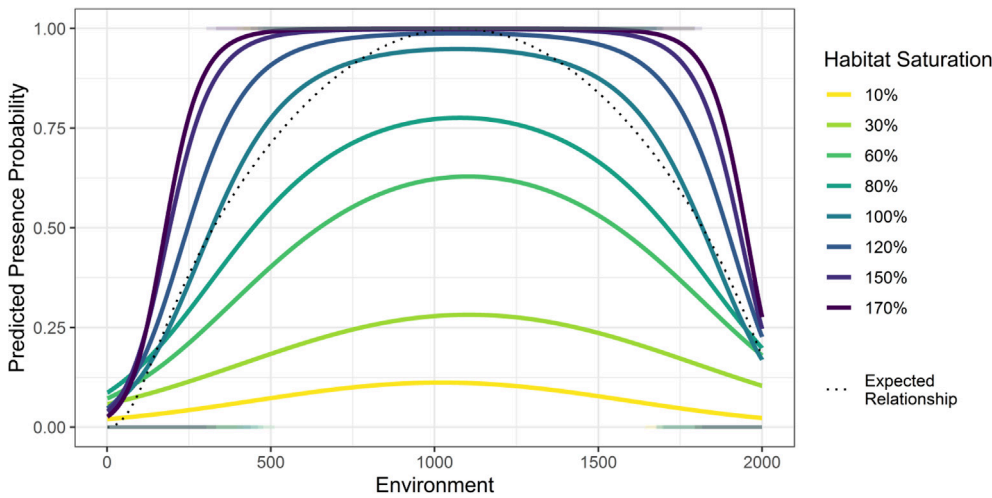


Fig. 2. Species expected and predicted presence probability with and without threshold. The solid curves are the predicted presence probabilities by the GLM used to model the presence of species. The dotted curve is the expected relationship given by the parameters of the species. Segments above and below respectively show predicted presences and absences using species-specific threshold.

the environmental values. However, binary predictions changed abruptly from absences to presences and from presences to absences for environment close to 500 and to 1500, respectively. On the contrary, the predicted presence probabilities did vary with habitat saturation (solid curves in the center). The greater the habitat saturation, the greater the maximum predicted probability. For example at 100% habitat saturation, the maximum predicted probability was close to 0.95, while at 70% saturation it was 0.7.

When comparing observed richness to probability-based richness and threshold-based richness (respectively green and purple points and curves on Fig. 3), we observed differences depending on habitat saturation. Across all habitat saturation levels, probability-based richness showed consistently lower RMSE and variance than threshold-based richness (Fig. 4). For habitat saturation below 100%, threshold-based richness was greater than observed richness, while probability-based

richness followed observed richness. Observed richness against probability-based richness followed the identity line closely with a slope not different from one (all $p > 0.5$, H_0 being that the slope is not different from one) and an intercept not different from zero, related to zero bias at all habitat saturation levels (Fig. 4 middle). The relationship between observed richness and threshold-based richness was not linear and did not follow the identity line whatever habitat saturation. Probability-based richness showed similar RMSE at all habitat saturation levels, while threshold-based richness reached its minimum RMSE when habitat saturation was 80%. When species under-saturated their habitats ($\beta < 100\%$), probability-based richness followed closely observed richness while threshold-richness almost always over-predicted richness. Threshold-based richness lowest RMSE at 80% habitat saturation can be explained by a balance between slight under-prediction when richness was smaller than 75 and over-prediction when richness was

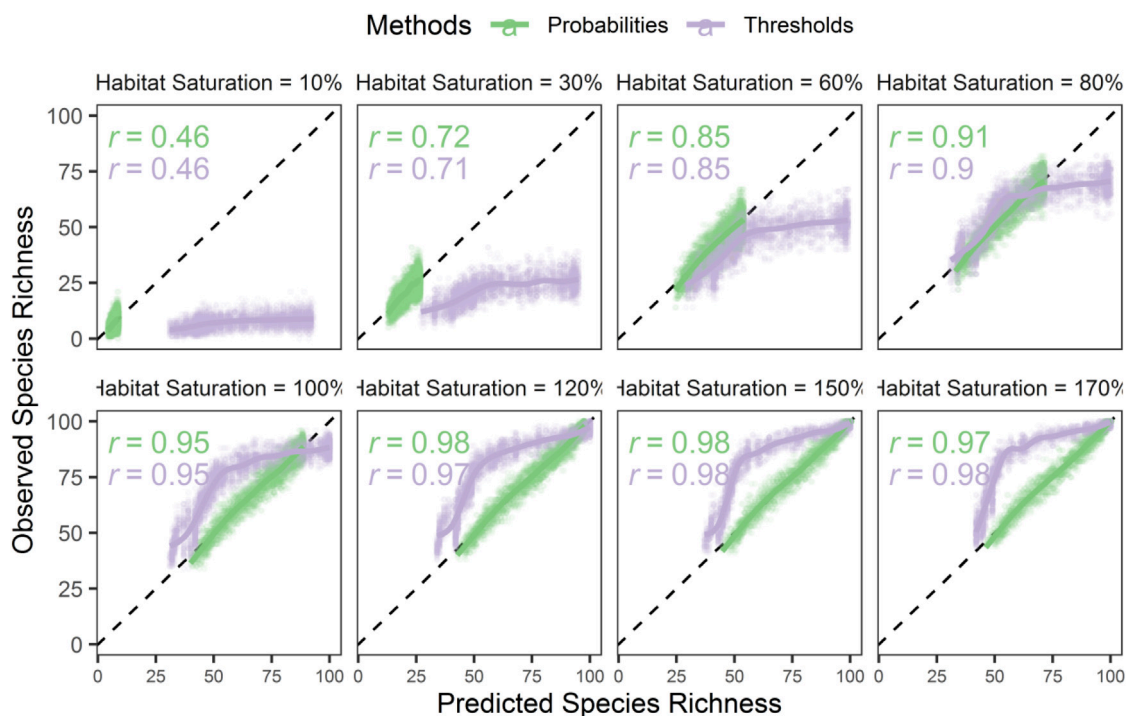


Fig. 3. Observed vs. predicted richness between two prediction methods as a function of habitat saturation. Each facet shows different species habitat saturation (see Material and Methods). The dashed line is the identity line ($y = x$), indicating perfect predictions. Green points are probability-based richness predictions; Purple points are threshold-based richness predictions. The corresponding colored lines are cubic splines smoothers trend lines. Spearman correlation coefficients are shown in the top left corner of each facet.

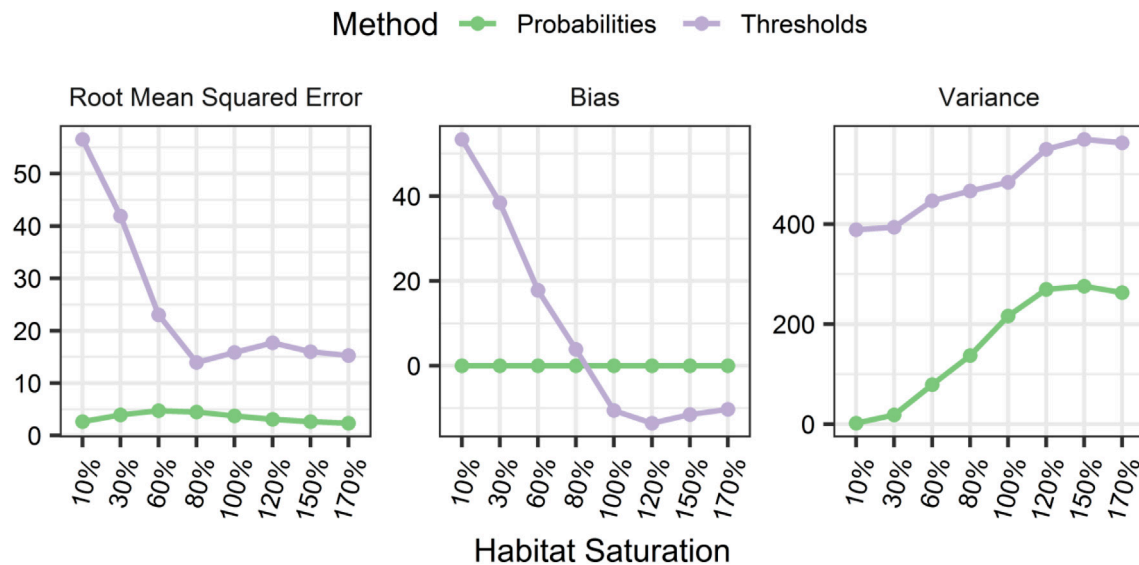


Fig. 4. Prediction accuracy of probability-based and threshold-based richness predictions in function of habitat saturation. Green points and lines: probability-based richness; purple points and lines: threshold-based richness. **(left)** Root Mean Square Error (RMSE) of predicted richness, the average error of richness prediction; **(middle)** Bias, the average difference across all assemblages between predicted and observed richness; **(right)** Variance, the variance of richness predictions across all assemblages.

greater than 75 (Fig. 3). When habitat saturation reached 100%, both types of predictions were close to observed richness (Fig. 3). At this habitat saturation, threshold-based richness showed slight over-prediction in richer communities (predicted richness around 90 species for sites containing 80 species) and slight under-prediction in poorer sites (predicted richness of around 30 for sites containing 45 species), and an average under-prediction (negative bias). At this habitat saturation, the RMSE of both methods was close to the one at 80% habitat saturation, but the variance in prediction increased for probability-based richness. When species over-saturated their habitats ($\beta > 100\%$), threshold-based richness strongly under-predicted richness in poorer communities (negative bias) while probability-based richness showed no bias on average (Fig. 4 middle). For example at 150% habitat saturation, for sites with observed richness around 75, threshold-based richness was around 30 while probability-based richness was 75.

4. Discussion

We designed a virtual experiment of species occurrences along an environmental gradient and performed binomial GLM-based species distribution modeling on these data. The binary threshold-based presence prediction represented the potential habitat of each species based on its fundamental niche (Guisan and Zimmermann, 2000), whatever its actual habitat saturation. On the contrary, the range and average values of predicted presence probabilities depended on habitat saturation, for a given fundamental niche. When summing the individual species predictions, the summed presence probabilities well fitted actual richness, as expected, while habitat saturation strongly affected the threshold-based richness. We thus recommend summing stacked-SDMs probabilities to predict richness. Still, threshold-based richness can also be a useful predictor of potential richness, as species threshold-based binary predictions can be used as a reference species pool for hypothesis testing and modeling of biodiversity dynamics.

In our simulations, probability-based richness on average followed observed richness whatever habitat saturation. This is in line with the fact that probability-based richness should provide the mathematical expectation of richness at a given site (Calabrese et al., 2014). Our results also showed that probability-based richness had a consistently lower RMSE than threshold-based richness, mostly because of its absence of bias. However, both methods had higher variance with higher

habitat saturation, as a consequence of a mean-variance relationships. Thus at high habitat saturation, both methods predict an unreliable richness.

Much emphasis has been put in species distribution modeling on providing binary occurrence prediction. Methods to define thresholds for reliable occurrence prediction have been extensively debated and alternative options have been proposed (Allouche et al., 2006; Freeman and Moisen, 2008; Liu et al., 2005, 2013). However, such a prediction does not grasp the inherently gradual response of species to environmental gradients (Hutchinson, 1957; Meynard and Kaplan, 2012), and tends to generate an artificial dichotomy. This “binarization” has two major caveats. First, it does not acknowledge the gradual variation of performance along the gradient, which increases under-prediction below the threshold and over-prediction above the threshold. Furthermore, the closer to the threshold the higher the prediction bias: just over/below the threshold, there is a greater chance to find a species present/absent than further away from the threshold. Second, it predicts only presences above the threshold and only absences below the threshold, which does not acknowledge the influence of habitat saturation irrespective of habitat suitability. In other words, threshold-based richness will always estimate richness as if species habitat saturation was 100%. Because threshold-based richness over-predicts richness for habitat saturation under 100% (or under-predicts when habitat saturation is over 100%), its accuracy regarding the prediction of species turnover may be low (D’Amen et al., 2015b; Dubuis et al., 2011). At coarser and larger scales, because niche preferences dominate the distribution of species (Pearson and Dawson, 2003), we expect a more deterministic response to the environment in a threshold-like fashion (Guisan and Thuiller, 2005). Species response to environmental gradients is thus highly scale-dependent, specific at local and fine scales and threshold-like at large and coarse scales (Meynard and Kaplan, 2013). The assumption that species distribution at large and coarse scales is in a threshold fashion (Guisan and Thuiller, 2005) has been difficult to prove (Boucher-Lalonde et al., 2014, 2012). Instead in birds, mammals and North American trees, a Gaussian distribution best explained the occurrence-environment relationship for most species (Boucher-Lalonde et al., 2014, 2012), while the threshold model was selected only 5% of the time. Only a fraction of species responds to broad environmental gradients in a binary way. Meynard and Kaplan (2013) also argued that threshold response of species observed in many

datasets could be the results of data aggregation over various spatial and temporal scales.

We defined habitat saturation as a coefficient (β) that affects environmental suitability of species: it increases ($\beta > 1$) or decreases ($\beta < 1$) habitat suitability. It has been shown in diverse taxa that most species do not saturate their habitat: they occupy less habitat than their potential habitat (Boucher-Lalonde et al., 2012; Munguia et al., 2008; Svenning and Skov, 2004). Several mechanisms can explain why a species under-saturates its habitat. For example, dispersal limitation due to slow recolonization of European trees from glacial refugia has led to habitat under-saturation (Svenning and Skov, 2004). Biotic interactions are often cited as an additional factor explaining habitat under-saturation (Svenning and Skov, 2004), as species close in traits can experience limiting similarity and competitively exclude one another. On the contrary, positive biotic interactions as well as source-sink dynamics can cause habitat over-saturation (Eriksson, 1996; Pulliam, 2000; Pulliam and Danielson, 1991). Positive interactions such as facilitation make facilitated species occupy less suitable habitat thanks to the presence of other species (Bertness and Callaway, 1994; Stachowicz, 2001). Source-sink theory explains how a species can be present in unsuitable habitat (sink) by continuously immigrating from a suitable habitat (source) (Pulliam and Danielson, 1991). Here we considered a single habitat saturation coefficient (β) used for all species across all assemblages. This coefficient does not take into account the variability of habitat saturation that may exist between species, where some species saturate more their habitat than others. Furthermore, the habitat saturation coefficient cannot take into account biotic interactions as it is not conditional to the presence of other species; nor that we expect source-sink dynamics to occur only close to the sources, which should lead to a context-dependent habitat saturation. Habitat saturation is also influenced by the extent to which it is measured. In very small areas (e.g., a single quadrat), species tend to fully saturate their suitable habitat, because they occupy the only micro-habitat available for them. For larger areas (e.g., several plots), the occurrence of species should be more stochastic due to dispersal limitation, limiting similarity and biotic interactions as stated above. For even larger areas (e.g., regional, continental or global), habitat saturation should increase again with the dominance of deterministic processes that influence occurrence. As such, we could use a species habitat saturation profile at different areas whose variation would show the change in main assembly processes. Further research is needed regarding habitat (un)saturation to understand its causes and consequences.

A recent study mentions a different but related concept of saturation (Mateo et al., 2017). Mateo et al. (2017) defined saturation as “environmental constraints [that] limit the number of species that can coexist in a community”. Here, we defined habitat saturation as a species-level pattern: it represents the proportion of suitable habitat that a species occupies, it is a species-level property not a community-level property. Community saturation, i.e. saturation *sensu* Mateo et al. (2017), depends on habitat saturation of species. If species have a limited habitat saturation, it imposes an upper bound to species richness. In our model, there are no strict limits on species richness, but on each species' capacity to saturate its habitat. The neutral theory imposes a limit on the number of individuals in any community (Hubbell, 2001), it is a subset of the individuals/species present in the species pool that encompasses a larger area. Changes in number of individuals per community, species regional abundances and/or immigration probability m from the species pool can cause changes in species habitat saturation. By choosing these three parameters, we can obtain a stable richness that can be interpreted as an upper bound, as if saturation was community-level process. However, in this case there is no direct bound on species richness. Richness results from the dynamic extinction-colonization equilibrium and fluctuates over time, it is not a property of the community *per se*. Mateo et al. (2017) focused mostly on S-SDMs that can be constrained with an explicit constraint on richness. As stated above, species richness is unlikely to be directly constrained and

thus modeling explicitly a richness constraint may not underline the true community assembly mechanisms that affect community composition.

Our model, while an interesting basis to test assumptions regarding SDM stacking, represents an ecologically idealistic situation that uses virtual species. We used a single linear environmental gradient, which is an over-simplification of environmental gradients. Indeed, species occurrences are jointly affected by multi-dimensional environmental gradients, which can be non-linear and lead to observed trait syndromes (Laughlin and Messier, 2015). In our simulations, all species have a single trait with a single optimum, however with multi-dimensional environmental gradients we could also expect multi-dimensional optima (Oksanen and Minchin, 2002). Our simulations do not consider biotic interactions, as we simulated the presence of species independently, while, as stated above, biotic interactions can strongly influence species habitat saturation (Pulliam, 2000). We used a species pool containing species with optima on the whole range of the environmental gradient. However, the distribution in species optima among the species pool can be asymmetrically distributed, which in turn can affect local community assembly dynamics (Patrick and Brown, 2018). Furthermore, because of the way the species were simulated, most species niche breadth covered around one third of the range of the environment. While real communities contain a mix of species with narrow and wide niches, with many of species having narrow niches and a few having wide niches (Brown, 1984). Thus, we could determine a ratio between species with wide and narrow environmental niches, based on observed communities, and simulate virtual communities accordingly. We also assumed that species' suitabilities had a quadratic response to the environment, while more complex relationships exist (Oksanen and Minchin, 2002) and could be used in our model. Our simulation setup can thus be made more complex for more investigations on factors that may influence S-SDMs richness predictions. Still, our simplified model can help gain insights about S-SDMs.

Depending on the scales considered, we can expect different shapes of species occurrence-environment relationships. At local scale, we expect many stochastic processes (e.g., demographic stochasticity, competitive exclusion, biotic interactions, microclimatic variations, etc.) to be at play and drive community assembly (Chase and Myers, 2011). Dominance of stochastic processes leads to blurred response to environmental variables, because species occurrence is then not only determined by environmental variables. Predicted presence probability can account for these processes, because they predict the parameter that governs the stochastic process leading to species occurrence such as a binomial trial (Pottier et al., 2013). Indeed, probability-based richness have been shown to estimate the richness of local assemblages well (Calabrese et al., 2014; D'Amen et al., 2015a,b; Guisan and Rahbek, 2011; Pellissier et al., 2013).

Threshold-based richness can be thought as the potential richness expected considering only abiotic deterministic processes. It can be a useful baseline to compare to models that consider a broader set of processes (Pouteau et al., 2019; Violle et al., 2011). Threshold-based richness defines a reference pool against which null models or hierarchical analyses can be performed. It can be considered as an additional method to define species pool (Carstensen et al., 2013; Lessard et al., 2012). Indeed, threshold-based richness would represent a species pool (Lessard et al., 2015) that considers only the response to environmental filtering for a large area. Without explicitly considering functional traits, threshold-based richness can also represent a functional species pool as species traits are filtered by the environment (de Bello et al., 2012). Threshold-based richness is nested in a hierarchy of models similar to the hierarchy of scales and processes that shape community assembly (Keil et al., 2013; Mackey and Lindenmayer, 2001; Mertes and Jetz, 2018; Meyer, 2007; Pearson and Dawson, 2003). Threshold-based predictions, because they consider environmental filtering only, are representative of coarse and large scales in this

hierarchy of models. In the SESAM framework (D'Amen et al., 2015a; Guisan and Rahbek, 2011), threshold richness is a reference richness before applying a cutoff in species presences to account for local variations. The use of threshold-based predictions can thus be compared to more mechanistic models, to know to what extent observed communities are mostly shaped by environmental filtering. Other models, to which they can be compared, can incorporate other important community assembly factors such as dispersal limitation, limiting similarity or biotic interactions (Chase and Myers, 2011; Munoz et al., 2017; Pouteau et al., 2019; van der Plas et al., 2015). In summary, using both threshold-based richness and probability-based richness in succession—first threshold-based richness as a pure environmental prediction then compare it to probability-based richness—can shed light on community assembly processes. When both agree, environmental filtering dominates community assembly. If not, habitat saturation can strongly change threshold-based richness and/or other processes may affect community assembly. Threshold-based richness and probability-based can further be compared to other predictions using process-based models that consider additional processes. Essential Biodiversity Variables can be measured using multiple methods (Pereira et al., 2017), and there is no clear recommendation on which method should be prioritized to predict EBVs. Probability-based species richness could be used as a reliable method to predict taxonomic diversity in the EBV framework, while threshold-based richness can be a useful tool to assess community assembly processes (Pouteau et al., 2019).

CRediT authorship contribution statement

Matthias Grenié: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing. **Cyrille Violle:** Conceptualization, Supervision, Writing - review & editing. **François Munoz:** Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Pierre Denelle and Christine Meynard for helpful discussions. MG was supported by the ENS de Lyon. This study was supported by the European Research Council (ERC) Starting Grant Project 'ecophysiological and biophysical constraints on domestication in crop plants' (grant ERC-StG-2014-639706-CONSTRAINTS) and by the French Foundation for Research on Biodiversity (FRB; < www.fondationbiodiversite.fr >) in the context of the CESAB project 'causes and consequences of functional rarity from local to global scales' (FREE).

References

- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS): assessing the accuracy of distribution models. *J. Appl. Ecol.* 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>.
- Bertness, M.D., Callaway, R., 1994. Positive interactions in communities. *Trends Ecol. Evol.* 9, 191–193. [https://doi.org/10.1016/0169-5347\(94\)90088-4](https://doi.org/10.1016/0169-5347(94)90088-4).
- Boucher-Lalonde, V., Morin, A., Currie, D.J., 2014. A consistent occupancy Climate relationship across birds and mammals of the Americas. *Oikos* 123, 1029–1036. <https://doi.org/10.1111/oik.01277>.
- Boucher-Lalonde, V., Morin, A., Currie, D.J., 2012. How are tree species distributed in climatic space? A simple and general pattern. *Glob. Ecol. Biogeogr.* 21, 1157–1166. <https://doi.org/10.1111/j.1466-8238.2012.00764.x>.
- Brown, J.H., 1984. On the relationship between abundance and distribution of species. *Am. Nat.* 124, 255–279. <https://doi.org/10.1086/284267>.
- Calabrese, J.M., Certain, G., Kraan, C., Dormann, C.F., 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models: stacking species distribution models. *Glob. Ecol. Biogeogr.* 23, 99–112. <https://doi.org/10.1111/geb.12102>.
- Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., Naeem, S., 2012. Biodiversity loss and its impact on humanity. *Nature* 486, 59–67. <https://doi.org/10.1038/nature11148>.
- Carstensen, D.W., Lessard, J.-P., Holt, B.G., Krabbe Borregaard, M., Rahbek, C., 2013. Introducing the biogeographic species pool. *Ecography* 36, 1310–1318. <https://doi.org/10.1111/j.1600-0587.2013.00329.x>.
- Chase, J.M., Myers, J.A., 2011. Disentangling the importance of ecological niches from stochastic processes across scales. *Philos. Trans. R. Soc. B: Biol. Sci.* 366, 2351–2363. <https://doi.org/10.1098/rstb.2011.0063>.
- Chown, S.L., van Rensburg, B.J., Gaston, K.J., Rodrigues, A.S.L., van Jaarsveld, A.S., 2003. Energy, species richness, and human population size: conservation implications at a national scale. *Ecol. Appl.* 13, 1233–1241. <https://doi.org/10.1890/02-5105>.
- D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L., Guisan, A., 2015a. Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *J. Biogeogr.* 42, 1255–1266. <https://doi.org/10.1111/jbi.12485>.
- D'Amen, M., Pradervand, J.-N., Guisan, A., 2015b. Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Glob. Ecol. Biogeogr.* 24, 1443–1453. <https://doi.org/10.1111/geb.12357>.
- de Bello, F., Price, J.N., Münkemüller, T., Liira, J., Zobel, M., Thuiller, W., Gerhold, P., Götzberger, L., Lavergne, S., Lepš, J., Zobel, K., Pärtel, M., 2012. Functional species pool framework to test for biotic effects on community assembly. *Ecology* 93, 2263–2273. <https://doi.org/10.1890/11-1394.1>.
- Dodson, S., 1992. Predicting crustacean zooplankton species richness. *Limnol. Oceanogr.* 37, 848–856. <https://doi.org/10.4319/lo.1992.37.4.0848>.
- Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P., Guisan, A., 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Divers. Distrib.* 17, 1122–1131. <https://doi.org/10.1111/j.1472-4642.2011.00792.x>.
- Eriksson, O., 1996. Regional dynamics of plants: a review of evidence for remnant, source-sink and metapopulations. *Oikos* 77, 248–258. <https://doi.org/10.2307/3546063>.
- Ferrier, S., Guisan, A., 2006. Spatial modelling of biodiversity at the community level. *J. Appl. Ecol.* 43, 393–404. <https://doi.org/10.1111/j.1365-2664.2006.01149.x>.
- Freeman, E.A., Moisen, G.G., 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecol. Model.* 217, 48–58. <https://doi.org/10.1016/j.ecolmodel.2008.05.015>.
- Gavish, Y., Marsh, C.J., Kuemmerlen, M., Stoll, S., Haase, P., Kunin, W.E., 2017. Accounting for biotic interactions through alpha-diversity constraints in stacked species distribution models. *Methods Ecol. Evol.* 8, 1092–1102. <https://doi.org/10.1111/2041-210X.12731>.
- GBIF, 2019. What is GBIF?
- Graham, C.H., Hijmans, R.J., 2006. A comparison of methods for mapping species ranges and species richness. *Glob. Ecol. Biogeogr.* 15, 578–587. <https://doi.org/10.1111/j.1466-8238.2006.00257.x>.
- Guisan, A., Rahbek, C., 2011. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages: predicting spatio-temporal patterns of species assemblages. *J. Biogeogr.* 38, 1433–1444. <https://doi.org/10.1111/j.1365-2699.2011.02550.x>.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.
- Hirzel, A.H., Helfer, V., Metral, F., 2001. Assessing habitat-suitability models with a virtual species. *Ecol. Model.* 145, 111–121. [https://doi.org/10.1016/S0304-3800\(01\)00396-9](https://doi.org/10.1016/S0304-3800(01)00396-9).
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*, Monographs in Population Biology. Princeton University Press, Princeton.
- Hurlbert, A.H., Stegen, J.C., 2014. When should species richness be energy limited, and how would we know? *Ecol. Lett.* 17, 401–413. <https://doi.org/10.1111/ele.12240>.
- Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* 22, 415–427. <https://doi.org/10.1101/SQB.1957.022.01.039>.
- Jiménez-Valverde, A., Lobo, J.M., 2007. Threshold criteria for conversion of probability of species presence to either or presence absence. *Acta Oecologica* 31, 361–369. <https://doi.org/10.1016/j.actao.2007.02.001>.
- Keil, P., Belmaker, J., Wilson, A.M., Unitt, P., Jetz, W., 2013. Downscaling of species distribution models: a hierarchical approach. *J. Appl. Ecol.* 82–94. [https://doi.org/10.1111/j.2041-210x.2012.00264.x@10.1111/\(ISSN\)1365-2664.INTERNATIONAL](https://doi.org/10.1111/j.2041-210x.2012.00264.x@10.1111/(ISSN)1365-2664.INTERNATIONAL).
- Laughlin, D.C., Messier, J., 2015. Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends Ecol. Evol.* 30, 487–496. <https://doi.org/10.1016/j.tree.2015.06.003>.
- Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F., 2016. Virtualspecies, an R package to generate virtual species distributions. *Ecography* 39, 599–607. <https://doi.org/10.1111/ecog.01388>.
- Lessard, J.-P., Belmaker, J., Myers, J.A., Chase, J.M., Rahbek, C., 2012. Inferring local ecological processes amid species pool influences. *Trends Ecol. Evol.* 27, 600–607. <https://doi.org/10.1016/j.tree.2012.07.006>.
- Lessard, J.-P., Weinstein, B.G., Borregaard, M.K., Marske, K.A., Martin, D.R., McGuire, J.A., Parra, J.L., Rahbek, C., Graham, C.H., 2015. Process-based species pools reveal the hidden signature of biotic interactions amid the influence of temperature filtering. *Am. Nat.* 187, 75–88. <https://doi.org/10.1086/684128>.

- Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28, 385–393. <https://doi.org/10.1111/j.0906-7590.2005.03957.x>.
- Liu, C., White, M., Newell, G., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.* 40, 778–789. <https://doi.org/10.1111/jbi.12058>.
- Mackey, B.G., Lindenmayer, D.B., 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. *J. Biogeogr.* 28, 1147–1166. <https://doi.org/10.1046/j.1365-2699.2001.00626.x>.
- Mateo, R.G., Mokany, K., Guisan, A., 2017. Biodiversity models: what if unsaturation is the rule? *Trends Ecol. Evol.* <https://doi.org/10.1016/j.tree.2017.05.003>.
- Mazel, F., Guilhaumon, F., Mouquet, N., Devictor, V., Gravel, D., Renaud, J., Cianciaruso, M.V., Loyola, R., Diniz-Filho, J.A.F., Mouillot, D., Thuiller, W., 2014. Multifaceted diversity area relationships reveal global hotspots of mammalian species, trait and lineage diversity. *Glob. Ecol. Biogeogr.* 23, 836–847. <https://doi.org/10.1111/geb.12158>.
- Mertes, K., Jetz, W., 2018. Disentangling scale dependencies in species environmental niches and distributions. *Ecography* 41, 1604–1615. <https://doi.org/10.1111/ecog.02871>.
- Meyer, C.B., 2007. Does scale matter in predicting species distributions? Case study with the marbled murrelet. *Ecol. Appl.* 17, 1474–1483. <https://doi.org/10.1890/06-1410.1>.
- Meynard, C.N., Kaplan, D.M., 2013. Using virtual species to study species distributions and model performance. *J. Biogeogr.* 40, 1–8. <https://doi.org/10.1111/jbi.12006>.
- Meynard, C.N., Kaplan, D.M., 2012. The effect of a gradual response to the environment on species distribution modeling performance. *Ecography* 35, 499–509. <https://doi.org/10.1111/j.1600-0587.2011.07157.x>.
- Meynard, C.N., Leroy, B., Kaplan, D.M., 2019. Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? *Ecography*. <https://doi.org/10.1111/ecog.04385>.
- Munguia, M., Peterson, A.T., Sanchez-Cordero, V., 2008. Dispersal limitation and geographical distributions of mammal species. *J. Biogeogr.* 35, 1879–1887. <https://doi.org/10.1111/j.1365-2699.2008.01921.x>.
- Munoz, F., Grenié, M., Denelle, P., Taudiere, A., Laroche, F., Tucker, C., Violle, C., 2017. Ecottery: Simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution* in press.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. <https://doi.org/10.1038/35002501>.
- Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L., Bennett, D.J., Choimes, A., Collen, B., Day, J., De Palma, A., Díaz, S., Echeverri-Londoño, S., Edgar, M.J., Feldman, A., Garon, M., Harrison, M.L.K., Alhussaini, T., Ingram, D.J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Correia, D.L.P., Martin, C.D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H.R.P., Purves, D.W., Robinson, A., Simpson, J., Tuck, S.L., Weiher, E., White, H.J., Ewers, R.M., Mace, G.M., Scharlemann, J.P.W., Purvis, A., 2015. Global effects of land use on local terrestrial biodiversity. *Nature* 520, 45–50. <https://doi.org/10.1038/nature14324>.
- O'Brien, E., 1998. Water-energy dynamics, climate, and prediction of woody plant species richness: an interim general model. *J. Biogeogr.* 25, 379–398. <https://doi.org/10.1046/j.1365-2699.1998.252166.x>.
- Oksanen, J., Minchin, P.R., 2002. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecol. Model.* 157, 119–129. [https://doi.org/10.1016/S0304-3800\(02\)00190-4](https://doi.org/10.1016/S0304-3800(02)00190-4).
- Patrick, C.J., Brown, B.L., 2018. Species pool functional diversity plays a hidden role in generating β -diversity. *Am. Nat.* 191, E159–E170. <https://doi.org/10.1086/696978>.
- Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.* 12, 361–371. <https://doi.org/10.1046/j.1466-822X.2003.00042.x>.
- Pellissier, L., Esp ndola, A., Pradervand, J.-N., Dubuis, A., Pottier, J., Ferrier, S., Guisan, A., 2013. A probabilistic approach to niche-based community models for spatial forecasts of assemblage properties and their uncertainties. *J. Biogeogr.* 40, 1939–1946. <https://doi.org/10.1111/jbi.12140>.
- Pereira, H.M., Belnap, J., B hm, M., Brummitt, N., Garcia-Moreno, J., Gregory, R., Martin, L., Peng, C., Proen a, V., Schmeller, D., van Swaay, C., 2017. Monitoring essential biodiversity variables at the species level. In: Walters, M., Scholes, R.J. (Eds.), *The GEO Handbook on Biodiversity Observation Networks*. Springer International Publishing, Cham, pp. 79–105. https://doi.org/10.1007/978-3-319-27288-7_4.
- Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., H ft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. *Science* 339, 277–278. <https://doi.org/10.1126/science.1229931>.
- Pineda, E., Lobo, J.M., 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *J. Anim. Ecol.* 78, 182–190. <https://doi.org/10.1111/j.1365-2656.2008.01471.x>.
- Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P., Guisan, A., 2013. The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients: climate and species assembly predictions. *Glob. Ecol. Biogeogr.* 22, 52–63. <https://doi.org/10.1111/j.1466-8238.2012.00790.x>.
- Pouteau, R., Munoz, F., Birnbaum, P., 2019. Disentangling the processes driving tree community assembly in a tropical biodiversity hotspot (New Caledonia). *J. Biogeogr.* 46, 796–806. <https://doi.org/10.1111/jbi.13535>.
- Pulliam, H.R., 2000. On the relationship between niche and distribution. *Ecol. Lett.* 3, 349–361.
- Pulliam, H.R., Danielson, B.J., 1991. Sources, sinks, and habitat selection: a landscape perspective on population dynamics. *Am. Nat.* 137, S50–S66. <https://doi.org/10.1086/285139>.
- Core Team, R., 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Scherrer, D., D'Amen, M., Fernandes, R.F., Mateo, R.G., Guisan, A., 2018. How to best threshold and validate stacked species assemblages? Community optimisation might hold the answer. *Methods Ecol. Evol.* <https://doi.org/10.1111/2041-210X.13041>.
- Schmitt, S., Pouteau, R., Robin, Justeau, Dimitri, Boissieu, Florian, Birnbaum, Philippe, Goiding, Nick, 2017. SSDM: an R package to predict distribution of species richness and composition based on stacked species distribution models. *Methods Ecol. Evol.* 8, 1795–1803. <https://doi.org/10.1111/2041-210X.12841>.
- Stachowicz, J.J., 2001. Mutualism, facilitation, and the structure of ecological communities positive interactions play a critical, but underappreciated, role in ecological communities by reducing physical or biotic stresses in existing habitats and by creating new habitats on which many species depend. *Bioscience* 51, 235–246. [https://doi.org/10.1641/0006-3568\(2001\)051\[0235:MEATSO\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0235:MEATSO]2.0.CO;2).
- Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: a citizen-based bird observation network in the biological sciences. *Biol. Conserv.* 142, 2282–2292. <https://doi.org/10.1016/j.biocon.2009.05.006>.
- Svenning, J.-C., Skov, F., 2004. Limited filling of the potential range in European tree species. *Ecol. Lett.* 7, 565–573. <https://doi.org/10.1111/j.1461-0248.2004.00614.x>.
- Tedesco, P.A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., Conti, L., Cornu, J.-F., Dias, M.S., Grenouillet, G., Hugueny, B., J'ez'equel, C., Leprieux, F., Brosse, S., Oberdorff, T., 2017. A global database on freshwater fish species occurrence in drainage basins. *Sci. Data* 4, 170141. <https://doi.org/10.1038/sdata.2017.141>.
- van der Plas, F., Janzen, T., Ordonez, A., Fokkema, W., Reinders, J., Etienne, R.S., Olff, H., 2015. A new modeling approach estimates the relative importance of different community assembly processes. *Ecology* 96, 1502–1515. <https://doi.org/10.1890/14-0454.1>.
- V clavik, T., Meentemeyer, R.K., 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Divers. Distrib.* 18, 73–83. <https://doi.org/10.1111/j.1472-4642.2011.00854.x>.
- Violle, C., Bonis, A., Plantegenest, M., Cudennec, C., Damgaard, C., Marion, B., C eur, D.L., Bouzill e, J.-B., 2011. Plant functional traits capture species richness variations along a flooding gradient. *Oikos* 120, 389–398. <https://doi.org/10.1111/j.1600-0706.2010.18525.x>.

2 DIMENSIONNALITÉ DES INDICES DE DIVERSITÉ FONCTIONNELLE

When more is less: adding more traits dilutes the Functional
Diversity-Environment Relationship

Matthias Grenié^{1*†}François Munoz^{2‡}Cyrille Violle^{1§}

2020-06-29

1: CEFE, Univ. Montpellier, CNRS, EPHE, IRD, Univ. Paul Valéry Montpellier 3, Montpellier, France

2: University Grenoble-Alpes, LECA, 2233 Rue de la Piscine, 38041 Grenoble Cedex 9, France

*corresponding author matthias.grenie@ens-lyon.fr

†<https://orcid.org/0000-0002-4659-7522>

‡<https://orcid.org/0000-0001-8776-4705>

§<https://orcid.org/0000-0002-2471-9226>

When more is less: adding more traits dilutes the Functional Diversity-Environment Relationship

Running Title: Dilution Effect blurs FD-Environment relationships

2020-06-29

5 Introduction

Organismal phenotypes are multidimensional by nature (Hutchinson, 1957; Laughlin & Messier, 2015). This is the reason why trait-based ecology has been interested in studying the nature and shape of species' multidimensional functional space (Winemiller *et al.*, 2015; Díaz *et al.*, 2016). In this perspective, major axes of trait variation have been identified for many taxa (Díaz *et al.*, 2016; Madin *et al.*, 2016; Pianka *et al.*, 2017; Céréghino *et al.*, 2018; Pigot *et al.*, 2020). However, whether and how these axes are associated to environmental gradients remains an unsolved question, partly due to the joint effects of myriad local and global factors at play (Bruehlheide *et al.*, 2018). Despite this complexity, multi-trait distance-based indices (functional diversity indices *sensu lato*) have been popularized for the last two decades (Mouchet *et al.*, 2010; Schleuter *et al.*, 2010; Mason *et al.*, 2013) for at least three reasons: (i) capturing the multidimensional nature of organismal phenotypes using single synthetic metrics (Dalerum, 2013; Maire *et al.*, 2015), (ii) capturing multiple niche-based processes (e.g. competitive interactions) that simultaneously act on several facets of organismal phenotypes (Mouillot *et al.*, 2013), (iii) identifying and deciphering the effects of environmental filters that shape the structure and dynamics of ecological communities (Mason *et al.*, 2013; Floury *et al.*, 2017). This led to a growing number of studies that examined the changes in functional diversity along local and regional gradients, without a prior hypothesis regarding the influence of particular environmental factors. In the meantime, the huge efforts of trait measurement standardization and more recently of databasing (Jones *et al.*, 2009; Kattge *et al.*, 2011; Madin *et al.*, 2016; Oliveira *et al.*, 2017; Chapman *et al.*, 2019), notably in plants, have catalyzed the computation of functional diversity indices with increasingly more trait dimensions. These have favored a more complete characterization of phenotypes. An emblematic example is a more systematic incorporation of root traits in plant trait-based ecology (Iversen *et al.*, 2017). However, incorporating more and more traits in functional diversity indices can in turn alter the detection of the effect of environmental filters.

Multiple indices are needed to estimate the variability of multi-dimensional organismal phenotypes (Laughlin & Messier, 2015). Functional diversity is indeed multi-faceted and different indices estimate different facets

30 (Villéger *et al.*, 2008; Mouchet *et al.*, 2010; Mason *et al.*, 2013). For example Functional Richness (FRic) estimates the total amount of diversity of traits in a multi-dimensional space, while Functional Dispersion (FDis) describes the abundance-weighted centroid of the functional space and tells us about the evenness of abundance in the functional space (Villéger *et al.*, 2008). Because the phenotype is multidimensional, multiple traits should be used to describe the different facets of functional diversity. A given community
35 assembly process can play on a given trait without playing on other traits. We expect that, if a single trait is affected by a given process, the use of additional unaffected traits to compute functional diversity should dilute the expected change in functional diversity. Biologically, however, the dilution effect is not as straightforward, because individuals are affected through their multidimensional phenotype (Laughlin & Messier, 2015). Multiple traits should then covary or be affected by similar community assembly processes,
40 counteracting the dilution effect.

Community assembly processes affect which species occur in ecological communities (Keddy, 1992). Different community assembly processes can affect the different functional diversity facets. For example, environmental filtering tends to reduce functional richness (Keddy, 1992; Kraft *et al.*, 2015a). A reduction of functional richness along the environment compared to a null expectation is thus generally interpreted as environmental
45 filtering (Kraft *et al.*, 2015a). Neutral processes, on the other hand —because they affect species irrespective of traits— tend to increase variation in functional diversity (Schwilk & Ackerly, 2005). The dilution effect of unaffected traits on functional diversity should be clear if environmental filtering only affect few traits. Neutral processes should make the dilution effect less clear as they increase the stochasticity of functional diversity. Community assembly processes should thus shape the expected dilution effect.

50 One factor that influences estimated functional diversity is the number of traits (Maire *et al.*, 2015; Legras *et al.*, 2019). For example, depending on the indices used, increasing number of traits can increase functional diversity for mathematical reasons (Legras *et al.*, 2019). Functional diversity is traditionally computed through species trait dissimilarity (Maire *et al.*, 2015). The estimated dissimilarities between species can be highly driven by the number of traits (or PCA axes) used to compute functional diversity indices (Maire
55 *et al.*, 2015). Using more traits in functional diversity should thus better detect such processes and lead to stronger diversity-environment relationships. However, because a process may affect traits non-linearly and some traits may be affected by different processes, we expect a dilution effect when adding unaffected traits. More generally, the number of traits used in functional diversity is known to relate to estimated phylogenetic diversity (Kraft *et al.*, 2007; Tucker *et al.*, 2018).

60 An absence of relationship between functional diversity and the environment using multi-dimensional functional diversity indices does not preclude for an absence of environmental filtering (Spasojevic & Suding,

2013). One possibility to obtain a constant multi-dimensional diversity is to have two traits opposingly affected by environmental filtering along the environmental gradient (Bernard-Verdier *et al.*, 2012). One trait could converge while the other trait would diverge along the environmental gradient. In this case, uni-dimensional functional diversity indices based on these two traits would show opposed trends. Two-dimensional functional diversity indices using the two traits would on the contrary be constant along the gradient, obscuring the processes affecting single traits. This example would go beyond the simplest scenario where a single trait is filtered and would show to what extent the dilution of additional trait can be changed if they are also affected by environmental filtering.

Linking functional diversity patterns to specific community assembly processes from observed data can be quite difficult as a number of processes can obscure the expected patterns (Cornwell & Ackerly, 2009; McPherson *et al.*, 2018). Numerical simulations, as virtual experiments, can be used to measure the influence of ecological processes (Peck, 2004; Zurell *et al.*, 2010). Virtual ecology has indeed helped better calibrate species distribution models (Hirzel *et al.*, 2001; Meynard *et al.*), design sampling scheme to best represent biodiversity (Thomas *et al.*, 2018), and disentangle community assembly processes (Botta-Dukát & Czúcz, 2015; Münkemüller & Gallien, 2015; Munoz *et al.*, 2017). Through simulation, we can manipulate the strength of neutral processes and environmental filtering independently (Kraft *et al.*, 2007). Simulations also let us simulate species with independent traits, and generate environmental filtering affecting single or multiple trait(s). With these simulations we can quantify the expected dilution effect due to additional traits in the functional diversity-environment, while it would be impossible to do so with observed data.

This study aims to estimate the dilution effect when using additional traits that can fade away the functional diversity-environment relationship. We used a simulation framework to test the effect of both environmental filtering and neutral dynamics on the relationship between environment and functional diversity metrics. We tested two scenarios: the first one, where a single trait was filtered along an environmental gradient, additional traits were not affected by environmental filtering; the second one had two traits affected by environmental filtering in opposed ways along the environmental gradient, the first trait converged while the second one diverged, other simulated traits were not affected by environmental filtering.

Material and Methods

Simulations

90 Community Assembly

To test the dilution effect of traits on the functional diversity-environment relationship, we simulate communities using a community assembly model. In our framework, each individual immigrates from the species pool in the local community. The immigration probability of each individual depends on the local environment g and the trait value t of the species. We used a Gaussian function with parameters t_g and σ_g 95 to characterize the environment. t_g represents the optimal trait value in the local environment such that a species with $t = t_g$ has a probability of one of immigrating in the local environment. While σ_g represents the width of environmental filtering, the smaller σ_g the narrower the environmental filtering, and the closer the species have to be to t_g to immigrate and maintain in the local community. In addition, the smaller σ_g the less functional diverse a community is (for this particular trait).

100 We simulated community assembly using `ecolottery` (Munoz *et al.*, 2017) (function `coalesc()`), a fast coalescent-based model which considers both neutral dynamics and environmental filtering. We generated 11 uncorrelated uniform traits between 0 and 1 for 200 species. The regional species pool hosted 500 individuals with identical trait values for each species. We used a Gaussian environmental filter that affected species immigration probability from the pool in the community as follows:

$$p_{i, g} = \exp \left[-\frac{(t_i - t_g)^2}{2\sigma_g^2} \right] (\#eq : envfilter) \quad (1)$$

105 with $p_{i, g}$ the probability of successful immigration by species i in environment g , t_i its associated trait, t_g the optimal trait values in environment g and σ_g^2 the variance of the environmental filter in this environment. We created a gradient of 10 environments with corresponding optimal traits values t_g ranging from 0.1 to 0.9. We wanted to test to what extent the number of traits used in functional diversity indices affected the detection of a trend along the environmental gradient. Thus, we assigned decreasing σ_g along the 10 110 environments from 0.2 to 0.03. Because of the property of the Gaussian function, 95% of the distribution is within $t_g \pm 1.96 \times \sigma_g$. Thus at the beginning of the gradient, the filter is weak and we expect a high functional diversity for the filtered trait (many species away from the optimum can maintain themselves), while it is harsh at the end of the gradient (only species close to the optimum can maintain themselves). Our environmental gradient consists of 10 environments with increasing optimal trait values t_g from 0.1 to 115 0.9 and, at the same time, decreasing variance σ_g from 0.2 to 0.03. For each environment, we simulated

independently 10 communities of 500 individuals using `coalesc()` with the migration parameter `m` set to 1. In order to test to what extent neutral dynamics would affect the functional diversity-environment relationship, we additionally simulated communities with `m = 0.1`.

Because regional pool size has been shown to influence phylogenetic diversity patterns (Kraft *et al.*, 2007), we simulated similar communities as above with a migration rate of 1, but instead of using a pool that contained 500 individuals of each species, i.e. 100 000 individuals total, we used a pool that contained only 20 individuals of each species, i.e. 4000 individuals total.

To test the effect of traits being affected in opposite directions we also simulated 10 communities with a first environmental filter set up as above, and a second environmental filter taken into account a second filtered trait with a variance linearly increasing in the opposite direction as the first environmental filter (Supplementary Figure S 1). Both environmental filters had the same optimal trait value t_g .

Functional Diversity Indices

For each n-uplets of traits we computed two indices of functional diversity: Functional Richness (FRic) and Functional Dispersion (FDis). Functional Richness quantifies the proportion of the trait space occupied by all the species in the assemblage. It is the volume of the minimum convex hull encompassing all the species in the assemblage (Villéger *et al.*, 2008). Because FRic is linked to the number of traits used to compute it (Legras *et al.*, 2019), we standardized FRic between 0 and 1, where 1 corresponds to the volume of all the species in the species pool. Functional Dispersion quantifies the evenness of trait distribution along the multiple dimension (Laliberté & Legendre, 2010): if all species are close together FDis will be low, while it will be high if species are further apart. Using presence-absence data, FDis is computed as follows:

$$\text{FDis}_k = \frac{\sum_{i=1}^{S_k} a_i z_{ik}}{\sum_{i=1}^{S_k} a_i} (\#eq : fdis) \quad (2)$$

where FDis_k is the Functional Dispersion of site k , S_k the total number of species in site k , a_i the abundance of species i in site k and z_{ik} the distance between species traits and the weighted centroid of traits of all species present at this site in the trait space. The two indices were computed using the function `fdisp()` in the `FD` R package v1.0-12 (Laliberté *et al.*, 2014) and using the function `convhulln()` in the `geometry` package v0.4.5 (Habel *et al.*, 2019).

Statistical Analyses

All analyses were made using R (R Core Team, 2019).

Null models

To test to what extent the observed functional diversity patterns were due to the effect of environmental filtering we used null models to compare observed communities to null communities (Gotelli & Graves, 1996). We shuffled 100 times trait combinations among all species. We recomputed functional diversity indices using shuffled traits.

We computed Standard Effect Size (SES) as follows:

$$\text{SES} = \frac{\text{FD}_{\text{obs}} - \widehat{\text{FD}}_{\text{null}}}{\sigma_{\text{FD}_{\text{null}}^2}} \quad (3)$$

where FD_{obs} is the observed functional diversity of a site, $\widehat{\text{FD}}_{\text{null}}$ the average of null functional diversity at this site and $\sigma_{\text{FD}_{\text{null}}^2}$ the standard deviation of null functional diversity at the same site. If the null functional diversity distribution is approximately normal, we can use SES values of -1.96 and 1.96 as significance thresholds (Gotelli & Graves, 1996).

Statistical power computation

When adding non-filtered traits to the trait used to compute functional diversity, the functional diversity environment relationship fades away. We however want to distinguish trends of functional diversity indices along the environment when they are computed using combinations of traits that contain vs that do not contain the filtered trait. To avoid accepting the null hypothesis when the combination contains the filtered trait, we computed the statistical power when varying the SES thresholds. We computed statistical power $1 - \beta$ as follows: for a fixed SES threshold and a fixed number of traits, we counted the proportions of sites along the gradient with an SES above the threshold (significantly different from null models) when the traits used to compute functional diversity included the filtered trait. Statistical power is our ability to correctly reject the null hypothesis ($H_0 =$ functional diversity is not different than random). We computed power for all functional diversity indices, all number of traits with various SES thresholds. We then use these power estimates to adjust the SES threshold depending on the number of traits and the index used to get a fixed power level.

Results

As expected from the design of the simulations, FRic decreases along the environmental gradient when using the filtered trait (see Figure 1). With additional traits, including the filtered trait, FRic still shows a decrease along the environmental gradient, but the decrease is less pronounced as the number of traits increases. When using a single non-filtered trait, FRic does not change along the environmental gradient. However, when considering two or more non-filtered traits, FRic decreases with increasing environmental filter, even though the filtered trait is not among the considered traits. This decrease in FRic is due to the increase in dimensionality, because we standardized FRic to its maximum theoretical value if all traits varied independently, and the higher dimensionality, the emptier the functional space. Still, the combinations of non-filtered traits shows higher FRic at all environments than combinations that contained the filtered trait.

The Standard Effect Sizes (SES) from that quantify departures from the null expectations were within -1.96 and 1.96 significance limits for all trait combinations that did not contain the filtered trait, while they were all under -1.96 for trait combinations containing the filtered trait whatever the number of traits considered. This means that FRic was lower than expected by the null model when including the filtered trait. Even if much lower than the significance threshold -1.96, the SES values for the combination of 11 traits were on average -3.4. Thus adding more traits could make the SES not significant, even if the combination contains the filtered trait.

FDis decreases along the environmental gradient when using the filtered trait (Figure 1)), while it shows no trend when using non-filtered traits. When using combinations of two traits including the filtered trait there is still a decrease in FDis, while there is no more when using combinations of 3 or more traits. The greater the number of traits used, the closer the FDis values between combinations with and without the filtered trait. The SES for FDis remains between -1.96 and 1.96 for all combinations of non-filtered traits, while it stays well below -1.96 for almost all combinations containing the filtered trait; except at the extreme of the gradient, environment 0.9 for five, six and seven traits.

Simulations with low migration rate show higher variability in both FRic and FDis observed values (Figure 2), with smaller differences between combinations of non-filtered traits and filtered trait than cases with higher migration rate. For SESs, the effect of a decrease in migration rate was similar, lower migration rate simulations show higher variability in SES and smaller differences between combinations containing only non-filtered traits and with the filtered trait.

At fixed SES threshold, the statistical power decreased with increasing number of traits (Figure 3). However, the power decreased more for FDis reaching about 80% for 7 traits at 1.96 SES, than for FRic where the

power did not decreased when using the 1.96 SES threshold. FRic showed almost no change in power whatever the SES threshold used and the number of traits, except with the six or seven traits with more stringent threshold around 2.5.

200 As expected, both FRic and FDis showed no trend along the environment when considering a non-filtered trait, while filtered trait 1 and trait 2 showed opposing trends along the environment (Figure 4 **A** and **B**, respectively green and blue lines). For FRic, all combinations of two or more traits without filtered traits showed a humped-shaped relationship with the environment. FRic being lower on the margins of the environment and higher across all sites in the center. FRic decreased as the number of traits increased. All
205 combinations containing either trait 1 or trait 2 showed similar trend to the single trait case with decreasing (respectively increasing) trend of FRic along the environment. Combinations of traits with both trait 1 and trait 2 showed lower FRic than any other combination for all environments, with a humped shape of higher FRic at the middle of the environmental gradient and lower FRic at the extremes. FRic SES showed similar trends, with combinations containing either trait 1 or trait 2 showing opposing trend along the environment
210 well below -1.96, combinations without trait 1 nor trait 2 being between -1.96 and 1.96 and combinations containing both trait 1 and trait 2 showing lower FRic SES than combinations with either ones below -1.96.

FDis showed trends similar to FRic. Combinations of non-filtered traits showed no trend along the environment. Combination with either trait 1 or trait 2 showed decreasing (respectively increasing). Combinations with both trait 1 and trait 2 showed no trend in FDis, they however exhibit lower FDis than combinations
215 with neither or either one. FDis increased with increasing number of traits. FDis SESs were similar to FDis trends along the environment. All combinations of non-filtered traits had SES values between -1.96 and 1.96, while combinations with either trait 1 or trait 2 showed gradients with SES values lower than -1.96. Combinations with both traits showed lower SES than other combinations for all traits.

Discussion

220 Through a simulation approach we showed that increasing the number of traits not affected by the community assembly process can blur the functional diversity-environment relationship. We tested various situations with different migration rates and different traits with opposed filtering along the environmental variables. We proposed a method to compute the statistical power of each functional diversity the SES threshold to maximize the trend detection depending on the number of traits. This threshold adjustment should be
225 dependent on the functional diversity index used and the main community assembly process hypothesized. Our results show that additional non-filtered traits weaken the functional diversity-environment relationship

compared to the filtered trait.

Our simulations underline the importance of trait selection when computing functional diversity to avoid the dilution effect seen when adding unaffected traits. Among all the possible measured traits, the recommendation is to focus on the trait that are the most linked to the studied process (Violle *et al.*, 2007; Shipley *et al.*, 2016). At smaller-scale the knowledge of the main community assembly drivers help selecting the most relevant traits to detect this effect (Shipley *et al.*, 2016). At this scale, the link between trait and fitness can be more straightforward. However, at biogeographical scales, when the knowledge of both traits and the environment is coarser, trait-environment relationship are less tight, and the risk of dilution effect should be greater (Swenson *et al.*, 2012; Borgy *et al.*, 2017). The context-dependency of traits as well as the fact that we are missing measurements for relevant traits increase the risk for dilution. To account for the dilution effect, we proposed to compute statistical power when testing trends in functional diversity-environment relationships. It shows how when using seven or more traits the dilution effect was strong. This number of trait echoes the dimensionality of trait space recently shown in plants (Laughlin, 2014).

The difference of sensitivity of FDis and FRic to the dilution effect of additional traits, underlines how different functional diversity indices assess distinct aspects of Functional Diversity (Villéger *et al.*, 2008; Mouchet *et al.*, 2010; Schleuter *et al.*, 2010; Pavoine & Bonsall, 2011; Mason *et al.*, 2013). Different functional indices reflect distinct aspect of functional diversity (McPherson *et al.*, 2018). For example Functional Richness (FRic), being defined as the convex hull of species present in the functional space, is particularly sensitive to the presence of outliers and represents the trait space occupied by all species present (Villéger *et al.*, 2008). Functional Dispersion (FDis), on the other hand, is the abundance-weighted centroid in trait space. FDis reflects to what extent most abundant species are at the border or in the center of trait space. It is used to detect limiting similarity (Villéger *et al.*, 2008). The decrease in power with additional traits was stronger in FDis than in FRic in our simulations. It suggests that adjusting the SES threshold to get a fixed statistical power may be index-specific. Our method has to be adapted to additional combination of functional diversity indices and community assembly processes, to determine the power reduction associated with increasing number of traits.

Our simulation setup only considered environmental filtering and neutral dynamics while many other assembly processes can affect functional diversity-environment relationships (Münkemüller *et al.*, 2012; Mason *et al.*, 2013). For example limiting similarity also affects trait-environment distribution and thus functional diversity-environment relationships, it increases the expected functional diversity and counteracts environmental filtering. Limiting similarity is expected to spread and increase the evenness of species in the trait space and thus to increase FRic and FDis along the environment (Cornwell & Ackerly, 2009). But it has

been recently shown that competitive hierarchy could lead to similar patterns than environmental filtering
260 (Mayfield & Levine, 2010; Bernard-Verdier *et al.*, 2012; HilleRisLambers *et al.*, 2012; Perronne *et al.*, 2017).
Furthermore, depending on the scale at which we consider functional diversity, distinct processes can af-
fect different functional diversity facets (de Bello *et al.*, 2013). Indeed, de Bello *et al.* (2013) found that
environmental filtering at large scales mainly affect FRic while at finer-scale it affected FDis and FDiv.

Neutral processes increase functional diversity variability along the environment. We showed in our simu-
265 lation how lower migration rate made observed functional diversity more variable and lower—in absolute—
SEs. Indeed, low migration rate implies that less individuals and, in turn, less species can reach the observed
communities from the species pool, making functional diversity more variable (Münkemüller *et al.*, 2012).
This can hinder our ability to correctly identify a functional diversity-environment relationship with more
than one trait. Dispersal limitation can also decrease observed functional diversity along the environment
270 because of low migration rate (Spasojevic *et al.*, 2014). Thus our simulation approach could be completed by
the integration of additional community assembly processes that also affect functional diversity-environment
relationship, similarly to what has been done on phylogenetic diversity indices (Kraft *et al.*, 2007).

When considering single trait functional diversity-environment relationships, our simulations distinguish
between the filtered trait and all the other traits, while the distinction gets more difficult with additional
275 traits. Indeed, single-trait approaches have been suggested to delve into the details of community assembly,
while multi-trait approaches aggregate many processes (Butterfield & Suding, 2013; Spasojevic & Suding,
2013; Lefcheck *et al.*, 2015). From our simulations, we estimated the dilution effect due to the use of
unaffected traits in functional diversity. If a trait is singled out by showing a distinct relationship than
the other traits with the environment, then it may be filtered. However, our simulation setup is a simplified
280 situation where a single trait respond to the environment while others are uncorrelated to it, while trait
correlations are widespread in real datasets (Westoby *et al.*, 2002). Trait correlation can cause multiple
traits to have similar relationships with the environment while only one is affected by the process (Laughlin
& Messier, 2015; Dwyer & Laughlin, 2017). Different trait combinations can also exhibit similar fitness
along the environment if environmental filtering is not uni-modal without any trait correlations (Marks &
285 Lechowicz, 2006; Laughlin & Messier, 2015; Corrêa Dias *et al.*, 2019). Single traits can also be influenced
by multiple processes simultaneously and thus reach far from the optimum value when considering a single
process (Sack & Buckley, 2019).

Multi-traits functional diversity indices can however unravel patterns that cannot be found using single-trait
indices (Cornwell & Ackerly, 2009; Kraft *et al.*, 2015b). For example, disturbances can have effect that
290 cannot be resumed looking at single trait indices (Mouillot *et al.*, 2013), but revealed by multi-traits metrics

(Villéger *et al.*, 2011). Similarly, ecosystem multifunctionality is better predicted when accounting for multiple traits than using single traits (Mouillot *et al.*, 2011). Multiple traits need to be taken to account when explain multiple species stable coexistence (Kraft *et al.*, 2015b). Indeed, only using single-trait index can be informative as stated above, but multiple independent functional dimensions cannot be reduced by
295 uni-dimensional indices. Furthermore, in the absence of a priori knowledge of trait importance for community assembly, multidimensional functional diversity indices can help understand broad patterns (Lefcheck *et al.*, 2015).

In the absence of a priori knowledge regarding the relationship between individual traits and environmental gradients, single-trait indices may be impractical (Lefcheck *et al.*, 2015). We suggested a method to reliably
300 identify functional diversity-environment relationship. We showed in our simulations that to obtain 90% power with 7 traits, the SES threshold should be decreased to 1.5 instead of the commonly used 1.96. Our simulation approach let us run statistical power analyses to avoid not detecting functional diversity trends. The SES significance threshold needed to be adjusted to keep the same statistical power for FDis but not FRic in our simulations. Thus the change in significance threshold we suggest should not be applied blindly but
305 properly estimated depending on the expected dominant community assembly processes and the functional diversity indices used.

Acknowledgments

We would like to thank Pierre Denelle for helpful discussions. MG was supported by the ENS de Lyon. This study was supported by the European Research Council (ERC) Starting Grant Project ‘ecophysiological and
310 biophysical constraints on domestication in crop plants’ (grant ERC-StG-2014-639706-CONSTRAINTS) and by the French Foundation for Research on Biodiversity (FRB; <www.fondationbiodiversite.fr>) and EDF in the context of the CESAB project ‘causes and consequences of functional rarity from local to global scales’ (FREE).

References

- 315 Bernard-Verdier, M., Navas, M.-L., Vellend, M., Violle, C., Fayolle, A. & Garnier, E. (2012) Community assembly along a soil depth gradient: Contrasting patterns of plant trait convergence and divergence in a Mediterranean rangeland. *Journal of Ecology*, **100**, 1422–1433.
- Borgy, B., Violle, C., Choler, P., Garnier, E., Kattge, J., Loranger, J., Amiaud, B., Cellier, P., Debarros, G., Denelle, P., Diquelou, S., Gachet, S., Jolivet, C., Lavorel, S., Lemauviel-Lavenant, S., Mikolajczak, A.,

- 320 Munoz, F., Olivier, J. & Viovy, N. (2017) Sensitivity of community-level trait–environment relationships to data representativeness: A test for functional biogeography. *Global Ecology and Biogeography*, **26**, 729–739.
- Botta-Dukát, Z. & Czúcz, B. (2015) Testing the ability of functional diversity indices to detect trait convergence and divergence using individual-based simulation. *Methods in Ecology and Evolution*, n/a–n/a.
- Bruehlheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B., Hennekens, S.M., Botta-Dukát, Z.,
 325 Chytrý, M., Field, R., Jansen, F., Kattge, J., Pillar, V.D., Schrodte, F., Mahecha, M.D., Peet, R.K., Sandel, B., Bodegom, P. van, Altman, J., Alvarez-Dávila, E., Khan, M.A.S.A., Attorre, F., Aubin, I., Baraloto, C., Barroso, J.G., Bauters, M., Bergmeier, E., Biurrun, I., Bjorkman, A.D., Blonder, B., Čarni, A., Cayuela, L., Černý, T., Cornelissen, J.H.C., Craven, D., Dainese, M., Derroire, G., Sanctis, M.D., Díaz, S., Doležal, J., Farfan-Rios, W., Feldpausch, T.R., Fenton, N.J., Garnier, E., Guerin, G.R., Gutiérrez, A.G., Haider, S.,
 330 Hattab, T., Henry, G., Hérault, B., Higuchi, P., Hölzel, N., Homeier, J., Jentsch, A., Jürgens, N., Kačák, Z., Karger, D.N., Kessler, M., Kleyer, M., Knollová, I., Korolyuk, A.Y., Kühn, I., Laughlin, D.C., Lens, F., Loos, J., Louault, F., Lyubenova, M.I., Malhi, Y., Marcenò, C., Mencuccini, M., Müller, J.V., Munzinger, J., Myers-Smith, I.H., Neill, D.A., Niinemets, Ü., Orwin, K.H., Ozinga, W.A., Penuelas, J., Pérez-Haase, A., Petřík, P., Phillips, O.L., Pärtel, M., Reich, P.B., Römermann, C., Rodrigues, A.V., Sabatini, F.M.,
 335 Sardans, J., Schmidt, M., Seidler, G., Espejo, J.E.S., Silveira, M., Smyth, A., Sporbert, M., Svenning, J.-C., Tang, Z., Thomas, R., Tsiripidis, I., Vassilev, K., Violle, C., Virtanen, R., Weiher, E., Welk, E., Wesche, K., Winter, M., Wirth, C. & Jandt, U. (2018) Global trait–environment relationships of plant communities. *Nature Ecology & Evolution*, **2**, 1906.
- Butterfield, B.J. & Suding, K.N. (2013) Single-trait functional indices outperform multi-trait indices in
 340 linking environmental gradients and ecosystem services in a complex landscape. *Journal of Ecology*, **101**, 9–17.
- Céréghino, R., Pillar, V.D., Srivastava, D.S., de Omena, P.M., MacDonald, A.A.M., Barberis, I.M., Corbara, B., Guzman, L.M., Leroy, C., Bautista, F.O., Romero, G.Q., Trzcinski, M.K., Kratina, P., Debastiani, V.J., Gonçalves, A.Z., Marino, N.A.C., Farjalla, V.F., Richardson, B.A., Richardson, M.J., Dézerald, O., Gilbert,
 345 B., Petermann, J., Talaga, S., Piccoli, G.C.O., Jocqué, M. & Montero, G. (2018) Constraints on the functional trait space of aquatic invertebrates in bromeliads. *Functional Ecology*, **32**, 2435–2447.
- Chapman, A.S.A., Beaulieu, S.E., Colaço, A., Gebruk, A.V., Hilario, A., Kihara, T.C., Ramirez-Llodra, E., Sarrazin, J., Tunnicliffe, V., Amon, D.J., Baker, M.C., Boschen-Rose, R.E., Chen, C., Cooper, I.J., Copley, J.T., Corbari, L., Cordes, E.E., Cuvelier, D., Duperron, S., Du Preez, C., Gollner, S., Horton, T., Hourdez,
 350 S., Krylova, E.M., Linse, K., LokaBharathi, P.A., Marsh, L., Matabos, M., Mills, S.W., Mullineaux, L.S.,

- Rapp, H.T., Reid, W.D.K., Rybakova (Goroslavskaya), E., A. Thomas, T.R., Southgate, S.J., Stöhr, S., Turner, P.J., Watanabe, H.K., Yasuhara, M. & Bates, A.E. (2019) sFDvent: A global trait database for deep-sea hydrothermal-vent fauna. *Global Ecology and Biogeography*, **28**, 1538–1551.
- Cornwell, W.K. & Ackerly, D.D. (2009) Community assembly and shifts in plant trait distributions across
 355 an environmental gradient in coastal California. *Ecological Monographs*, **79**, 109–126.
- Corrêa Dias, A.T., Rosado, B.H.P., de Bello, F., Pistón, N. & de Mattos, E.A. (2019) Alternative plant designs: Consequences for community assembly and ecosystem functioning. *Annals of Botany*, mcz180.
- Dalerum, F. (2013) Phylogenetic and functional diversity in large carnivore assemblages. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**, 20130049.
- 360 de Bello, F., Lavorel, S., Lavergne, S., Albert, C.H., Boulangeat, I., Mazel, F. & Thuiller, W. (2013) Hierarchical effects of environmental filters on the functional structure of plant communities: A case study in the French Alps. *Ecography*, **36**, 393–402.
- Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Prentice, I.C., Garnier, E., Bönisch, G., Westoby, M., Poorter, H., Reich, P.B., Moles, A.T., Dickie, J.,
 365 Gillison, A.N., Zanne, A.E., Chave, J., Wright, S.J., Sheremet'ev, S.N., Jactel, H., Baraloto, C., Cerabolini, B., Pierce, S., Shipley, B., Kirkup, D., Casanoves, F., Joswig, J.S., Günther, A., Falczuk, V., Rüger, N., Mahecha, M.D. & Gorné, L.D. (2016) The global spectrum of plant form and function. *Nature*, **529**, 167–171.
- Dwyer, J.M. & Laughlin, D.C. (2017) Constraints on trait combinations explain climatic drivers of biodiversity: The importance of trait covariance in community assembly. *Ecology Letters*, **20**, 872–882.
- 370 Floury, M., Souchon, Y. & Looy, K.V. (2017) Climatic and trophic processes drive long-term changes in functional diversity of freshwater invertebrate communities. *Ecography*, n/a–n/a.
- Gotelli, N.J. & Graves, G.R. (1996) *Null models in ecology*, Smithsonian Institution Press, Washington.
- Habel, K., Grasman, R., Gramacy, R.B., Mozharovskiy, P. & Sterratt, D.C. (2019) *Geometry: Mesh generation and surface tessellation*,
- 375 HilleRisLambers, J., Adler, P.B., Harpole, W.S., Levine, J.M. & Mayfield, M.M. (2012) Rethinking Community Assembly through the Lens of Coexistence Theory. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 227–248.
- Hirzel, A.H., Helfer, V. & Metral, F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, **145**, 111–121.

380 Hutchinson, G.E. (1957) Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

Iversen, C.M., McCormack, M.L., Powell, A.S., Blackwood, C.B., Freschet, G.T., Kattge, J., Roumet, C., Stover, D.B., Soudzilovskaia, N.A., Valverde-Barrantes, O.J., van Bodegom, P.M. & Violle, C. (2017) A global Fine-Root Ecology Database to address below-ground challenges in plant ecology. *New Phytologist*,
385 **215**, 15–26.

Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L., Safi, K., Sechrest, W., Boakes, E.H., Carbone, C., Connolly, C., Cutts, M.J., Foster, J.K., Grenyer, R., Habib, M., Plaster, C.A., Price, S.A., Rigby, E.A., Rist, J., Teacher, A., Bininda-Emonds, O.R.P., Gittleman, J.L., Mace, G.M. & Purvis, A. (2009) PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently
390 extinct mammals. *Ecology*, **90**, 2648–2648.

Kattge, J., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P.B., Wright, I.J., Cornelissen, J.H.C., Violle, C., Harrison, S.P., Van BODEGOM, P.M., Reichstein, M., Enquist, B.J., Soudzilovskaia, N.A., Ackerly, D.D., Anand, M., Atkin, O., Bahn, M., Baker, T.R., Baldocchi, D., Bekker, R., Blanco, C.C., Blonder, B., Bond, W.J., Bradstock, R., Bunker, D.E., Casanoves, F., Cavender-Bares, J., Chambers, J.Q., Chapin Iii, F.S., Chave, J., Coomes, D., Cornwell, W.K., Craine, J.M., Dobrin, B.H., Duarte, L., Durka, W., Elser, J., Esser, G., Estiarte, M., Fagan, W.F., Fang, J., Fernández-Méndez, F., Fidelis, A., Finegan, B., Flores, O., Ford, H., Frank, D., Freschet, G.T., Fyllas, N.M., Gallagher, R.V., Green, W.A., Gutierrez, A.G., Hickler, T., Higgins, S.I., Hodgson, J.G., Jalili, A., Jansen, S., Joly, C.A., Kerkhoff, A.J., Kirkup, D., Kitajima, K., Kleyer, M., Klotz, S., Knops, J.M.H.,
400 Kramer, K., Kühn, I., Kurokawa, H., Laughlin, D., Lee, T.D., Leishman, M., Lens, F., Lenz, T., Lewis, S.L., Lloyd, J., Llusà, J., Louault, F., Ma, S., Mahecha, M.D., Manning, P., Massad, T., Medlyn, B.E., Messier, J., Moles, A.T., Müller, S.C., Nadrowski, K., Naeem, S., Niinemets, Nöllert, S., Nüske, A., Ogaya, R., Oleksyn, J., Onipchenko, V.G., Onoda, Y., Ordoñez, J., Overbeck, G., Ozinga, W.A., Patiño, S., Paula, S., Pausas, J.G., Peñuelas, J., Phillips, O.L., Pillar, V., Poorter, H., Poorter, L., Poschlod, P., Prinzing, A., Proulx, R., Rammig, A., Reinsch, S., Reu, B., Sack, L., Salgado-Negret, B., Sardans, J., Shiodera, S., Shipley, B., Siefert, A., Sosinski, E., Soussana, J.-F., Swaine, E., Swenson, N., Thompson, K., Thornton, P., Waldram, M., Weiher, E., White, M., White, S., Wright, S.J., Yguel, B., Zaehle, S., Zanne, A.E. & Wirth, C. (2011) TRY a global database of plant traits. *Global Change Biology*, **17**, 2905–2935.

Keddy, P.A. (1992) Assembly and response rules: Two goals for predictive community ecology. *Journal of*
410 *Vegetation Science*, **3**, 157–164.

- Kraft, N.J.B., Adler, P.B., Godoy, O., James, E.C., Fuller, S. & Levine, J.M. (2015a) Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, **29**, 592–599.
- Kraft, N.J.B., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2007) Trait Evolution, Community Assembly, and the Phylogenetic Structure of Ecological Communities. *The American Naturalist*, **170**, 271–283.
- 415 Kraft, N.J.B., Godoy, O. & Levine, J.M. (2015b) Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences*, **112**, 797–802.
- Laliberté, E. & Legendre, P. (2010) A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, **91**, 299–305.
- Laliberté, E., Legendre, P. & Shipley, B. (2014) *FD: Measuring functional diversity from multiple traits, and*
420 *other tools for functional ecology*,
- Laughlin, D.C. (2014) The intrinsic dimensionality of plant traits and its relevance to community assembly. *Journal of Ecology*, **102**, 186–193.
- Laughlin, D.C. & Messier, J. (2015) Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends in Ecology & Evolution*, **30**, 487–496.
- 425 Lefcheck, J.S., Bastazini, V.A.G. & Griffin, J.N. (2015) Choosing and using multiple traits in functional diversity research. *Environmental Conservation*, **42**, 104–107.
- Legras, G., Loiseau, N., Gaertner, J.-C., Poggiale, J.-C. & Gaertner-Mazouni, N. (2019) Assessing functional diversity: The influence of the number of the functional traits. *Theoretical Ecology*.
- Madin, J.S., Anderson, K.D., Andreasen, M.H., Bridge, T.C.L., Cairns, S.D., Connolly, S.R., Darling, E.S.,
430 Diaz, M., Falster, D.S., Franklin, E.C., Gates, R.D., Hoogenboom, M.O., Huang, D., Keith, S.A., Kosnik, M.A., Kuo, C.-Y., Lough, J.M., Lovelock, C.E., Luiz, O., Martinelli, J., Mizerek, T., Pandolfi, J.M., Pochon, X., Pratchett, M.S., Putnam, H.M., Roberts, T.E., Stat, M., Wallace, C.C., Widman, E. & Baird, A.H. (2016) The Coral Trait Database, a curated database of trait information for coral species from the global oceans. *Scientific Data*, **3**, sdata201617.
- 435 Maire, E., Grenouillet, G., Brosse, S. & Villéger, S. (2015) How many dimensions are needed to accurately assess functional diversity? A pragmatic approach for assessing the quality of functional spaces. *Global Ecology and Biogeography*, **24**, 728–740.
- Marks, C.O. & Lechowicz, M.J. (2006) Alternative Designs and the Evolution of Functional Diversity. *The American Naturalist*, **167**, 55–66.

- 440 Mason, N.W.H., de Bello, F., Mouillot, D., Pavoine, S. & Dray, S. (2013) A guide for using functional diversity indices to reveal changes in assembly processes along ecological gradients. *Journal of Vegetation Science*, **24**, 794–806.
- Mayfield, M.M. & Levine, J.M. (2010) Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, **13**, 1085–1093.
- 445 McPherson, J.M., Yeager, L.A. & Baum, J.K. (2018) A simulation tool to scrutinise the behaviour of functional diversity metrics. *Methods in Ecology and Evolution*, **9**, 200–206.
- Meynard, C.N., Leroy, B. & Kaplan, D.M. Testing methods in species distribution modelling using virtual species: What have we learnt and what are we missing? *Ecography*, **0**.
- Mouchet, M.A., Villéger, S., Mason, N.W.H. & Mouillot, D. (2010) Functional diversity measures: An
450 overview of their redundancy and their ability to discriminate community assembly rules: Functional diversity measures. *Functional Ecology*, **24**, 867–876.
- Mouillot, D., Graham, N.A.J., Villéger, S., Mason, N.W.H. & Bellwood, D.R. (2013) A functional approach reveals community responses to disturbances. *Trends in Ecology & Evolution*, **28**, 167–177.
- Mouillot, D., Villéger, S., Scherer-Lorenzen, M. & Mason, N.W.H. (2011) Functional Structure of Biological
455 Communities Predicts Ecosystem Multifunctionality. *PLoS ONE*, **6**.
- Munoz, F., Grenié, M., Denelle, P., Taudiere, A., Laroche, F., Tucker, C. & Violle, C. (2017) Ecolottery: Simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution*, **in press**.
- Münkemüller, T., de Bello, F., Meynard, C.N., Gravel, D., Lavergne, S., Mouillot, D., Mouquet, N. &
460 Thuiller, W. (2012) From diversity indices to community assembly processes: A test with simulated data. *Ecography*, **35**, 468–480.
- Münkemüller, T. & Gallien, L. (2015) VirtualCom: A simulation model for eco-evolutionary community assembly and invasion. *Methods in Ecology and Evolution*, **6**, 735–743.
- Oliveira, B.F., São-Pedro, V.A., Santos-Barrera, G., Penone, C. & Costa, G.C. (2017) AmphiBIO, a global
465 database for amphibian ecological traits. *Scientific Data*, **4**, 170123.
- Pavoine, S. & Bonsall, M.B. (2011) Measuring biodiversity to explain community assembly: A unified approach. *Biological Reviews*, **86**, 792–812.
- Peck, S.L. (2004) Simulation as experiment: A philosophical reassessment for biological modeling. *Trends*

in *Ecology & Evolution*, **19**, 530–534.

470 Perronne, R., Munoz, F., Borgy, B., Reboud, X. & Gaba, S. (2017) How to design trait-based analyses of community assembly mechanisms: Insights and guidelines from a literature review. *Perspectives in Plant Ecology, Evolution and Systematics*, **25**, 29–44.

Pianka, E.R., Vitt, L.J., Pelegrin, N., Fitzgerald, D.B. & Winemiller, K.O. (2017) Toward a Periodic Table of Niches, or Exploring the Lizard Niche Hypervolume. *The American Naturalist*, **190**, 601–616.

475 Pigot, A.L., Sheard, C., Miller, E.T., Bregman, T.P., Freeman, B.G., Roll, U., Seddon, N., Trisos, C.H., Weeks, B.C. & Tobias, J.A. (2020) Macroevolutionary convergence connects morphological form to ecological function in birds. *Nature Ecology & Evolution*, **4**, 230–239.

R Core Team (2019) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

480 Sack, L. & Buckley, T.N. (2019) Trait multi-functionality in plant stress response. *Integrative and Comparative Biology*, icz152.

Schleuter, D., Daufresne, M., Massol, F. & Argillier, C. (2010) A user's guide to functional diversity indices. *Ecological Monographs*, **80**, 469–484.

Schwilk, D.W. & Ackerly, D.D. (2005) Limiting similarity and functional diversity along environmental
485 gradients. *Ecology Letters*, **8**, 272–281.

Shipley, B., Bello, F.D., Cornelissen, J.H.C., Laliberté, E., Laughlin, D.C. & Reich, P.B. (2016) Reinforcing loose foundation stones in trait-based plant ecology. *Oecologia*, 1–9.

Spasojevic, M.J., Copeland, S. & Suding, K.N. (2014) Using functional diversity patterns to explore meta-community dynamics: A framework for understanding local and regional influences on community structure.
490 *Ecography*, **37**, 939–949.

Spasojevic, M.J. & Suding, K.N. (2013) Inferring community assembly mechanisms from functional diversity patterns: The importance of multiple assembly processes. *Journal of Ecology*, 652–661.

Swenson, N.G., Enquist, B.J., Pither, J., Kerkhoff, A.J., Boyle, B., Weiser, M.D., Elser, J.J., Fagan, W.F., Forero-Montaña, J., Fyllas, N., Kraft, N.J.B., Lake, J.K., Moles, A.T., Patiño, S., Phillips, O.L., Price, C.A.,
495 Reich, P.B., Quesada, C.A., Stegen, J.C., Valencia, R., Wright, I.J., Wright, S.J., Andelman, S., Jørgensen, P.M., Jr, T.E.L., Monteagudo, A., Núñez-Vargas, M.P., Vasquez-Martínez, R. & Noltling, K.M. (2012) The biogeography and filtering of woody plant functional diversity in North and South America. *Global Ecology*

and *Biogeography*, **21**, 798–808.

Thomas, F.M., Vesk, P.A. & Hauser, C.E. (2018) A field ecologist's adventures in the virtual world: Using
 500 simulations to design data collection for complex models. *Ecological Applications*, **28**, 2130–2141.

Tucker, C.M., Davies, T.J., Cadotte, M.W. & Pearse, W.D. (2018) On the relationship between phylogenetic
 diversity and trait diversity. *Ecology*, **99**, 1473–1479.

Villéger, S., Novack-Gottshall, P.M. & Mouillot, D. (2011) The multidimensionality of the niche reveals
 functional diversity changes in benthic marine biotas across geological time. *Ecology Letters*, **14**, 561–568.

505 Villéger, S., Mason, N.W.H. & Mouillot, D. (2008) New Multidimensional Functional Diversity Indices for a
 Multifaceted Framework in Functional Ecology. *Ecology*, **89**, 2290–2301.

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I. & Garnier, E. (2007) Let the
 concept of trait be functional! *Oikos*, **116**, 882–892.

Westoby, M., Falster, D.S., Moles, A.T., Vesk, P.A. & Wright, I.J. (2002) PLANT ECOLOGICAL STRATE-
 510 GIES: Some Leading Dimensions of Variation Between Species. *Annual Review of Ecology and Systematics*,
33, 125–159.

Winemiller, K.O., Fitzgerald, D.B., Bower, L.M. & Pianka, E.R. (2015) Functional traits, convergent evo-
 lution, and periodic tables of niches. *Ecology Letters*, **18**, 737–751.

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel,
 515 J., Reineking, B., Schröder, B. & Grimm, V. (2010) The virtual ecologist approach: Simulating data and
 observers. *Oikos*, **119**, 622–635.

Figures

**Figure 1: Functional diversity vs. environmental filtering strength in the case of a single
 filtered trait in function of the number of traits included to compute functional dissimilarities.**

520 In each panel the top row is the observed value and the bottom row is the Standard Effect Size (SES) from
 null models. All indices were computed based on presence-absence data only. **A:** Functional Richness. **B:**
 Functional Dispersion. Each point represent the functional diversity computed for a single community. **red**
 points shows combination which did not contained the filtered trait, **blue** points shows combination that
 contained the filtered trait. Solid lines are significant linear regressions. The dashed lines represents the line
 525 where the SES equals zero. Above it, it means that the observed functional diversity exceeds the functional

diversity in null models. Under it, observed functional diversity is below functional diversity in null models. Dotted lines shows lines of -1.96 and 1.96 SES representing quantiles 2.5% and 97.5% assuming a normal distribution.

Figure 2: Functional Diversity vs. environment with varying immigration rate from the same species pool. **A:** Functional Richness (FRic), **B:** Functional Dispersion (FDis). Each point represent the functional diversity computed for a single community. **red** points shows combination which did not contained the filtered trait, **blue** points shows combination that contained the filtered trait. Solid lines are significant linear regressions. The dashed lines represents the line where the SES equals zero. Above it, it means that the observed functional diversity exceeds the functional diversity in null models. Under it, observed functional diversity is below functional diversity in null models. Dotted lines shows lines of -1.96 and 1.96 SES representing quantiles 2.5% and 97.5% assuming a normal distribution. The panels differentiate between **observed** values of functional diversity and **Standard Effect Sizes (SESS)**, with migration rate of 0.1 ($m = 0.1$) or of 1 ($m = 1$).

Figure 3: Power analyses of functional diversity with varying SES thresholds. Evolution of statistical power with varying SES threshold for (left) Functional Dispersion and (right) Functional Richness. The colored lines showed the relationship for different trait numbers. The ticked line shows the “classical” threshold of 1.96.

Figure 4: Functional diversity vs. environmental filtering strength in the case of a two filtered traits, considering increasing number of traits. In each panel the top row is the observed value and the bottom row is the Standard Effect Size (SES) from null models. All indices were computed based on presence-absence data only. **A:** Functional Richness. **B:** Functional Dispersion. **red** points shows combination which did not contained either of the filtered traits, **green** points combination which contained trait 1, **blue** points combination which contained trait 2, **purple** points combination which contained both filtered traits. The dashed lines represents the line where the SES equals zero. Above it, it means that the observed functional diversity exceeds the functional diversity in null models. Under it, observed functional diversity is below functional diversity in null models. Dotted lines shows lines of -1.96 and 1.96 SES representing quantiles 2.5% and 97.5% assuming a normal distribution.

Supporting Information

Supplementary Figure S1: Conceptual Figure showing the simulation process. (1) We first define Gaussian environmental filter for a single trait along a gradient that increase in mean regularly and

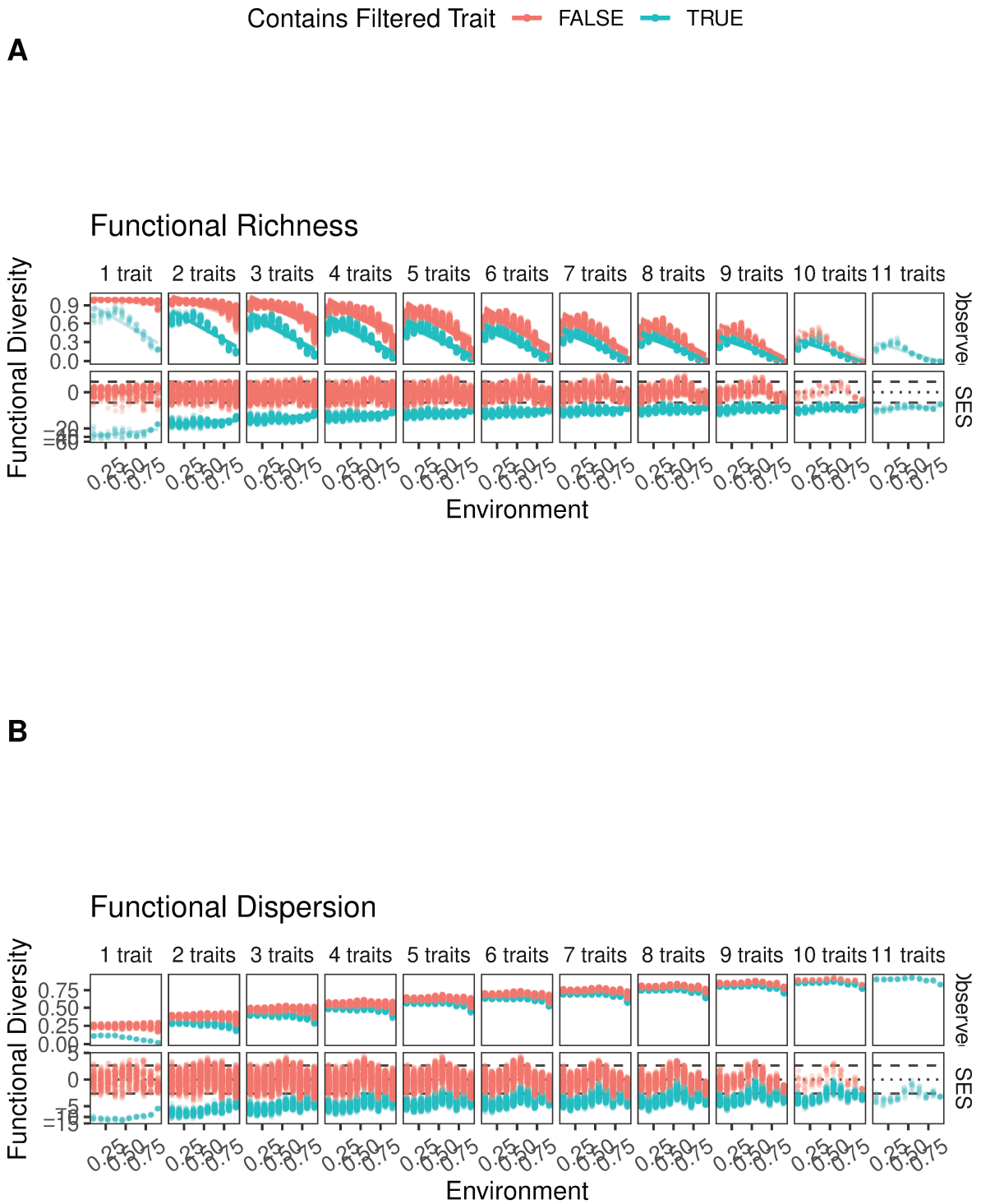


Figure 1

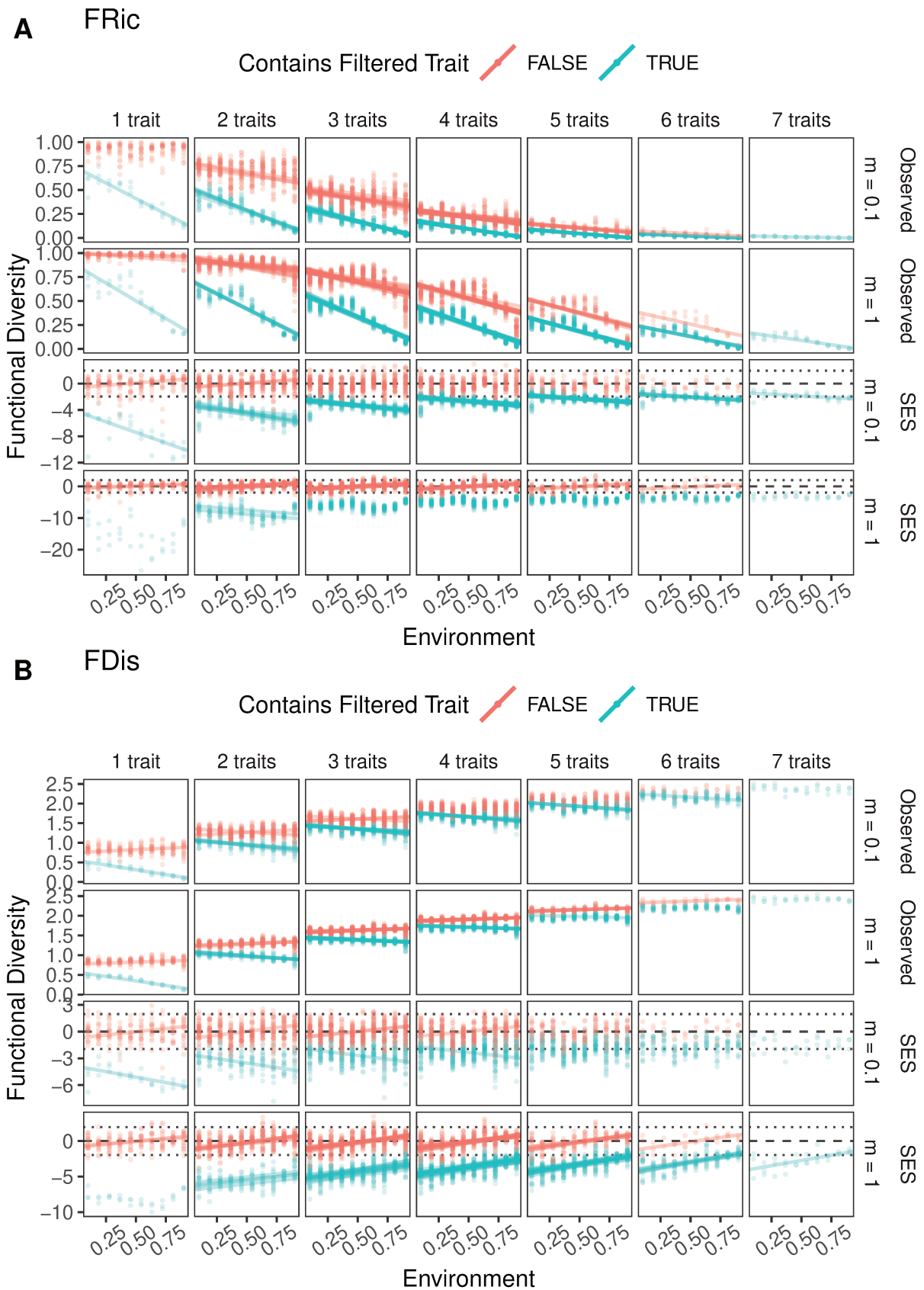


Figure 2

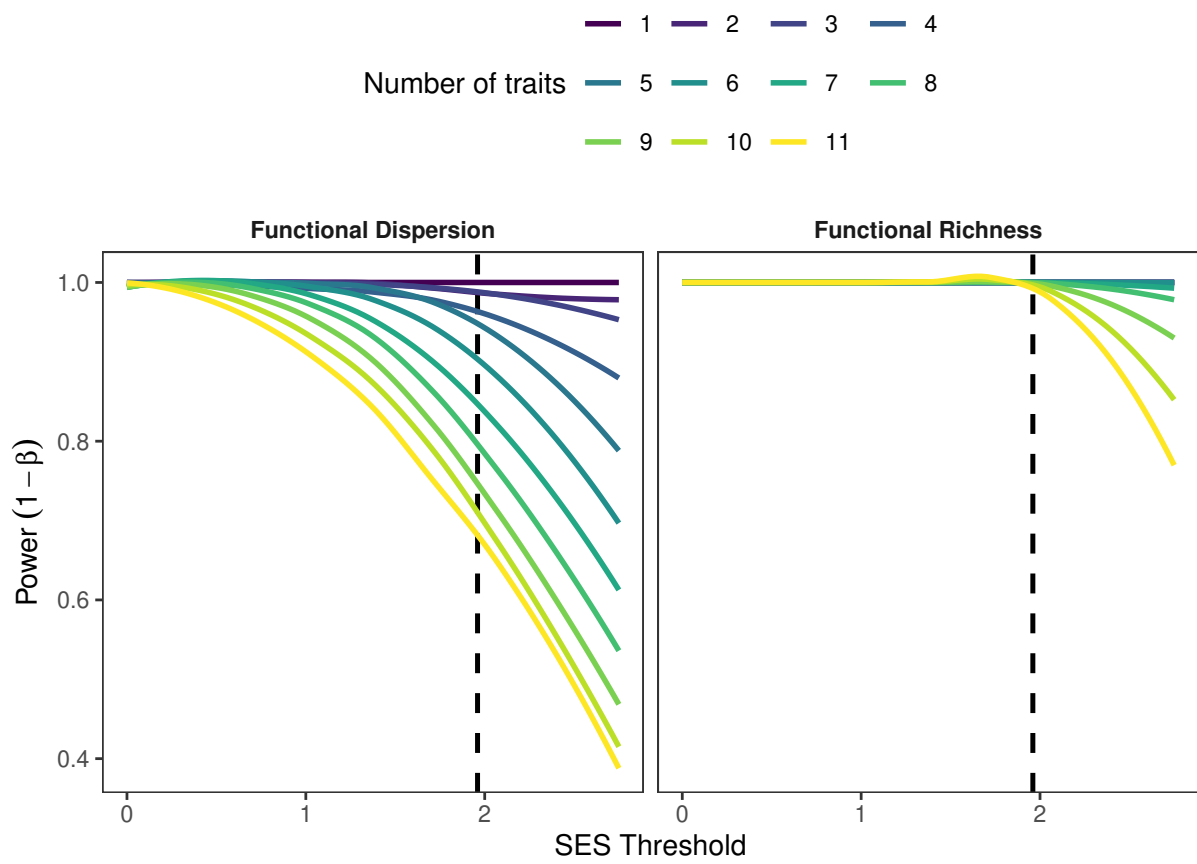


Figure 3

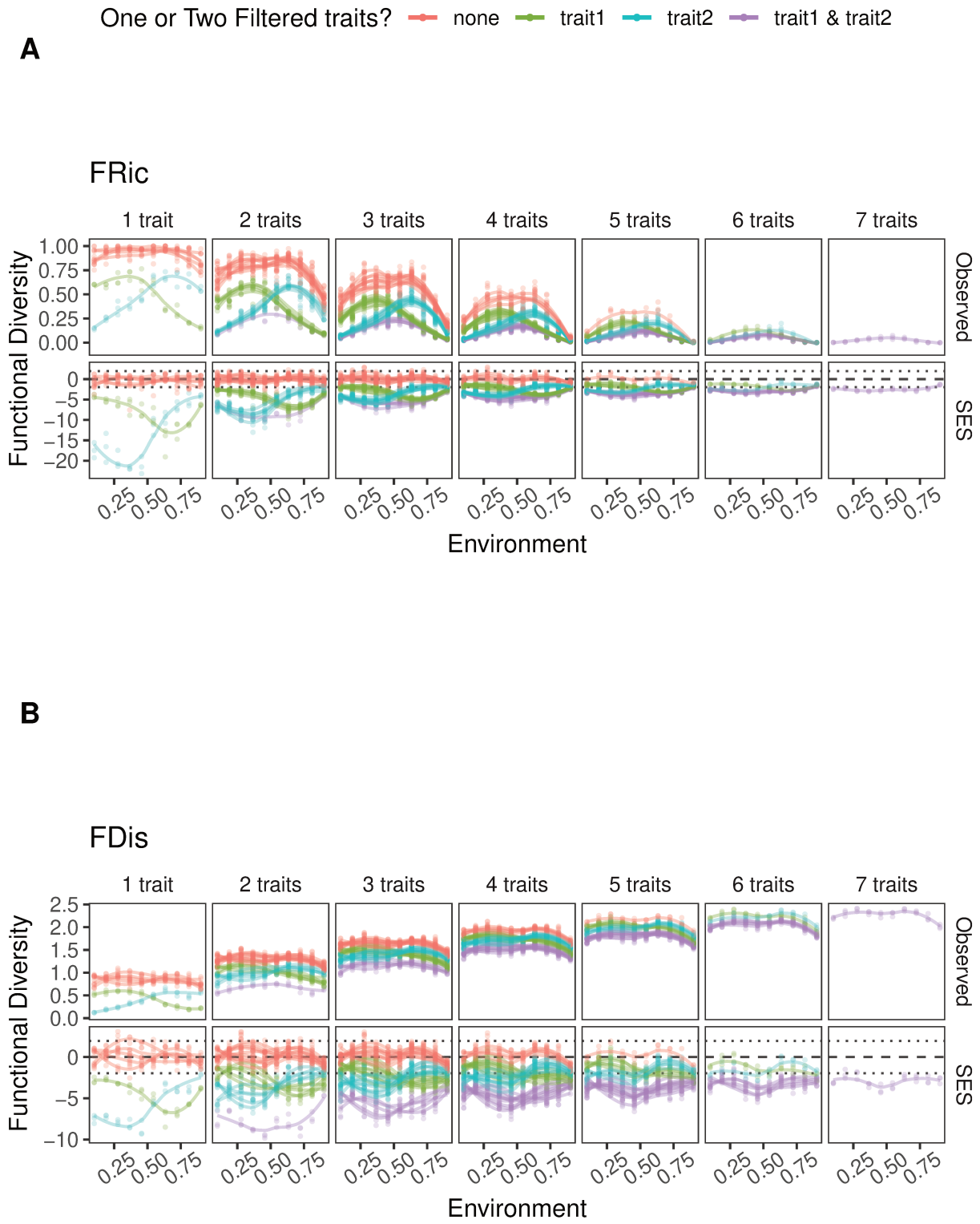
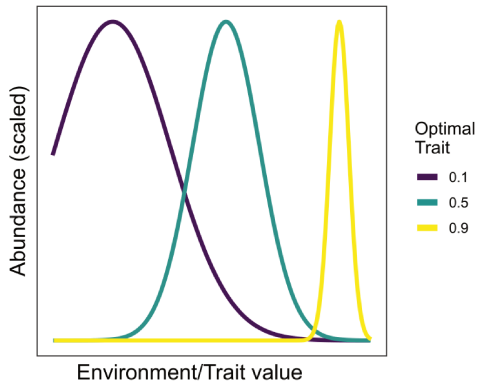


Figure 4

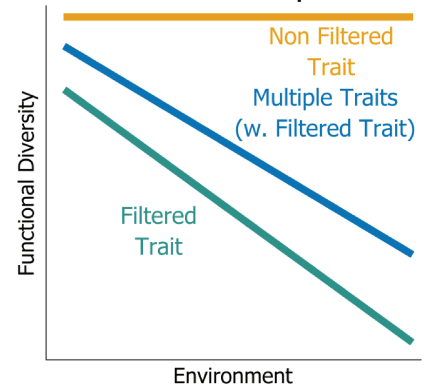
1 Single Filtered Trait

Environmental Filtering



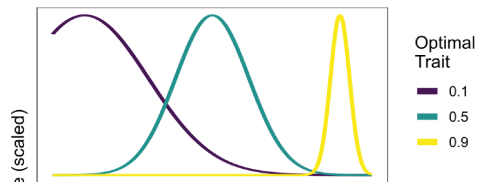
Simulate Communities
ecolottery

Expected FD-Environment relationships

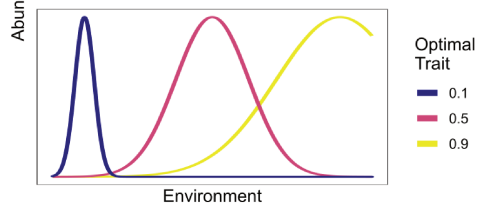


2 Opposing Filtered Traits

Trait 1

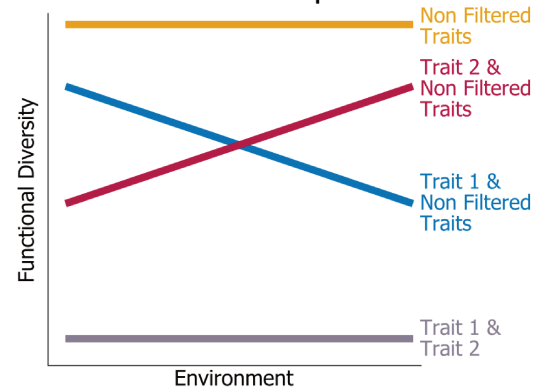


Trait 2



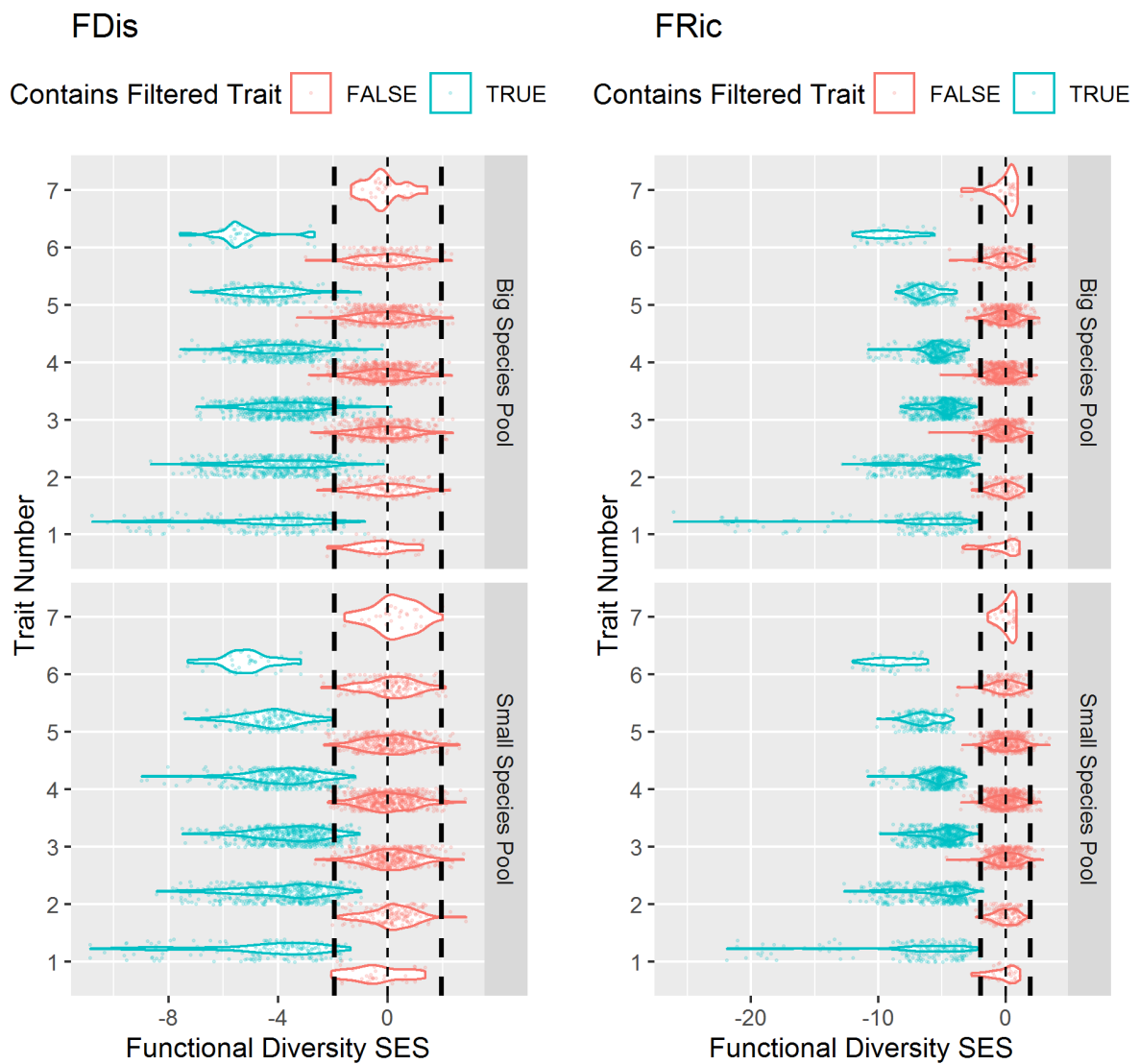
Simulate Communities
ecolottery

Multi Traits FD Expectations

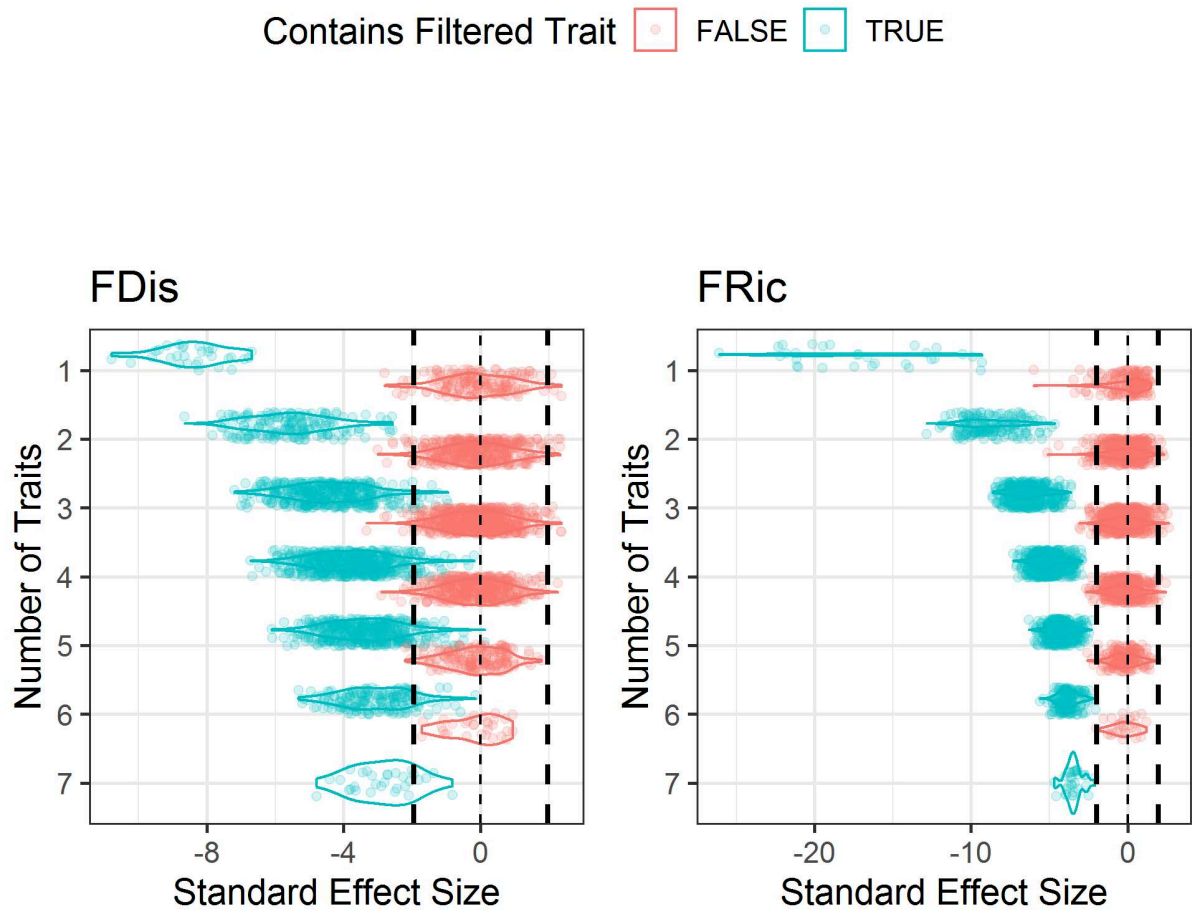


Supplementary Figure S1

decrease in variance. The environmental filter only applies to a single trait. With these filters and a uniform pool of species, we simulate communities along the environmental gradient using ecolottery (Munoz *et al.*, 2017). We then compute functional diversity indices for each community and all combinations of one to seven traits. We expect functional diversity to decrease along the gradient when computed using the filtered trait (green line) while it should be constant without (orange line); any combination of traits containing the filtered trait should also exhibit decreasing functional diversity along the environment (blue line). **(2)** For our second simulation setup we use same environmental filters, with one in similar direction on a first trait while the other one is in the opposite direction on a second trait. We do similarly by simulating communities then computing functional diversity indices with all possible combinations of traits. We expect opposing functional diversity-environment relationships when considering each filtered trait (blue and red lines), while they should be no trend in functional diversity when considering both traits at the same time (gray line). However we expect a lower functional diversity when considering both traits (gray line) than a random combinations of non-filtered traits (orange line).



Supplementary Figure S 2: Functional diversity SES distributions with varying pool sizes



Supplementary Figure S 3: **Functional Diversity Standard Effect Size (SES) distribution in function of the number of traits for trait combinations that contained and did not contained the filtered trait.** (left) SES of Functional dispersion (right) SES of Functional Richness. The SES distribution corresponds to a standardized difference between null models and observed distribution. Vertical dashed lines show SES of 0 (thin dashed line) and -1.96 & 1.96 SES values (thick dashed lines) which is a classically used threshold value to define significant deviation from null models.

3

Les espèces fonctionnellement originales dévient
de l'optimum environnemental

Dans le paradigme de l'optimalité fonctionnelle, on se concentre sur l'étude du trait moyen pondéré dans la communauté. Ce trait ne décrit pourtant que le trait le plus abondant de la communauté. Au contraire certaines espèces possèdent des traits particuliers et sont à la marge de la communauté en termes de traits, elles sont originales fonctionnellement. L'objectif de ce chapitre est de présenter des outils permettant de quantifier l'originalité fonctionnelle dans les communautés et d'en évaluer certains déterminants.

Pour évaluer l'originalité fonctionnelle, on peut utiliser des indices proposés par Violle et collab. (2017). La première partie de ce chapitre présente un paquet R qui permet de calculer ces indices, à différentes échelles (locale et régionale) et en considérant les différentes dimensions (originalité et fréquence). Cet article a fait l'objet d'une publication dans *Diversity and Distributions*.


Si le cadre proposé pour étudier l'originalité fonctionnelle propose des catégories discrètes. Les indices d'originalité fonctionnelle en proposent une quantification continue, qui permet d'ordonner les espèces selon leur originalité. De même si le cadre conceptuel propose d'évaluer l'originalité fonctionnelle à deux échelles spatiales (régionale et locale), il est possible de calculer l'originalité fonctionnelle à toutes les échelles spatiales considérées. Cette partie précise le contexte d'utilisation de l'originalité fonctionnelle et les questions qu'elle permet de traiter. Cet article a fait l'objet d'une publication dans *Trends in Ecology and Evolution*.

Certains facteurs environnementaux favorisent certaines stratégies. Il peut donc y avoir des environnements qui favorisent les espèces originales fonctionnellement à l'échelle régionale. Nous proposons dans la troisième partie de ce chapitre, d'évaluer des déterminants probables de l'excès d'espèces originales pour les plantes des Amériques. Cet article est en cours de préparation pour *Ecography*.

I UN OUTIL POUR CALCULER LA RARETÉ FONCTIONNELLE

DOI: 10.1111/ddi.12629

BIODIVERSITY LETTER

WILEY Diversity and Distributions A Journal of Community Biogeography**funrar: An R package to characterize functional rarity**Matthias Grenié¹  | Pierre Denelle¹ | Caroline M. Tucker^{1,2} | François Munoz^{3,4} |
Cyrille Violle¹¹Centre d'Écologie Fonctionnelle et Évolutive (UMR 5175), CNRS – Université de Montpellier – Université Paul Valéry Montpellier, EPHE, Montpellier, France²Department of Biology, University of North Carolina at Chapel Hill, Coker HallCB #3280 120 South Road, Chapel Hill, NC 27599-3280 USA³University of Montpellier, AMAP, Montpellier, France⁴French Institute of Pondicherry, Pondicherry, India**Correspondence**Matthias Grenié, Centre d'Écologie Fonctionnelle et Évolutive (UMR 5175), CNRS – Université de Montpellier – Université Paul Valéry Montpellier, EPHE, Montpellier, France.
Email: matthias.grenie@cefe.cnrs.fr**Funding information**

H2020 European Research Council, Grant/Award Number: ERC-StG-2014-639706-CONSTRAINTS and H2020-MSCA-IF-2014-657951

Editor: Cory Merow

Abstract

Emphasis has been put in recent ecological research on investigating phylogenetic, functional and taxonomic facets of biological diversity. While a flourishing number of indices have been proposed for assessing functional diversity, surprisingly few options are available to characterize functional rarity. Functional rarity can play a key role in community and ecosystem dynamics. We introduce here the *funrar* R package to quantify functional rarity based on species trait differences and species frequencies at local and regional scales. Because of the increasing availability of big datasets in macroecology and biogeography, we optimized *funrar* to work with large datasets of thousands of species and sites. We illustrate the use of the package to investigate the functional rarity of North and Central American mammals.

KEYWORDS

biodiversity, biodiversity indices, functional biogeography, functional trait, R package, rarity

1 | A TOOLKIT FOR THE DEFINITION AND QUANTIFICATION OF FUNCTIONAL RARITY

Biodiversity is multifaceted (Cardoso, Rigal, Borges, & Carvalho, 2014; Safi et al., 2011), and many indices have been proposed to summarize the taxonomic, phylogenetic and functional composition of ecological assemblages (Jarzyna & Jetz, 2016; Mazel et al., 2014). Such indices are used to investigate the influence of ecological, biogeographical and evolutionary processes at local and regional scales (McGill, Enquist, Weiher, & Westoby, 2006; Violle, Reich, Pacala, Enquist, & Kattge, 2014; Weiher et al., 2011). Many indices, such as community-weighted moments, emphasize the contribution of abundant taxa because they are expected to make significant contributions to community and ecosystem functioning (Grime, 1998; Enquist et al. 2015), while the role of rare taxa is less addressed.

Rarity relates to biodiversity dynamics at multiple scales of geographical and niche space. Rabinowitz (1981) defined rarity based on the geographical range, habitat specificity and local population

size of taxa, yielding seven forms of rarity. More recently, Violle et al. (2017) extended the scope of Rabinowitz's (1981) classification to further incorporate differences in functional traits among taxa, defining a new component—functional rarity. In this perspective, a species (or an individual) can be rare because of the uncommonness of its trait values and/or because of its low abundance at the local scale (Pavoine, Ollier, & Dufour, 2005). Indices of Functional Distinctiveness and Taxon Scarcity were proposed to quantify those two aspects at the local scale, respectively. A species can also be functionally rare at the regional scale because its functional characteristics are unique given the pool of species and/or because it is spatially restricted. Functional Uniqueness and Taxon Restrictedness respectively assess these two aspects at the regional scale. The four indices together provide a framework for characterizing functional rarity (Figures 1 and 2). Because functional rarity is expected to play a major role in ecosystem and biodiversity dynamics, the indices can be used to assess the influences of rare trait values on local and regional dynamics (Ricotta et al., 2016; Violle et al., 2017). We here introduce an R (R

DOI: 10.1111/ddi.12629

BIODIVERSITY LETTER

WILEY

Diversity and
DistributionsA Journal of
Community
Biogeography

funrar: An R package to characterize functional rarity

Matthias Grenié¹  | Pierre Denelle¹ | Caroline M. Tucker^{1,2} | François Munoz^{3,4} |
Cyrille Violle¹

¹Centre d'Écologie Fonctionnelle et Évolutive (UMR 5175), CNRS – Université de Montpellier – Université Paul Valéry Montpellier, EPHE, Montpellier, France

²Department of Biology, University of North Carolina at Chapel Hill. Coker HallCB #3280 120 South Road, Chapel Hill, NC 27599-3280 USA

³University of Montpellier, AMAP, Montpellier, France

⁴French Institute of Pondicherry, Pondicherry, India

Correspondence

Matthias Grenié, Centre d'Écologie Fonctionnelle et Évolutive (UMR 5175), CNRS – Université de Montpellier – Université Paul Valéry Montpellier, EPHE, Montpellier, France. Email: matthias.grenie@cefe.cnrs.fr

Funding information

H2020 European Research Council, Grant/Award Number: ERC-StG-2014-639706-CONSTRAINTS and H2020-MSCA-IF-2014-657951

Editor: Cory Merow

Abstract

Emphasis has been put in recent ecological research on investigating phylogenetic, functional and taxonomic facets of biological diversity. While a flourishing number of indices have been proposed for assessing functional diversity, surprisingly few options are available to characterize functional rarity. Functional rarity can play a key role in community and ecosystem dynamics. We introduce here the *funrar* R package to quantify functional rarity based on species trait differences and species frequencies at local and regional scales. Because of the increasing availability of big datasets in macroecology and biogeography, we optimized *funrar* to work with large datasets of thousands of species and sites. We illustrate the use of the package to investigate the functional rarity of North and Central American mammals.

KEYWORDS

biodiversity, biodiversity indices, functional biogeography, functional trait, R package, rarity

1 | A TOOLKIT FOR THE DEFINITION AND QUANTIFICATION OF FUNCTIONAL RARITY

Biodiversity is multifaceted (Cardoso, Rigal, Borges, & Carvalho, 2014; Safi et al., 2011), and many indices have been proposed to summarize the taxonomic, phylogenetic and functional composition of ecological assemblages (Jarzyna & Jetz, 2016; Mazel et al., 2014). Such indices are used to investigate the influence of ecological, biogeographical and evolutionary processes at local and regional scales (McGill, Enquist, Weiher, & Westoby, 2006; Violle, Reich, Pacala, Enquist, & Kattge, 2014; Weiher et al., 2011). Many indices, such as community-weighted moments, emphasize the contribution of abundant taxa because they are expected to make significant contributions to community and ecosystem functioning (Grime, 1998; Enquist et al. 2015), while the role of rare taxa is less addressed.

Rarity relates to biodiversity dynamics at multiple scales of geographical and niche space. Rabinowitz (1981) defined rarity based on the geographical range, habitat specificity and local population

size of taxa, yielding seven forms of rarity. More recently, Violle et al. (2017) extended the scope of Rabinowitz's (1981) classification to further incorporate differences in functional traits among taxa, defining a new component—functional rarity. In this perspective, a species (or an individual) can be rare because of the uncommonness of its trait values and/or because of its low abundance at the local scale (Pavoine, Ollier, & Dufour, 2005). Indices of Functional Distinctiveness and Taxon Scarcity were proposed to quantify those two aspects at the local scale, respectively. A species can also be functionally rare at the regional scale because its functional characteristics are unique given the pool of species and/or because it is spatially restricted. Functional Uniqueness and Taxon Restrictedness respectively assess these two aspects at the regional scale. The four indices together provide a framework for characterizing functional rarity (Figures 1 and 2). Because functional rarity is expected to play a major role in ecosystem and biodiversity dynamics, the indices can be used to assess the influences of rare trait values on local and regional dynamics (Ricotta et al., 2016; Violle et al., 2017). We here introduce an R (R

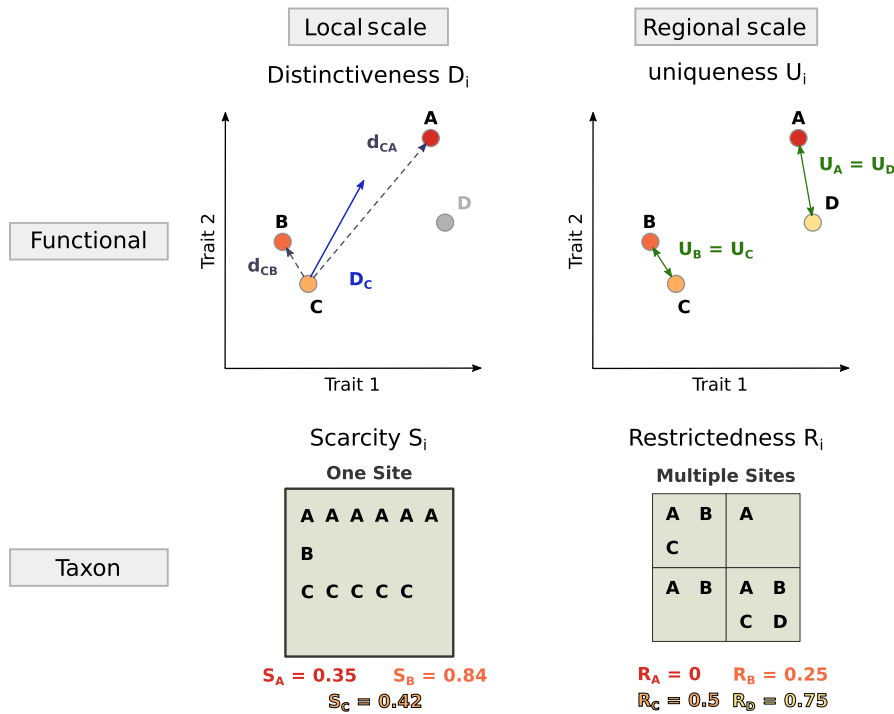


FIGURE 1 Basic patterns of Functional Rarity. Four species A to D are illustrated. Functional indices are represented in top figures: Functional Distinctiveness (left) is the average functional distance of a species to the other species in the community, species D is absent from this community, thus, the Distinctiveness of species C D_C is simply the average of distance of species C to species A, d_{CA} , and to species B, d_{CB} ; Functional Uniqueness (top right) is the functional distance of a species to its nearest neighbour in a regional species pool (see Equation 3); here, A and D are nearest neighbours as well as B and C. Taxon Scarcity (bottom left), where S_i denotes the Scarcity of species i , it is inversely proportional to the abundance of species i (see Equation 4), because species D is absent from the community its scarcity cannot be computed; Taxon Restrictedness (bottom right) is assessed from the occurrences of species across four sites (four tiles) and R_i denotes the Restrictedness of species i , it equals one minus the number of times a species across all sites over the total number of sites (see Equation 5), species A is present in all four sites, thus its Restrictedness R_A equals zero. [Colour figure can be viewed at wileyonlinelibrary.com]

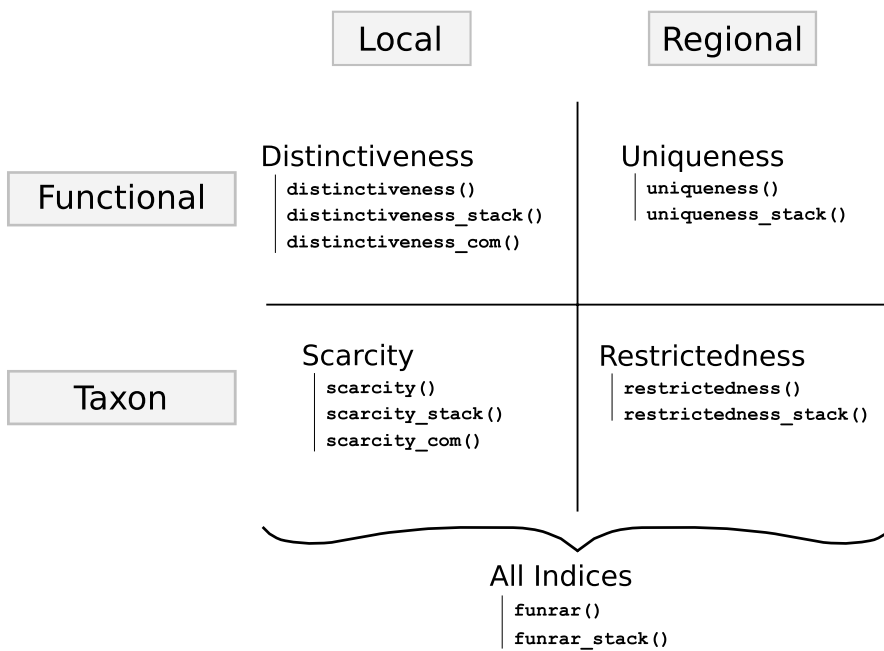


FIGURE 2 Functions available in *funrar* to compute the different facets of functional rarity. Functions handle two formats of site composition, the default one assumes that the input dataset is a site-species matrix, while the `_stack()` versions use “tidy” format; `_com()` functions provided for Functional Distinctiveness and Scarcity take a single community as input. Note that regional-level indices—Restrictedness and Functional Uniqueness—are computed using the complete dataset, giving a single index per species. The site-level indices—Functional Distinctiveness and Scarcity—are computed for each site-species combination, giving one value per site-species combination

Core Team, 2016) package named *funrar*, to quantify functional rarity based on abundance or occurrence data and trait data. The *funrar* package, available through the Comprehensive R Archive Network

(CRAN), computes Functional Distinctiveness, Functional Uniqueness, Taxon Scarcity and Taxon Restrictedness and is optimized to handle high-dimensional data (large number of sites and/or large number of

species) using sparse matrix algebra. We illustrate the application of this package for examining functional rarity using data on North and Central American mammals (Lawing, Eronen, Blois, Graham, & Polly, 2016a) (the code to run the analyses is available on Github at https://github.com/Rekyt/mamm_funrar archived on Zenodo <https://doi.org/10.5281/zenodo.375605>).

2 | QUANTIFYING FUNCTIONAL RARITY USING FUNRAR FUNCTIONS

Users must provide a site-by-species matrix of community composition across sites, with either the presence-absence or relative abundances of species. *funrar* functions can handle site-by-species data in any of three formats: site-species matrix (with sites as rows and species as columns, see Figure 1 for available functions); “tidy” format (Wickham, 2014), with each row coding the observation of a single species at a given site (the function has a `_stack` suffix); or as a single community (the function has a `_com` suffix). Abundance or occurrence information can be based on population or community census and possibly account for imperfect detection (Dénes, Silveira, & Beissinger, 2015; Iknayan, Tingley, Furnas, & Beissinger, 2014; Jarzyna & Jetz, 2016).

Functional distances or dissimilarities are used in the calculation of functional rarity indices (Violle et al., 2017). In *funrar*, a functional distance matrix can be calculated from a table of one or several traits measured for each taxa with `compute_dist_matrix()`. Different kinds of traits (continuous, categorical, ordinal) can be scaled or weighted in various ways when combined (Pavoine, Vallet, Dufour, Gachet, & Daniel, 2009). By default `compute_dist_matrix()` computes the unweighted Gower's (1971) distance because it covers classes of trait data and makes them comparable, but the user can also specify `euclidean` or `manhattan` distances. It is possible to scale traits using the `center` and `scale` arguments when traits are continuous.

Functional Distinctiveness and Functional Uniqueness are computed from the functional distance and species composition matrices. Functional Distinctiveness (D_i , `distinctiveness()`) of a species, that is the uncommonness of a species' traits compared to other species' traits in an assemblage (Figure 1 bottom left), weighted or not by species' relative abundances (Violle et al., 2017) is:

$$D_i = \frac{\sum_{j=1, j \neq i}^N d_{ij} A_j}{\sum_{j=1, j \neq i}^N A_j}, \quad (1)$$

with d_{ij} the functional dissimilarity between species i and species j , N the total number of species in the given assemblage, A_j the relative abundance of species j in the given assemblage. D_i is scaled between zero, if the focal species is identical to all the other species, and one when the focal species is most dissimilar to the other species. If only the presences-absences are provided, $A_j = 1/N$ for all j and D_i simplifies as:

$$D_i = \frac{\sum_{j=1, j \neq i}^N d_{ij}}{N-1}, \quad (2)$$

D_i is the mean dissimilarity of a focal species as defined in Ricotta et al. (2016), that is the mean pairwise functional dissimilarity from a focal species to all the others. Functional Distinctiveness then relates to functional redundancy in an assemblage: the larger the index value, the more distant (less redundant) a species (or an individual) is to the average functional position of the assemblage in the functional space, that is the centroid.

Functional Uniqueness (U_i , `uniqueness()`) is the functional distance of a focal species i to its nearest neighbour in a set of assemblages (Figure 1 bottom right):

$$U_i = \min(d_{ij}), \quad (3)$$

with d_{ij} the functional dissimilarity between species i and species j , for all pairs of species considered across the site-species matrix with $j \neq i$. It quantifies how isolated a species is in the functional space without considering abundances: the higher the index value, the more distant a species is to its closest neighbour in the functional space.

As emphasized by Violle et al. (2017), a species can be functionally distinct (high D_i) in a given community but not functionally unique in an entire region (small U_i). In this regard, Distinctiveness and Uniqueness are used to uncover scale-dependent biodiversity dynamics: by default, *funrar* provides the former at local site level while the latter is computed at regional scale (whole site-species matrix). At local scale, community dynamics involve all coexisting species and their relative abundance is expected to convey the signature of assembly processes. Between-species dissimilarities and Functional Distinctiveness are thus relevant to assess the role of functional originality in community assembly. At regional scale, Functional Uniqueness can represent how taxa depart from a regional pool due to specific biogeographical and evolutionary legacies and should then be estimated based on the whole site-species matrix. Nevertheless, each index can be computed at both scales to grasp the different aspects of functional rarity (examples in the help of `distinctiveness()` and `uniqueness()`).

Because Distinctiveness and Uniqueness are computed using multiple traits, it can be difficult to disentangle if a species exhibits high values because of a single extreme trait value or because it has several rare trait values. The `uniqueness_dimensions()` and `distinctiveness_dimensions()` functions respectively compute Uniqueness and Distinctiveness values from the traits taken one by one as well as altogether. The former outputs a table with the value of Uniqueness for each trait and for all the traits considered together, while the latter outputs a list of site-species matrices of computed Distinctiveness values, one matrix per trait and one for all the traits considered together.

The second set of functions deals with the taxon component of functional rarity. Two indices estimate it: Taxon Scarcity (`scarcity()`) in an assemblage and Taxon Restrictedness (`restrictedness()`) in a set of assemblages. Taxon Scarcity (`scarcity()`) of a species in a given assemblage gets close to one when the species has low abundance in the site and gets close to zero when it dominates the assemblage:

$$S_i = \exp(-NA_i \ln 2), \quad (4)$$

where N is the number of species and A_i the relative abundance of species i at the focal site. When species are equally abundant in the assemblage, with $1/N$ relative abundances, S_i equals 0.5. Scarcity cannot be computed with only the presence–absence data in the site-by-species matrix.

Taxon Restrictedness is an index between zero and one. It increases when a species is present in less sites of the site-species matrix. Restrictedness nearly equals one when a species is present in a single site (examples on Figure 1):

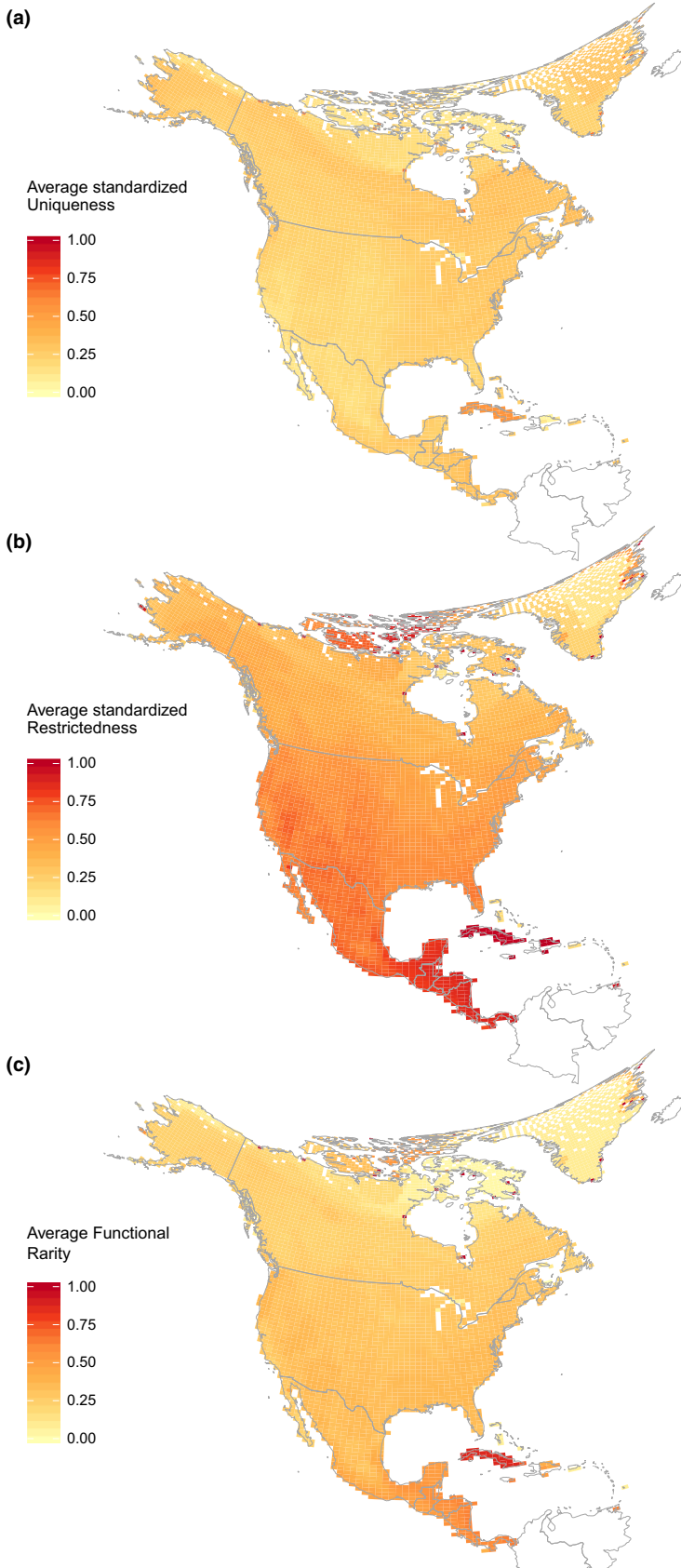


FIGURE 3 Maps of functional rarity indices averaged per site in North and Central America using a subset of the dataset of North and Central American Mammals from Lawing et al. (2016a). All indices have been scaled per site between 0 and 1. (a) Functional Uniqueness; (b) Taxon Restrictedness; (c) Functional Rarity, the average of Functional Uniqueness and Taxon Restrictedness per site. The geographical projection of maps is Albers Equal Area (ESRI:102008). [Colour figure can be viewed at wileyonlinelibrary.com]

$$R_i = 1 - \frac{K_i}{K_{\text{tot}}}, \quad (5)$$

where K_i is the number of sites where species i occurs and K_{tot} the total number of sites in the dataset. R_i equals one when the species is completely absent from the dataset. Restrictedness can also be computed for predicted species distributions from ecological models (Guisan & Thuiller, 2005). A threshold of the predicted probabilities of occurrence (Jiménez-Valverde & Lobo, 2007; Liu, Berry, Dawson, & Pearson, 2005) is then used to derive the matrix of species occurrences per pixel.

Because of the increasing availability of large-scale and intensive datasets in ecology (Hampton et al., 2013), a site-species matrix can contain thousands of sites and thousands of species. However, as not all species are everywhere, site-species matrices are usually filled with many zeroes. Sparse matrices allow storing only the position of non-zero cells, saving memory. *funrar* performs sparse matrix calculations using the *Matrix* package for quicker and memory-efficient computations (Bates & Maechler, 2016). For more details, see the vignette included in the package.

3 | FUNCTIONAL RARITY OF NORTH AND CENTRAL AMERICAN MAMMALS

We used *funrar* to analyse a subset of the dataset of North and Central American Mammals from Lawing et al. (2016a), Lawing,

Eronen, Blois, Graham, and Polly (2016b). We selected 265 species out of 558 for which trait information was available. We used six traits relevant to mammal ecology (body mass, litter size, diet breadth, trophic level, habitat breadth and terrestriality, see Jones et al. (2009) for detailed trait explanation). The dataset comprises the presence-absence information for the 265 species across 9699 50 km x 50 km cells. We asked whether there are “hotspots” of Functional Uniqueness in North America and Central America for the six aforementioned traits; whether species that are functionally unique are geographically restricted; what the most functionally distinct and unique mammal species in the dataset are; and whether there are more functionally rare species in temperate, tropical or boreal areas.

For each species, we calculated Functional Uniqueness and Taxon Restrictedness indices and averaged them across species by grid cell. Because those two indices are regional-level indices, each species had a unique value, and the variation in averaged indices among grid cells thus reflects change in species composition. We produced maps of the average values for indices in North and Central America (Figure 3). We also computed functional rarity—the average of Functional Uniqueness and Taxon Restrictedness—where each is scaled between zero and one. The most functionally unique cells were in Cuba (Figure 3a), meaning that they hosted, on average, species that are quite functionally unique compared to the species pool of North and Central America. This pattern may be due to the tropical climate present in Cuba, which is less present

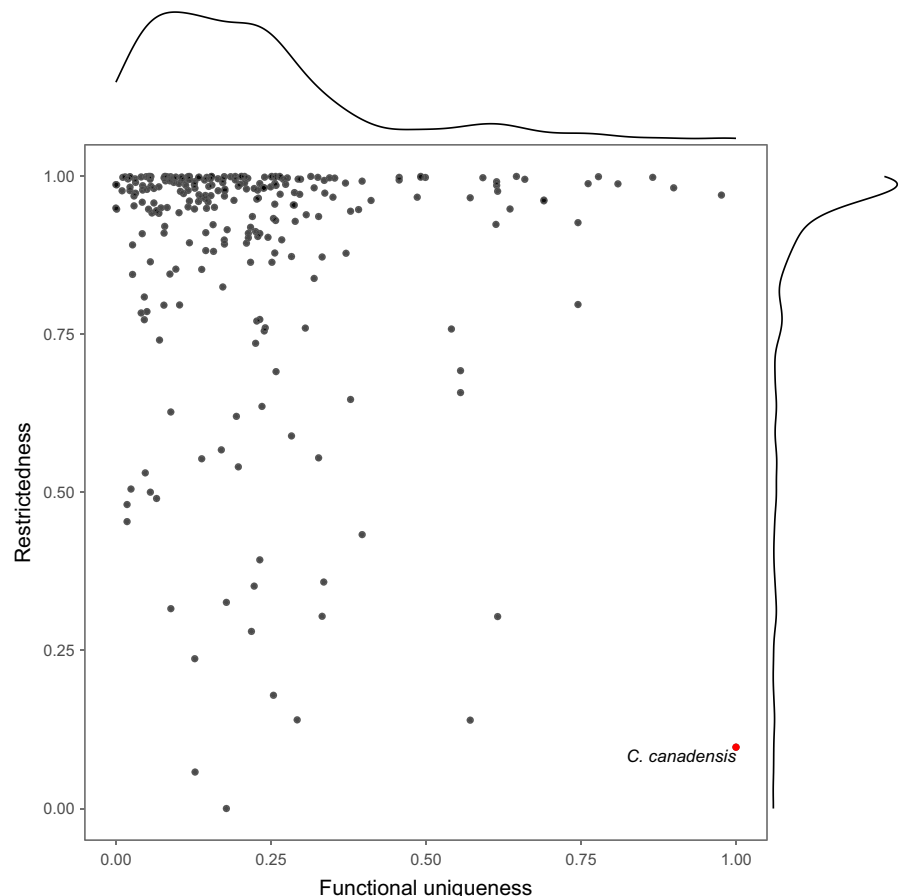


FIGURE 4 Biplot of Functional Uniqueness against Taxon Restrictedness per species across the whole dataset ($N = 265$ species). Note that both Functional Uniqueness and Taxon Restrictedness have been scaled between zero and one for easier comparison (Spearman's $\rho = -.06$, $p = .323$, $S = 3290600$). The red dot indicates the position of *Castor canadensis*. Marginal distributions are indicated on the sides of the graph. [Colour figure can be viewed at wileyonlinelibrary.com]

across the rest of the dataset. We also identified a latitudinal gradient in Taxon Restrictedness (Figure 3b): sites at low latitude hosted more restricted species on average than sites in temperate and boreal regions, a pattern that complies with Rapoport's rule (Gaston, Blackburn, & Spicer, 1998; Rapoport, 1982). Altogether, combining the two facets into a single index highlighted Cuba as a hotspot of functional rarity (Figure 3c).

At the species level, Functional Uniqueness and Taxon Restrictedness were not correlated (Spearman's $\rho = -.06$, $p = .323$) (Figure 4). Most species were geographically restricted (with many values around 1, meaning that they were present in a few grid cells only) but functionally redundant, that is with low values of Functional Uniqueness. Conversely, almost no species were both geographically widespread and functionally unique, apart from *Castor canadensis* (red dot in the bottom right corner of Figure 4), which occupies a very specific habitat (fossorial and ground dwelling) and is distributed continent-wide. Functional Distinctiveness and Functional Uniqueness were correlated (Spearman's $\rho = .37$, $p < .001$), indicating that species that were locally functionally distinct tended to be regionally functionally unique. Even though local functional rarity and regional functional rarity were quite correlated, the weak correlation implies that rarity should be estimated both at the local and regional scale because it contrasts different types of rarity.

In summary, North and Central American mammals display a biogeographical gradient of functional rarity, such as species at low latitude show higher Functional Uniqueness and Taxon Restrictedness (Figure 3), even though most mammal species have low Functional Uniqueness and high Taxon Restrictedness (top left corner in Figure 4). The absence of correlation of Functional Uniqueness and Taxon Restrictedness shows that these components of functional rarity provide complementary information. Therefore, both components should be considered when mapping rarity and defining priority targets in conservation programs.

4 | CONCLUSION

Conservation biology has historically focused on the protection of rare species in terms of taxon occurrences and abundances (Prendergast, Quinn, Lawton, Eversham, & Gibbons, 1993). Although it can be decomposed in local, regional and habitat dimensions (Rabinowitz, 1981), this approach has emphasized taxonomic rarity and neglected the originality of functional attributes. Recently though, assessing species originality in terms of phylogenetic rarity (Cadotte & Jonathan Davies, 2010; Isaac, Turvey, Collen, Waterman, & Baillie, 2007; Rosauer, Laffan, Crisp, Donnellan, & Cook, 2009) and functional rarity (Mouillot et al., 2013; Umaña, Zhang, Cao, Lin, & Swenson, 2015; Violle et al., 2017) has gained momentum. It underlines the need to characterize patterns of rarity through the ecological and evolutionary attributes that influence biodiversity dynamics at multiple scales. Uncorrelated Functional Uniqueness and Restrictedness in North and Central American Mammals suggest that the functional component of rarity should be considered for a more

comprehensive assessment of biodiversity dynamics and a better design of conservation strategies. Such integrated view on rarity echoes Winter, Devictor, and Schweiger (2013) suggestion to "[include] other facets of diversity" for conservation. The *funrar* package contributes to the growing toolbox available for researchers to study and quantify the various dimensions of biodiversity and rarity. Adding the functional rarity string would strengthen the bow of diversity and rarity facets.

ACKNOWLEDGEMENTS

This research is supported by the French Foundation for Research on Biodiversity (FRB; www.fondationbiodiversite.fr) in the context of the CESAB project "Causes and consequences of functional rarity from local to global scales" (FREE). PD and CV were supported by the European Research Council (ERC) Starting Grant Project "Ecophysiological and biophysical constraints on domestication of crop plants" (Grant ERC-StG-2014-639706-CONSTRAINTS). CMT acknowledges the European Commission for the Marie Curie IIF (H2020-MSCA-IF-2014-657951).

DATA ACCESSIBILITY

funrar stable version is available on CRAN (via `install.package("funrar")`) or <https://cran.r-project.org/package=funrar> and the development version is on Github at <https://github.com/Rekyt/funrar>. The code necessary to reproduce the analyses is on Github at https://github.com/Rekyt/mamm_funrar, an archived version is accessible on Zenodo <https://doi.org/10.5281/zenodo.375605>

ORCID

Matthias Grenié  <http://orcid.org/0000-0002-4659-7522>

REFERENCES

- Bates, D., & Maechler, M. (2016). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-11, URL <http://cran.r-project.org/package=Matrix>
- Cadotte, M. W., & Jonathan Davies, T. (2010). Rarest of the rare: Advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales: Conservation phylo-biogeography. *Diversity and Distributions*, 16, 376–385.
- Cardoso, P., Rigal, F., Borges, P. A. V., & Carvalho, J. C. (2014). A new frontier in biodiversity inventory: A proposal for estimators of phylogenetic and functional diversity. *Methods in Ecology and Evolution*, 5, 452–461.
- Dénes, F. V., Silveira, L. F., & Beissinger, S. R. (2015). Estimating abundance of unmarked animal populations: Accounting for imperfect detection and other sources of zero inflation. *Methods in Ecology and Evolution*, 6, 543–556.
- Enquist, B. J., Norberg, J., Bonser, S. P., Violle, C., Webb, C. T., Henderson, A., Sloat, L. L., & Savage, V. M. (2015). Scaling from Traits to Ecosystems. In: Guy Woodward (Ed.), *Advances in Ecological Research* (pp. 249–318). Oxford, UK: Elsevier. <https://doi.org/10.1016/bs.aecr.2015.02.001>
- Gaston, K. J., Blackburn, T. M., & Spicer, J. I. (1998). Rapoport's rule: Time for an epitaph? *Trends in Ecology & Evolution*, 13, 70–74.

- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27, 857–871.
- Grime, J. P. (1998). Benefits of plant diversity to ecosystems: Immediate, filter and founder effects. *Journal of Ecology*, 86, 902–910.
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8, 993–1009.
- Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., ... Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, 11, 156–162.
- Ikhnayan, K. J., Tingley, M. W., Furnas, B. J., & Beissinger, S. R. (2014). Detecting diversity: Emerging methods to estimate species diversity. *Trends in Ecology & Evolution*, 29, 97–106.
- Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. M. (2007). Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny. *PLoS ONE*, 2, e296.
- Jarzyna, M. A., & Jetz, W. (2016). Detecting the multiple facets of biodiversity. *Trends in Ecology & Evolution*, 31, 527–538.
- Jiménez-Valverde, A., & Lobo, J. M. (2007). Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecologica*, 31, 361–369.
- Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., Teacher, A., Bininda-Emonds, O. R. P., Gittleman, J. L., Mace, G. M., & Purvis, A. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90, 2648.
- Lawing, A. M., Eronen, J. T., Blois, J. L., Graham, C. H., & Polly, P. D. (2017). Community functional trait composition at the continental scale: the effects of non-ecological processes. *Ecography*, 40, 651–663.
- Lawing, A. M., Eronen, J. T., Blois, J. L., Graham, C. H., & Polly, P. D. (2016). Data from: Community functional trait composition at the continental scale: the effects of non-ecological processes. *Ecography*, 40, 651–663.
- Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28, 385–393.
- Mazel, F., Guilhaumon, F., Mouquet, N., Devictor, V., Gravel, D., Renaud, J., ... Thuiller, W. (2014). Multifaceted diversity–area relationships reveal global hotspots of mammalian species, trait and lineage diversity. *Global Ecology and Biogeography*, 23, 836–847.
- McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, 21, 178–185.
- Mouillot, D., Bellwood, D. R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., ... Thuiller, W. (2013). Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. *PLoS Biology*, 11, e1001569.
- Pavoine, S., Ollier, S., & Dufour, A.-B. (2005). Is the originality of a species measurable? *Ecology Letters*, 8, 579–586.
- Pavoine, S., Vallet, J., Dufour, A.-B., Gachet, S., & Daniel, H. (2009). On the challenge of treating various types of variables: Application for improving the measurement of functional diversity. *Oikos*, 118, 391–402.
- Prendergast, J. R., Quinn, R. M., Lawton, J. H., Eversham, B. C., & Gibbons, D. W. (1993). Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature*, 365, 335–337.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rabinowitz, D. (1981). Seven forms of rarity. In: H. Synge (Ed.), *The biological aspects of rare plants conservation* (pp. 205–217). New York: Wiley.
- Rapoport, E. H. (1982). 1982: *Areography: geographical strategies of species*. Oxford: Pergamon Press.
- Ricotta, C., de Bello, F., Moretti, M., Caccianiga, M., Cerabolini, B. E. L., & Pavoine, S. (2016). Measuring the functional redundancy of biological communities: A quantitative guide. *Methods in Ecology and Evolution*, 7, 1386–1395.
- Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, S. C., & Cook, L. G. (2009). Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology*, 18, 4061–4072.
- Safi, K., Cianciaruso, M. V., Loyola, R. D., Brito, D., Armour-Marshall, K., & Diniz-Filho, J. A. F. (2011). Understanding global patterns of mammalian functional and phylogenetic diversity. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366, 2536–2544.
- Umaña, M. N., Zhang, C., Cao, M., Lin, L., & Swenson, N. G. (2015). Commonness, rarity, and intraspecific variation in traits and performance in tropical tree seedlings. *Ecology letters*, 18, 1329–1337.
- Violle, C., Reich, P. B., Pacala, S. W., Enquist, B. J., & Kattge, J. (2014). The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences*, 111, 13690–13696.
- Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N. J. B., Cadotte, M. W., Livingstone, S. W., & Mouillot, D. (2017). Functional Rarity: The Ecology of Outliers. *Trends in Ecology & Evolution*, 32, 356–367.
- Weiher, E., Freund, D., Bunton, T., Stefanski, A., Lee, T., & Bentivenga, S. (2011). Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366, 2403–2413.
- Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59, 1–23.
- Winter, M., Devictor, V., & Schweiger, O. (2013). Phylogenetic diversity and nature conservation: Where are we? *Trends in Ecology & Evolution*, 28, 199–204.

BIOSKETCH

Matthias Grenié is a PhD student at the University of Montpellier and at the Centre d'Écologie Fonctionnelle et Évolutive in Montpellier. He studies the common determinants of diversity across different taxonomic groups and has broad interests in biogeography, community ecology and functional ecology.

Author contributions: M.G. and P.D. designed the package; C.T. reviewed it; M.G. wrote the first draft of the manuscript; and all the authors edited and commented it.

How to cite this article: Grenié M, Denelle P, Tucker CM, Munoz F, Violle C. *funrar*: An R package to characterize functional rarity. *Divers Distrib*. 2017;23:1365–1371. <https://doi.org/10.1111/ddi.12629>

2 L'ORIGINALITÉ FONCTIONNELLE PEUT SE CALCULER À DE MULTIPLES ÉCHELLES SPATIALES ET BIOLOGIQUES

Trends in Ecology & Evolution

CellPress
REVIEWS

authors defined functional uniqueness as an indicator of species functional redundancy. However, their proposed index (functional distance to the nearest neighbor) depends only on a single species. In this sense, indices of redundancy considering more species seem more adequate alternatives [2,3]. Moreover, estimations of distances between species should be considered carefully, including the possibility of combining functional and phylogenetic information [4]. For example, estimating functional dissimilarities via Gower distances or standardized Euclidean distances can make estimations of rarity not comparable across species pools [5]. We suggest that most of these limitations can be overcome by applying the trait probability density (TPD) approach to estimate functional diversity ([2]; Figure 1). The main three advantages of the TPD approach are that species abundance is explicitly considered in these functions; that they can be expressed at any spatial scale or organizational level; and that results can be directly compared across species pools [2,6]. This allows for seamless transitions and comparisons across scales (species within habitats or regions as in Violle *et al.* [1], but also habitats within landscapes or regions, regions within countries, biogeographical domains within the world, or any combination of these) using a single, probabilistic, and scale-independent definition (Figure 1).

Finally, future developments will need to establish clear connections between extinction risk and functional rarity. It is not straightforward to assume that rarity always implies higher extinction risk. While extinction is generally expected not to be a random process, common, instead of rare species, can be lost if their traits make them more susceptible to environmental changes [7]. Previous studies applying species loss simulations on functional diversity/ecosystem functions

have assumed extinction orders according to metacommunity nestedness patterns and species response traits (e.g., body size [8] or species palatability to livestock [9]), and not necessarily on their rarity in a community or region [10]. Ecological consequences of species loss are better evaluated in a continuous fashion, using estimations of vulnerability. Such estimations combine functional trait information of species (preferably based on multiple traits, as a proxy for overall functioning) with information on species extinction risk and their expected response to environmental changes [11,12].

¹Institute of Ecology and Earth Sciences, University of Tartu, Lai 40, 51005, Tartu, Estonia

²Department of Botany, Faculty of Science, University of South Bohemia, Branišovská 31, České Budějovice, Czech Republic

³Institute of Botany, Czech Academy of Sciences, Treboň, Czech Republic

⁴Graduate School of Environment and Information Sciences, Yokohama National University, 79-7 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan

*Correspondence:
perezcarmonacarlos@gmail.com (C.P. Carmona).
<https://doi.org/10.1016/j.tree.2017.09.010>

References

- Violle, C. *et al.* (2017) Functional rarity: the ecology of outliers. *Trends Ecol. Evol.* 32, 356–367
- Carmona, C.P. *et al.* (2016) Traits without borders: integrating functional diversity across scales. *Trends Ecol. Evol.* 31, 382–394
- Bennett, J.A. and Pärtel, M. (2017) Predicting species establishment using absent species and functional neighborhoods. *Ecol. Evol.* 7, 2223–2237
- de Bello, F. *et al.* (2017) Decoupling phylogenetic and functional diversity to reveal hidden signals in community assembly. *Methods Ecol. Evol.* <http://dx.doi.org/10.1111/2041-210X.12735> Published online March 20, 2017
- de Bello, F. *et al.* (2013) Which trait dissimilarity for functional diversity: trait means or trait overlap? *J. Veg. Sci.* 24, 807–819
- Carmona, C.P. *et al.* (2016) The density awakens: a reply to Blonder. *Trends Ecol. Evol.* 31, 667–669
- Laliberté, E. *et al.* (2010) Land-use intensification reduces functional redundancy and response diversity in plant communities. *Ecol. Lett.* 13, 76–86
- Larsen, T.H. *et al.* (2005) Extinction order and altered community structure rapidly disrupt ecosystem functioning. *Ecol. Lett.* 8, 538–547
- Verón, S.R. *et al.* (2011) Grazing-induced losses of biodiversity affect the transpiration of an arid ecosystem. *Oecologia* 165, 501–510
- Sasaki, T. *et al.* (2017) Differential responses and mechanisms of productivity following experimental species loss scenarios. *Oecologia* 183, 785–795

11. Sasaki, T. *et al.* (2014) Vulnerability of moorland plant communities to environmental change: consequences of realistic species loss on functional diversity. *J. Appl. Ecol.* 51, 299–308

12. Carmona, C.P. *et al.* (2017) Assessing vulnerability of functional diversity to species loss: a case study in Mediterranean agricultural systems. *Funct. Ecol.* 31, 427–435

Letter

A Common Toolbox to Understand, Monitor or Manage Rarity? A Response to Carmona *et al.*

Cyrille Violle,^{1,*} Wilfried Thuiller,² Nicolas Mouquet,³ François Munoz,^{2,4} Nathan J.B. Kraft,⁵ Marc W. Cadotte,^{6,7} Stuart W. Livingstone,⁸ Matthias Grenie,¹ and David Mouillot^{3,9}

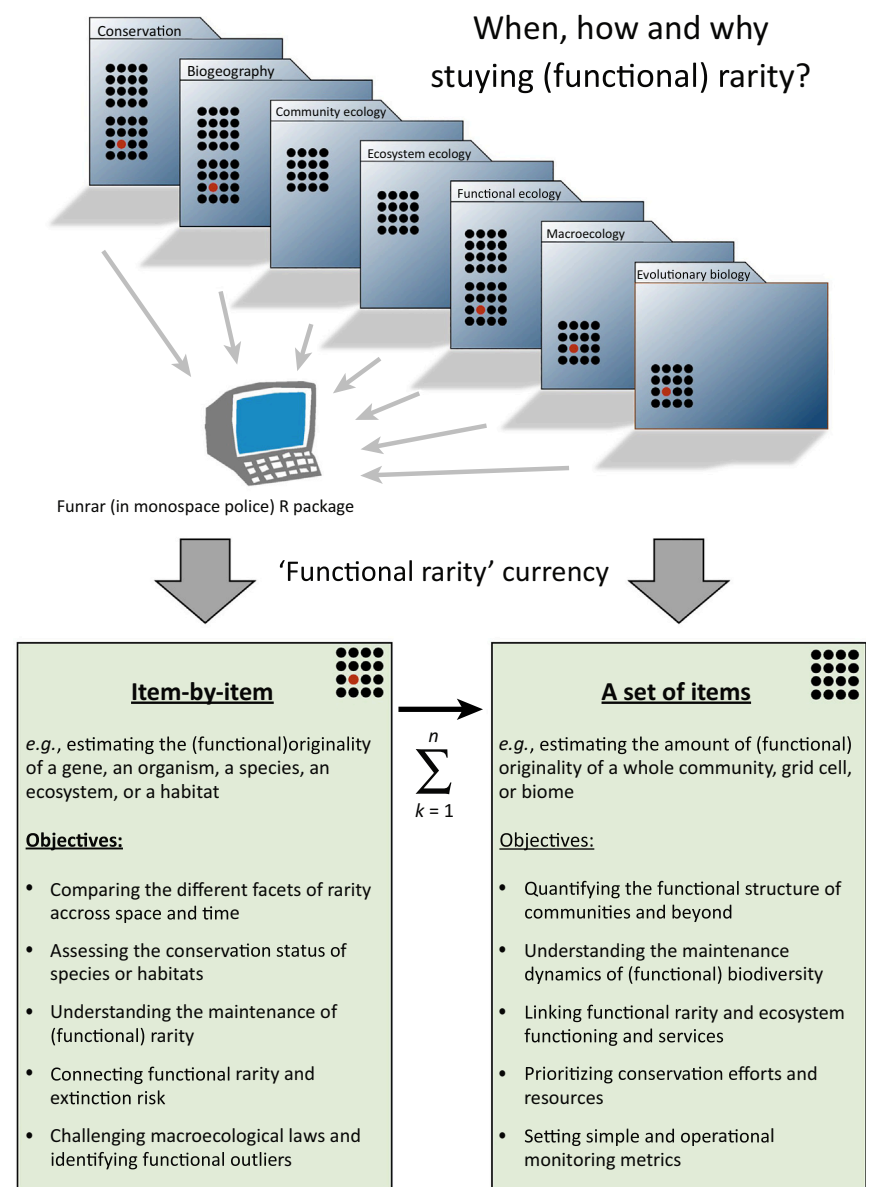
Carmona *et al.* [1] highlight a probabilistic approach to functional rarity as an extension of our integrated framework to functional rarity [2]. The authors argue that it could be considered as a common toolbox for rarity. While we certainly agree with the authors about the necessity to unify the quantification of biodiversity and rarity in a community ecology and biogeography perspective [3], we call for a more operational and pragmatic quantification in a conservation perspective.

Carmona *et al.* [1] question the relevance of categorizing functional rarity instead of providing a continuous quantification. In fact, the integrated view of functional rarity we proposed [2] is not categorized by principle or design. We proposed a set of indices that are continuous and in line with the probabilistic approach promoted by Carmona *et al.* [1] (see also [4]). We defined local and regional scales for the sake of simplicity but the delimitation

between these two scales is not fixed; calculating the indices for various scales can be easily achieved. This set of functional rarity indices can now be calculated using the R package, *funrar*, available on CRAN (<https://cran.r-project.org/web/packages/funrar/index.html>) [5]. We encourage everyone to complement or improve functions available in *funrar*. Implementing the framework of Carmona *et al.* [4] in *funrar* appears to be a natural perspective.

As a byproduct of our quantitative framework for assessing functional rarity, we originally proposed to categorize it through 12 forms of functional rarity, echoing Rabinowitz's seven forms of (taxon) rarity [6]. We argue that this two-step assessment (quantification and then categorization) of functional rarity should not be minimized or ignored. In a conservation perspective, it is essential to keep in mind that any new tool for quantifying different facets of biodiversity will be in the hands of resource managers and decision-makers. Providing continuous and sophisticated metrics for rarity may be mathematically appealing, but will likely be counterproductive in some cases if interpretation of values is unclear. There are strong arguments in favour of treating rarity as a discontinuous or categorical variable rather than as a continuous variable in the conservation literature: 'because for legal and conservation purposes species often need to be categorized as rare or otherwise, a more pragmatic approach is often desirable' [7]. Our 12 forms of functional rarity follow this recommendation. More broadly, there is a growing effort to identify simple and operational metrics to facilitate the monitoring and management of biodiversity and rarity. We consider that the basic categories of functional rarity that we proposed could be easily added to the list of Essential Biodiversity Variables (EBV) [8]. For instance, our approach allows the identification of species that possess

the highest degree of functional rarity (e.g., top 5 or 10%), potentially useful information for monitoring and managing biodiversity and rarity worldwide. Bridging theoretical and applied ecology is not a new challenge for the field, but we



Trends in Ecology & Evolution

Figure 1. Cookbook for the Application of Violle *et al.*'s Framework to Functional Rarity: When, How and Why Study Functional Rarity? Several fields of (theoretical and applied) biology needs a unified framework to quantify rarity. The *funrar* R package [5] implements the framework proposed in Violle *et al.* [2]. Depending on the objectives of the research question in each field, one can be interested to: (i) assess the originality of a single item (e.g., species or habitat) (the red point is original – i.e., rare – compared to the black ones); or (ii) quantify the whole functional rarity of a set of items (e.g., a community or a grid cell) (the sum, or any other integration functions of functional rarity values of each item). In the latter case, functional rarity and functional redundancy are the two sides of the same coin and Carmona *et al.*'s framework [4] can be easily implemented.

question whether the complexification of science, the emergence of big data and sophisticated approaches to analyse them, although necessary, may separate instead of bridging both sides of ecology.

The concept of (functional) rarity is multifaceted by nature [2], and its application involves at least two forms: an item-by-item (Ibl) analysis versus a set-of-items (Sol) analysis (Figure 1). In an Ibl perspective, the interest is to characterise the functional rarity of a species or any item of lower or higher organisation level (e.g., community, habitat, or biome) compared to other items of the same type (e.g., analysing the functional rarity of a given plant species relatively to all other land plant species). In this case, the main questions are: what causes functional rarity; what are the consequences of functional rarity; and what is the link between species' extinction risk and functional rarity? These are crucial questions for conservation ecology, and also for more theoretical fields like functional ecology, macroecology, and evolutionary biology. From a functional ecology or macroecology perspective, an appealing research frontier would be to identify outliers from 'universal laws' of functioning and phenotypic diversification (e.g., the leaf economics spectrum in plants, or allometric relationships in both plants and animals), their causes of persistence in nature, and the reasons why theoretical laws can be violated. As a consequence, beyond the need in conservation, Ibl analyses appear also particularly relevant in many fields of biology (Figure 1).

Sol analyses are specifically relevant in a community ecology and biogeography perspective [3,9]. Sol refers to the amount of functional rarity that does exist in a given assemblage, for example, in a community or a biome. Sol functional rarity indices can be compiled by, for example, averaging or summing species-based Ibl

indices [5]. It is important to note that Sol indices and functional redundancy analyses are the two sides of the same coin [10] and thus can address complementary topics. Is the functional space of a community saturated? What are the causes of the maintenance of rare phenotypes in a community? Are ecosystem functioning and stability driven by the functions supported by some rare phenotypes, or by functional redundancy? As a unified analytical framework, the trait probability density (TPD) [1,4] approach can be relevant for Sol analyses. Nevertheless, there are some practical limitations in the application of this framework given that it requires ideal and precise descriptions of continuous trait distributions (within species, communities, etc), which are rarely available.

There have been many attempts to mathematically unify and integrate the different facets of biodiversity and rarity [4,11]. This is valuable given that single biodiversity (rarity) metrics can be provided. We proposed one of them through a multiplicative framework [2,5], echoing abundance-weighted evolutionary distinctiveness scores [12]. TPD can be used for assemblage-level analyses. In any case, we call for simplicity and pragmatism when navigating the jungle of rarity indices, so as to remain useful for the conservation and monitoring of biodiversity whose objectives are tightly linked to rarity issues within the global context of extinction of both species and functions.

Acknowledgements

This work is supported by the French Foundation for Research on Biodiversity (FRB; www.fondationbiodiversite.fr) in the context of the CESAB project 'Causes and Consequences of Functional Rarity from Local to Global Scales' (FREE), and by the European Research Council (ERC) Starting Grant Project 'Ecophysiological and Biophysical Constraints on Domestication of Crop Plants' (Grant ERC-StG-2014-639706-CONSTRAINTS).

¹CEFE UMR 5175, CNRS – Université de Montpellier – Université Paul-Valéry Montpellier – EPHE, 1919 route de Mende, F-34293 Montpellier, CEDEX 5, France

²LECA, University Grenoble Alpes, CNRS, F-38000 Grenoble, France

³MARBEC, UMR IRD-CNRS-UM-IFREMER 9190, Université de Montpellier, 34095 Montpellier Cedex, France

⁴French Institute of Pondicherry, 11 St. Louis Street, Pondicherry 605001, India

⁵Department of Ecology and Evolutionary Biology, University of California, Los Angeles, 621 Charles E. Young Drive South, Los Angeles, CA, 90095, USA

⁶Department of Biological Sciences, University of Toronto–Scarborough, 1265 Military Trail, Toronto, ON, M1C 1A4, Canada

⁷Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, ON, M5S 3B2, Canada

⁸Department of Physical and Environmental Science, University of Toronto–Scarborough, 1265 Military Trail, Toronto, ON, M1C 1A4, Canada

⁹Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia

*Correspondence: cyrille.violle@cefe.cnrs.fr (C. Violle).

<https://doi.org/10.1016/j.tree.2017.10.001>

References

1. Carmona, C. et al. (2017) Towards a common toolbox for rarity: a response to Violle et al. *Trends Ecol. Evol.* 32, 889–891
2. Violle, C. et al. (2017) Functional rarity: the ecology of outliers. *Trends Ecol. Evol.* 32, 356–367
3. Violle, C. et al. (2014) The emergence and promise of functional biogeography. *Proc. Natl. Acad. Sci. U. S. A.* 111, 13690–13696
4. Carmona, C.P. et al. (2016) Traits without borders: integrating functional diversity across scales. *Trends Ecol. Evol.* 31, 382–394
5. Grenie, M. et al. (2017) funrar: an R package to characterize functional rarity. *Diversity Distrib.* <http://dx.doi.org/10.1111/ddi.12629> Published online September 11, 2017
6. Rabinowitz, D. (1981) Seven forms of rarity. In *The Biological Aspects of Rare Plant Conservation* (Synge, H., ed.), pp. 205–217. Wiley
7. Gaston, K. (1994) *Rarity*, Chapman & Hall
8. Pereira, H.M. et al. (2013) Essential biodiversity variables. *Science* 339, 277–278
9. Enquist, B.J. et al. (2015) Scaling from traits to ecosystems: developing a general Trait Driver Theory via integrating trait-based and metabolic scaling theories. *Adv. Ecol. Res.* 52, 249–318
10. Ricotta, C. et al. (2016) Measuring the functional redundancy of biological communities: a quantitative guide. *Methods Ecol. Evol.* 7, 1386–1395
11. Cadotte, M.W. et al. (2013) The ecology of differences: integrating evolutionary and functional distances. *Ecol. Lett.* 16, 1234–1244
12. Cadotte, M. and Davies, T. (2010) Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. *Divers. Distrib.* 16, 376–385

3 DÉTERMINANTS ENVIRONNEMENTAUX DE L'ORIGINALITÉ FONCTIONNELLE GLOBALE ET RELATIONS AVEC LES AUTRES RARETÉS

Distribution and environmental determinants of plant ecological
originality the New World

Matthias Grenié François Munoz FREE Consortium Cyrille Violle

Introduction

5 Biodiversity hotspots, defined as places of unusually high biodiversity, have attracted much attention in biogeography and conservation (Marchese, 2015; Myers et al., 2000). The of idea of “bright spots” were recently suggested to identify places that show an excess of outliers compared to the expected values (Cinner et al., 2016). While hotspots identify absolute maximum, bright spots identify sites with an excess of outlier species. The underlying postulate is that the loss of original species under biodiversity erosion scenarios will
10 lead to the loss of irreplaceable functions in the ecosystems (Chapman et al., 2018; Jain et al., 2014; Mouillot et al., 2013, 2014). Surprisingly though, the ecological nature of outliers has long been overlooked in ecology (Violle et al., 2017) despite a theoretical corpus stemming from functional ecology (Espeland and Emam, 2011; Lavergne et al., 2003; Murray et al., 2002; Pavoine et al., 2005). With the recent unprecedented rise of global trait databases in plants (Enquist et al., 2016; Kattge et al., 2020), it is now possible to characterize
15 more precisely whether and how a species is ecologically distinct from others in a functional trait space (Diaz et al., 2016; Violle et al., 2017). From this premise, putting original species on the map should reveal bright spots of particular interest for both biodiversity conservation and ecosystem functioning (Cooke et al., 2020).

The way original plant species are maintained in ecosystems remains unexplored. This is a critical question because ongoing climate changes should deeply affect biodiversity and, while original species can provide
20 lifebelts and improve resilience ecosystem functioning on the long term (Cardinale et al., 2012; Naeem et al., 2012). While most ecological and evolutionary theories predict a purge of original phenotypes at equilibrium (Kingsolver and Pfennig, 2007), many biogeographical, historical and ecological factors can act against this purge and participate to the maintenance of these outliers. Climate refugia are well known players in the maintenance of non-equilibrium populations and communities acting as museums for relicts (Ordóñez and
25 Svenning, 2016), but their role in the maintenance of phenotypic outliers is unknown. More broadly, past

Distribution and environmental determinants of plant ecological originality the New World

Matthias Grenié François Munoz FREE Consortium Cyrille Violle

Introduction

5 Biodiversity hotspots, defined as places of unusually high biodiversity, have attracted much attention in biogeography and conservation (Marchese, 2015; Myers et al., 2000). The idea of “bright spots” were recently suggested to identify places that show an excess of outliers compared to the expected values (Cinner et al., 2016). While hotspots identify absolute maximum, bright spots identify sites with an excess of outlier species. The underlying postulate is that the loss of original species under biodiversity erosion scenarios will lead to the loss of irreplaceable functions in the ecosystems (Chapman et al., 2018; Jain et al., 2014; Mouillot et al., 2013, 2014). Surprisingly though, the ecological nature of outliers has long been overlooked in ecology (Violle et al., 2017) despite a theoretical corpus stemming from functional ecology (Espeland and Emam, 2011; Lavergne et al., 2003; Murray et al., 2002; Pavoine et al., 2005). With the recent unprecedented rise of global trait databases in plants (Enquist et al., 2016; Kattge et al., 2020), it is now possible to characterize more precisely whether and how a species is ecologically distinct from others in a functional trait space (Diaz et al., 2016; Violle et al., 2017). From this premise, putting original species on the map should reveal bright spots of particular interest for both biodiversity conservation and ecosystem functioning (Cooke et al., 2020).

The way original plant species are maintained in ecosystems remains unexplored. This is a critical question because ongoing climate changes should deeply affect biodiversity and, while original species can provide lifebelts and improve resilience ecosystem functioning on the long term (Cardinale et al., 2012; Naeem et al., 2012). While most ecological and evolutionary theories predict a purge of original phenotypes at equilibrium (Kingsolver and Pfennig, 2007), many biogeographical, historical and ecological factors can act against this purge and participate to the maintenance of these outliers. Climate refugia are well known players in the maintenance of non-equilibrium populations and communities acting as museums for relicts (Ordonez and Svenning, 2016), but their role in the maintenance of phenotypic outliers is unknown. More broadly, past

climates have shaped extent biodiversity, and they could have also acted on ecological originality. Stable climates promote trait differentiation and would lead to a greater number of original species (Harrison and Noss, 2017; Ordonez and Svenning, 2016). Climate instability could however increase speciation and spark rapid adaptive radiation (Harrison and Noss, 2017). This apparent contradiction underlines the lack of
30 knowledge regarding the role of past climate on the distribution of ecological outliers.

Similarly, extreme environments exert strong biological constraints suggesting that they should host specifically adapted species. Indeed, extreme environments in terms of stress or resource availability can select only a few adapted species located in a limited region of the functional trait space (Bjorkman et al., 2018; Lhotsky et al., 2016; Šímová et al., 2017; Zanne et al., 2018). Conversely, more diverse strategies should
35 be possible and more species can occur in less stressful environments (Šímová et al., 2017; Zanne et al., 2018). Therefore, we expect more original strategies in species occurring in extreme environments than in less extreme environments. Climate varies in its spatial frequency, and only few species will be exposed to infrequent climates (Fournier et al., 2019). As such, we expect a greater concentration of original strategies in extreme and frequent climates. While the lower exposure of species to infrequent climates means that
40 species have lower chance of undergoing adaptive radiation to these climates. As such infrequent climate should host a lower number of ecologically original species (but see Ohlemüller et al., 2008). Because the diversity of local environmental conditions promotes the diversity of species strategies (Stein et al., 2014), greater environmental heterogeneity should go with greater number of ecological outliers. Overall, we expect bright spots to be situated in stable, extreme, frequent and heterogeneous climates. Combinations of
45 environmental factors could reinforce their influence on the number of ecological outliers, such as a stable and extreme climate that should favor even greater adaptive radiations leading to an even greater number of outliers. The extent at which these factors (climate stability, climate frequency, and climate extremity) interplay is still unknown.

Even if global plant functional space has been well described (Diaz et al., 2016), a description of plant
50 ecological outliers is still missing. The wide variety of strategies observed among plants could lead to distinct ways of being an outlier. For example, *Lodoicea maldivica* the double coconut is an enormous ecological outlier because of its enormous seed mass, but other outliers could differ for other traits. Ecological outliers could be lumped together in the functional space if there was a leading dimension in originality. Depending on their traits, ecological outliers may suffer from double jeopardy if they are rare (Gaston, 1998). But the link,
55 in plant, between ecological originality and rarity has not been well assessed yet especially at a continental level (Boulangeat et al., 2012). Similarly, ecological original species could also have a specific evolutionary history, making them phylogenetically original.

We here assessed (i) the distribution of global ecological originality in plant trait space, (ii) the spatial distribution of ecologically original species, (iii) the extent at which originality hotspots are congruent with diversity and rarity hotspots, and (iv) to what extent current climate, past-climate stability, climatic frequency, and environmental heterogeneity affected the excess or deficit in original species. We used a global database of plant occurrences and corresponding functional trait values (BIEN) (Maitner et al., 2018), comprising tens of thousands of species in the Americas. We characterized the ecological originality of New World plants based on three cores traits, Leaf, Height and Seed Mass (LHS), capturing plant metabolic, competitive and reproductive strategies (Westoby, 1998). Specific Leaf Area is a proxy for photosynthesis efficiency, and its variation across plants reflects a trade-off between resource acquisition and resource conservation (Westoby et al., 2002). Plant height represents a proxy of plant competitive ability for light (Westoby, 1998). Seed mass relates to plant dispersal strategy and through reserve amount to seedling survival (Westoby, 1998).

Material and Methods

Data

BIEN range data

We used the function `BIEN_range_range()` in the R package `BIEN` (Maitner et al., 2018, v. 1.2.3) to get geographic ranges for 98829 plants species.

We defined a regular grid over the Americas including 100km by 100km cells in Lambert Equal Area Azimuthal projection, with 13°N as its latitude of origin and -75°W as its central meridian. We focused on terrestrial vegetation and selected cells with over 30% of land area based on a map of the Americas. We overlaid the species ranges to obtain potential assemblages in the 4168 pixels. We removed all species that occurred in less than 10 pixels.

BIEN trait data

We characterized the ecological strategies of plants in the dataset based on the three traits of LHS schema (Westoby, 1998), Specific Leaf Area (SLA), maximum plant height (PH), and seed mass (SM). We retrieved publicly available trait data using the function `BIEN_trait_trait()` and completed BIEN traits with public data from TRY trait database (Kattge et al., 2020), for all 98829 species with range information. We retrieved the cultivation status of species in BIEN database and removed 151 species that were only cultivated in the Americas. We ended with 2093 non-cultivated species with both trait (LHS) and range

information (see Supplementary Figure S1).

Functional Distinctiveness

We computed pairwise Euclidean distances of species for the three log-transformed and scaled LHS traits,
 90 and derived global functional distinctiveness D_i of species i :

$$D_i = \frac{\sum_{j=1}^S d_{ij}}{S-1}, \quad (1)$$

with d_{ij} the dissimilarity between species i and j , and S the total number of species (Violle et al., 2017). We used the R package `funrar` to compute functional distinctiveness (Grenié et al., 2017, p. v1.4.1). We also computed functional distinctiveness trait by trait, with the same formula.

BIEN Phylogenetic Data and Evolutionary Distinctiveness

95 We extracted 100 inferred phylogenetics tree of the 2093 selected species using the `BIEN_phylogeny_complete()` function in the BIEN package. We computed the average Evolutionary Distinctiveness (ED) of each species for each tree following Isaac et al. (2007) using the `evol.distinct()` function in R package `picante` (Kembel et al., 2010, v. 1.8) with the `equal.splits` option (Redding and Mooers, 2006). We averaged species ED across all trees to get an average ED per species.

100 Environmental Variables and Environmental Heterogeneity

We retrieved rasters of mean annual temperature, mean annual precipitation, temperature seasonality and precipitation seasonality from CHELSA from 1973-2013 at 30" resolution (CHELSA, Karger et al., 2017). We obtained the global soil water balance from CGIAR (Trabucco and Zomer, 2019) and elevation data from Worldclim (Fick and Hijmans, 2017). All environmental variables had a resolution of 1km by 1km. We
 105 calculated average environmental values for each 100km by 100km pixel.

We also assessed environmental heterogeneity within cells. We performed a Principal Component Analysis (PCA) of environmental variables, kept the first four axes explaining 93% of total variance, and calculated the average distance among the 1km by 1km sub-cells making each 100km by 100km pixel. The greater the average distance, the more heterogeneous the environmental conditions within a pixel (Fournier et al., 2019).

110 **Paleo-climate environmental variables**

We used PaleoClim v1.0 data to calculate the stability in mean annual temperature and mean annual precipitation from the Pliocene (Brown et al., 2018). We computed climate stability indices following Owens and Guralnick (2019), as the standard deviation between values at successive time intervals, divided by the length of the interval. We used the following sequence : Pliocene M2, 3.3Ma bp; mid-Pliocene warm period; 115 3.205Ma bp; Pleistocene MIS19, 787ka bp; Pleistocene Last Interglacial, 130ka bp; Pleistocene Heinrich Stadial 1, 17-14ka bp; Pleistocene Bølling-Allerød, 14-12.9ka bp; Pleistocene Younger Dryas Stadial, 12.9-11.7ka bp; early-Holocene, 11.7-8.326ka bp; mid-Holocene, 8.326-4200 ka bp; late-Holocene 4.2-0.3ka bp. We then calculated the mean deviation through all intervals, giving us the average deviation of the environmental variable through the whole time slice. We took the inverse of this value to get a paleo-climate stability index.

120 We computed the average climate stability for mean annual temperature and mean annual precipitation in each 100km by 100km pixel. To get a single metric of paleo-climate stability we logged the stability indices, before scaling both variables, with then computed the mean stability between temperature and precipitation indices.

Climate frequency index and Environmental Distinctiveness

125 We used the climate frequency index (CFI) defined by Fournier et al. (2019) to characterize the environmental rarity of each site. The climate frequency index represents the frequency of environmental conditions at a pixel compared to all other pixels, based on a kernel density distribution in a multi-dimensional climate space. The greater the index, the more frequent the environment. We computed CFI of each pixel based on the environmental PCA.

130 In addition, we computed environmental distinctiveness quantifying how extreme is environment in a pixel relative to others. We used similar formula as functional distinctiveness but based on the pixel scores in environmental PCA.

Environmental distinctiveness and CFI assess different facets of environmental originality and were thereby moderately correlated (Pearson's $r = -0.25$, 95% CI [-0.28, -0.22], $t(4166) = -16.68$, $p < 0.001$. An extreme 135 environment could be highly infrequent, while an infrequent environment could be mild.

Number of top species per pixel and restricted species

For each pixel we counted the number of top ranked species for ecological originality and evolutionary distinctiveness. We counted it for the top 1%, 5%, and 10% ranked species for each metric. Because range

size is a common metric of species rarity (Gaston, 1997), we also evaluated the number of restricted species
 140 through the same logic but ranking species in increasing range sizes.

Statistical Analyses

All analyses were performed with R (R Core Team, 2019, p. v3.6.3).

Spatial Congruence of Biodiversity Hotspots

We quantified the spatial congruence of biodiversity (and rarity) hotspots while controlling for spatial non-
 145 independence using the `cor.spatial()` function in the `SpatialPack` R package (Osorio et al., 2014). The
 test adjusts the degrees of freedom depending on the spatial auto-correlation of variables.

Null models and empirical p-values

We used the `nullmodel()` function in the `vegan` package (Oksanen et al., 2018, pp. v2.5–6), with the
 curveball algorithm to resample the presence-absence matrix while preserving the total per pixel and the
 150 number of pixels where each species occurred. We computed 1500 null simulations with 200 as the thinning
 parameter. We derived the number of top X% original species in pixels (with X = 1%, 5%, and 10%)
 under the null model and the empirical p-value of the observed number of top original species in the null
 distribution (Botta-Dukát, 2018). When the pixel showed an excess of original species –compared to the null
 distribution– then its empirical p-value was close to one. On the contrary, when a pixel showed a deficit in
 155 original species then its empirical p-value was close to zero.

Using the same null models, we computed the empirical p-values for the number of top evolutionary distinct
 species per pixel as well as the number of top restricted species.

We assessed whether and how the empirical p-values depended on environmental predictors using beta
 regression (Cribari-Neto and Zeileis, 2010; Douma and Weedon, 2019, with `betareg()` function in `betareg`
 160 package v3.1-3). Because we sometimes observed empirical p-values of exactly 0 and 1, we transformed the
 data following Smithson and Verkuilen (2006), with p' the transformed p-value $p' = (p(n - 1) + 0.5)/n$ with
 p the original empirical p-value and $n = 4168$ the number of observations.

The full regression model included all environmental variables (mean annual temperature, temperature
 seasonality, mean annual precipitation, precipitation seasonality, soil water balance, and average elevation)
 165 with both linear and quadratic effects, as well as linear effects of environmental heterogeneity, environmental
 distinctiveness, and paleo-climatic stability.

We performed model selection with the full beta regression model using the `dredge()` function in the `MuMIn` package (Barton, 2020, p. v1.14.17). To test the prediction ability of alternative models, we used model averaging with the `model.avg()` function on models that had a cumulative sum of 95% of Akaike's weight.

170 Results

We located the most original species in the functional space defined with LHS traits. The first two axes of a Principal Component analysis accounted for 89.2% of the variance of LHS traits. Consistent with the result of Díaz et al. (2016), the first axis differentiated tall species with big seeds and small species with small seeds, while the second axis differentiated species according to SLA. Ecologically original species were spread over
 175 the borders of the functional space, meaning that species could be original with all combinations of extreme trait values (high/low SLA, seed mass, and/or plant height). *Pinus monophylla* (Fig. 1A, bottom left) had very low SLA, quite high plant height and small seed mass, while *Prioria copaifera* had a quite regular SLA, high stature, and enormous seed mass; *Chloris gayana* showed a low SLA, small stature, and very low seed mass (see Table 1 for a list of the top 25 most ecologically original species).

180 We examined how ecological originality related to other rarity indices. Species evolutionary distinctiveness was moderately correlated with ecological originality (Fig. 1B, Spearman's $\rho = 0.19$, $p < 0.001$), meaning that original species tended to have a long unshared evolutionary history with other species (even when accounting for angiosperms and gymnosperms separately see Supplementary Figure S2). Species range size was weakly negatively correlated with ecological originality (Fig. 1C, Spearman's $\rho = -0.12$, $p < 0.001$),
 185 indicating that original species have smaller range size than common species (even when accounting for angiosperms and gymnosperms separately see Supplementary Figure S2), but with great variability around this average relationship. Trait-specific originality was not correlated with range size but weakly for seed mass (Spearman's $\rho = -0.035$, $p < 0.001$, Supplementary Figure S3).

The distribution of top 5% original species was not homogeneous across the Americas (Figure 2A). Compared
 190 to null expectations (Figure 2D), we found an excess of top 5% original species in assemblages of North of South America and Central America while assemblages in temperate North America and in South America showed a deficit of original species. In addition, the hotspots and coldspots of top original species differed depending on the trait considered. When using all LHS traits, we found hotspots of original species in California in the North of the Andes with over 30 top 5% original species. Conversely, coldspots (<3 top
 195 5% distinct species) were located close to Greenland, in central Canada as well as the southern Andes. The observed patterns were similar with other fraction of top original species (Supplementary Figures S4 and

S5).

We compared the excess and deficit in top 5% original species to top 5% evolutionary distinct (ED) and top 5% restricted species (Figure 2E and 2F). We used the same null models to compute empirical p-values
200 for each index. Central US showed similar deficits in top 5% ED and top 5% restricted species. Florida and Western Mexico also showed similar excess in top ED species and top restricted species. The empirical p-values for ED and restricted species showed weak correlation (Corrected $r = 0.19$, $F = 14.1$, $p < 0.001$). While for ED and original species there were no clear correlation (Corrected $r = 0.11$, $F = 1.12$, $p = 0.29$). The excess and deficits of original and restricted species showed moderate correlation (Corrected $r = 0.35$,
205 $F = 10.33$, $p = 0.002$).

To assess the environmental causes of ecological originality we fitted a model with environmental predictors on the empirical p-values of the top 5% original species per pixel (Figure 3). The highest absolute estimates were temperature seasonality (-1.65), mean annual temperature (linear = -0.75, quadratic = 0.48), precipitation seasonality (0.23) and mean elevation (-0.21). Environmental distinctiveness had a small non-significant effect
210 (0.07), paleo-climate stability had no detectable effect, while both climatic frequency index and environmental heterogeneity had negative effects (respectively -0.07 and -0.11). When considering the main effects, with all other factors kept constants, (Supplementary Figure S6), pixels with low mean annual temperature, low temperature seasonality and low annual precipitation all tended to show an excess of original species. Pixels with high annual precipitation and high temperature seasonality tended to show a deficit in original species.

215 Discussion

Bright spots in ecological originality were located in the North of South America and Central America, while dark spots, showing a deficit in original species, were located in the center of North America and the South of South America. Current climate and environmental heterogeneity strongly determined the location of bright and dark spots, while we observed little influence of other climatic factors (climate stability, climate
220 frequency, and extreme climates).

Ecological originality bright spots were incongruent with evolutionary distinctiveness bright spots but somewhat congruent with restricted species bright spots. The incongruence between ecological originality and evolutionary distinctiveness bright spots has been observed in other taxa (Grenié et al., 2018; Vasconcelos et al., 2019), this suggest that different processes shape these different rarity facets. Evolutionary distinctiveness bright spots are “museums” of diversity where a long evolutionary history is preserved (Mace et al., 2003) while ecological originality bright spots may instead represent “cradles” where biodiversity emerge

(Mace et al., 2003). We however, observed some congruence between evolutionary distinctiveness and ecological originality in some regions (North of Canada, Florida, Atlantic Brazilian coast). Different process may drive bright spots dynamics in different locations and at different scales. The incongruence of bright spots
230 call for innovative conservation actions that take into account phylogenetic, functional, and rarity facets altogether in prioritization scheme (Arponen, 2012; Guilhaumon et al., 2015; Pollock et al., 2017).

Bright spots of ecologically original species showed high precipitation seasonality, low temperature seasonality, and low mean annual precipitation. We found no effect of past-climate stability, a small negative effect of climate frequency and environmental heterogeneity, as well as small positive effect of climate extremity. Thus, all else equal, we showed that extreme, infrequent and homogeneous environments hosted more
235 ecologically original species than expected according to their species richness solely. This is in line with our hypothesis of extreme environments selecting for particular combinations of traits. (Zanne et al., 2018) Indeed among the most ecologically original species are species adapted to extreme environments such as deserts and alpine environments (Table 1). For example, the most distinct species of our dataset is *Aster*
240 *alpinus* a small herbaceous plant adapted to alpine environments, which are quite extreme and infrequent at the scale of the Americas. Other highly original species such as *Juniperus osteosperma* and *Pinus monophylla* are adapted to deserts. More striking was the negative effect of environmental heterogeneity as the diversity of environmental conditions is supposed to drive the diversity of species strategies locally (Stein et al., 2014). As environmental heterogeneity alone cannot explain bright spots, we expect that a combination of factors
245 such as both homogeneous and extreme environments favour the specific adaptations of species. In summary, contrary to our hypotheses we found no effect of climate stability on bright spots, and found little effects of climate frequency, environmental heterogeneity and climate extremity.

Current environmental conditions were the main effects for bright spots of ecological originality. Different climates may thus show different excess or deficit in global original plants species. Our study highlighted a
250 latitudinal gradient in the excess of original species with the Tropics harboring the greater excess in original species. Of course this in line with the latitudinal gradient in species richness (Hillebrand, 2004). We would expect by pure sampling effect we expect greater number of original species in species-rich environments (Hillebrand, 2004). Through our null model analyses we explicitly controlled for the expected richness in the community, meaning that the latitudinal gradient we observe is a gradient of excess of original species.
255 The excess of original species in Tropics maybe due to increased specialization at these latitudes (Antonelli and Sanmartín, 2011; Brown, 2014). There is a latitudinal gradient in habitat specialization with a greater number of habitat specialists in the tropics because of species lower environmental tolerances (Gaston and Blackburn, 2000). The originality gradient we observe may thus be due to this specialization gradient as

we expect original species to be specialized of a given environment. However, we also observe an excess of
260 original species in the North of Canada. These bright spots of originality may be caused by greater species
adaptations to the reigning extreme environmental conditions (Bjorkman et al., 2018; Zanne et al., 2018).
Different environmental factors may thus drive bright spots located in distinct regions.

We found a slight negative correlation between ecological originality and range size, underlining that original
species are globally more restricted than common species at continental scale. This negative relationship has
265 also been observed at the biome-scale (Echeverría-Londoño et al., 2018). Original species may be specialized
to specific environmental conditions making them quite restricted in range. While common species are more
generalists, adapted to a wide range in environmental conditions (Denelle et al., 2020). Because there is
considerable variation over this relationship some species can show a restricted range while being ecologically
common such as *Dirca occidentalis* (a shrub endemic from San Francisco Bay Area) or have a wide range and
270 being ecologically original such as *Chloris gayana* (a grass that is extremely drought tolerant). The negative
relationship identified here however contrasts with the absence of (or the positive) relationship between
ecological originality and range size at global scale observed in other organisms (Chapman et al., 2018;
Cooke et al., 2020; Grenié et al., 2018). The low ecological originality observed between widespread species
of plants maybe linked to the similarity of traits needed to occur in a wide range of environmental conditions
275 (Díaz et al., 2016; Westoby et al., 2002). While for other species, there may not be a single combination
of traits that allows wide environmental tolerances. The observed negative relationship between range size
and ecological originality call for specific conservation actions regarding ecologically original and restricted
species as they may contribute uniquely to ecosystem functioning (Jain et al., 2014; Leitão et al., 2016).

Ecological originality exhibited a moderate positive correlation with evolutionary distinctiveness. This is
280 inline with the predictions that evolutionary distinct species are likely to have more unique features (Faith,
1992; Redding and Mooers, 2015). However, this result depends on the considered taxonomic group (Supple-
mentary Figure S2), indeed Angiosperms showed similar positive correlation, while Gymnosperms showed a
stronger negative relationship between ecological originality and evolutionary distinctiveness (Cornwell et al.,
2014). These results are distinct from what has been observed on mammals (Violle et al., 2017), hydrother-
285 mal vents organisms (Chapman et al., 2018), coral reef fishes (Grenié et al., 2018), and birds (Cooke et al.,
2020), where no relationship between ecological original and evolutionary distinctiveness has been found. We
found a moderate relationship with considerable variation along it. These studies also used a greater number
of traits which can average out the evolutionary signal of a few traits (Tucker et al., 2018), while we focused
only on three traits that may show similar signal as they are physiologically highly constrained (Díaz et al.,
290 2016) Evolutionary distinctiveness has been used for a long time to prioritize species conservation and we

suggested adding ecological originality to prioritization scheme (Isaac et al., 2007; Vane-Wright et al., 1991; but see criticisms in Tucker et al., 2019).

Our study is the first to describe ecological originality and its drivers in plants in the Americas. Current environmental conditions were the main drivers of the excess of original species. Our results question the role of historical factors such as climate stability in shaping the distribution of original species in space, but also in trait space. More investigations are needed on the links between ecological originality and species evolutionary history to understand the effect of macro-evolutionary processes in determining current original species.

Acknowledgements

We would like to thank Pierre Denelle for helpful discussions. MG was supported by the ENS de Lyon. This study was supported by the European Research Council (ERC) Starting Grant Project 'ecophysiological and biophysical constraints on domestication in crop plants' (grant ERC-StG-2014-639706-CONSTRAINTS) and by the French Foundation for Research on Biodiversity (FRB; <www.fondationbiodiversite.fr>) and EDF in the context of the CESAB project 'causes and consequences of functional rarity from local to global scales' (FREE).

Data Availability Statement

The code and data to reproduce these analyses will be made available on Zenodo (cite DOI), while the latest development version can be accessed on GitHub.

Figures

Table 1: Table of the top 25 functionally distinct species according to their LHS traits

Group	Order	Family	Species	Functional Distinctiveness	SLA (mm ² .mg ⁻¹)	Plant Height (m)	Seed Mass (mg)
flowering plants	Asterales	Asteraceae	Aster alpinus	3.83	119.64	0.08	1.17
flowering plants	Sapindales	Meliaceae	Carapa guianensis	3.81	10.93	60.00	17685.65

Group	Order	Family	Species	Functional Distinctiveness	SLA (mm ² ·mg ⁻¹)	Plant Height (m)	Seed Mass (mg)
flowering plants	Fagales	Casuarinaceae	Casuarina cristata	4.35	1.57	11.67	7.65
flowering plants	Poales	Poaceae	Chloris gayana	4.10	1.79	0.78	0.35
flowering plants	Brassicales	Capparaceae	Cynophalla flexuosa	4.33	1.62	6.62	67.00
flowering plants	Poales	Poaceae	Digitaria eriantha	4.20	1.63	0.63	0.87
flowering plants	Poales	Poaceae	Eragrostis curvula	4.32	1.60	0.53	0.21
flowering plants	Poales	Poaceae	Eragrostis superba	3.81	2.13	0.48	0.96
flowering plants	Poales	Juncaceae	Juncus maritimus	3.83	2.59	0.75	0.03
gymnosperms (conifers)	Cupressales	Cupressaceae	Juniperus monosperma	4.66	1.28	8.46	26.62
gymnosperms (conifers)	Cupressales	Cupressaceae	Juniperus occidentalis	3.96	2.21	14.28	47.83
gymnosperms (conifers)	Cupressales	Cupressaceae	Juniperus osteosperma	4.46	1.55	7.87	112.34
flowering plants	Poales	Poaceae	Panicum coloratum	3.85	2.13	0.36	0.62
gymnosperms (conifers)	Pinales	Pinaceae	Pinus edulis	3.97	2.44	14.06	239.39
gymnosperms (conifers)	Pinales	Pinaceae	Pinus jeffreyi	3.87	3.08	50.25	117.16
gymnosperms (conifers)	Pinales	Pinaceae	Pinus monophylla	4.87	1.32	11.02	474.91

Group	Order	Family	Species	Functional Distinctiveness	SLA (mm ² ·mg ⁻¹)	Plant Height (m)	Seed Mass (mg)
gymnosperms (conifers)	Pinales	Pinaceae	Pinus resinosa	4.19	1.95	27.30	8.88
flowering plants	Fabales	Fabaceae	Prioria copaifera	3.98	11.05	38.83	50864.95
flowering plants	Malpighiales	Rhizophoraceae	Rhizophora mangle	3.95	4.65	20.60	10100.00
flowering plants	Poales	Poaceae	Setaria sphacelata	3.85	2.05	0.66	0.75
flowering plants	Poales	Poaceae	Sporobolus pyramidalis	4.42	1.50	0.72	0.16
flowering plants	Caryophyllales	Caryophyllaceae	Stellaria media	4.27	176.28	0.34	0.37
flowering plants	Fabales	Fabaceae	Swartzia polyphylla	3.85	8.40	25.62	30236.00
flowering plants	Poales	Poaceae	Tragus berteronianus	4.12	2.14	0.08	0.28
flowering plants	Santalales	Ximeniaceae	Ximenia americana	4.03	2.37	5.71	694.31

A

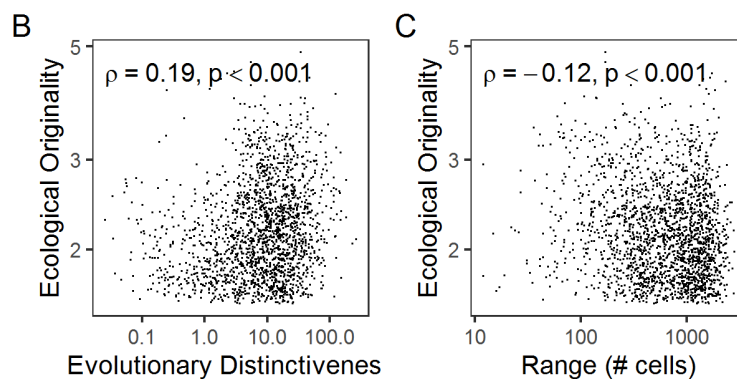
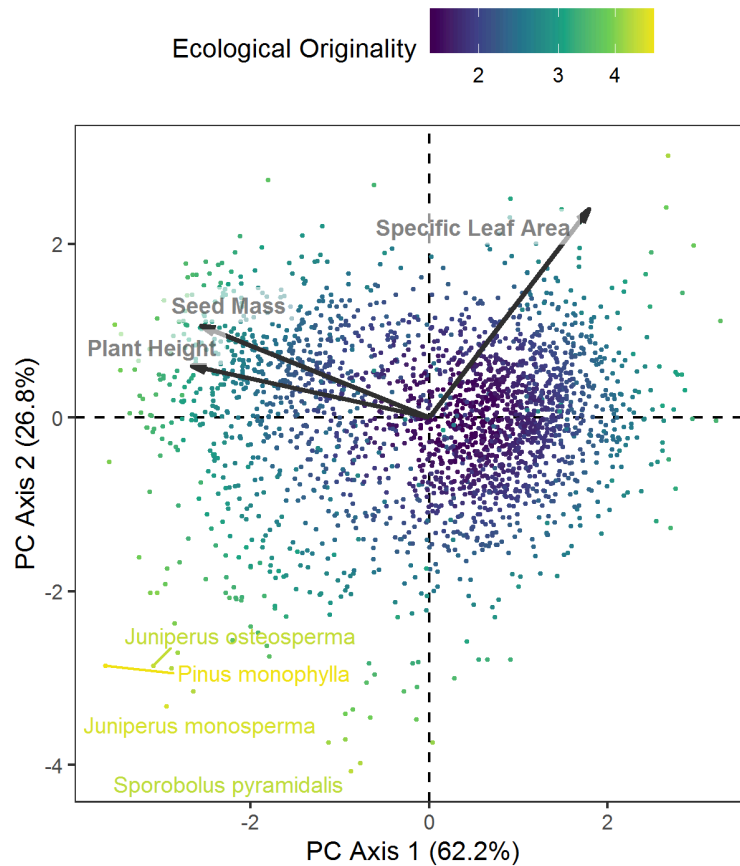


Figure 1: **Relationship between ecological originality and trait, evolutionary distinctiveness, and species range.** (A) Principal component analysis of the traits of all 2093 species, with points colored in function of their ecological originality; arrows show the projection of trait on the PCA axes; the 25 top distinct species are labelled. (B) Relationship between ecological originality and evolutionary distinctiveness, the correlation coefficient is indicated at the top of the plot, both x- and y- axes are in log₁₀ scale. (C) Relationship between ecological originality and species ranges for species occurring in more than 10 sites, the correlation coefficient is indicated at the top of the plot, both x- and y- axes are in log₁₀ scale.

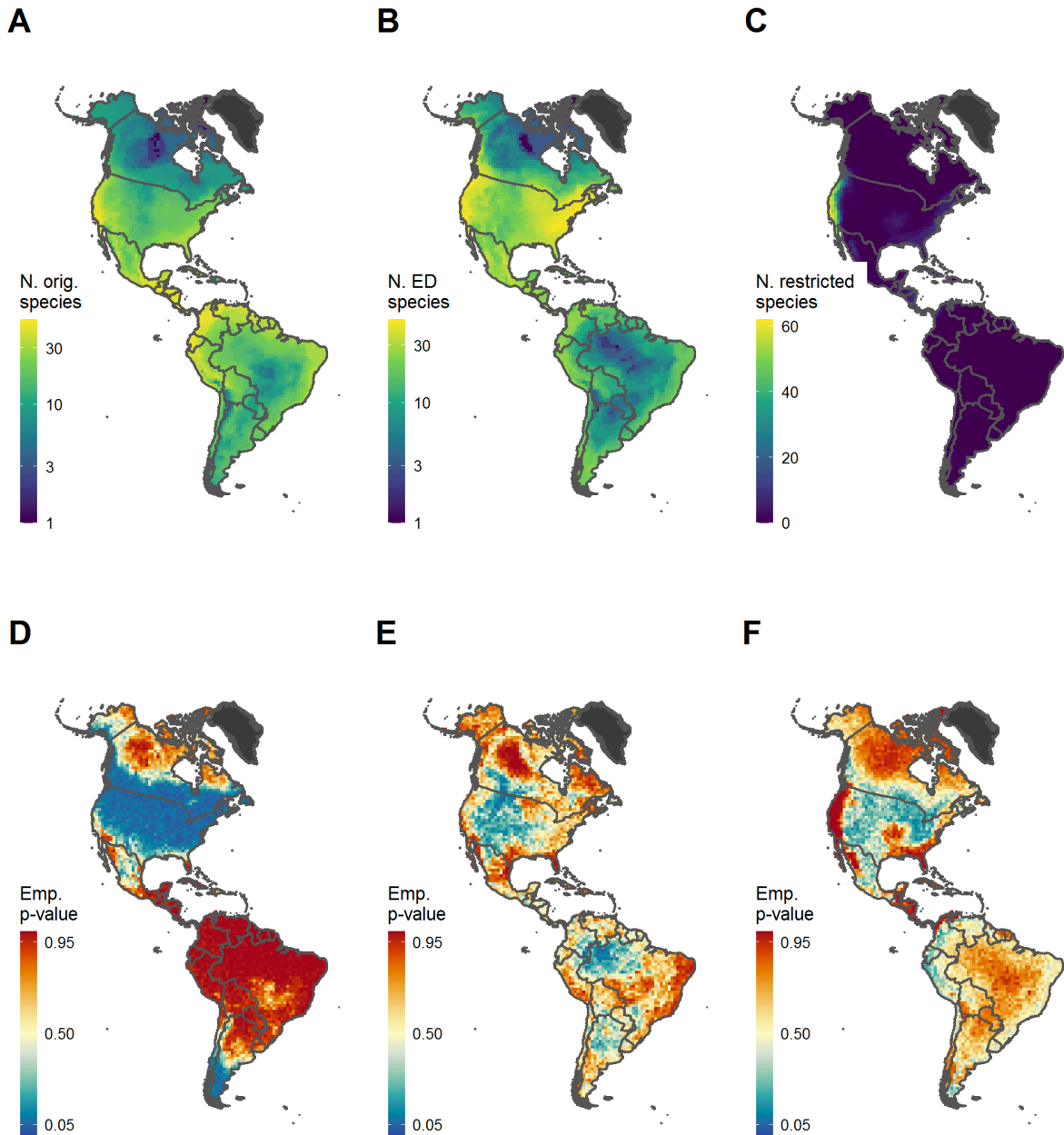


Figure 2: Number of top 5% original, evolutionary distinct, and restricted species per 100km by 100km assemblage and corresponding empirical p-value based on null models. (A, B, C) each map shows the number of top 5% species for originality, evolutionary distinct, and restricted species respectively. (D, E, F) the corresponding map of empirical two-sided p-values. If the value is under 0.05 then the pixel shows a deficit in top species, while when the value is over 0.95 the pixel shows an excess in top species. The maps are shown in an equal area projection.

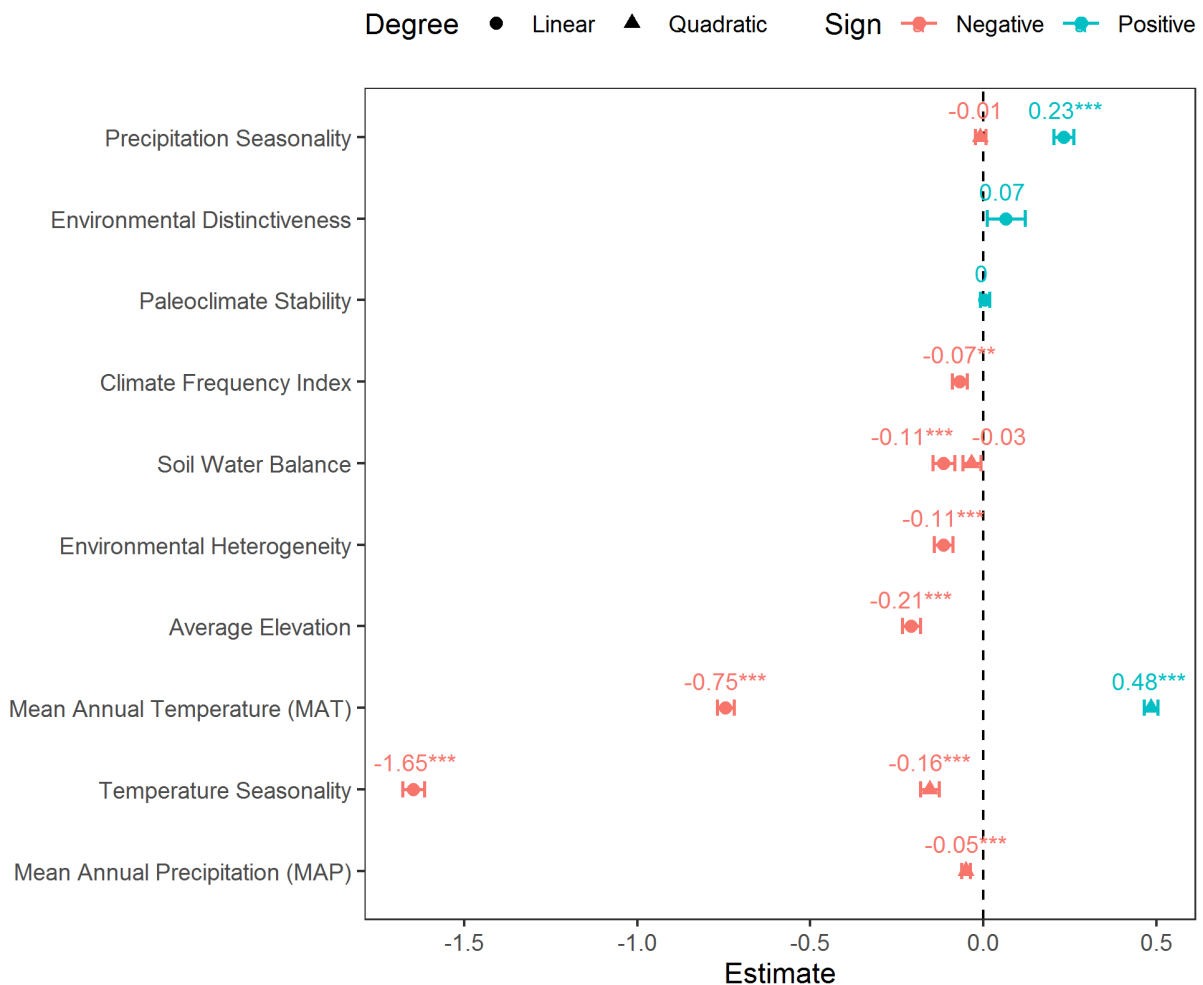


Figure 3: Estimates of standardized environmental predictors from averaged beta regression model of the empirical p-value of top 5% distinct species considering LHS traits. Points indicate the estimated value while the error bar represents the standard error. The values of the estimates are reported on top of the points with corresponding p-values coded as follow: ***= $p < 0.001$, **= $p < 0.01$, *= $p < 0.05$, .= $p < 0.1$, otherwise $p > 0.1$. Predictors are ranked by decreasing average linear estimates. Predictors in blue show predictors with positive average estimates, while predictors in red show negative average estimates. Circles show the value of the linear effect of predictors, while square show the value for quadratic effect of predictors.

310 **Supplementary material**

Supplementary Figure S1: Species Richness of subset of species included in the dataset.

Supplementary Figure S2: Relationship between Evolutionary Distinctiveness and Ecological Originality, and Range Size and Ecological Originality in function of taxonomic group. (A) Evolutionary distinctiveness vs. ecological originality. (B) Range size vs. ecological originality. The Spearman's correlation coefficients are indicated for each group in the top left of each (ferns $n = 4$; flowering plants $n = 2287$; gymnosperms $n = 76$). Both evolutionary distinctiveness and ecological originality are computed using all groups together. Note that both x and y axes are in log10 scale.

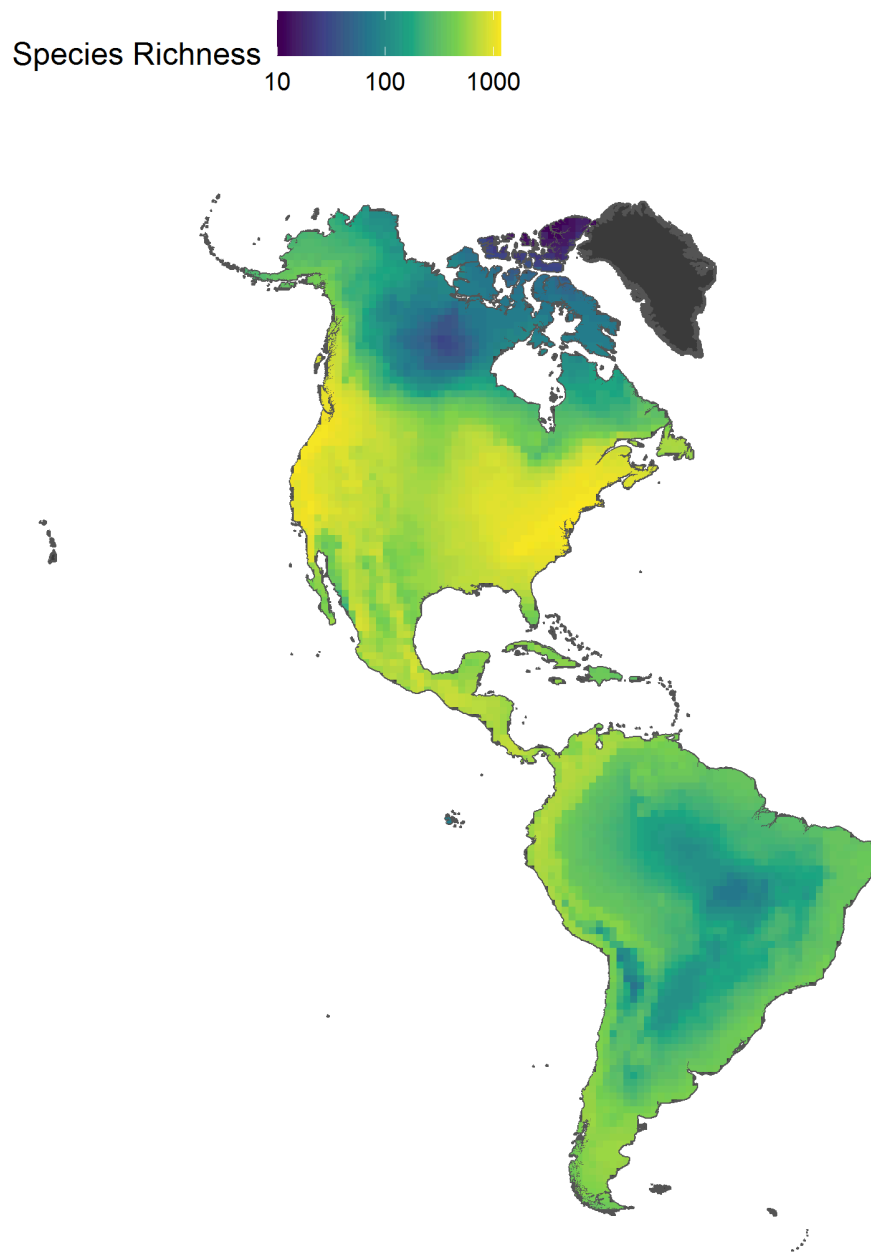
Supplementary Figure S3: Single trait ecological originality in function of species range. (A) Specific Leaf Area. (B) Maximum Plant Height. (C) Seed Mass. Each panel indicates the Spearman correlation coefficient and the corresponding p-values.

Supplementary Figure S4: Maps of number of top X% species ecologically original species based on all three LHS traits, or each one independently. (A-D) Number of top 1% ecologically original species, (E-H) number of top 5% ecologically original species, (I-L) number of top 10% ecologically original species. (A, E, I) All three LHS traits, (B, F, J) specific leaf area, (C, G, K) plant height, (D, H, L) seed mass.

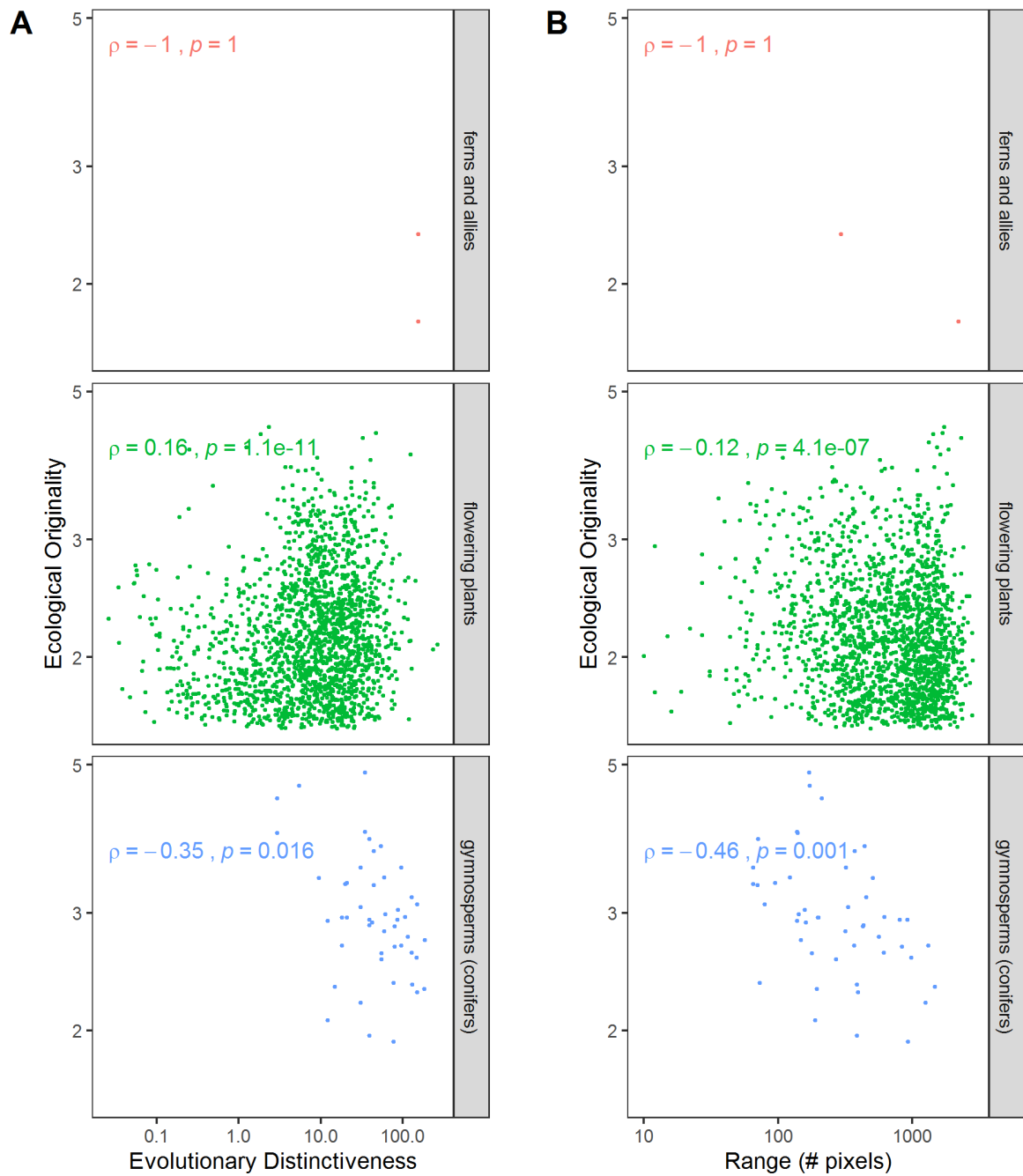
Supplementary Figure S5: Maps of empirical p-value of top X% distinct species depending on the trait used and the percentage of species considered.

Supplementary Figure S6 Partial residuals plots from beta-regression models of empirical p-value against environmental predictors.

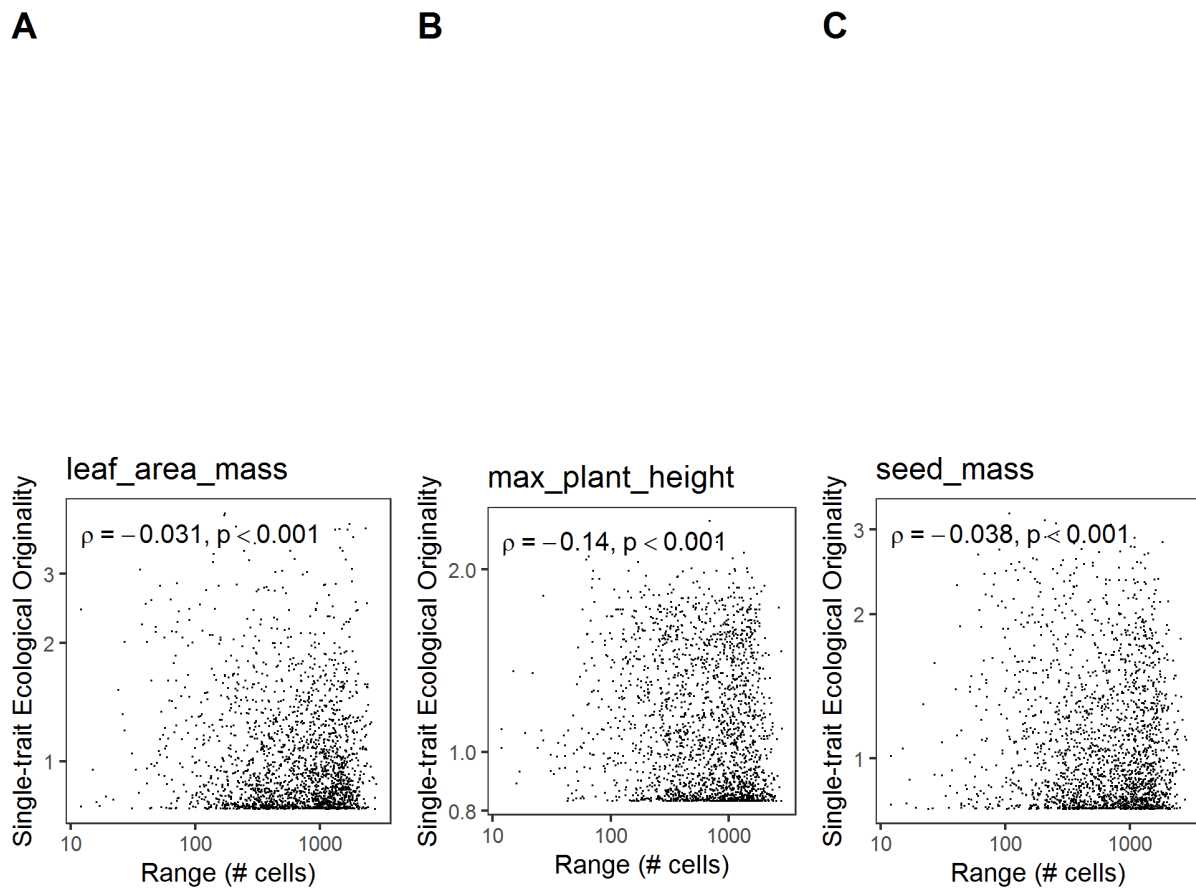
Supplementary Figure S7: (A) Map of climatic frequency index; (B) map of environmental distinctiveness.



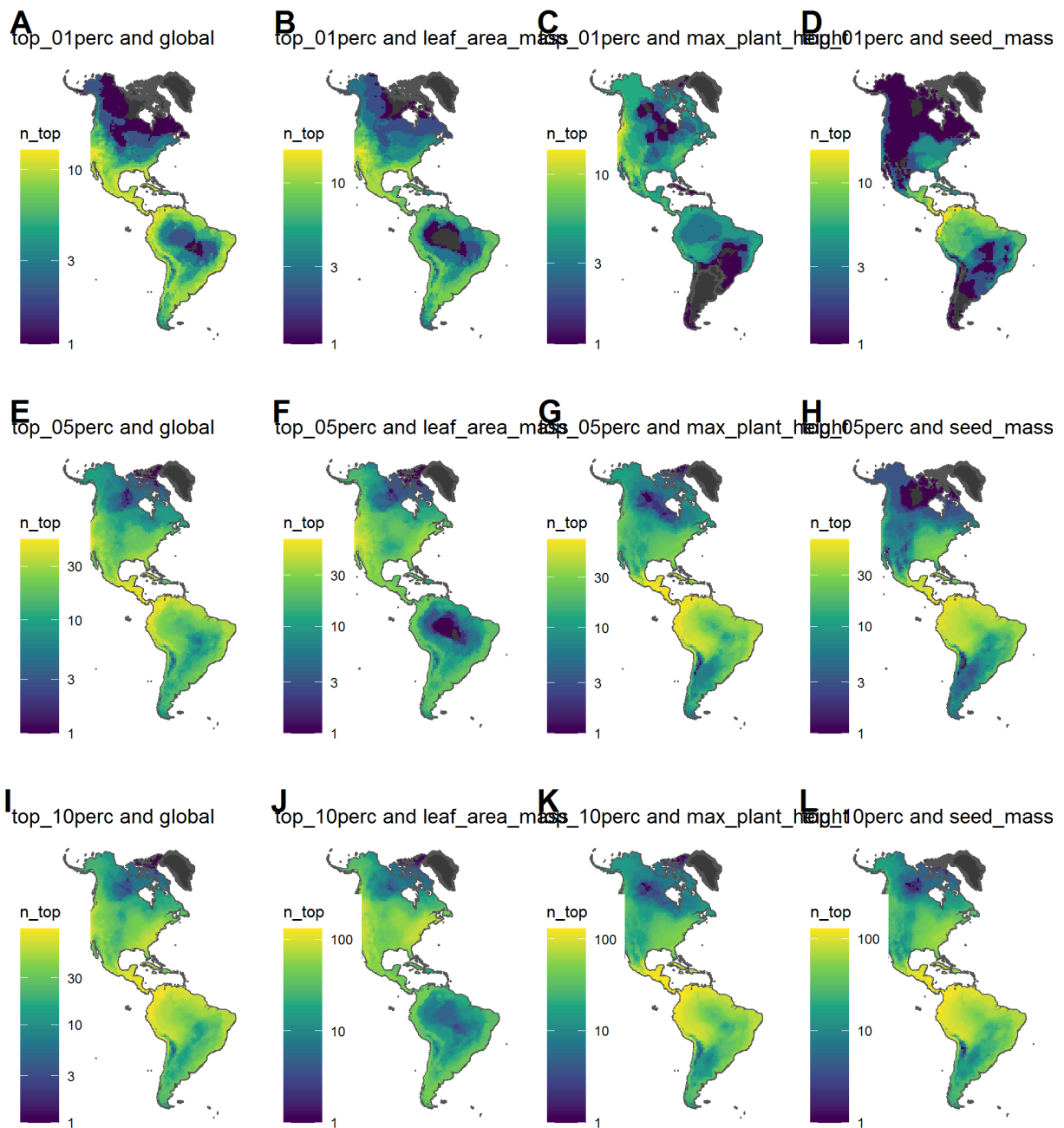
Supplementary Figure S1



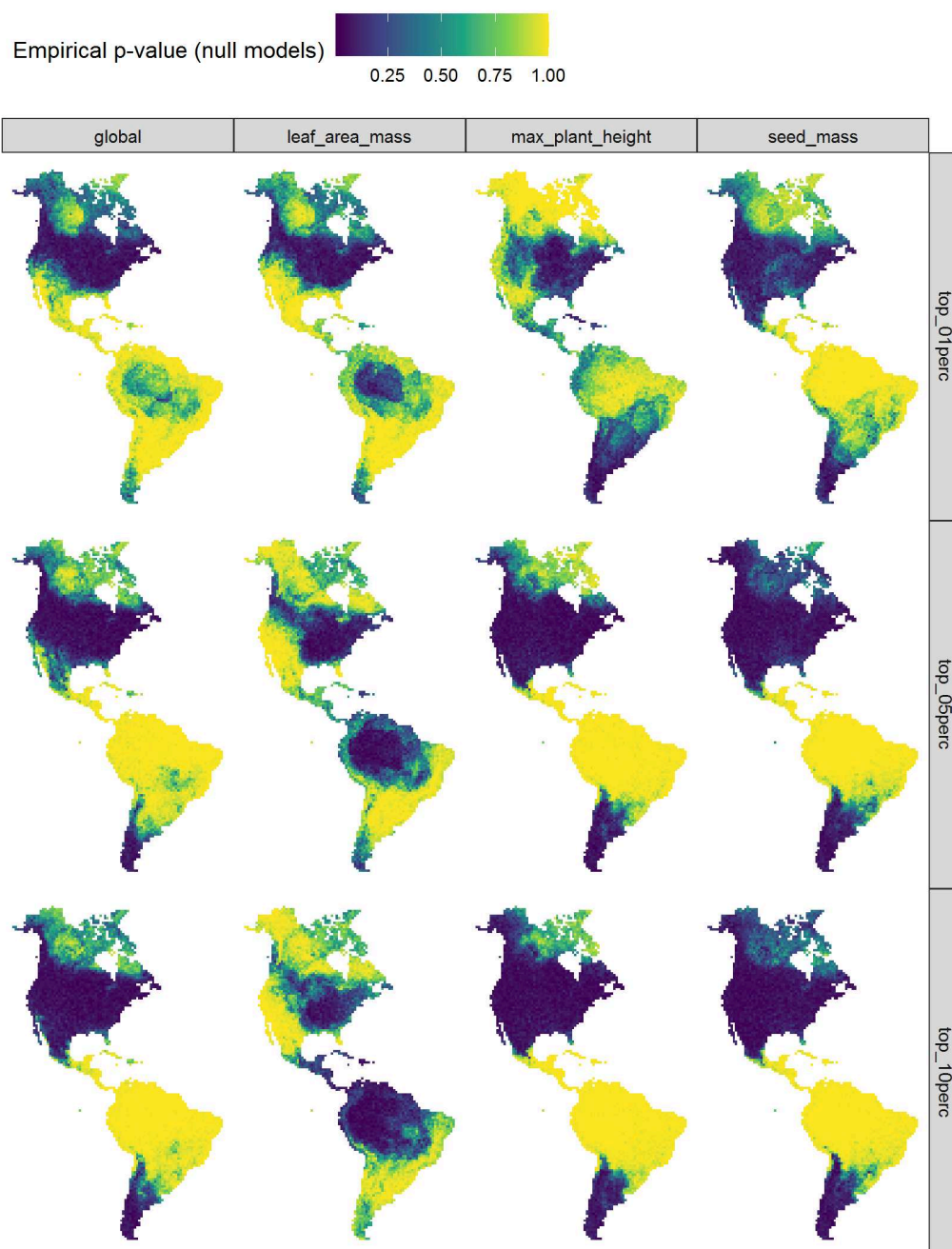
Supplementary Figure S2



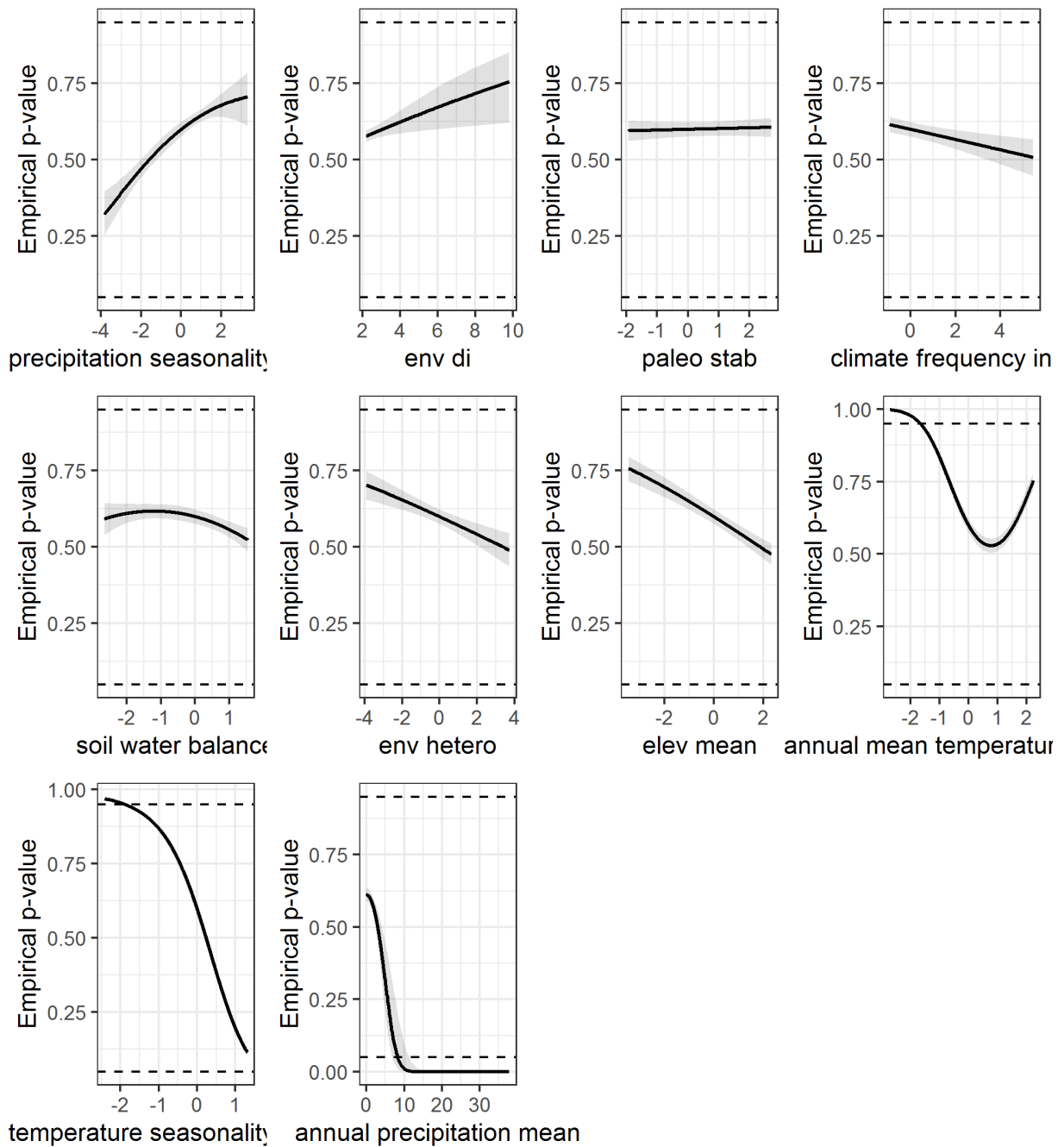
Supplementary Figure S3



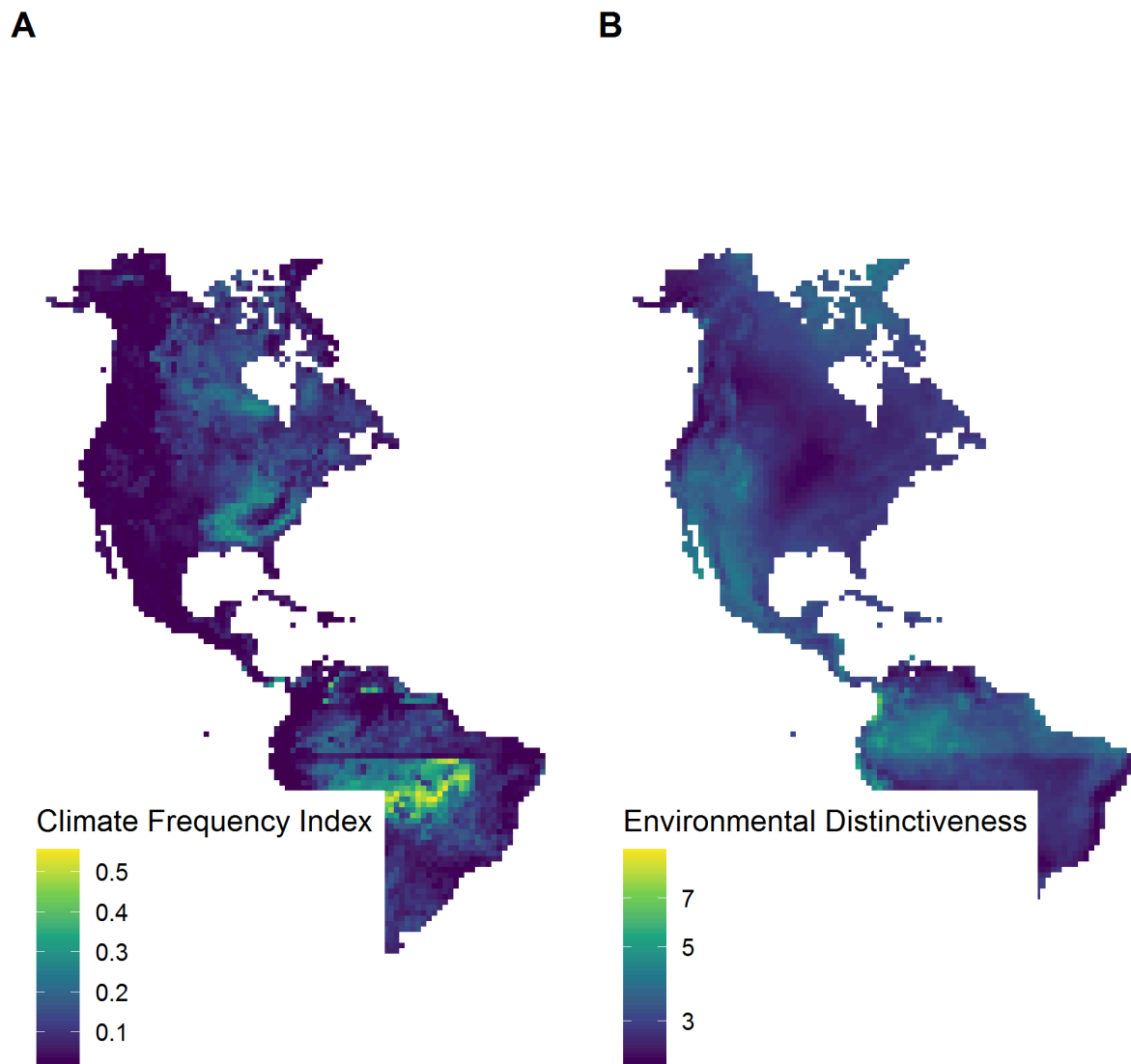
Supplementary Figure S4



Supplementary Figure S5



Supplementary Figure S6



Supplementary Figure S7

References

- Antonelli, A., Sanmartín, I., 2011. Why are there so many plant species in the Neotropics? *TAXON* 60, 403–414. <https://doi.org/10.1002/tax.602010>
- 335 Arponen, A., 2012. Prioritizing species for conservation planning. *Biodiversity and Conservation* 21, 875–893. <https://doi.org/10.1007/s10531-012-0242-1>
- Barton, K., 2020. MuMIn: Multi-model inference.
- Bjorkman, A.D., Myers-Smith, I.H., Elmendorf, S.C., Normand, S., Rüger, N., Beck, P.S.A., Blach-Overgaard, A., Blok, D., Cornelissen, J.H.C., Forbes, B.C., Georges, D., Goetz, S.J., Guay, K.C., Henry, G.H.R., HilleRisLambers, J., Hollister, R.D., Karger, D.N., Kattge, J., Manning, P., Prevéy, J.S., Rixen, C., 340 Schaepman-Strub, G., Thomas, H.J.D., Vellend, M., Wilmking, M., Wipf, S., Carboognani, M., Hermanutz, L., Lévesque, E., Molau, U., Petraglia, A., Soudzilovskaia, N.A., Spasojevic, M.J., Tomaselli, M., Vowles, T., Alatalo, J.M., Alexander, H.D., Anadon-Rosell, A., Angers-Blondin, S., te Beest, M., Berner, L., Björk, R.G., Buchwal, A., Buras, A., Christie, K., Cooper, E.J., Dullinger, S., Elberling, B., Eskelinen, A., Frei, 345 E.R., Grau, O., Grogan, P., Hallinger, M., Harper, K.A., Heijmans, M.M.P.D., Hudson, J., Hülber, K., Iturrate-Garcia, M., Iversen, C.M., Jaroszynska, F., Johnstone, J.F., Jørgensen, R.H., Kaarlejärvi, E., Klady, R., Kuleza, S., Kulonen, A., Lamarque, L.J., Lantz, T., Little, C.J., Speed, J.D.M., Michelsen, A., Milbau, A., Nabe-Nielsen, J., Nielsen, S.S., Ninot, J.M., Oberbauer, S.F., Olofsson, J., Onipchenko, V.G., Rumpf, S.B., Semenchuk, P., Shetti, R., Collier, L.S., Street, L.E., Suding, K.N., Tape, K.D., Trant, A., 350 Treier, U.A., Tremblay, J.-P., Tremblay, M., Venn, S., Weijers, S., Zamin, T., Boulanger-Lapointe, N., Gould, W.A., Hik, D.S., Hofgaard, A., Jónsdóttir, I.S., Jorgenson, J., Klein, J., Magnusson, B., Tweedie, C., Wookey, P.A., Bahn, M., Blonder, B., van Bodegom, P.M., Bond-Lamberty, B., Campetella, G., Cerabolini, B.E.L., Chapin, F.S., Cornwell, W.K., Craine, J., Dainese, M., de Vries, F.T., Díaz, S., Enquist, B.J., Green, W., Milla, R., Niinemets, Onoda, Y., Ordoñez, J.C., Ozinga, W.A., Penuelas, J., Poorter, H., 355 Poschlod, P., Reich, P.B., Sandel, B., Schamp, B., Sheremetev, S., Weiher, E., 2018. Plant functional trait change across a warming tundra biome. *Nature* 562, 57–62. <https://doi.org/10.1038/s41586-018-0563-7>
- Botta-Dukát, Z., 2018. Cautionary note on calculating standardized effect size (SES) in randomization test. *Community Ecology* 19, 77–83. <https://doi.org/10.1556/168.2018.19.1.8>
- Boulangeat, I., Lavergne, S., Es, J.V., Garraud, L., Thuiller, W., 2012. Niche breadth, rarity and ecological 360 characteristics within a regional flora spanning large environmental gradients. *Journal of Biogeography* 39, 204–214. <https://doi.org/10.1111/j.1365-2699.2011.02581.x>

Brown, J.H., 2014. Why are there so many species in the tropics? *Journal of biogeography* 41, 8–22.

Brown, J.L., Hill, D.J., Dolan, A.M., Carnaval, A.C., Haywood, A.M., 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* 5, 1–9. <https://doi.org/10.1038/sdata.2018.254>

365

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., Naeem, S., 2012. Biodiversity loss and its impact on humanity. *Nature* 486, 59–67. <https://doi.org/10.1038/nature11148>

370

Chapman, A.S.A., Tunnicliffe, V., Bates, A.E., 2018. Both rare and common species make unique contributions to functional diversity in an ecosystem unaffected by human activities. *Diversity and Distributions* 24, 568–578. <https://doi.org/10.1111/ddi.12712>

Cinner, J.E., Huchery, C., MacNeil, M.A., Graham, N.A.J., McClanahan, T.R., Maina, J., Maire, E., Kittinger, J.N., Hicks, C.C., Mora, C., Allison, E.H., D'Agata, S., Hoey, A., Feary, D.A., Crowder, L., Williams, I.D., Kulbicki, M., Vigliola, L., Wantiez, L., Edgar, G., Stuart-Smith, R.D., Sandin, S.A., Green, A.L., Hardt, M.J., Beger, M., Friedlander, A., Campbell, S.J., Holmes, K.E., Wilson, S.K., Brokovich, E., Brooks, A.J., Cruz-Motta, J.J., Booth, D.J., Chabanet, P., Gough, C., Tupper, M., Ferse, S.C.A., Sumaila, U.R., Mouillot, D., 2016. Bright spots among the world's coral reefs. *Nature* 535, 416–419. <https://doi.org/10.1038/nature18607>

380

Cooke, R.S.C., Eigenbrod, F., Bates, A.E., 2020. Ecological distinctiveness of birds and mammals at the global scale. *Global Ecology and Conservation* e00970. <https://doi.org/10.1016/j.gecco.2020.e00970>

Cornwell, W.K., Westoby, M., Falster, D.S., FitzJohn, R.G., O'Meara, B.C., Pennell, M.W., McGlinn, D.J., Eastman, J.M., Moles, A.T., Reich, P.B., Tank, D.C., Wright, I.J., Aarssen, L., Beaulieu, J.M., Kooyman, R.M., Leishman, M.R., Miller, E.T., Niinemets, Oleksyn, J., Ordonez, A., Royer, D.L., Smith, S.A., Stevens, P.F., Warman, L., Wilf, P., Zanne, A.E., 2014. Functional distinctiveness of major plant lineages. *Journal of Ecology* 102, 345–356. <https://doi.org/10.1111/1365-2745.12208>

385

Cribari-Neto, F., Zeileis, A., 2010. Beta regression in R. *Journal of Statistical Software* 34, 1–24.

Denelle, P., Violle, C., DivGrass Consortium, Munoz, F., 2020. Generalist plants are more competitive and more functionally similar to each other than specialist plants: Insights from network analyses. *Journal of Biogeography* n/a. <https://doi.org/10.1111/jbi.13848>

390

Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth,

- C., Prentice, I.C., Garnier, E., Bönisch, G., Westoby, M., Poorter, H., Reich, P.B., Moles, A.T., Dickie, J., Gillison, A.N., Zanne, A.E., Chave, J., Wright, S.J., Sheremet'ev, S.N., Jactel, H., Baraloto, C., Cerabolini, B., Pierce, S., Shipley, B., Kirkup, D., Casanoves, F., Joswig, J.S., Günther, A., Falczuk, V., Rüger, N.,
 395 Mahecha, M.D., Gorné, L.D., 2016. The global spectrum of plant form and function. *Nature* 529, 167–171. <https://doi.org/10.1038/nature16489>
- Douma, J.C., Weedon, J.T., 2019. Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology and Evolution* 10, 1412–1430. <https://doi.org/10.1111/2041-210X.13234>
- 400 Echeverría-Londoño, S., Enquist, B.J., Neves, D.M., Violle, C., Boyle, B., Kraft, N.J.B., Maitner, B.S., McGill, B., Peet, R.K., Sandel, B., Smith, S.A., Svenning, J.-C., Wisser, S.K., Kerkhoff, A.J., 2018. Plant Functional Diversity and the Biogeography of Biomes in North and South America. *Frontiers in Ecology and Evolution* 6. <https://doi.org/10.3389/fevo.2018.00219>
- Enquist, B.J., Condit, R., Peet, R.K., Schildhauer, M., Thiers, B.M., 2016. Cyberinfrastructure for an
 405 integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity (No. e2615v2). *PeerJ Inc.* <https://doi.org/10.7287/peerj.preprints.2615v2>
- Espeland, E.K., Emam, T.M., 2011. The value of structuring rarity: The seven types and links to reproductive ecology. *Biodiversity and Conservation* 20, 963–985. <https://doi.org/10.1007/s10531-011-0007-2>
- Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61, 1–10.
 410 [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3)
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>
- Fournier, B., Vázquez-Rivera, H., Clappe, S., Donelle, L., Braga, P.H.P., Peres-Neto, P.R., 2019. The spatial frequency of climatic conditions affects niche composition and functional diversity of species assemblages:
 415 The case of Angiosperms. *Ecology Letters*. <https://doi.org/10.1111/ele.13425>
- Gaston, K., Blackburn, T., 2000. *Pattern and Process in Macroecology*, 1 edition. ed. Wiley-Blackwell, Oxford.
- Gaston, K.J., 1998. Rarity as double jeopardy. *Nature* 394, 229–230. <https://doi.org/10.1038/28288>
- Gaston, K.J., 1997. What is rarity?, in: Kunin, W.E., Gaston, K.J. (Eds.), *The Biology of Rarity, Population and Community Biology Series*. Springer Netherlands, pp. 30–47. https://doi.org/10.1007/978-94-011-5874-9_3

- Grenié, M., Denelle, P., Tucker, C.M., Munoz, F., Violle, C., 2017. Funrar: An R package to characterize functional rarity. *Diversity and Distributions* 23, 1365–1371. <https://doi.org/10.1111/ddi.12629>
- Grenié, M., Mouillot, D., Villéger, S., Denelle, P., Tucker, C.M., Munoz, F., Violle, C., 2018. Functional
425 rarity of coral reef fishes at the global scale: Hotspots and challenges for conservation. *Biological Conservation* 226, 288–299. <https://doi.org/10.1016/j.biocon.2018.08.011>
- Guilhaumon, F., Albouy, C., Claudet, J., Velez, L., Ben Rais Lasram, F., Tomasini, J.-A., Douzery, E.J.P., Meynard, C.N., Mouquet, N., Troussellier, M., Araújo, M.B., Mouillot, D., 2015. Representing taxonomic, phylogenetic and functional diversity: New challenges for Mediterranean marine-protected areas. *Diversity*
430 and *Distributions* 21, 175–187. <https://doi.org/10.1111/ddi.12280>
- Harrison, S., Noss, R., 2017. Endemism hotspots are linked to stable climatic refugia. *Annals of Botany* 119, 207–214. <https://doi.org/10.1093/aob/mcw248>
- Hillebrand, H., 2004. On the Generality of the Latitudinal Diversity Gradient. *The American Naturalist* 163, 192–211. <https://doi.org/10.1086/381004>
- 435 Isaac, N.J.B., Turvey, S.T., Collen, B., Waterman, C., Baillie, J.E.M., 2007. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny. *PLOS ONE* 2, e296. <https://doi.org/10.1371/journal.pone.0000296>
- Jain, M., Flynn, D.F.B., Prager, C.M., Hart, G.M., DeVan, C.M., Ahrestani, F.S., Palmer, M.I., Bunker, D.E., Knops, J.M.H., Jouseau, C.F., Naeem, S., 2014. The importance of rare species: A trait-based
440 assessment of rare species contributions to functional diversity and possible ecosystem function in tall-grass prairies. *Ecology and Evolution* 4, 104–112. <https://doi.org/10.1002/ece3.915>
- Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4, 170122. <https://doi.org/10.1038/sdata.2017.122>
- 445 Kattge, J., Bönsch, G., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Tautenhahn, S., Werner, G.D.A., Aakala, T., Abedi, M., al, 2020. TRY plant trait database enhanced coverage and open access. *Global Change Biology* 26, 119–188. <https://doi.org/10.1111/gcb.14904>
- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464.
- 450 Kingsolver, J.G., Pfennig, D.W., 2007. Patterns and Power of Phenotypic Selection in Nature. *BioScience* 57, 561–572. <https://doi.org/10.1641/B570706>

- Lavergne, S., Garnier, E., Debussche, M., 2003. Do rock endemic and widespread plant species differ under the LeafHeightSeed plant ecology strategy scheme? *Ecology Letters* 6, 398–404. <https://doi.org/10.1046/j.1461-0248.2003.00456.x>
- 455 Leitão, R.P., Zuanon, J., Villéger, S., Williams, S.E., Baraloto, C., Fortunel, C., Mendonça, F.P., Mouillot, D., 2016. Rare species contribute disproportionately to the functional structure of species assemblages. *Proceedings of the Royal Society B: Biological Sciences* 283, 20160084. <https://doi.org/10.1098/rspb.2016.0084>
- Lhotsky, B., Kovács, B., Ónodi, G., Csecserits, A., Rédei, T., Lengyel, A., Kertész, M., Botta-Dukát, Z.,
460 2016. Changes in assembly rules along a stress gradient from open dry grasslands to wetlands. *Journal of Ecology* 104, 507–517. <https://doi.org/10.1111/1365-2745.12532>
- Mace, G.M., Gittleman, J.L., Purvis, A., 2003. Preserving the Tree of Life. *Science* 300, 1707–1709. <https://doi.org/10.1126/science.1085510>
- Maitner, B.S., Boyle, B., Casler, N., Condit, R., Donoghue, J., Durán, S.M., Guaderrama, D., Hinchliff,
465 C.E., Jørgensen, P.M., Kraft, N.J.B., McGill, B., Merow, C., Morueta-Holme, N., Peet, R.K., Sandel, B., Schildhauer, M., Smith, S.A., Svenning, J.-C., Thiers, B., Violle, C., Wiser, S., Enquist, B.J., 2018. The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods in Ecology and Evolution* 9, 373–379. <https://doi.org/10.1111/2041-210X.12861>
- Marchese, C., 2015. Biodiversity hotspots: A shortcut for a more complicated concept. *Global Ecology and*
470 *Conservation* 3, 297–309. <https://doi.org/10.1016/j.gecco.2014.12.008>
- Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., Kulbicki, M., Lavergne, S., Lavorel, S., Mouquet, N., Paine, C.E.T., Renaud, J., Thuiller, W., 2013. Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. *PLOS Biology* 11, e1001569. <https://doi.org/10.1371/journal.pbio.1001569>
- 475 Mouillot, D., Villéger, S., Parravicini, V., Kulbicki, M., Arias-González, J.E., Bender, M., Chabanet, P., Floeter, S.R., Friedlander, A., Vigliola, L., Bellwood, D.R., 2014. Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. *PNAS* 111, 13757–13762. <https://doi.org/10.1073/pnas.1317625111>
- Murray, B.R., Thrall, P.H., Gill, A.M., Nicotra, A.B., 2002. How plant life-history and ecological traits
480 relate to species rarity and commonness at varying spatial scales. *Austral ecology* 27, 291–310.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots

- for conservation priorities. *Nature* 403, 853–858. <https://doi.org/10.1038/35002501>
- Naeem, S., Duffy, J.E., Zavaleta, E., 2012. The Functions of Biological Diversity in an Age of Extinction. *Science* 336, 1401–1406. <https://doi.org/10.1126/science.1215855>
- 485 Ohlemüller, R., Anderson, B.J., Araújo, M.B., Butchart, S.H.M., Kudrna, O., Ridgely, R.S., Thomas, C.D., 2008. The coincidence of climatic and species rarity: High risk to small-range species from climate change. *Biology Letters* 4, 568–572. <https://doi.org/10.1098/rsbl.2008.0097>
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2018. *Vegan: Community*
 490 *Ecology Package*.
- Ordóñez, A., Svenning, J.-C., 2016. Strong paleoclimatic legacies in current plant functional diversity patterns across Europe. *Ecology and Evolution* 6, 3405–3416. <https://doi.org/10.1002/ece3.2131>
- Osorio, F., Vallejos, R., 2014. *SpatialPack: Package for analysis of spatial data*.
- Owens, H.L., Guralnick, R., 2019. *climateStability: An R package to estimate climate stability from time-*
 495 *slice climatologies*. *Biodiversity Informatics* 14, 8–13. <https://doi.org/10.17161/bi.v14i0.9786>
- Pavoine, S., Ollier, S., Dufour, A.-B., 2005. Is the originality of a species measurable? *Ecology Letters* 8, 579–586. <https://doi.org/10.1111/j.1461-0248.2005.00752.x>
- Pollock, L.J., Thuiller, W., Jetz, W., 2017. Large conservation gains possible for global biodiversity facets. *Nature* 546, 141–144. <https://doi.org/10.1038/nature22368>
- 500 R Core Team, 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Redding, D.W., Mooers, A.O., 2015. Ranking Mammal Species for Conservation and the Loss of Both Phylogenetic and Trait Diversity. *PLoS ONE* 10, e0141435. <https://doi.org/10.1371/journal.pone.0141435>
- Redding, D.W., Mooers, A. Ø., 2006. Incorporating Evolutionary Measures into Conservation Prioritization. *Conservation Biology* 20, 1670–1678. <https://doi.org/10.1111/j.1523-1739.2006.00555.x>
 505
- Smithson, M., Verkuilen, J., 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11, 54–71. <https://doi.org/10.1037/1082-989X.11.1.54>
- Stein, A., Gerstner, K., Kreft, H., 2014. Environmental heterogeneity as a universal driver of species richness
 510 across taxa, biomes and spatial scales. *Ecology Letters* 17, 866–880. <https://doi.org/10.1111/ele.12277>

- Šímová, I., Rueda, M., Hawkins, B.A., 2017. Stress from cold and drought as drivers of functional trait spectra in North American angiosperm tree assemblages. *Ecology and Evolution* 7, 7548–7559. <https://doi.org/10.1002/ece3.3297>
- Trabucco, A., Zomer, R.J., 2019. Global High-Resolution Soil-Water Balance. <https://doi.org/10.6084/m9.figshare.7707605.v3>
- 515 Tucker, C.M., Aze, T., Cadotte, M.W., Cantalapiedra, J.L., Chisholm, C., Díaz, S., Grenyer, R., Huang, D., Mazel, F., Pearse, W.D., Pennell, M.W., Winter, M., Mooers, A.O., 2019. Assessing the utility of conserving evolutionary history. *Biological Reviews* 94, 1740–1760. <https://doi.org/10.1111/brv.12526>
- Tucker, C.M., Davies, T.J., Cadotte, M.W., Pearse, W.D., 2018. On the relationship between phylogenetic diversity and trait diversity. *Ecology* 99, 1473–1479. <https://doi.org/10.1002/ecy.2349>
- 520 Vane-Wright, R.I., Humphries, C.J., Williams, P.H., 1991. What to protect? Systematics And the agony of choice. *Biological Conservation* 55, 235–254. [https://doi.org/10.1016/0006-3207\(91\)90030-D](https://doi.org/10.1016/0006-3207(91)90030-D)
- Vasconcelos, T.S., da Silva, F.R., dos Santos, T.G., Prado, V.H.M., Provete, D.B., 2019. Geographical Patterns of Functional Diversity of South American Anurans, in: Vasconcelos, T.S., da Silva, F.R., dos Santos, T.G., Prado, V.H.M., Provete, D.B. (Eds.), *Biogeographic Patterns of South American Anurans*. Springer International Publishing, Cham, pp. 107–123. https://doi.org/10.1007/978-3-030-26296-9_5
- 525 Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N.J.B., Cadotte, M.W., Livingstone, S.W., Mouillot, D., 2017. Functional Rarity: The Ecology of Outliers. *Trends in Ecology & Evolution* 32, 356–367. <https://doi.org/10.1016/j.tree.2017.02.002>
- 530 Westoby, M., 1998. A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil* 199, 213–227. <https://doi.org/10.1023/A:1004327224729>
- Westoby, M., Falster, D.S., Moles, A.T., Vesk, P.A., Wright, I.J., 2002. PLANT ECOLOGICAL STRATEGIES: Some Leading Dimensions of Variation Between Species. *Annual Review of Ecology and Systematics* 33, 125–159. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150452>
- 535 Zanne, A.E., Pearse, W.D., Cornwell, W.K., McGlinn, D.J., Wright, I.J., Uyeda, J.C., 2018. Functional biogeography of angiosperms: Life at the extremes. *New Phytologist* 218, 1697–1709. <https://doi.org/10.1111/nph.15114>

4

Raretés & originalités écologiques dans la conservation

Dans le chapitre précédent, nous avons présenté le cadre d'originalité fonctionnelle et étudié certains de ces déterminants. Les espèces originales fonctionnellement ont une contribution importante au fonctionnement des écosystèmes. Pour garantir la stabilité de ceux-ci, il est donc important de protéger les espèces originales. Les objectifs de ce chapitre sont donc (i) d'évaluer dans quelle mesure les politiques de conservation sont déjà liées, implicitement, à l'originalité fonctionnelle; (ii) d'identifier des points chauds d'originalité fonctionnelle qui pourraient attirer l'attention des politiques de conservation dans les cas où il différerait des points chauds de richesse spécifique; (iii) enfin d'étudier si des menaces pèsent sur les espèces originales en particulier, par rapport aux espèces communes.

La première partie de ce chapitre propose de décliner les deux premiers objectifs en étudiant une base de données mondiale de poissons de récifs coralliens. Cette partie a fait l'objet d'une publication dans *Biological Conservation*.

La deuxième partie répond aux trois objectifs du chapitre, en se concentrant sur les mammifères et les oiseaux en utilisant un jeu de données mondial. Dans cette partie, nous évaluons les menaces qui peuvent peser sur les espèces originales, en évaluant différents facteurs d'impact humain. Nous quantifions ensuite l'impact probable du changement climatique sur les espèces originales par rapport aux espèces communes. L'article issu de cette partie est en révision pour *Nature Ecology and Evolution*.

I POINTS CHAUDS DE RARETÉ FONCTIONNELLE ET PRISE EN COMPTE DE L'ORIGINALITÉ FONCTIONNELLE DANS LA CONSERVATION

Biological Conservation 226 (2018) 288–299



ELSEVIER

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Functional rarity of coral reef fishes at the global scale: Hotspots and challenges for conservation

Matthias Grenié^{a,*}, David Mouillot^b, Sébastien Villéger^b, Pierre Denelle^a, Caroline M. Tucker^c, François Munoz^d, Cyrille Violle^a^a CEFE, Univ. Montpellier, CNRS, EPHE, IRD, Univ. Paul Valéry Montpellier 3, Montpellier, France^b MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Montpellier, France^c Department of Biology, University of North Carolina at Chapel Hill, Coker Hall, CB #3280 120 South Road, Chapel Hill, NC 27599-3280, United States of America^d University Grenoble-Alpes, LECA, 2233 Rue de la Piscine, 38041 Grenoble Cedex 9, France

ARTICLE INFO

Keywords:

Functional distinctiveness
Evolutionary distinctiveness
Coral triangle
Funrar
Biodiversity facet

ABSTRACT

Characterizing functional diversity has become central in ecological research and for biodiversity assessment. Understanding the role of species with rare traits, i.e. functionally rare species, in community assembly, ecosystem dynamics and functioning has recently gained momentum. However, functional rarity is still ignored in conservation strategies.

Here, we quantified global functional and evolutionary rarity for 2073 species of coral reef fishes and compared the rarity values to IUCN Red List status. Most species were functionally common but geographically rare. However, we found very weak correlation between functional rarity and evolutionary rarity. Functional rarity was highest for species classified as not evaluated or threatened by the IUCN Red List. The location of functional rarity hotspots (Tropical Eastern Pacific) did not match hotspots of species richness and evolutionary distinctiveness (Indo-Australian Archipelago), nor the currently protected areas. We argue that functional rarity should be acknowledged for both species and site prioritization in conservation strategies.

1. Introduction

Current conservation plans and policy decisions largely rely on the evaluation of taxonomic diversity, at both local and large scales (Davies et al., 2017; Hidas-Neto et al., 2013; Rosenfeld, 2002; Thuiller et al., 2015). For example, current protected areas mostly match with biodiversity hotspots defined by species number for a given area (Brum et al., 2017; Davies et al., 2017; Devictor et al., 2010; Pollock et al., 2017). At the species level, one major conservation tool is the Red List of the International Union for Conservation of Nature (IUCN) (Rodrigues et al., 2006), which still ignores the diversity of biological attributes (traits and genes) supported by species (Vié et al., 2009). However, assessing the diversity of organismal evolutionary history has become central to biodiversity monitoring and analysis (Cadotte et al., 2013), as well as accounting for organismal traits that are linked to species performance and niche axes (Díaz and Cabido, 2001; Stuart-Smith et al., 2013; Violle et al., 2007; Violle and Jiang, 2009), so to species coexistence and ecosystem dynamics (McGill et al., 2006; Violle et al., 2007) but also to species functional distinctiveness or uniqueness (Chase, 2013; Godet et al., 2015; Jain et al., 2014; Mouillot et al., 2013; Murray

et al., 2002; Pavoine et al., 2005; Violle et al., 2017b). In this context, it appears urgent to bring organismal functions – e.g., using functional traits as proxies (Díaz and Cabido, 2001; Violle et al., 2007) – into the global conservation agenda (Cadotte et al., 2011; Naeem et al., 2012; Pollock et al., 2017).

Conservation efforts usually focus on subsets of emblematic or threatened species and habitats due to limited resources (Arponen, 2012). With the ongoing extinction crisis (Ceballos et al., 2017), much emphasis has been put on prioritizing the protection of rare species and their habitat (Cofre and Marquet, 1999). Species' population size, geographical range and habitat breadth together define several forms of rarity reflecting the potential vulnerability of species to demographic extinctions and habitat alteration (Rabinowitz, 1981). More recently, phylogenetic relationships have been included in species rarity assessment (Isaac et al., 2007; Redding et al., 2010). In the same vein, Violle et al. (2017b) have recently proposed a framework to define functional rarity, accounting for both distinctiveness of species trait values and species geographical range size. After considering a pool of species and a spatial scale, the functional rarity of a species can be quantified based on both the distinctiveness of its trait values compared to the other

* Corresponding author.

E-mail address: matthias.grenie@cefe.cnrs.fr (M. Grenié).

species of the pool and its geographical range compared to the geographic extent of the study. In this framework, a species can be geographically rare because of its small range size, but at the same time common in terms of its traits. Conversely, a species can be geographically common with a large range size, but rare in terms of traits. The same framework can also apply to define phylogenetic rarity which is already included in species rarity assessment (Isaac et al., 2007; Redding et al., 2010). Functional rarity as defined by Violle and colleagues mirrors the concept of functional redundancy (Ricotta et al., 2016; Yachi and Loreau, 1999). Functional redundancy is the fact that a group of individuals (or species) at a given site share similar functions, thus being redundant. While, functional redundancy focuses at the site-level on shared functions, functional rarity concentrates on unshared functions at the species-level.

Previous works only examined the amount of phylogenetic rarity in biodiversity hotspots covered by protected areas (Mouillot et al., 2016), as it represents a crucial evolutionary heritage and potential for adaptation (Forest et al., 2007), but ignored functional rarity. When they did include functional rarity, phylogenetic and functional rarity did not necessarily match between species (Violle et al., 2017b; Winter et al., 2013). For example, carnivores at a global scale have concordant functional and phylogenetic rarity values (Dalerum, 2013), allowing consistent strategies to conserve both dimensions. Yet, other studies have reported a decoupling of taxonomic, phylogenetic and functional rarities across space (Gonçalves-Souza et al., 2014; Hidas-Neto et al., 2013, 2015). If a decoupling between the different dimensions of rarity emerges at a global scale, we need to reassess whether current conservation practices, and especially protected areas, cover the hotspots of functional, phylogenetic and taxonomic rarity (Pollock et al., 2017). In addition to local and regional conservation initiatives, a global species-based assessment of functional rarity can pinpoint areas or taxa that need further protection (Mouillot et al., 2016; Pollock et al., 2017).

Coral reef ecosystems are the most diverse marine systems, containing about 30% of all multicellular marine species (Fisher et al., 2015). They provide important ecosystems functions and services, including nitrogen cycling and larvae nurseries, as well as fisheries and shoreline protection (Harborne et al., 2006; Moberg and Folke, 1999). Coral reef ecosystems are currently facing multiple threats (Bellwood et al., 2004; Hoegh-Guldberg et al., 2007; Hughes et al., 2018). They notably experience massive, rapid and repeated coral bleaching events (Hughes et al., 2018; van Hooidonk et al., 2014; Wilson et al., 2006). These disturbances lead to a decrease in fish diversity and abundance potentially leading to a loss of key ecosystem functions and services (Pratchett et al., 2011; Rogers et al., 2018). Indeed, some fish species have unique combinations of traits (Mouillot et al., 2013) or perform unique functions on reefs (Bellwood et al., 2006), and as such deserve conservation attention. Furthermore, coral reef fishes show high functional vulnerability (Mouillot et al., 2014; Parravicini et al., 2014), because many combinations of functional traits are supported by only few species. However, the global geographic distribution of functional rarity is virtually unknown for coral reef fishes. The density of functionally rare species may simply match with species richness hotspots, such as the Coral Triangle, owing to a sampling effect, or it may peak in other areas where coral reef fishes have an original evolutionary history. Here we examined the global distribution of functional rarity in coral reef fishes and how it relates to their evolutionary rarity, to IUCN Red List Status and conservation efforts.

2. Material and methods

2.1. Fish occurrences

We used a worldwide database of reef fish presence-absence across tropical oceans (Parravicini et al., 2013), aggregated from almost 500 references (Kulbicki et al., 2013). The initial data contained both tropical and coral reefs, however, we limited our study to locations

containing coral reefs for easier comparison between sites, ending with 259 cells of 5° by 5° (corresponding to ~555 km by 555 km at the Equator, see Fig. A1 for details on data selection). We selected 2073 species (over 6300 reported coral reef fish species) for which functional trait information and the position on a phylogenetic tree were available. These species belong to 19 fish families (over 169 occurring on coral reefs) that are dominant on coral reefs (Kulbicki et al., 2013). Our sub-sampling of species compared to the full range of species did not change the relative species richness across cells (see Fig. A2, Spearman's rho = 0.99, p < 0.001).

2.2. Trait data

We considered six categorical functional traits representing the ability of tropical-reef fish species to forage resources and to reproduce (Mouillot et al., 2014; Villéger et al., 2017, see Appendix B for detailed links between traits and functions). Fish size (total body length) was coded using six ordered categories: 0–7 cm, 7.1–15 cm, 15.1–30 cm, 30.1–50 cm, 50.1–80 cm, and > 80 cm of body length. Mobility was coded using three ordered categories: sedentary (including territorial species), mobile within a reef, and mobile between reefs. The period of activity was coded using three ordered categories: diurnal, both diurnal and nocturnal, and nocturnal. Schooling was coded using five ordered categories: solitary, pairing, or living in small (3–20 individuals), medium (20–50 individuals), or large (> 50 individuals) groups. Vertical position in the water column was coded using three ordered categories: benthic, benthopelagic, and pelagic. We defined seven trophic categories based on the main items consumed by each species: herbivorous-detritivorous (i.e., fish feeding on turf or filamentous algae and/or undefined organic material), macro-algal herbivorous (i.e., fish eating large fleshy algae and/or seagrass), invertivorous targeting sessile invertebrates (i.e., corals, sponges, ascidians), invertivorous targeting mobile invertebrates (i.e., benthic species such as crustaceans), planktivorous (i.e., fish eating small organisms in the water column), piscivorous (including fish and cephalopods), and omnivorous (i.e., fish for which both vegetal and animal material are important in their diet).

2.3. Fish IUCN status

We used the `taxize` package v.0.7.8 (Chamberlain et al., 2016; Chamberlain and Szöcs, 2013) to retrieve up-to-date IUCN status for fishes (IUCN, 2017). For easier interpretation, we grouped species into four categories depending on their threat level: Critically Endangered (CR), Endangered (EN) and Vulnerable (VU) species as “Threatened” (n = 56); Least Concern (LC) and Near Threatened (NT) species as “Not Threatened” (n = 1011); Data Deficient (DF) species as “Data Deficient” (n = 120) and species without known status as “Not Evaluated” species (n = 886). We found no difference (p = 0.11, Fisher's exact test for count data) in the proportion of species in each threat group between our subset of 2073 species and the 6300 known species of coral reef fishes (Fig. A3). The proportions of species in each status differed between the Atlantic and Indo-Pacific realms ($\chi^2 = 121$, d.f. = 3, p < 0.001) (Fig. A4).

2.4. Functional rarity components

We computed Gower's pairwise distances between species because we had both ordinal and nominal traits (Gower, 1971). We used the `dist.ktab()` function in `ade4` v.1.7-6 to compute the distances (Dray and Dufour, 2007; Pavoine et al., 2009). We computed rarity components based on the framework of Violle et al. (2017b) and the `funrar` package v.1.2.0 (Grenié et al., 2017; Violle et al., 2017a). We estimated functional distinctiveness D_i of species i using functional dissimilarities, representing how rare the traits of a given species are compared to all the other species globally as:

$$D_i = \frac{\sum_{j=1, j \neq i}^N d_{ij}}{N-1}, \quad (1)$$

where d_{ij} was the functional distance between species i and j and N the total number of species in the species pool. The functional distances d_{ij} were scaled between 0 and 1. D_i captured how different are the traits, on average, of a given species compared to the rest of the species pool. D_i was equal to one when species i was maximally different to the other species. We tested the influence of the species pool definition from which functional distinctiveness was extracted. To this aim, we split the cells into two realms—Atlantic and Pacific—and estimated species distinctiveness values separately in each realm. There was a strong correlation between realm-specific and global distinctiveness values (see Fig. A5; Spearman's rho = 0.82, $p < 0.001$ for Atlantic; Spearman's rho = 1, $p < 0.001$ for Indo-Pacific).

We performed a sensitivity analysis to check whether specific traits had strong influences on functional distinctiveness values: we left out each trait one by one and recomputed functional distinctiveness values using the five remaining traits. The recomputed functional distinctiveness values correlated strongly with functional distinctiveness obtained from all the traits (Fig. A6, Spearman's rho > 0.86 for all the combinations).

We defined geographical restrictedness of species i , R_i , as:

$$R_i = 1 - \frac{K_i}{K_{tot}}, \quad (2)$$

where K_i was the number of cells where species i was present and K_{tot} the total number of cells. R_i was close to one for a species present at a single site only, and was equal to zero for a species present in all sites over the whole geographic range considered (here all coral reefs cells). Because each cell did not contain the same habitat area, we also computed restrictedness acknowledging relative reef area, i.e., K_i was then the cumulated habitat area in all the cells where species i was present and K_{tot} the sum of habitat area in all cells. The correlation between geographic restrictedness estimated on coral reefs and on coral plus tropical reefs was high (Fig. A7, Spearman's rho = 0.98, $p < 0.001$). Estimating the range of species based on coral reef cells only marginally changed the estimate of relative range size between species.

We combined the two indices into a synthetic index of functional rarity (FR), to rank the functionally rarest species:

$$FR_i = \frac{D_{i,sc} + R_{i,sc}}{2}, \quad (3)$$

where $D_{i,sc}$ was the functional distinctiveness of species i scaled between zero and one and $R_{i,sc}$ the restrictedness of species scaled between zero and one. FR_i weighs equally species with very distinct traits (high D_i) and species present in a few cells (high R_i). A species with FR_i close to one was both very restricted in geographic distribution (high $R_{i,sc}$) and had very distinct traits compared to the rest of the species pool (high D_i).

2.5. Evolutionary distinctiveness

We used the supertree of coral reef fish taxa provided by Leprieur et al. (2016). Even though this tree comprises a large number of polytomies, it is one of the most exhaustive phylogenetic tree available for coral reef fishes, and as such, is well-suited to study global patterns of fish functional rarity. Based on the phylogenetic relationships among taxa, we computed their Evolutionary Distinctiveness (Isaac et al., 2007). The Evolutionary Distinctiveness of species i , ED_i , is high when the species has a long unshared branch length with all the other species. The more “isolated” a species is in a phylogenetic tree, the higher its evolutionary distinctiveness. We computed ED using the `picante` package (Kembel et al., 2010) v.1.6-2 and the fair proportion option following Isaac et al. (2007).

We computed an evolutionary equivalent of FR_i for species i ,

denoted as Evolutionary Rarity ER_i :

$$ER_i = \frac{ED_{i,sc} + R_{i,sc}}{2}, \quad (4)$$

where $ED_{i,sc}$ was the scaled evolutionary distinctiveness between zero and one of species i , and $R_{i,sc}$ the scaled geographic restrictedness between zero and one of species i . ER_i weighs equally species that are highly evolutionary distinct (high ED_i) and species present in a small number of cells (high R_i). A species with ER_i close to one was both very restricted in geographic distribution (high $R_{i,sc}$) and very distinct in phylogenetic history, having a very long branch length separating it from the rest of the species pool (high $ED_{i,sc}$). A species with intermediate values of both ED_i and R_i could also have a high ER_i value.

2.6. Statistical analyses and maps

All analyses were carried out using R (R Core Team, 2017) v.3.4.1.

A Principal Coordinates Analysis (PCoA) identified orthogonal dimensions determining the variation in functional distances (`pcoa()` function in the `ape` package v.4.1, Paradis et al., 2004). We assessed the contribution of the elementary traits to these dimensions, and represented how the functional rarity components varied along the dimensions. The axes were selected as their explained variance was greater to the null expectation of a broken stick model (Legendre and Legendre, 2012). We found no systematic bias in terms of occupation of the total functional trait space between our subset of species and the global set of 6316 tropical reef species (see Fig. A8, Spearman's = 0.98, $p < 0.001$).

To illustrate rarity hotspots and coldspots on the map we ranked all the species according to each above-mentioned index, and we counted the number of species per cell that were in the global top 10% (highest 10%) for each index. Because choosing a 10% threshold is still arbitrary, we also performed the analyses with 5% and 15% thresholds, but it did not change the results qualitatively (see Fig. A9). The number of top 10% evolutionary rare species at a site can be related to phylogenetic endemism (PE) as defined by Rosauer et al. (2009). We did find a strong correlation between this number of top 10% ER species and PE (see Fig. A11).

In order to assess to what extent the differences in the number of top-ranked species were primarily driven by species richness differences between cells we used null models (Gotelli and Graves, 1996). We used the curveball algorithm (Strona et al., 2014) from the `nullmodel()` function in the `vegan` package v.2.5-2 (Oksanen et al., 2018). The curveball algorithm randomizes the presences of species across cells while keeping the same species richness at each cell, and the same number of occurrences for each species. We computed 2000 permutations and for each of them we counted the number of top 10% species at each cell. For each cell we then compared the observed values to the randomized ones using the Standard Effect Size score (SES) computed as the difference between the observed value and the mean of predicted values by the null model, divided by the standard deviation of predicted values:

$$SES_j(I) = \frac{I_{j,Obs} - \hat{I}_{j,Pred}}{sd(I_{j,Pred})} \quad (5)$$

where j is the cell, I the index of interest, $I_{j,Obs}$ the observed index at cell j , $\hat{I}_{j,Pred}$ the average of predicted indices at cell j and $sd(I_{j,Pred})$ the standard deviation of the predicted indices. The SES score measures how much the observed value is far from the values predicted by the null model accounting for species richness. If the number of top species in the cell was driven by species richness only, SES scores should be close to 0. On the contrary, an SES score with a great absolute value means that this null hypothesis is less probable (Gotelli and Graves, 1996).

We compared the obtained maps using a modified version of the t -

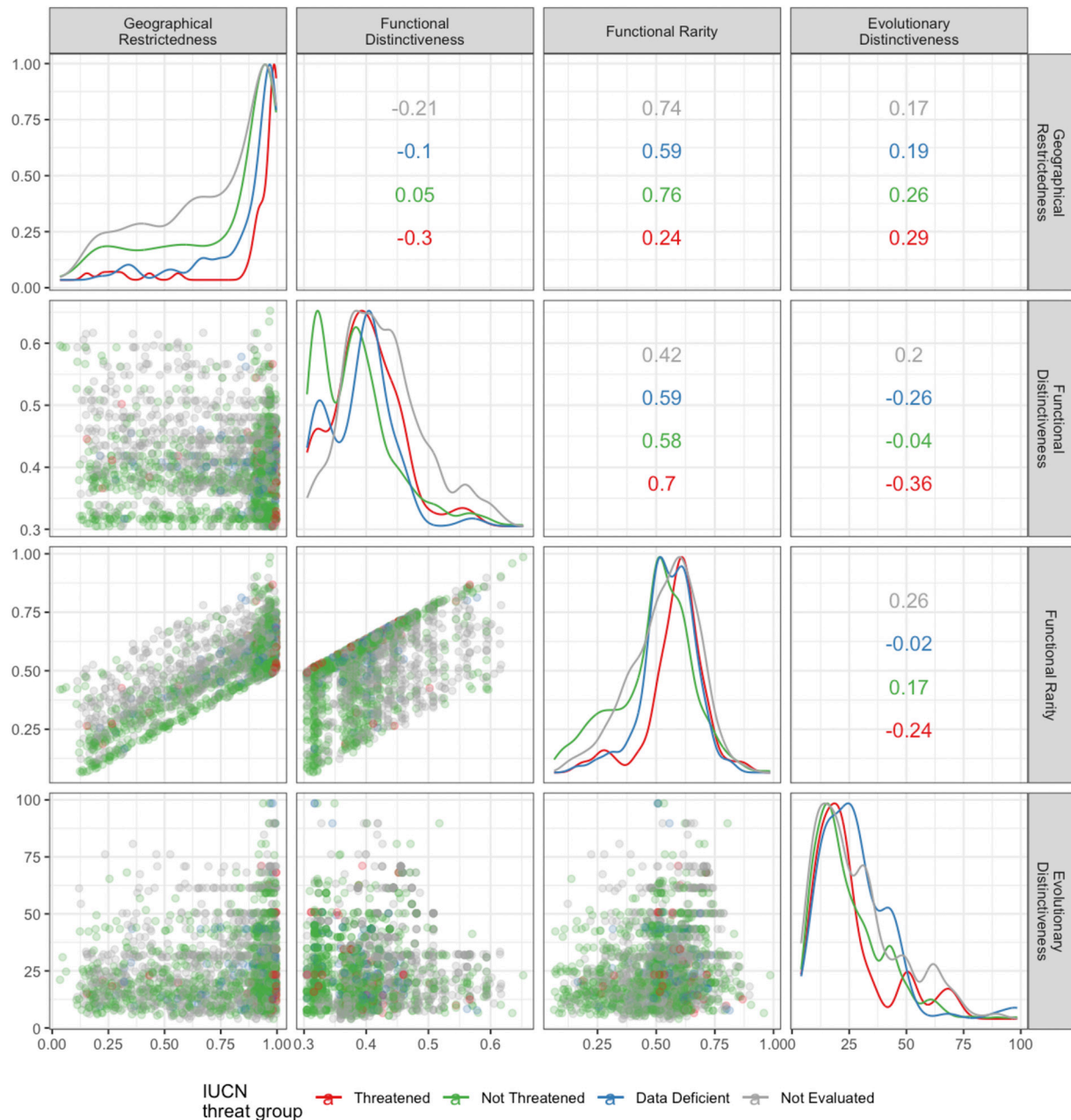


Fig. 1. Biplots showing the relationships between the various components of species rarity by IUCN status. The plots in the upper right triangle above the diagonal indicates the Spearman correlation between components for each IUCN threat group. The diagonal shows the density plot for each component per IUCN threat group. Plots in the lower left triangle under the diagonal are scatterplots linking two components with each dot representing a single species with color representing its IUCN threat group.

test to account for spatial autocorrelation (Dutilleul et al., 1993). We used the `SpatialPack` package v.0.2-3 (Osorio and Vallejos, 2014).

We performed a multiple comparison Kruskal-Wallis rank-sum test to compare rarity components across IUCN groups. We used the `kruskal()` function of the `agricolae` package v.1.2-6 (de Mendiburu, 2017). We also performed post-hoc tests using Fisher's least significant difference to differentiate between groups.

We calculated the Pagel's λ value of phylogenetic signal of species rarity components, using the `fitContinuous()` function in the `geiger` package v.2.0.6 (Harmon et al., 2008). Because Pagel's λ works only on binary trees, we resolved the polytomies randomly 30 times using the `multi2di()` function in the `ape` package v.4.1 (Paradis et al., 2004): the different λ values changed only by $< 0.1\%$.

3. Results

The distributions of functional distinctiveness and restrictedness were highly skewed (Fig. 1, panels on the diagonal), with most species being functionally common and geographically restricted, while evolutionary rarity showed a more symmetrical distribution. Functional distinctiveness was significantly but weakly correlated with both geographic restrictedness (Spearman's $\rho = -0.09$, $p < 0.001$) and evolutionary distinctiveness (Spearman's $\rho = -0.08$, $p < 0.001$). Geographic restrictedness showed the strongest correlation with evolutionary distinctiveness (Spearman's $\rho = 0.21$, $p < 0.001$).

The first two axes of the functional space explained 51.6% of trait variance among fishes (Fig. 2). The first axis differentiated big, large schooling and planktivorous species from small, solitary and

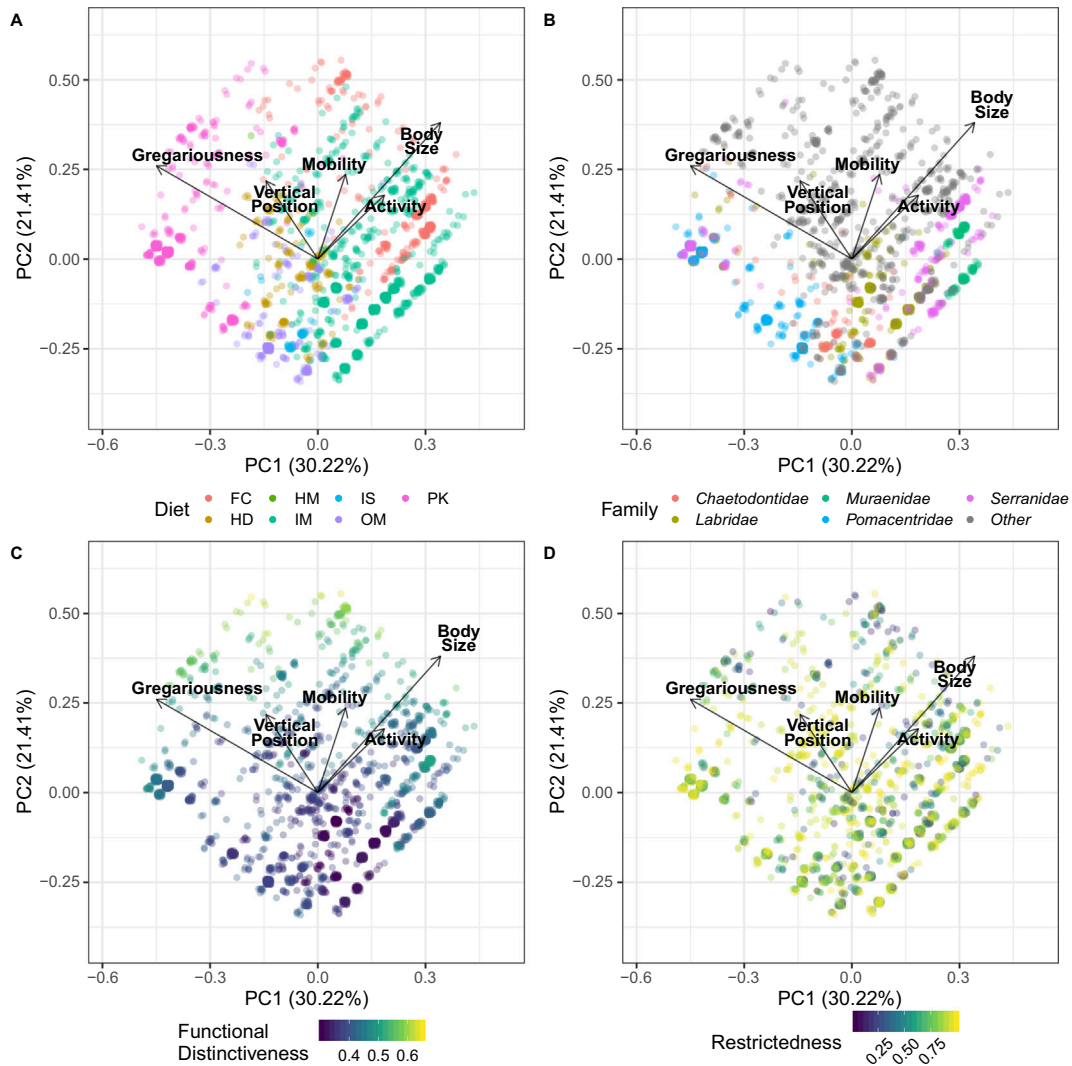


Fig. 2. Principal Coordinates Analysis (PCoA) representing the functional space of coral reef fishes based on their traits. The first two axes explain 51.6% of the variance in the distance matrix. (A) Colored by diet: FC, piscivorous; HM, macroalgal herbivorous; IS, invertivorous on sessile organisms; HD, herbivorous invertivorous; IM, invertivorous on mobile organism; OM, omnivorous; PK, planktivores (B) Colored with the five most numerous families (C) Colored by functional distinctiveness (D) Colored by geographical restrictedness.

omnivorous species. The second axis differentiated mobile species with highly specialized diet (planktivorous or piscivorous) from sedentary generalist (invertivorous, omnivorous) species (correlations between traits and axes in Table A1). We mapped the scores of the five most species-rich families on these axes (Fig. 2B). Because most species of these families were sedentary and benthic, they were located in the lower half of the functional space. Functional distinctiveness was correlated with the second axis (Fig. 2C) (Spearman's $\rho = 0.71$, $p < 0.001$), but not with the first axis (Spearman's $\rho = -0.02$, $p > 0.1$), i.e., species with large schools and large body size were functionally more distinct. Geographic restrictedness was weakly correlated with the two axes (Fig. 2D) (Axis 1: Spearman's $\rho = -0.05$, $p = 0.03$; Axis 2: Spearman's $\rho = -0.10$, $p < 0.001$), while functional rarity was weakly correlated with the first axis but more strongly with the second one (Axis 1: Spearman's $\rho = -0.09$, $p < 0.001$; Axis 2: Spearman's $\rho = -0.36$, $p < 0.001$). Evolutionary distinctiveness was weakly correlated with the first axis (Axis 1: Spearman's $\rho = 0.15$, $p < 0.001$; Axis 2: Spearman's $\rho = 0.03$, $p = 0.11$), while evolutionary rarity was correlated with neither axis (Axis 1: Spearman's $\rho = 0.01$, $p = 0.68$; Axis 2: Spearman's $\rho = -0.07$, $p = 0.002$).

Across IUCN threat groups, functional distinctiveness was the highest for Not Evaluated species, followed by Threatened and Data

Deficient species (Fig. 3A) (Kruskal-Wallis $\chi^2 = 262$, d.f. = 3, $p < 0.001$). Not Threatened species had lowest functional distinctiveness on average. Geographic restrictedness was the highest for Threatened species (Fig. 3B), followed by Data Deficient species, Not Threatened species and finally Not Evaluated species (Kruskal-Wallis $\chi^2 = 93$, d.f. = 3, $p < 0.001$). This was expected as geographic restrictedness is directly related to species range size which is a criterion for the IUCN assessment. Functional rarity (Fig. 3C) was the highest for Threatened and Data Deficient species (Kruskal-Wallis $\chi^2 = 58.9$, d.f. = 3, $p < 0.001$). Not Threatened species had the lowest functional rarity. Evolutionary distinctiveness was the highest for Data Deficient and Not Evaluated species (Kruskal-Wallis $\chi^2 = 18$, d.f. = 3, $p < 0.001$). Evolutionary rarity (Fig. 3D) was the highest for Threatened and Data Deficient species and the lowest for Not Threatened and Not Evaluated species (Kruskal-Wallis $\chi^2 = 18$, d.f. = 3, $p < 0.001$).

Bivariate relationships between rarity components showed very contrasted correlation coefficients between IUCN threat groups (Fig. 1). Functional distinctiveness and evolutionary distinctiveness were negatively correlated for Threatened and Data Deficient groups (Spearman's $\rho = -0.36$, $p = 0.006$; and Spearman's $\rho = -0.26$, $p = 0.004$ respectively), while this correlation was positive for Not Evaluated species (Spearman's $\rho = 0.2$, $p < 0.001$) and Not Threatened species showed

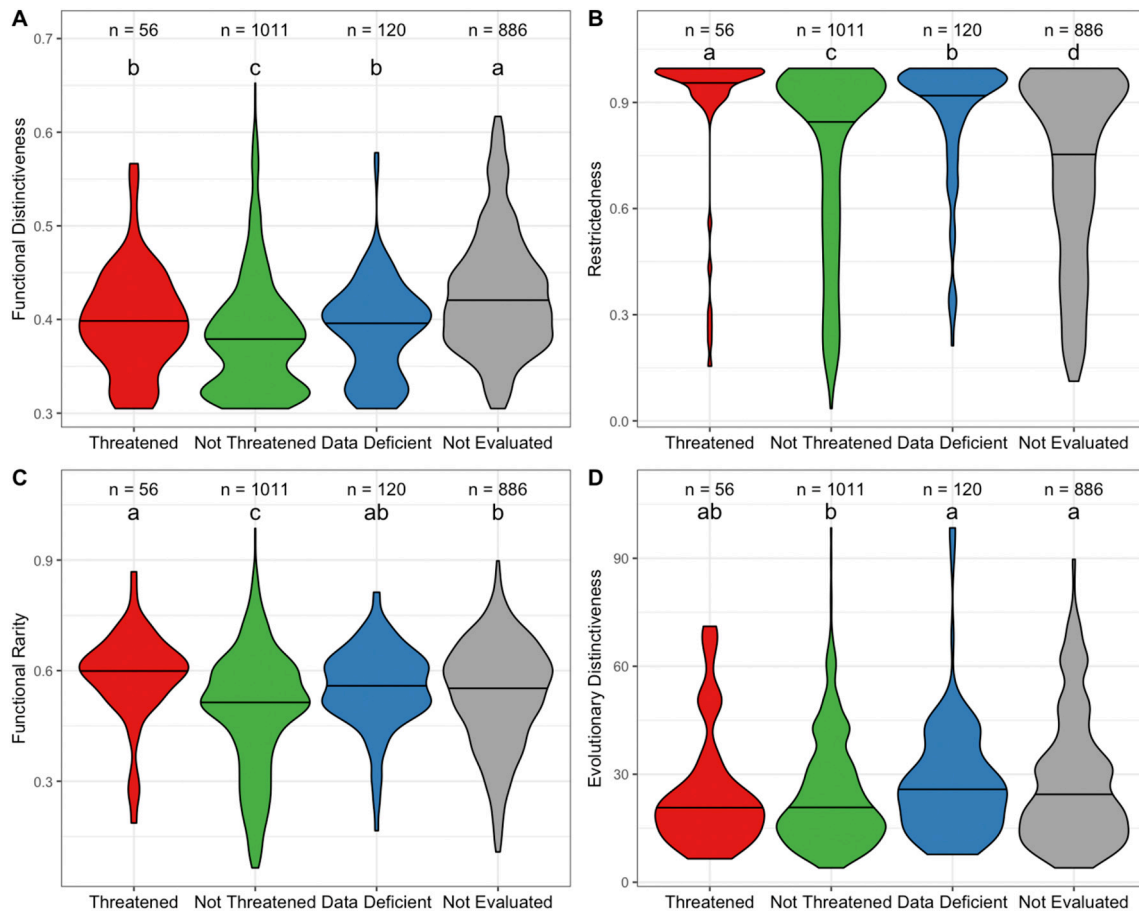


Fig. 3. Violin plots showing the smoothed distribution of species rarity components within each IUCN threat group: (A) Functional distinctiveness (B) Restrictedness (C) Functional rarity (D) Evolutionary distinctiveness. The solid line indicates the group median value. Letters indicate statistical sub-groups from a Kruskal-Wallis rank-sum test. Two groups that are significantly different have two different lower-case letters. Numbers above the violins indicate the number of species in each IUCN threat group.

no correlation (Spearman's $\rho = -0.04$, $p = 0.27$). Functional distinctiveness and geographic restrictedness also exhibited different correlation coefficients between groups: Threatened and Not Evaluated species had a moderate negative correlation (Spearman's $\rho = -0.30$, $p = 0.03$; Spearman's $\rho = -0.21$, $p < 0.001$), while Data Deficient and Not Evaluated species had no clear correlation (Spearman's $\rho = -0.10$, $p = 0.27$; Spearman's $\rho = 0.05$, $p = 0.11$).

Functional distinctiveness had a higher Pagel's λ value ($\lambda = 0.85$) than functional rarity and restrictedness ($\lambda = 0.70$ and $\lambda = 0.68$, respectively), while the phylogenetic structuring did not depart from a Brownian model of trait evolution. The randomly resolved polytomies had little effect on λ values. The two families with the highest functional rarity value were the *Caesionidae* and the *Carangidae*. *Caesionidae* had a mean functional rarity of 0.66, while other families had a mean of 0.4 (see Fig. 4). These families were functionally rare because most of their species live in large schools, have specialized diets (*Carangidae* are piscivores, while *Caesionidae* are planktivores), and are very mobile across reefs.

We counted the number of top 10% rare species in each 5° by 5° cell and examined the spatial congruence of rarity hotspots. Functional rarity hotspots poorly matched with evolutionary distinctiveness hotspots (Corrected Pearson's $\rho = 0.19$, $F = 1.18$, $p = 0.29$, see all tests in Table A2). The Indo-Australian Archipelago (IAA), as well as Eastern Africa, showed the highest number of top 10% functionally distinctive species (Fig. 5A), while the top 10% evolutionarily distinctive species (Fig. 5E) were concentrated in the IAA. However, the highest number of top 10% functionally rare species was located in the Tropical Eastern Pacific (Fig. 5C). The number of most evolutionarily rare species peaked

in the Caribbean as well as East China Sea and the South of Japan (Fig. 5G). When using null models, the IAA had a lower than expected number of top 10% functionally and evolutionarily rare species while reefs in the Tropical Eastern Pacific or along the East African coast host more top rarest species than expected given the number of species. The Arabic Peninsula, as well as the Tropical Eastern Pacific exhibited SES values over 3 for all indices, i.e. higher number of top 10% species than expected from the null models, while the Caribbean had SES values under -3 only for functional distinctiveness and around 3 for all the other indices.

There was a negative relationship between the protection percentage on a grid cell (using strict protection definition, see Mouillot et al., 2016) and the number of top 10% functionally distinct species (corrected Pearson's $\rho = -0.20$, $F = 7.86$, $p = 0.006$). Grid cells with the highest number of top 10% functionally distinct species were in areas with low protection coverage, such as the Tropical Eastern Pacific. Furthermore, there was no relation between protection coverage and functional rarity at the grid cell scale (corrected Pearson's $\rho = 0.02$, $F = 0.05$, $p = 0.83$).

4. Discussion

We used a global dataset on coral reef fishes to assess the geographic distribution of their functional and phylogenetic rarity, and the degree to which rarest species were under IUCN conservation status. Functional rarity correlates with specific trait values such as large body size and large schools (> 50 individuals). Functional rarity also reveals novel hotspots in areas that are less speciose and far from the Indo-

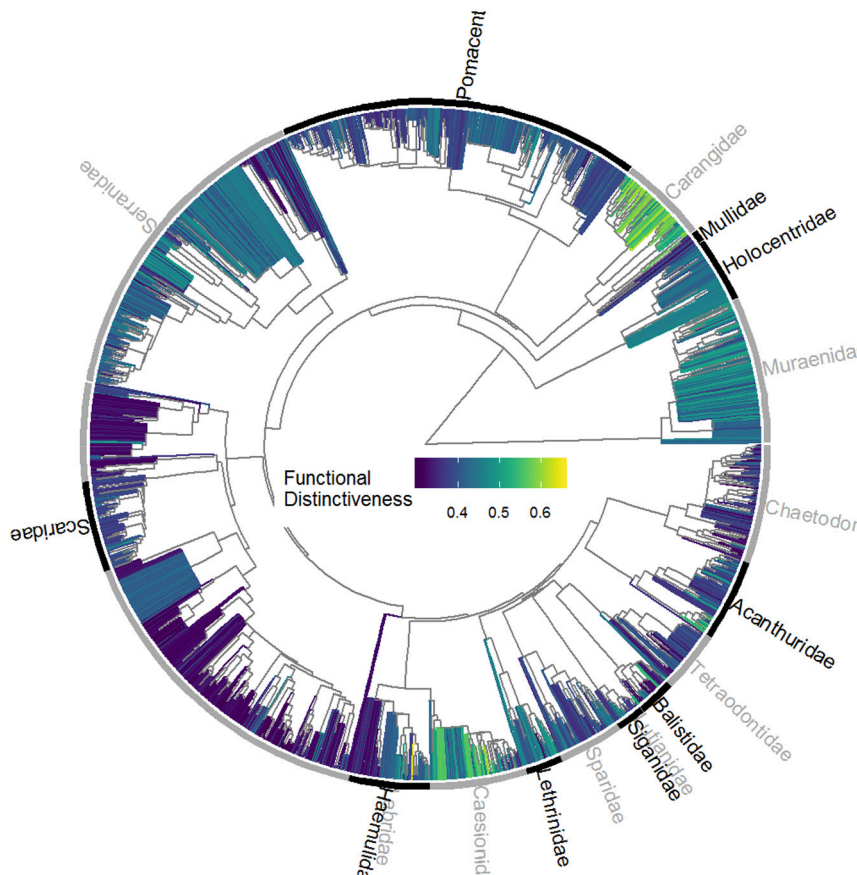


Fig. 4. Phylogenetic tree of coral reef fishes with functional distinctiveness values on terminal branches representing species. Family names are indicated on the outside ring. The colors of the families are only a visual cue to read the phylogenetic tree, they have no meaning per se. Grey and black labels may overlap because of non-monophyletic families. The extent of families is indicated by the colored rings. *Caesionidae* and *Carangidae* both have a higher mean functional distinctiveness than other families.

Australian Archipelago.

4.1. Functional and evolutionary rarity hotspots

Functional rarity hotspots are mainly located in the Tropical Eastern Pacific region and in the Caribbean, while evolutionary rarity hotspots are concentrated in the Indo-Australian Archipelago. The rarity hotspots, are consistent with what has already been shown according to the history of coral reef fish diversification and the influence of connectivity among regions (Cowman and Bellwood, 2013). Indeed, highly connected regions, such as the Indo-Australian Archipelago, host species that are less restricted in their distribution (low restrictedness) and thus exhibit lower functional rarity value. On the contrary, less connected regions, such as the Tropical Eastern Pacific or the Atlantic, have been isolated for longer (Pellissier et al., 2014). In addition, the extinction process in isolated regions is not balanced by colonization and induces lower diversity (Pellissier et al., 2014). Assuming random extinctions among all species, this lower diversity tends to increase functional distinctiveness, with less species having the same functional traits. All these combined factors make functional rarity to be higher in less connected regions, as in the Eastern Pacific. We must also take into account the isolation from coral reef Quaternary refugia, which explains patterns of diversification in the IAA (Pellissier et al., 2014). This refugia is located near the IAA, region exhibiting the highest species diversity, and thus lower overall functional differentiation, because of hosting more species in a similar trait space. Functional rarity hotspots could thus be explained by the history and biogeography of coral reef fishes.

4.2. Evolutionarily and functionally distinct species

Functional and evolutionary distinctiveness are correlated differently depending on the IUCN threat group. Furthermore, functional

rarity does not depart from a Brownian model of trait evolution across the phylogeny, which suggests that closely related species share a more similar functional rarity value than distantly related species.

The relationship between evolutionary (ED) and functional distinctiveness depends on the mode of trait evolution, the rates of speciation and extinction in a clade, and the interaction between these processes (Thuiller et al., 2015). Species lacking close relatives are likely to exhibit more distinct combinations of traits due to some unique evolutionary history. However, this is far from being a strict rule, as highlighted by the presence of both quite functionally ordinary (e.g. *Omylodon* rabbit) and quite distinct (e.g. northern hairy-nosed wombat) species on the mammal EDGE list (Isaac et al., 2007). Particular traits can impact evolutionary distinctiveness if they affect diversification rates. For example, some traits have been identified as driving higher extinction rates, making surviving taxa to be evolutionarily more distinct. For example, body size is often correlated with extinction risk (Davies et al., 2008; Lee and Jetz, 2011). Trait values can also be associated with differences in speciation rates between clades – for example, body size was found to be positively correlated with speciation rates in fishes (Rabosky et al., 2013). The influence of factors such as convergent evolution and niche conservatism on the relationship between ED and functional distinctiveness is still poorly explored. A comprehensive perspective on how functional and evolutionary distinctiveness relate across different taxa is necessary to guide conservation prioritization.

We used a recently published super-tree comprising over 2000 species of coral reef fishes (Leprieur et al., 2016) to assess ED. However, because this super-tree contains many polytomies at the genus level, the estimation of evolutionary distinctiveness may be biased. Indeed, polytomies induce high long terminal branches at the genus level, giving high ED scores to the sub-tending species. Fully resolved phylogenies would provide a better estimation of the link between functional and evolutionary distinctiveness but are not yet available.

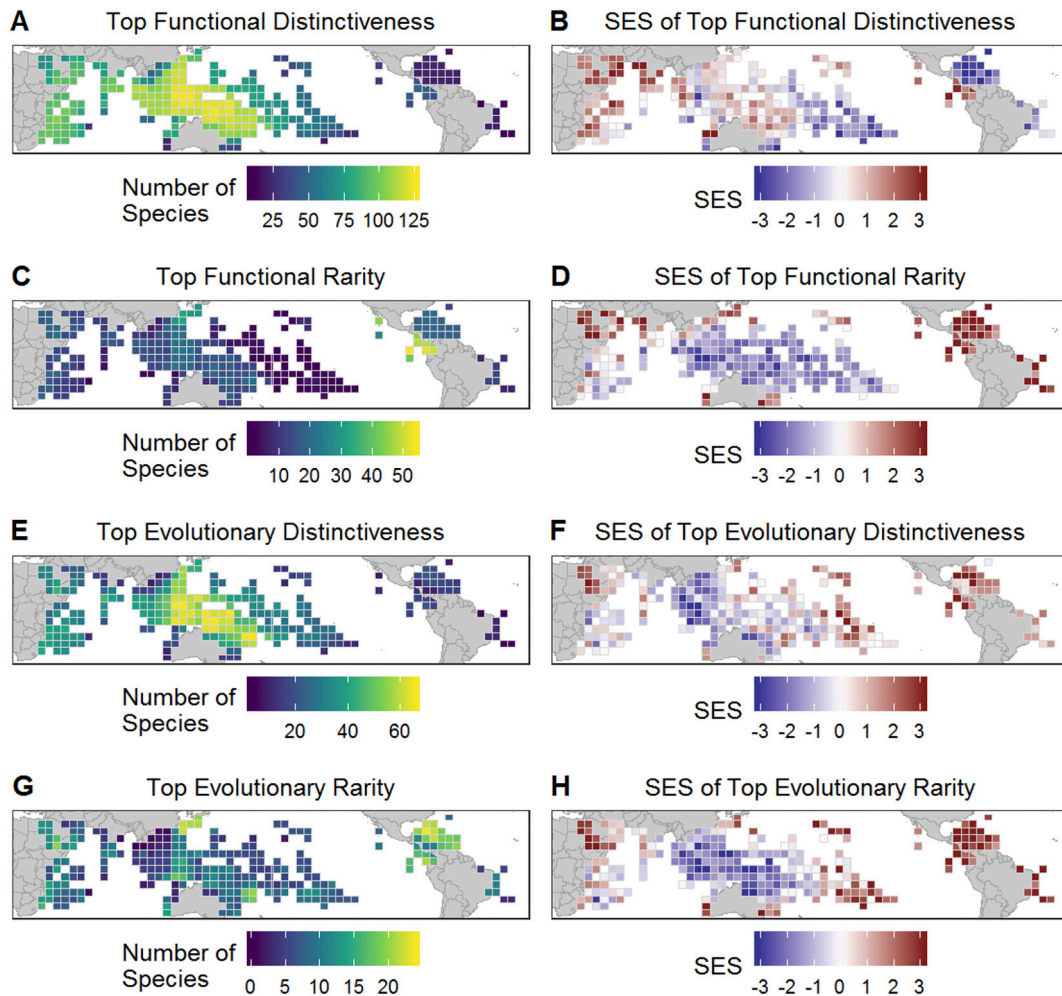


Fig. 5. Maps showing the global distribution of the number of top 10% coral reef fishes per cell for each species rarity component and the corresponding Standard Effect Size (SES) from a null model correcting for species richness. (A) Functional distinctiveness, (B) SES for functional distinctiveness; (C) Functional rarity, (D) SES for functional rarity; (E) Evolutionary distinctiveness sensu [Isaac et al., 2007](#), (F) SES for evolutionary distinctiveness; (G) Evolutionary rarity, (H) SES for evolutionary rarity. For the number of species in the top 10% (left column) for each component color scales linearly with the values and independently in each panel while color scale for SES values similarly across panels (right column).

4.3. Do we currently protect the functional rarity of coral reef fishes?

Geographic restrictedness is logically highest for Threatened species and lowest for Not Evaluated taxa. It confirms the importance of the Extent Of Occurrence (EOO) criteria when defining the threat status in the IUCN Red list. However, we also show that species not yet Evaluated by IUCN have higher Functional Rarity than species listed as “Not Threatened”, while Not Evaluated taxa have higher Distinctiveness than other categories. In addition, Data Deficient species is the second functionally rarest group (see [Fig. 3](#)). Since functionally rare species differ in traits compared to other species (low redundancy), they may support rare ecosystem functions that would not be supported by other species ([Mouillot et al., 2013](#)). Then some species among the Not Evaluated and Data Deficient groups are certainly ecologically unique in their roles while being ignored in conservation strategies. Functional rarity could then become a future prioritization criterion for IUCN within these groups ([Arponen, 2012](#)), see for example the top 20 functionally rarest species that are Data Deficient ([Table 2](#)).

Not Threatened species consistently show lower values of all considered rarity components than Threatened and Data Deficient species. However, there are several outlier taxa in the Not Threatened group. Eleven out of twenty species with the highest functional rarity come from the Not Threatened group ([Table 1](#)), and would deserve particular

conservation attention because of their low functional redundancy and possibly specific role. This finding validates the use of functional rarity as a new facet to identify species that are not threatened in terms of demography and geographic distribution but are still likely to support unique functions on coral reefs.

Threatened species tend to exhibit high functional rarity values, although it is not yet a conservation criterion. Furthermore, we find a negative relationship between the protection percentage on a given grid cell (using strict protection definition, see [Mouillot et al., 2016](#)) and the number of top 10% functionally distinct species present in that cell. Altogether these results suggest a spatial mismatch between conservation efforts and functional rarity. For instance, current conservation efforts cannot efficiently protect some of the functionally rarest coral reef fish families such as *Carangidae* or *Caesionidae*, since being highly mobile predators ([D'agata et al., 2016a](#)). Therefore these families would need alternative specific conservation measures like giant protected areas (e.g. Chagos), quotas or gear restrictions ([Graham et al., 2017](#); [Singleton and Roberts, 2014](#)).

Functional rarity is an appealing concept to study both applied and theoretical ecology but its computation must be handled with care. The significance of functional rarity critically depends on the traits and the spatial scale ([Violle et al., 2017a](#)). We selected functional traits representing key ecological strategies of coral reef fishes ([Mouillot et al., 2014](#); see Appendix B1 for trait-function details). We used a distance

Table 1
Top 20 species with highest functional rarity value. “–” means the species has no IUCN status.

Species	Functional rarity	Functional distinctiveness	Geographic restrictedness	IUCN threat group	IUCN status
<i>Xenistius californiensis</i>	0.986	0.652	0.969	Not Threatened	LC
<i>Lutjanus peru</i>	0.959	0.635	0.965	Not Threatened	LC
<i>Caranx hippos</i>	0.904	0.616	0.911	Not Threatened	LC
<i>Carangoides bartholomaei</i>	0.898	0.616	0.900	Not Evaluated	–
<i>Seriola quinqueradiata</i>	0.897	0.593	0.961	Not Evaluated	–
<i>Trachurus trachurus</i>	0.868	0.566	0.981	Threatened	VU
<i>Seriola peruana</i>	0.862	0.566	0.969	Not Threatened	LC
<i>Carangoides ruber</i>	0.859	0.589	0.900	Not Evaluated	–
<i>Caranx crysos</i>	0.859	0.589	0.900	Not Threatened	LC
<i>Pseudocaranx wrighti</i>	0.858	0.566	0.961	Not Evaluated	–
<i>Seriola hippos</i>	0.857	0.593	0.884	Not Evaluated	–
<i>Caesio suevica</i>	0.856	0.562	0.969	Not Evaluated	–
<i>Lutjanus novemfasciatus</i>	0.854	0.562	0.965	Not Threatened	LC
<i>Lutjanus viridis</i>	0.853	0.563	0.961	Not Threatened	LC
<i>Chloroscombrus chrysurus</i>	0.852	0.580	0.911	Not Threatened	LC
<i>Caranx caballus</i>	0.850	0.566	0.946	Not Threatened	LC
<i>Caesio striata</i>	0.848	0.558	0.965	Not Evaluated	–
<i>Decapterus punctatus</i>	0.842	0.562	0.942	Not Threatened	LC
<i>Pseudocaranx dinjerra</i>	0.841	0.544	0.988	Not Evaluated	–
<i>Caranx latus</i>	0.837	0.571	0.907	Not Threatened	LC

metric (Gower) to scale ordinal traits appropriately, functional dissimilarities—used to compute functional rarity indices—should be adapted to the nature of traits used (Pavoine et al., 2009). Also, the estimation of functional rarity depends on the spatial scale considered, here, the functional dissimilarities were considered only at global scale (but see Fig. A5), but finer scales (realms, regions or local) could be used and as such affect the results.

Furthermore, we studied coral reef fishes using presence-absence data, but if available, functional rarity components could be computed with abundance data (Violle et al., 2017a, 2017b). We also did a sensitivity analysis to ensure that the patterns of functional rarity we observed were not the result of a single trait driving the entire signal (see Fig. A6). Whether functional rarity is used in species assessments for conservation planning or for investigating fish biodiversity dynamics, the number and the nature of selected traits should be justified.

4.4. Functional rarity and insurance

Our study extends previous works on the lack of functional insurance in species-poor areas (Bender et al., 2017). Functional insurance describes the potential resilience in ecosystem functions that

results from redundancy, the fact that several species share the same characteristics in a given area and in the functional composition of a given ecosystem (Yachi and Loreau, 1999). Functional redundancy is a key component of functional insurance (Nyström, 2006). Functional redundancy is directly linked to functional distinctiveness in a community, because the higher the functional redundancy the lower the community-level functional distinctiveness (Ricotta et al., 2016). The most functionally distinct species are the least functionally redundant compared to the species pool. Thus sites harboring a high number of functionally distinct species may show the least functional redundancy such as in the Tropical Eastern Pacific. The functional groups with the least number of species may support non-redundant ecosystem functions across the Indo-Pacific region (D'agata et al., 2016b). Furthermore, vulnerability of fish species depends on their specific combination of traits; for example, large fishes are more susceptible to human impact (Mellin et al., 2016), or they can be more targeted by fishing depending on their trophic status, diet or size (Micheli et al., 2014). The vulnerability of ecosystem functions could also suffer from an ecological “inactivation threshold” under which, if a species has a lower abundance than the threshold, the function is not supported or active anymore (Soliveres et al., 2016). Non-linear effects between functional

Table 2
Top 20 Data Deficient species with the highest functional rarity value.

Species	Functional rarity	Functional distinctiveness	Geographic restrictedness	IUCN threat group	IUCN status
<i>Naso tuberosus</i>	0.813	0.578	0.842	Data Deficient	DD
<i>Naso reticulatus</i>	0.800	0.562	0.861	Data Deficient	DD
<i>Liopropoma longilepis</i>	0.737	0.471	0.992	Data Deficient	DD
<i>Orthopristis forbesi</i>	0.722	0.460	0.992	Data Deficient	DD
<i>Orthopristis cantharinus</i>	0.718	0.460	0.985	Data Deficient	DD
<i>Prionurus scalprum</i>	0.717	0.467	0.965	Data Deficient	DD
<i>Orthopristis lethopristis</i>	0.715	0.456	0.992	Data Deficient	DD
<i>Myripristis formosa</i>	0.713	0.458	0.981	Data Deficient	DD
<i>Boops lineatus</i>	0.690	0.451	0.958	Data Deficient	DD
<i>Gymnothorax serratidens</i>	0.686	0.436	0.992	Data Deficient	DD
<i>Uropterygius polystictus</i>	0.686	0.436	0.992	Data Deficient	DD
<i>Sparisoma griseorubrum</i>	0.682	0.431	0.996	Data Deficient	DD
<i>Epinephelus posteli</i>	0.678	0.446	0.950	Data Deficient	DD
<i>Cirrhilabrus brunneus</i>	0.671	0.438	0.958	Data Deficient	DD
<i>Paralabrax auroguttatus</i>	0.664	0.419	0.996	Data Deficient	DD
<i>Semicossyphus reticulatus</i>	0.660	0.427	0.965	Data Deficient	DD
<i>Stegastes otophorus</i>	0.658	0.420	0.981	Data Deficient	DD
<i>Canthigaster flavoreticulata</i>	0.658	0.417	0.988	Data Deficient	DD
<i>Paralabrax humeralis</i>	0.657	0.419	0.985	Data Deficient	DD
<i>Argyrops megalommatius</i>	0.655	0.422	0.969	Data Deficient	DD

rarity and ecosystem function are likely (Soliveres et al., 2016) and further research on the relationship between functional rarity and functional insurance is thus required. The functional redundancy literature could be revisited to highlight results on functional rarity instead.

4.5. Concluding remarks – Functionally rare coral reef fishes: rare birds or red herrings?

We highlight that functional rarity hotspots are spatially distinct from other previously known species richness or rarity hotspots. Adding functional rarity to the other classical biodiversity facets should offer a more integrated view for conservation prioritization (Davies and Cadotte, 2011; Pollock et al., 2017). Here we show that functional rarity hotspots spatially mismatch with evolutionary rarity hotspots. If we include different biodiversity facets to prioritize which areas we should protect first, then we should go beyond protecting the Coral Triangle area. The Tropical Eastern Pacific, for instance, hosts a high number of functionally rare species, and would deserve more protection coverage. However, given the fact that different biodiversity facets (taxonomic, phylogenetic or functional richness or rarities) provide different messages and highlight different areas to protect (Guilhaumon et al., 2015), choosing the “best” set of candidate areas is not a trivial issue. Our results suggest protecting areas everywhere, but as the resources are limited, i.e. we are facing an “agony of choice” (Vane-Wright et al., 1991), it underlines the limits of a hotspot-based strategy integrating multiple facets (Schmitt, 2011). As recently applied to birds and mammals (Pollock et al., 2017), functional rarity could be included as a maximization criterion: choosing protected areas that maximize representativeness of functionally rare species. This maximization should also take into account high connectivity between protected areas to allow better population persistence (Andrello et al., 2015), while also taking social and economic costs into account (Andrello et al., 2017).

Acknowledgments

The authors would like to thank M. Kulbicki, J. Belmaker and three anonymous reviewers for their insightful comments that deeply improved the manuscript.

This work benefited of one database from the GASPAR program financed by French Fondation pour la Recherche sur la Biodiversité (FRB; www.fondationbiodiversite.fr). This research is supported by the FRB in the context of the CESAB project ‘Causes and consequences of functional rarity from local to global scales’ (FREE). P.D. and C.V. were supported by the European Research Council (ERC) Starting Grant Project “Ecophysiological and biophysical constraints on domestication of crop plants” (Grant ERC-StG-2014-639706-CONSTRAINTS). C.M.T. acknowledges the European Commission for the Marie Curie IIF (H2020-MSCA-IF-2014-657951). M.G. was supported by a PhD funding from the ENS de Lyon.

Appendices. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.biocon.2018.08.011>.

References

Andrello, M., Jacobi, M.N., Manel, S., Thuiller, W., Mouillot, D., 2015. Extending networks of protected areas to optimize connectivity and population growth rate. *Ecography* 38, 273–282. <https://doi.org/10.1111/ecog.00975>.
 Andrello, M., Guilhaumon, F., Albouy, C., Parravicini, V., Scholtens, J., Verley, P., Barange, M., Sumaila, U.R., Manel, S., Mouillot, D., 2017. Global mismatch between fishing dependency and larval supply from marine reserves. *Nat. Commun.* 8, 16039. <https://doi.org/10.1038/ncomms16039>.
 Arponen, A., 2012. Prioritizing species for conservation planning. *Biodivers. Conserv.* 21, 875–893. <https://doi.org/10.1007/s10531-012-0242-1>.
 Bellwood, D.R., Hughes, T.P., Hoey, A.S., 2006. Sleeping functional group drives coral-

reef recovery. *Curr. Biol.* 16, 2434–2439. <https://doi.org/10.1016/j.cub.2006.10.030>.
 Bellwood, D.R., Hughes, T.P., Folke, C., Nyström, M., 2004. Confronting the coral reef crisis. *Nature* 429, 827–833. <https://doi.org/10.1038/nature02691>.
 Bender, M.G., Leprieur, F., Mouillot, D., Kulbicki, M., Parravicini, V., Pie, M.R., Barneche, D.R., Oliveira-Santos, L.G.R., Floeter, S.R., 2017. Isolation drives taxonomic and functional nestedness in tropical reef fish faunas. *Ecography* 40, 425–435. <https://doi.org/10.1111/ecog.02293>.
 Brum, F.T., Graham, C.H., Costa, G.C., Hedges, S.B., Penone, C., Radeloff, V.C., Rondinini, C., Loyola, R., Davidson, A.D., 2017. Global priorities for conservation across multiple dimensions of mammalian diversity. *Proc. Natl. Acad. Sci.* 201706461. <https://doi.org/10.1073/pnas.1706461114>.
 Cadotte, M.W., Carscadden, K., Mirotchnick, N., 2011. Beyond species: functional diversity and the maintenance of ecological processes and services. *J. Appl. Ecol.* 48, 1079–1087. <https://doi.org/10.1111/j.1365-2664.2011.02048.x>.
 Cadotte, M., Albert, C.H., Walker, S.C., 2013. The ecology of differences: assessing community assembly with trait and evolutionary distances. *Ecol. Lett.* 16, 1234–1244. <https://doi.org/10.1111/ele.12161>.
 Ceballos, G., Ehrlich, P.R., Dirzo, R., 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl. Acad. Sci.* 201704949. <https://doi.org/10.1073/pnas.1704949114>.
 Chase, J., 2013. An Inordinate Fondness of Rarity. *PLoS Biol.* 11, e1001573. <https://doi.org/10.1371/journal.pbio.1001573>.
 Chamberlain, S., Szöcs, E., 2013. Taxize: taxonomic search and retrieval in R [version 2; referees: 3 approved]. *F1000Research* 2, 191. <https://doi.org/10.12688/f1000research.2-191.v2>.
 Chamberlain, S., Szöcs, E., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., Foster, Z., O'Donnell, J., 2016. Taxize: Taxonomic Information From Around the Web.
 Cofre, H., Marquet, P.A., 1999. Conservation status, rarity, and geographic priorities for conservation of Chilean mammals: an assessment. *Biol. Conserv.* 88, 53–68.
 Cowman, P.F., Bellwood, D.R., 2013. The historical biogeography of coral reef fishes: global patterns of origination and dispersal. *J. Biogeogr.* 40, 209–224. <https://doi.org/10.1111/jbi.12003>.
 D'agata, S., Mouillot, D., Wantiez, L., Friedlander, A.M., Kulbicki, M., Vigliola, L., 2016a. Marine reserves lag behind wilderness in the conservation of key functional roles. *Nat. Commun.* 7, 12000. <https://doi.org/10.1038/ncomms12000>.
 D'agata, S., Vigliola, L., Graham, N.A.J., Wantiez, L., Parravicini, V., Villéger, S., Mou-Tham, G., Frolla, P., Friedlander, A.M., Kulbicki, M., Mouillot, D., 2016b. Unexpected high vulnerability of functions in wilderness areas: evidence from coral reef fishes. *Proc. R. Soc. B* 283, 20160128. <https://doi.org/10.1098/rspb.2016.0128>.
 Dalerum, F., 2013. Phylogenetic and functional diversity in large carnivore assemblages. *Proc. R. Soc. Lond. B Biol. Sci.* 280, 20130049. <https://doi.org/10.1098/rspb.2013.0049>.
 Davies, T.J., Cadotte, M.W., 2011. Quantifying biodiversity: does it matter what we measure? In: *Biodiversity Hotspots*. Springer, Berlin, Heidelberg, pp. 43–60. https://doi.org/10.1007/978-3-642-20992-5_3.
 Davies, T.J., Fritz, S.A., Grenyer, R., Orme, C.D.L., Bielby, J., Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., Gittleman, J.L., Mace, G.M., Purvis, A., 2008. Phylogenetic trees and the future of mammalian biodiversity. *Proc. Natl. Acad. Sci.* 105, 11556–11563. <https://doi.org/10.1073/pnas.0801917105>.
 Davies, T.E., Maxwell, S.M., Kaschner, K., Garilao, C., Ban, N.C., 2017. Large marine protected areas represent biodiversity now and under climate change. *Sci. Rep.* 7, 9569. <https://doi.org/10.1038/s41598-017-08758-5>.
 Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W., Mouquet, N., 2010. Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecol. Lett.* 13, 1030–1040. <https://doi.org/10.1111/j.1461-0248.2010.01493.x>.
 Díaz, S., Cabido, M., 2001. Vive la difference: plant functional diversity matters to ecosystem processes. *Trends Ecol. Evol.* 16, 646–655. [https://doi.org/10.1016/S0169-5347\(01\)02283-2](https://doi.org/10.1016/S0169-5347(01)02283-2).
 Dray, S., Dufour, A.B., 2007. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22, 1–20.
 Dutilleul, P., Clifford, P., Richardson, S., Hemon, D., 1993. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics* 49, 305. <https://doi.org/10.2307/2532625>.
 Fisher, R., O'Leary, R.A., Low-Choy, S., Mengersen, K., Knowlton, N., Brainard, R.E., Caley, M.J., 2015. Species richness on coral reefs and the pursuit of convergent global estimates. *Curr. Biol.* 25, 500–505. <https://doi.org/10.1016/j.cub.2014.12.022>.
 Forest, F., Grenyer, R., Rouget, M., Davies, T.J., Cowling, R.M., Faith, D.P., Balmford, A., Manning, J.C., Procheş, Ş., Bank, M. van der, Reeves, G., Hedderson, T.A.J., Savolainen, V., 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445, 757–760. <https://doi.org/10.1038/nature05587>.
 Godet, L., Gaiüzere, P., Jiguet, F., Devictor, V., 2015. Dissociating several forms of commonness in birds sheds new light on biotic homogenization. *Glob. Ecol. Biogeogr.* 24, 416–426. <https://doi.org/10.1111/geb.12266>.
 Gonçalves-Souza, T., Diniz-Filho, J.A.F., Romero, G.Q., 2014. Disentangling the phylogenetic and ecological components of spider phenotypic variation. *PLoS One* 9, e89314. <https://doi.org/10.1371/journal.pone.0089314>.
 Gotelli, N.J., Graves, G.R., 1996. *Null Models in Ecology*. Smithsonian Institution Press, Washington.
 Gower, J.C., 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27, 857–871. <https://doi.org/10.2307/2528823>.
 Graham, N.A.J., McClanahan, T.R., MacNeil, M.A., Wilson, S.K., Cinner, J.E., Huchery, C., Holmes, T.H., 2017. Human disruption of coral reef trophic structure. *Curr. Biol.* 27, 231–236. <https://doi.org/10.1016/j.cub.2016.10.062>.
 Grenié, M., Denelle, P., Tucker, C.M., Munoz, F., Violle, C., 2017. funrar: An R package to

M. Grenié et al.

Biological Conservation 226 (2018) 288–299

- characterize functional rarity. *Divers. Distrib.* 23, 1365–1371. <https://doi.org/10.1111/ddi.12629>.
- Guilhaumon, F., Albouy, C., Claudet, J., Velez, L., Ben Rais Lasram, F., Tomasini, J.-A., Douzery, E.J.P., Meynard, C.N., Mouquet, N., Troussellier, M., Araújo, M.B., Mouillot, D., 2015. Representing taxonomic, phylogenetic and functional diversity: new challenges for Mediterranean marine-protected areas. *Divers. Distrib.* 21, 175–187. <https://doi.org/10.1111/ddi.12280>.
- Harborne, A.R., Mumby, P.J., Micheli, F., Perry, C.T., Dahlgren, C.P., Holmes, K.E., Brumbaugh, D.R., 2006. The Functional Value of Caribbean Coral Reef, Seagrass and Mangrove Habitats to Ecosystem Processes. In: *Advances in Marine Biology*. Academic Press, pp. 57–189. [https://doi.org/10.1016/S0065-2881\(05\)50002-6](https://doi.org/10.1016/S0065-2881(05)50002-6).
- Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E., Challenger, W., 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24, 129–131. <https://doi.org/10.1093/bioinformatics/btm538>.
- Hidasi-Neto, J., Loyola, R.D., Cianciaruso, M.V., 2013. Conservation actions based on red lists do not capture the functional and phylogenetic diversity of birds in Brazil. *PLoS One* 8, e73431. <https://doi.org/10.1371/journal.pone.0073431>.
- Hidasi-Neto, J., Loyola, R., Cianciaruso, M.V., 2015. Global and local evolutionary and ecological distinctiveness of terrestrial mammals: identifying priorities across scales. *Divers. Distrib.* 21, 548–559. <https://doi.org/10.1111/ddi.12320>.
- Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E., Harvell, C.D., Sale, P.F., Edwards, A.J., Caldeira, K., Knowlton, N., Eakin, C.M., Iglesias-Prieto, R., Muthiga, N., Bradbury, R.H., Dubi, A., Hatziolos, M.E., 2007. Coral Reefs Under Rapid Climate Change and Ocean Acidification. *Science* 318, 1737–1742. <https://doi.org/10.1126/science.1152509>.
- Hughes, T.P., Anderson, K.D., Connolly, S.R., Heron, S.F., Kerry, J.T., Lough, J.M., Baird, A.H., Baum, J.K., Berumen, M.L., Bridge, T.C., 2018. Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. *Science* 359, 80–83.
- Isaac, N.J.B., Turvey, S.T., Collen, B., Waterman, C., Baillie, J.E.M., 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2, e296. <https://doi.org/10.1371/journal.pone.0000296>.
- IUCN, 2017. The IUCN Red List of Threatened Species. Version 2017-1.
- Jain, M., Flynn, D.F.B., Prager, C.M., Hart, G.M., DeVan, C.M., Ahrestani, F.S., Palmer, M.I., Bunker, D.E., Knops, J.M.H., Jouseau, C.F., Naeem, S., 2014. The importance of rare species: a trait-based assessment of rare species contributions to functional diversity and possible ecosystem function in tall-grass prairies. *Ecol. Evol.* 4, 104–112. <https://doi.org/10.1002/ece3.915>.
- Kemmel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464.
- Kulbicki, M., Parravicini, V., Bellwood, D.R., Arias-González, E., Chabanet, P., Floeter, S.R., Friedlander, A., McPherson, J., Myers, R.E., Vigliola, L., Mouillot, D., 2013. Global biogeography of reef fishes: a hierarchical quantitative delineation of regions. *PLoS One* 8, e81847. <https://doi.org/10.1371/journal.pone.0081847>.
- Lee, T.M., Jetz, W., 2011. Unravelling the structure of species extinction risk for predictive conservation science. *Proc. R. Soc. Lond. B Biol. Sci.* 278, 1329–1338. <https://doi.org/10.1098/rspb.2010.1877>.
- Legendre, P., Legendre, L.F.J., 2012. *Numerical Ecology*. Elsevier.
- Leprieur, F., Colosio, S., Descombes, P., Parravicini, V., Kulbicki, M., Cowman, P.F., Bellwood, D.R., Mouillot, D., Pellissier, L., 2016. Historical and contemporary determinants of global phylogenetic structure in tropical reef fish faunas. *Ecography* 39, 825–835. <https://doi.org/10.1111/ecog.01638>.
- McGill, B.J., Enquist, B.J., Weiher, E., Westoby, M., 2006. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* 21, 178–185. <https://doi.org/10.1016/j.tree.2006.02.002>.
- Mellin, C., Mouillot, D., Kulbicki, M., McClanahan, T.R., Vigliola, L., Bradshaw, C.J.A., Brainard, R.E., Chabanet, P., Edgar, G.J., Fordham, D.A., Friedlander, A.M., Parravicini, V., Sequeira, A.M.M., Stuart-Smith, R.D., Wantiez, L., Caley, M.J., 2016. Humans and seasonal climate variability threaten large-bodied coral reef fish with small ranges. *Nat. Commun.* 7, 10491. <https://doi.org/10.1038/ncomms10491>.
- de Mendiburu, F., 2017. *agricolae: Statistical Procedures for Agricultural Research*. Micheli, F., Mumby, P.J., Brumbaugh, D.R., Broad, K., Dahlgren, C.P., Harborne, A.R., Holmes, K.E., Kappel, C.V., Litvin, S.Y., Sanchirico, J.N., 2014. High vulnerability of ecosystem function and services to diversity loss in Caribbean coral reefs. *Biol. Conserv.* 171, 186–194. <https://doi.org/10.1016/j.biocon.2013.12.029>.
- Moberg, F., Folke, C., 1999. Ecological goods and services of coral reef ecosystems. *Ecol. Econ.* 29, 215–233. [https://doi.org/10.1016/S0921-8009\(99\)00009-9](https://doi.org/10.1016/S0921-8009(99)00009-9).
- Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., Kulbicki, M., Lavergne, S., Lavorel, S., Mouquet, N., Paine, C.E.T., Renaud, J., Thuiller, W., 2013. Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol.* 11, e1001569. <https://doi.org/10.1371/journal.pbio.1001569>.
- Mouillot, D., Villéger, S., Parravicini, V., Kulbicki, M., Arias-González, J.E., Bender, M., Chabanet, P., Floeter, S.R., Friedlander, A., Vigliola, L., Bellwood, D.R., 2014. Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. *Proc. Natl. Acad. Sci.* 111, 13757–13762. <https://doi.org/10.1073/pnas.1317625111>.
- Mouillot, D., Parravicini, V., Bellwood, D.R., Leprieur, F., Huang, D., Cowman, P.F., Albouy, C., Hughes, T.P., Thuiller, W., Guilhaumon, F., 2016. Global marine protected areas do not secure the evolutionary history of tropical corals and fishes. *Nat. Commun.* 7, 10359. <https://doi.org/10.1038/ncomms10359>.
- Murray, B.R., Thrall, P.H., Gill, A.M., Nicotra, A.B., 2002. How plant life-history and ecological traits relate to species rarity and commonness at varying spatial scales. *Austral Ecol.* 27, 291–310.
- Naeem, S., Duffy, J.E., Zavaleta, E., 2012. The functions of biological diversity in an age of extinction. *Science* 336, 1401–1406. <https://doi.org/10.1126/science.1215855>.
- Nyström, M., 2006. Redundancy and Response Diversity of Functional Groups: Implications for the Resilience of Coral Reefs. *AMBIO J. Hum. Environ.* 35, 30–35. <https://doi.org/10.1579/0044-7447-35-1.30>.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2018. *Vegan: Community Ecology Package*.
- Osorio, F., Vallejos, R., 2014. *SpatialPack: Package for Analysis of Spatial Data*.
- Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.
- Parravicini, V., Kulbicki, M., Bellwood, D.R., Friedlander, A.M., Arias-Gonzalez, J.E., Chabanet, P., Floeter, S.R., Myers, R., Vigliola, L., D'Agata, S., Mouillot, D., 2013. Global patterns and predictors of tropical reef fish species richness. *Ecography* 36, 1254–1262. <https://doi.org/10.1111/j.1600-0587.2013.00291.x>.
- Parravicini, V., Villéger, S., McClanahan, T.R., Arias-González, J.E., Bellwood, D.R., Belmaker, J., Chabanet, P., Floeter, S.R., Friedlander, A.M., Guilhaumon, F., Vigliola, L., Kulbicki, M., Mouillot, D., 2014. Global mismatch between species richness and vulnerability of reef fish assemblages. *Ecol. Lett.* 17, 1101–1110. <https://doi.org/10.1111/ele.12316>.
- Pavoine, S., Ollier, S., Dufour, A.-B., 2005. Is the originality of a species measurable? *Ecol. Lett.* 8, 579–586. <https://doi.org/10.1111/j.1461-0248.2005.00752.x>.
- Pavoine, S., Vallet, J., Dufour, A.-B., Gachet, S., Daniel, H., 2009. On the challenge of treating various types of variables: application for improving the measurement of functional diversity. *Oikos* 118, 391–402. <https://doi.org/10.1111/j.1600-0706.2008.16668.x>.
- Pellissier, L., Leprieur, F., Parravicini, V., Cowman, P.F., Kulbicki, M., Litsios, G., Olsen, S.M., Wisz, M.S., Bellwood, D.R., Mouillot, D., 2014. Quaternary coral reef refugia preserved fish diversity. *Science* 344, 1016–1019. <https://doi.org/10.1126/science.1249853>.
- Pollock, L.J., Thuiller, W., Jetz, W., 2017. Large conservation gains possible for global biodiversity facets. *Nature* 546, 141–144. <https://doi.org/10.1038/nature22368>.
- Pratchett, M.S., Hoey, A.S., Wilson, S.K., Messmer, V., Graham, N.A.J., 2011. Changes in Biodiversity and Functioning of Reef Fish Assemblages following Coral Bleaching and Coral Loss. *Diversity* 3, 424–452. <https://doi.org/10.3390/d3030424>.
- R Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rabinowitz, D., 1981. Seven forms of rarity. In: Sygne, H. (Ed.), *The Biological Aspects of Rare Plant Conservation*. Riley, pp. 205–217.
- Rabosky, D.L., Santini, F., Eastman, J., Smith, S.A., Sidlauskas, B., Chang, J., Alfaro, M.E., 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* 4, 2958. <https://doi.org/10.1038/ncomms2958>.
- Redding, D.W., Dewolf, C., Mooers, A.O., 2010. Evolutionary distinctiveness, threat status, and ecological oddity in primates. *Conserv. Biol.* 24, 1052–1058.
- Ricotta, C., de Bello, F., Moretti, M., Caccianiga, M., Cerabolini, B.E.L., Pavoine, S., 2016. Measuring the functional redundancy of biological communities: a quantitative guide. *Methods Ecol. Evol.* 7, 1386–1395. <https://doi.org/10.1111/2041-210X.12604>.
- Rodrigues, A.S.L., Pilgrim, J.D., Lamoreux, J.F., Hoffmann, M., Brooks, T.M., 2006. The value of the IUCN red list for conservation. *Trends Ecol. Evol.* 21, 71–76. <https://doi.org/10.1016/j.tree.2005.10.010>.
- Rogers, A., Blanchard, J.L., Newman, S.P., Dryden, C.S., Mumby, P.J., 2018. High refuge availability on coral reefs increases the vulnerability of reef-associated predators to overexploitation. *Ecology* 99, 450–463. <https://doi.org/10.1002/ecy.2103>.
- Rosauer, D., Laffan, S.W., Crisp, M.D., Donnellan, S.C., Cook, L.G., 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* 18, 4061–4072. <https://doi.org/10.1111/j.1365-294X.2009.04311.x>.
- Rosenfeld, J.S., 2002. Functional redundancy in ecology and conservation. *Oikos* 98, 156–162. <https://doi.org/10.1034/j.1600-0706.2002.980116.x>.
- Schmitt, C.B., 2011. A tough choice: approaches towards the setting of global conservation priorities. In: *Biodiversity Hotspots*. Springer, Berlin, Heidelberg, pp. 23–42. https://doi.org/10.1007/978-3-642-20992-5_2.
- Singleton, R.L., Roberts, C.M., 2014. The contribution of very large marine protected areas to marine conservation: giant leaps or smoke and mirrors? *Mar. Pollut. Bull.* 87, 7–10. <https://doi.org/10.1016/j.marpolbul.2014.07.067>.
- Soliveres, S., Manning, P., Prati, D., Gossner, M.M., Alt, F., Arndt, H., Baumgartner, V., Binkenstein, J., Birkhofer, K., Blaser, S., Blüthgen, N., Boch, S., Böhm, S., Borschig, C., Buscot, F., Dieckötter, T., Heinze, J., Hölzel, N., Jung, K., Klaus, V.H., Klein, A.-M., Kleinebecker, T., Klemmer, S., Krauss, J., Lange, M., Morris, E.K., Müller, J., Oelmann, Y., Overmann, J., Pašalić, E., Renner, S.C., Rilg, M.C., Schaefer, H.A., Schlöter, M., Schmitt, B., Schöning, I., Schruppf, M., Sikorski, J., Socher, S.A., Solly, E.F., Sonnemann, I., Sorkau, E., Steckel, J., Steffan-Wever, I., Stempfhuber, B., Tschapka, M., Türke, M., Venter, P., Weiner, C.N., Weisser, W.W., Werner, M., Westphal, C., Wilcke, W., Wolters, V., Wubet, T., Wurst, S., Fischer, M., Allan, E., 2016. Locally rare species influence grassland ecosystem multifunctionality. *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150269. <https://doi.org/10.1098/rstb.2015.0269>.
- Strona, G., Nappo, D., Boccacci, F., Fattorini, S., San-Miguel-Ayanz, J., 2014. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nat. Commun.* 5 (4114). <https://doi.org/10.1038/ncomms5114>.
- Stuart-Smith, R.D., Bates, A.E., Lefcheck, J.S., Duffy, J.E., Baker, S.C., Thomson, R.J., Stuart-Smith, J.F., Hill, N.A., Kininmonth, S.J., Airoldi, L., Becerro, M.A., Campbell, S.J., Dawson, T.P., Navarrete, S.A., Soler, G.A., Strain, E.M.A., Willis, T.J., Edgar, G.J., 2013. Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* 501, 539–542. <https://doi.org/10.1038/nature12529>.
- Thuiller, W., Maiorano, L., Mazel, F., Guilhaumon, F., Ficetola, G.F., Lavergne, S., Renaud, J., Roquet, C., Mouillot, D., 2015. Conserving the functional and phylogenetic trees of life of European tetrapods. *Philos. Trans. R. Soc. B* 370, 20140005.

M. Grenié et al.

Biological Conservation 226 (2018) 288–299

- <https://doi.org/10.1098/rstb.2014.0005>.
- van Hooijdonk, R., Maynard, J.A., Manzello, D., Planes, S., 2014. Opposite latitudinal gradients in projected ocean acidification and bleaching impacts on coral reefs. *Glob. Chang. Biol.* 20, 103–112. <https://doi.org/10.1111/gcb.12394>.
- Vane-Wright, R.I., Humphries, C.J., Williams, P.H., 1991. What to protect?—systematics and the agony of choice. *Biol. Conserv.* 55, 235–254. [https://doi.org/10.1016/0006-3207\(91\)90030-D](https://doi.org/10.1016/0006-3207(91)90030-D).
- Vié, J.-C., Hilton-Taylor, C., Pollock, C., Ragle, J., Smart, J., Stuart, S.N., Tong, R., 2009. The IUCN Red List: a key conservation tool. *Wildlife in a changing world—An analysis of the 2008 IUCN Red List of Threatened Species* 1.
- Villéger, S., Brosse, S., Mouchet, M., Mouillot, D., Vanni, M.J., 2017. Functional ecology of fish: current approaches and future challenges. *Aquat. Sci.* 79, 783–801. <https://doi.org/10.1007/s00027-017-0546-z>.
- Violle, C., Jiang, L., 2009. Towards a trait-based quantification of species niche. *J. Plant Ecol.* 2, 87–93. <https://doi.org/10.1093/jpe/rtp007>.
- Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., Garnier, E., 2007. Let the concept of trait be functional!. *Oikos* 116, 882–892. <https://doi.org/10.1111/j.0030-1299.2007.15559.x>.
- Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N.J.B., Cadotte, M.W., Livingstone, S.W., Grenié, M., Mouillot, D., 2017a. A common toolbox to understand, monitor or manage rarity? A response to Carmona et al. *Trends Ecol. Evol.* 32, 891–893. <https://doi.org/10.1016/j.tree.2017.10.001>.
- Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N.J.B., Cadotte, M.W., Livingstone, S.W., Mouillot, D., 2017b. Functional rarity: the ecology of outliers. *Trends Ecol. Evol.* 32, 356–367. <https://doi.org/10.1016/j.tree.2017.02.002>.
- Wilson, S.K., Graham, N.a.J., Pratchett, M.S., Jones, G.P., Polunin, N.V.C., 2006. Multiple disturbances and the global degradation of coral reefs: are reef fishes at risk or resilient? *Glob. Chang. Biol.* 12, 2220–2234. <https://doi.org/10.1111/j.1365-2486.2006.01252.x>.
- Winter, M., Devictor, V., Schweiger, O., 2013. Phylogenetic diversity and nature conservation: where are we? *Trends Ecol. Evol.* 28, 199–204. <https://doi.org/10.1016/j.tree.2012.10.015>.
- Yachi, S., Loreau, M., 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl. Acad. Sci.* 96, 1463–1468. <https://doi.org/10.1073/pnas.96.4.1463>.

2 MENACES SUR LES RARETÉS ÉCOLOGIQUES DANS L'ANTHROPOCÈNE

Target: Nature Ecology and Evolution

Ecological rarity of mammals and birds in the Anthropocene

Proposition :

Co-first: Loiseau Nicolas 1,2,3 & Mouquet Nicolas 1, 4

Second: Casajus Nicolas 4

Core: Alphabetic Order :Matthias Grenié 3, Maya Gueguen 2, Brian Maitner 5, David Mouillot 1-6, Annette Ostling 7, Julien Renaud 2, Caroline Tucker 8, Laure Velez 1

Co-last : Wilfried Thuiller 2 & Cyrille Violle 3

Affiliation : Please note your affiliations

1 MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Montpellier, France

2 Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine F-38000 Grenoble, France

3 CEFE UMR 5175, CNRS – Université de Montpellier – Université Paul-Valéry Montpellier – EPHE - IRD, Montpellier, France

4 FRB – CESAB, 34000 Montpellier, France

5 Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona

6 ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, 17 QLD, 4811, Australia

7 Ecology and Evolutionary Biology, University of Michigan, USA

8 Department of Biology, University of North Carolina at Chapel Hill, Coker Hall, CB #3280 120 South Road, Chapel Hill, NC 27599-3280, U.S.A.

Keywords: Functional diversity, distinctiveness, protected area, Climate change, vulnerability

Abstract max 150/150 WORD

Most species are geographically rare but not all have specific morphological or physiological characteristics that make them unique on Earth. Identifying those species that are both spatially restricted and functionally unique, i.e. ecologically rare, is thus of prime importance given their risk of extinction and their potential contribution to ecosystem functioning. We took advantage of the global distributions of birds and mammals to identify the ecological rare species, understand their characteristics, and identify hotspots where conservation efforts should take place. We found that ecological rare species have particular diet and body mass, are insufficiently covered by the current system of protected areas and are disproportionately sensitive to current and future threats. Put together, those results open the debate on whether conservation priority should be given to ecological rarity in respect to more widespread and functionally redundant species given future environmental conditions and their uncertainty.

Main max 4423/5000 words**Introduction**

The majority of species are geographically rare, and rarity has become one of the cornerstones of many ecological research studies and conservation strategies since decades (Rabinowitz 1981, Gaston & Kunin 1997, Pompa et al. 2011, Thuiller et al. 2015, Pollock et al. 2017). Rare species are of particular concern because they tend to have high extinction risk and may even face a “*double jeopardy*” (Gaston, 1998, Hughes et al. 2014) as low abundance species are also often narrowly distributed. While it has long been assumed that rare species do not contribute to community assembly, functional diversity and ecosystem functioning, recent studies have challenged this belief. Rare species may indeed contribute disproportionately to the diversity of traits within a community or within a region (Mouillot et al. 2013, Jain et al. 2014, Chapman et al. 2018). Since high trait diversity is usually assumed to enhance ecosystem functioning, these rare species thus bring specific traits or functions that might be irreplaceable (Winter et al. 2013, Thuiller et al. 2015). In other words, ecological rarity does not only relate to the mere abundance or geographic extent of species but also to their functional originality (Violle et al. 2017, Kondratyeva et al. 2019). If ecologically rare species are not redundant with other species and instead hold unique combination of traits, they will likely contribute to ecosystem

functioning and associated services (Dee et al. 2019). In this case, and in respect to the ever-increasing biodiversity crisis and uncertain environments, it is fundamental to understand their characteristics, to map their distribution and predict their vulnerability to current and future threats. Especially given the current rate of species loss (Johnson et al. 2017), a thorough evaluation of the spatial distribution of ecological rarity, its sensitivity to ongoing and future threats and its coverage by protected areas is urgently needed.

For a global assessment of ecological rarity, it appears relevant to build the functional space of species worldwide based on their relative position in the Eltonian niche space (Devictor et al. 2010, Brum et al. 2017, Mazel et al. 2018, Dehling & Stouffer, 2019), also known as the trophic niche space, which focuses on traits related to biotic interactions and resource–consumer dynamics (Elton 1927). To that end, worldwide data on Eltonian traits, including the characterization of diet and foraging activity, are relevant candidate features (e.g. in mammals and birds, Wilman et al. 2014). Recent efforts on mammals and birds (Wilman et al. 2014) indeed led to interesting findings about the shape of the functional space of these taxa and about the global distribution of their functional diversity, notably highlighting hotspots of functional redundancy (Safi et al. 2011, Mazel et al. 2014, Barnagaud et al. 2019, Cooke et al. 2019a). They also highlighted a spatial mismatch between protected areas and taxonomic, phylogenetic, and functional diversity (Pollock et al. 2017, Brum et al. 2017). Reversely, we lack a global assessment of ecological rarity in most taxa, while such assessment has recently been facilitated by the development of a novel theoretical corpus and associated methodology (Violle et al. 2017, Grenié et al. 2017).

Who is facing ongoing and future threats the most? Geographically rare species are *usual suspects* given their restricted distribution (Davies et al. 2004). It has indeed been shown that species with restricted geographic distribution have the highest risk of extinction under most of climate scenarios for the future (Ohlemüller et al. 2008). Alternative evaluation of ecological rarity further implies the need for a risk evaluation of the newly defined ecologically rare species (Grenié et al. 2018). Being at risk of extinction in the face of global change and not targeted by current conservation programs would be the worst-case scenario for these species since they would be under a stronger jeopardy than initially thought. In any case, the precautionary principle requires a sound assessment of the global distribution of ecological rarity to adapt or mitigate future global change impacts on this yet under-valued biodiversity aspect. Here, we characterized

the ecological rarity of mammals (4,654 species) and birds (9,287 species) that are largely distributed across the globe and support important ecological functions in ecosystems such as seed dispersal, trophic interaction and nutrient cycling (Cooke et al. 2019a, Barnagaud et al. 2019). For each taxa, we computed ecological rarity as a combination of geographical restrictiveness and functional distinctiveness (Supplementary Fig. 1, Violle et al. 2017). We defined *ecologically rare* and *common* species as having values of functional distinctiveness and geographical restrictiveness higher than 75% or lower than 25% of the entire species pool, respectively (Supplementary Fig. 1). Using that combined information, we investigated (a) who are the ecologically rare species and how they are distributed across the functional space and the tree of life?, (b) how are they spatially distributed and whether they follow general biodiversity patterns?, and finally (c) what current and future threats are they facing in the Anthropocene era and are they covered by current protection efforts?

Results

Global functional and phylogenetic position of ecologically rare species

We obtained two distributions based on (1) the functional distinctiveness of species trait values relative to the other species of the global pool (the average functional distance of a species to other species in the community) and (2) the geographical range of species compared to the geographic extent of all species in the global pool (Supplementary Fig. 1). Our analyses revealed that 237 (5%) of mammals and 573 (6%) of birds were ecologically rare, while 200 (4%) and 569 (6%) were ecologically common. We projected species into a global functional space based on their functional traits (see Methods). For both taxa, we found that the first axis (PC1) was not correlated to functional distinctiveness which was mostly explained by the second and third axes for mammals and by the second and fourth axes for birds (Supplementary Fig. 2, Supplementary Fig. 3). Ecologically rare mammals were mainly frugivorous, nocturnal with relatively large body size (e.g., bats, lemurs) or small invertebrate-eaters (e.g., rodents). Ecologically rare birds were mainly nocturnal frugivores or nectarivores (e.g., hummingbirds) or diurnal and piscivorous (e.g., large marine birds). Despite their small number, ecologically rare species occupy a large volume in the functional space for both taxa, and did not overlap with ecologically common species

which fill a much smaller volume of the functional space (Fig. 1, Supplementary Table 1 and Supplementary Table 2).

We plotted ecological rarity on the tree of life and tested phylogenetic signal of ecological rarity. We found that closely related species were not necessarily more similar in their degree of ecological rarity than distantly related species (Fig. 2, $D = 0.53$ and $D = 0.56$, respectively). However, we found a phylogenetic clustering of ecological rarity for several orders. The orders with the highest concentration of ecologically rare species were the Primate and Chiroptera for mammals and Psittaciformes, Caprimulgiformes and Apodiformes (Strisores) for birds. Finally we found that ecologically rare species were on average not more evolutionary distinct than other species (Supplementary Fig. 4).

Global distribution of ecological rarity

To identify the geographical ‘hotspots’ of ecological rarity, we mapped the number of ecologically rare species within each 50 by 50km cell around the world and examined its spatial congruence with ecological commonness and total species richness. We found that ecological rarity was aggregated on only 2.8% and 8.9% of the total grid cells for mammals and birds respectively (with maximum values of 12 and 28 species). Altogether only 1.1% of the global land hosted at least one ecologically rare species of both taxa. Ecological rarity of mammals predominantly occurred in the tropics and in the Southern Hemisphere, peaking in Indonesia islands, Madagascar and Costa Rica (Fig. 3). Ecological rarity of birds predominantly occurred in mountainous tropical and subtropical regions, peaking in New Guinea, Indonesia, the Andes and Central America (Fig. 3). Ecological rarity was over-represented on islands with 39% and 30% of grid cells containing ecologically rare mammals and birds being on islands. We also found a strong mismatch between the geographical distribution of ecological rarity and commonness for both taxa (Fig. 3, Pearson correlation corrected for spatial autocorrelation ((Dutilleul, 1993), $R^2 = 0.026$, $F = 1.03$, $P = 0.31$ for mammals and $R^2 = 0.012$, $F = 0.05$, $P = 0.82$ for birds).

To evaluate the link between the number of species and the number of ecologically rare species per cell, we simulated this link under the null expectation that ecologically rare species are randomly distributed among cells regardless of the number of species (see Methods). We expected more ecologically rare species in high species diversity areas. For both taxa, we found that standardized effect size (SES) was higher than expected at random for all cells hosting at

least one ecologically rare species (Supplementary Fig. 5), highlighting that these cells host more ecologically rare species than expected by chance. We also highlighted that the number of ecologically rare species is not trivially linked to the overall species richness within cells.

Ecological rarity under global threats

We classified species according to their IUCN status. We found that ecologically rare species were disproportionately packed in IUCN threatened categories for both mammals and birds and significantly more threatened than ecologically common species (71% and 44.2% against 2% and 0.5%, respectively, Fig. 4, $P < 0.001$). As expected, geographical restrictiveness, one of the main IUCN criteria to estimate vulnerability, is higher for threatened species (Supplementary Fig. 6, Supplementary Fig. 7). However, we also found that threatened species were functionally more distinct (Supplementary Fig. 6, Supplementary Fig. 7). A significant proportion of ecologically rare species were also considered as least concerned (13% for mammals and 52% for birds) or non-evaluated (16% for mammals and 3.8% for birds) by the IUCN.

For each ecologically rare species we evaluated exposure to human footprint, human development (HDI) and the number of conflicts, known to influence conservation outcomes. We found that ecologically rare mammals and birds were respectively 1.35 ± 1 and 1.2 ± 1 times more impacted by human footprint than ecologically common species (Fig. 4, $P < 0.001$). Ecologically rare mammals occurred in countries with a lower HDI than ecologically common species (Fig. 4, $P = 0.0032$). Ecologically rare birds occurred in countries with HDI similar to ecologically common species (Fig. 4, $P = 0.72$). Ecologically rare mammals and birds occurred in countries with a number of conflicts not different than for common species (Supplementary Fig. 8, Supplementary Fig. 9, $P > 0.05$). However, some countries with a high number of conflicts (e.g. Colombia, Indonesia) host at least five ecologically rare mammals. Several countries (e.g. Philippines) with a low HDI and a high number of conflicts are hotspots of ecological rarity (19 and 15 ecologically rare mammals and birds, Supplementary Fig. 8, Supplementary Fig. 9). On the other hand, some countries with high HDI and low number of conflicts are also hotspots of ecological rarity (e.g. Australia, respectively hosted 5 ecologically rare mammal and 10 bird species).

We then quantified the influence of climate change on ecologically rare and common species. We modelled the current and future climatic suitability for birds and mammals at the

global scale (Fig. 4). More specifically, we modelled the distribution of species as a function of current climate using species distribution models (SDM) and projected their future climatic suitability as a function of combinations of RCPs scenarios (see Methods). We selected only SDMs that reached high predictive accuracies ($TSS > 0.8$). Consequently, we kept SDMs outputs for 28% (67 species) and 59% (337 species) of ecologically rare mammal and bird species. We found winners and losers under future climates for both ecologically rare mammals and birds (Fig. 4). By the time horizon 2041-2060, 36% and 58% of modelled ecologically rare mammals and birds are projected to lose suitable areas (45% and 64% by horizon 2061-2080, Supplementary Fig. 10). Overall the ecological rare mammals will be less impacted (Fig. 4, $P = 2.4e-03$) while rare birds will be more impacted than common ones by climate change (Fig. 4, $P = 3.5e-04$).

Finally, to evaluate the potential benefits of conservation efforts on ecological rarity we estimated species-specific target achievements defined as proportions of geographic ranges covered by protected areas. These specific targets were related to species range sizes with the most restricted species needing more coverage (e.g. 100%) than widespread one (e.g. 10%) to avoid extinction. We found for both mammals and birds that target conservation achievement of ecologically rare species were lower than for common species (Fig. 4, $P < 0.001$). Average target achievement for mammals and birds were respectively 15% and 14% for ecologically rare species against 31% and 36% for ecologically common species.

Discussion

Who are ecologically rare species ?

Because it integrates functional traits and geographical restrictiveness, ecological rarity provides new insights for functional biogeography and conservation (Violle al. 2014, 2017). We found that both mammals and birds that are ecologically rare fill a much wider breadth of ecological strategies in the Eltonian niche space than ecological common species (Fig. 1). Mouillot et al. (2013) and Leitao et al. (2016) demonstrated that geographically and locally rare species usually bear distinct traits that could put ecosystems functioning at risk if they go extinct. Our findings extend this result as we find that the functional space filled by ecologically rare species does not overlap much with the functional space of common species for both mammals

and birds. In particular, we show that specific sets of traits were over-contributing to ecological rarity, echoing the results of Barnagaud et al. (2019) who found that nectarivory, carnivory and piscivory are rare diets in birds. A significant proportion of ecologically rare mammals and birds were indeed nectarivores or frugivores. These traits are important for ecosystem functioning as they are directly linked to pollination and seed dispersal, essential for plant survival (Fleming and Kress 2013, Estrada et al. 2017) and community stability (Wiszniewski et al. 2013). For instance, ecologically rare lemurs in Madagascar (such as *Eulemur macaco*, Birkinshaw 1999) display complex relationships with trees producing large seeds. The extinction of these species will likely amplify the ongoing decline of some tree species under other accelerating global changes in the region (Federman et al. 2016). Similarly, while most hummingbirds are ecologically rare, they are the main pollinators of ~7,000 plant species and thus maintain ecological networks and crucial ecosystem processes (Abrahamczyk & Renner, 2017). We also highlight that several seabirds are ecologically rare and play unique functions. By feeding in the open ocean these species transport large quantities of nutrients onto islands, these birds can enhance the productivity of island fauna and flora with a marked benefit for coral reef ecosystems (Graham et al. 2018, Benkwitt et al. 2019). Some predatory birds such as *Circus maillardi* also strongly participate to the regulation of small introduced mammals (Cheke, 1987). More generally, even at low abundance, ecologically rare mammal or bird predators can have disproportionate impacts on ecosystem functioning (e.g. through top-down controls along the trophic chain and the associated energy fluxes (Mateo-Tomás et al. 2017)) and services (e.g. bats, an order with a high proportion of ecologically rare species, are able to reduce arthropod abundance by 84% in agroforestry, Williams-Guillén et al. 2008).

Establishing a universal list of core traits in order to measure ecological rarity in absolute terms is a difficult task given the multitude and species-specific functions that organisms can perform in ecosystems. An alternative, based on trait conservatism, would be to use species evolutionary distinctiveness as a proxy for species functional distinctiveness (Cadotte et al. 2008, Srivastava et al. 2012, Winter et al. 2013). We found that ecological rarity was over-represented in some orders (Primate and Chiroptera for mammals and Psittaciformes, Caprimulgiformes and Apodiformes (Strisores) for birds). Some of these clades are well known to be highly threatened (Galán-Acedo et al. 2019). For instance, primates are highly sensitive to land-use changes with ~60% of the world's species having a high risk of extinction (Galán-Acedo et al. 2019). Yet, we

found that overall ecological rarity was widely distributed across the phylogeny for both taxa, in line with the widespread distribution of ecological rarity across the functional space, and that ecological rare species are not more evolutionarily distinct than other species (Fig. 2). This suggests that evolutionarily distinctiveness cannot be used to infer ecological status, such as ecological rarity, to prioritize global conservation strategy (Mazel et al. 2018, Kling et al. 2018). However, the relationship between ecological rarity and evolutionary distinctiveness will need to be more thoroughly examined in the future as the rates of speciation and species extinction coupled to trait diversification may blur the link between functional distinctiveness and evolutionary distinctiveness (Thuiller et al. 2015).

Where are ecologically rare species ?

The distribution of ecological rarity provides complementary information when compared to geographical rarity of vertebrates (Grenyer et al. 2006) and their functional redundancy (Cooke et al. 2019a). We found an aggregation of ecological rarity in a handful of hotspots (Fig. 3). Because geographical restrictiveness is a component of ecological rarity, hotspots of mammal ecological rarity are located in the major centers of endemism for Chiroptera, i.e. Southeast Asia, Taiwan, the Caribbean, the Neotropics and Madagascar (Rosauer & Jetz, 2015). Similarly, hotspots of bird ecological rarity are spatially congruent with endemism and the history of bird diversification (Sandel et al. 2011, Jetz et al. 2014, Quintero & Jetz, 2018). The Andes, for instance, are well known to host a high diversity of hummingbird species adapted to high altitude (Abrahamczyk & Renner, 2017). Ecologically rare species are also disproportionately represented on islands. This over-representation on islands could emerge from speciation and adaptation outcomes in the context of ecological opportunity produced by isolation (Losos & Ricklefs, 2009). Except hotspots which can reach up to 12 ecologically rare mammals and 28 birds, most areas hosted very few ecologically rare species, making difficult the establishment of a simple conservation strategy to protect ecological rarity. We also found that areas with ecologically rare species host more ecologically rare species than expected by chance, suggesting that particular environmental conditions (such as paleoclimate, Sandel et al. 2011) may shape the distribution of ecological rarity beyond the level of species richness. Besides hotspots, it is also important to note that we also found ecologically rare mammals and birds across many different regions suggesting that different combinations of paleoclimatic

fluctuations, environmental variables and plate tectonic processes (Leprieur et al. 2016) contribute to the emergence of ecological rarity.

What threats are ecologically rare species facing ?

We found that ecologically rare species are disproportionately packed in IUCN threatened groups indicating that ecological rarity is globally under threat (Fig. 4). Yet, 13% and 52% of ecologically rare mammals and birds were classified as “Not Threatened-Least Concern” and 16% and 3.8% were “Not Evaluated or Data Deficient” (such as the Ethiopian big-eared bat, *Plecotus balensis* or the coppery thorn-tail *Discosura letitia*). This finding suggests to use ecological rarity as a complementary conservation facet to identify species that are worthy of particular conservation attention because of their functional distinctiveness and possible specific role on long-term ecosystem functioning (Grenié et al. 2018) even if not threatened in terms of demography (IUCN status). Functional distinctiveness should thus be included as an additional indicator of the IUCN red list, along with population size and geographic range, to better inform conservation decision (Bland et al. 2015, Grenié et al. 2018).

Evidence is accumulating that ongoing climate changes are already affecting living organisms (Dawson et al. 2011, Verges et al. 2016, Pacifici et al. 2017). Taxa with limited geographic distributions are generally highly vulnerable to shifting environmental conditions and are likely to be the most affected (Harnik et al. 2012). Here, we show that ecologically rare species do not escape the rule. Climate change is likely to drastically reduce the geographic range of ecologically rare mammals and birds (Fig. 4), driving some of them towards global extinction. Ecologically rare birds should be more impacted by climate change than mammals due to their over-representation in tropical mountains. These ecosystems are classified as highly vulnerable to climate change impacts (Buytaert et al. 2011), mainly due to the narrow spatial distribution and environmental niche of many taxa but also the morphological and physiological adaptations to live in such stress-limiting environments. We can however clearly distinguish winners and losers among ecologically rare mammals under future climates. For instance *Sylvisorex konganensis*, a small shrew, could double its range extent while *Myrmecobius fasciatus* (a marsupial whose diet consists almost exclusively of termites, Berry et al. 2019) could lose 65% of its current range. Yet our scenarios are based on climate change only. Habitat transformation induced by human land-use is another important driver of the ongoing mammal and bird biodiversity loss but was

not considered in our models (Power & Jetz, 2019). For instance, Madagascar, which hosts 26 ecologically rare mammals have already lost 50% of forest cover between 1950 and 2005 (Harper et al. 2007). More generally, human footprint has a generally negative impact on restriction and fragmentation of species habitats (Betts et al. 2017) and should be included in conjunction with climate change to predict the future distribution of ecological rarity. Since we found that ecologically rare species occur in areas where human footprint is higher than in areas hosting common species (Fig. 4), they should indeed face a multiple jeopardy in the very near future (Powers & Jetz, 2019, Cooke et al. 2019b).

Political and socio-economic factors are vital in conservation strategies and outputs (Daskin & Pringle 2018). Poverty and conflicts contribute to poor governance, unsustainable bushmeat hunting, wildlife trade and anthropization of untouched habitats, thereby potentially increasing pressure on species and their habitats (Smith et al. 2003). Interestingly, we observe a continuum of political and socio-economic contexts across areas where ecological rarity occurs. Both countries with a low number of conflicts and a high human development index (HDI, such as Australia or the United States of America) and countries with a high number of conflicts and a low HDI (such as Indonesia or Madagascar) host ecologically rare species (Supplementary Fig. 8, Supplementary Fig. 9). Even if countries with many conflicts and low HDI host more ecologically rare species, our study suggests a shared and transboundary responsibility for the conservation of ecological rarity between the developing and the developed world (von Bieberstein et al. 2019). Distributing the responsibilities and efforts among countries or regions in a coordinated manner may enhance the cost-effectiveness of conservation strategies targeted towards ecological rarity (Kukkala et al. 2019).

Although the percentage of the land surface devoted to protected areas has globally and markedly increased and despite rarity has motivated research in conservation since decades (von Bieberstein et al. 2019), we show that ecologically rare species generally reach low conservation target achievement (Fig. 4), so are poorly covered by protected areas. This strongly contrasts with the high level of target achievement observed for ecologically common species. In the current global context of scarce resources allocated to conservation, there is now a general consensus that, beyond focusing on the number of species, other facets of biodiversity need to be considered (Pollock et al. 2017). Given the large number of ecologically rare species currently threatened and experiencing population declines, the biosphere will soon be facing a major functional

extinction crisis if effective and targeted actions are not soon implemented. Our results urge to include ecological rarity in priority conservation areas. Achieving the goal of protecting large fraction of ecological rarity seems feasible since we found that a relatively small fraction of the Earth concentrate a large number ecologically rare species (Pimm et al. 2018). A minor but focused increase of protection on these areas could trigger large conservation gains for the global functional diversity (Pollock et al. 2017). In parallel to new targets advocating the protection of 30% (IUCN 2016) or half of Earth (Pimm et al. 2018), there is thus an urgent need and potential reward in protecting the few ecological rarity hotspots.

Limitations and future directions

The estimation of ecological rarity obviously depends on the scale considered (Gaston, 1998, Grenié et al. 2018, Kondratyeva et al. 2019). We chose to work at a global scale because while a species might become locally extinct, its presence elsewhere in the region may prevent definitive extinction. Finer scales (realms, regions or local) and thus smaller species pools could be used and would certainly modify our results. This distinction between global and local scales is especially important since species influence on most ecosystem processes and services are provided locally (Pearson, 2016). Here, we found that functionally distinct mammals and birds (globally) remained strongly distinct at local scale (Supplementary Fig. 11), suggesting that these species potentially support unique functions within ecosystems (Dehling et al. 2016). In the same way, at the species level, we used only species geographic range to estimate geographical restrictiveness because this measure is critically related to global extinction risk (Purvis et al. 2000, Ohlemüller et al. 2008) but other metrics such as habitat specificity and local abundance could also define rarity (Gaston, 1998, Violle et al. 2017). In order to draw the best conservation strategies and preserve ecosystems functions, future directions will thus need to refine the measure of ecological rarity at different spatial scales and include different dimensions of taxonomic rarity. In the context of the ongoing rise in the number of functional trait databases, ecological rarity is an important divergent axis of diversity to consider in conservation while we need to evaluate more thoroughly the impact of ecologically rare species on ecosystem functioning and their contribution to people (Dee et al. 2019).

Conclusion

Our results bring new insights into conservation perspectives through profiling ecologically rare species, i.e. species both located at the margins of the global functional space and geographically restricted. The underlying rationale is that species bearing distinct combinations of traits compared to others and filling a large volume of the functional space should be prioritized in conservation programs owing to their unique potential contribution to ecosystem functioning. We show that the coverage of protected areas for ecological rarity was low, while human pressure and climate change could limit persistence of populations of ecological rare mammals and birds calling for a shared and transboundary responsibility for the conservation of these species.

Methods 1850/3000

All analyses were done using R (R Core Team, 2019) v.3.6.0 (specific functions within specific package are indicated in *italic*). All relevant R code is available from GitHub Repository (see section Data and Code availability).

Distribution data

We used the IUCN range maps for 4,787 terrestrial mammal species and the BirdLife range maps for 9,993 bird species. We removed extinct (EX) and extinct in the wild (EW) species. Given that trait and/or phylogenetic information were not available for all species, we restricted our analyses to 9,287 bird and 4,654 mammal species. Ranges were converted to 50x50 km equal-area grid cells. Mainland cells with > 70% water were excluded but all oceanic island cells with smaller land areas were kept.

Functional Traits

For mammals and birds, four traits (diet, body mass (log transformed), activity cycle and foraging height) were extracted from EltonTraits1.0 (Wilman et al. 2014). These traits are generally assumed to appropriately represent Eltonian niche dimensions of mammals or birds (Wilman et al. 2014, Cooke et al. 2019a). These traits have already been used to investigate community assembly rules and biogeography of both taxa (Barbet-Massin & Jetz 2015, Mazel et al. 2017). We assessed how functional traits relate to rarity indices through a Principal Coordinates Analysis (PCoA) of the functional distances (*pcoa()* function in the *ape* package v.5.2, Paradis et al. 2004). We then projected the traits, using appropriate scaling, onto the first four axes of the PCoA. The axes were selected as their explained variance was greater to the null expectation of a broken stick model (Legendre and Legendre, 2012).

Phylogenetic data

We used randomly selected 100 phylogenies from the posterior distribution of phylogenies from Bininda-Emonds et al. (2007) with updates (Fritz et al. 2009 & Kuhn et al. 2011) for mammals, and the phylogenies from Jetz et al. (2012) for birds.

Ecologically Rarity

We computed functional distances between species using the *dist.ktab()* function in *ade4* v.1.7-6 (Dray and Dufour, 2007, Pavoine et al. 2009) that includes ordinal and fuzzy traits in the distance metric computation. Using *funrar()* function in *funrar* v.1.3-0 (Grenié et al. 2017) we then computed functional distinctiveness D_i of species i in the global functional space (Violle et al. 2017, Grenié et al. 2017), representing how distinct the traits of a given species are compared to all the other species from the same taxa :

$$D_i = \frac{\sum_{j=1, j \neq i}^N d_{ij}}{N - 1} \quad \text{Equation. 1}$$

where d_{ij} the functional distance between species i and j , N the total number of species. The functional distances d_{ij} are scaled between 0 and 1. D_i is the average functional distance between the species of interest and all the other species of the pool. It captures how much the traits of the species of interest differ, on average, compared to the rest of the species pool. D_i is 0 when all species in the set have the same trait values (the functional distance between all species is 0), and 1 when species i is maximally different to other species.

We also computed geographical restrictedness (Violle et al. 2017, Grenié et al. 2017) using the grid cell by species matrix to measure how restricted the distribution of a species is,

$$R_i = 1 - \frac{K_i}{K_{tot}}, \quad \text{Equation. 2}$$

where R_i is the geographical restrictedness of species i , K_i the number of cells where species i is present and K_{tot} the total number of cells. R_i is close to 1 for a species present in a single cell and 0 for a species present in all cells.

Species were classified into three groups regarding their values and the quantile partitions of the bivariate space functional distinctiveness vs geographical restrictedness (Supplementary Fig. 1). We defined *ecologically rare* and *ecologically common* species as having values of functional distinctiveness and geographical restrictiveness either higher than 75% or lower than 25% of the entire species pool of interest. *Ecologically average* species have values of functional distinctiveness and geographical restrictiveness respectively, lower than 75% and higher than 25%.

To measure phylogenetic signal of ecological rarity we computed, on the 100 phylogenetic trees, the D index (Fritz & Purvis, 2010), using the *phylo.d()* function in the *caper*

v.2.0.6 package (Orme et al. 2012). The D index is equal to 1 if ecological rarity has a phylogenetic random distribution and 0 if ecological rarity is clumped into the phylogeny. We also computed Evolutionary Distinctiveness of species (Isaac et al. 2007). The Evolutionary Distinctiveness of species i , ED $_i$, is high when the species has a long unshared branch length with all the other species. The more “isolated” a species is in a phylogenetic tree, the higher its evolutionary distinctiveness. We computed ED using the `evol.distinct()` function from *picante* v.1.6-2 package (Kembel et al. 2010).

Null model

To distinguish the effects of the number of species on the number of ecologically rare species per cell, we simulated the distribution of ecologically rare species that would be expected under the null expectation that they are randomly distributed among cells, regardless of the total number of species. For both mammals and birds, we generated 1000 random assemblage matrices using the curveball algorithm using `nullmodel()` function from *vegan* v2.4-2 packages (Strona et al. 2014). This algorithm maintains species occurrence frequency and sample species richness while shuffling species co-occurrence across cells. Classes of ecological rarity (rare, common, average, others) were kept constant. For each randomization, we calculated the number of ecologically rare species per cells. Standardized effect sizes (SES) were obtained by comparing the observed number of ecologically rare species and the mean and standard deviation of the null distributions.

$$SES = \frac{\text{Observed} - \text{Mean}(\text{Null})}{sd(\text{Null})} \quad \text{Equation. 3}$$

Values above the null expectation indicate that the cell contains more ecologically rare species than expected given the number of species in the cell, and vice-versa.

IUCN status

We used the *taxize* package v.0.7.8 (Chamberlain et al. 2016, Chamberlain & Szocs, 2013) to retrieve up-to-date IUCN status for mammals and birds (IUCN, 2017). For easier interpretation, we grouped species in three categories: Critically Endangered (CR), Endangered (EN) and Vulnerable (VU) species as “Threatened” (TH); Least Concern (LC) and Near Threatened (NT) species as “Least Concern” (LC); Data Deficient (DF) and Not Evaluated

species as “Not Evaluated” (NE). We performed a multiple comparison Kruskal-Wallis rank-sum test to compare indices across IUCN groups. We used the *kruskal()* function of the *agricolae* package v.1.2-6 (Mendiburu, 2014).

Exposure to human pressure

For each species, we assigned three indicators of exposure to human pressure. (a) Global human footprint, that measure the cumulative impact of direct human pressures (extent of built environments, crop and pasture lands, population density, night-time lights, railways, roads and navigable water-ways) on nature between 1993 and 2009 (Venter et al. 2016). (b) Human development Index (HDI), that capture elements of life expectancy, education and wealth for the year 2017 (Human Development Indicators and Indices : 2018). (c) Human conflicts which sum the years of conflicts in each country between 1946 and 2015 (UCDP/PRIO Armed Conflict Dataset v. 4-2016 from the Version 2.2-2016). Each dataset was rescaled at 50km x 50km resolution and we computed the average value over each species spatial distribution.

Global biodiversity scenarios

We used a similar approach as Thuiller *et al.* (2019) and assessed potential climate change impacts on mammal and bird species under climate change scenarios using an ensemble projection framework. More specifically, we related species distribution to four climate variables describing current climate (1979-2013) derived from the CHELSA dataset (Karger *et al.* 2017): annual mean temperature, annual temperature range, annual sum of precipitation and precipitation seasonality.

Four Species Distribution Models (SDM) algorithms available in the *biomod2* package (Thuiller *et al.* 2009) were used to estimate these species-climate relations: Generalized Linear Model, Generalized Additive Model, Boosting Regression Trees and Random Forest. Models were calibrated using 80% of the initial data and evaluated against the remaining 20% of data using the True Skill Statistic (TSS, Allouche *et al.* 2006). Data were randomly assigned to each sub-dataset and this step was repeated four times in order to perform a robust cross-validation. Only models with a TSS > 0.8 were kept and projected into future conditions. Consequently, we kept 28% (67 species) and 59% (337 species) of ecologically rare mammals and birds, respectively. To consider realistic species dispersion, we selected absences in 3000 and 4000 km

buffer around mammal and bird species range, respectively (see Thuiller *et al.* (2019) for further details).

These models were then projected under future climate change scenarios derived from five Global Circulation Models (GCM) run under the Representative Concentration Pathway (RCP) 8.5 from the Coupled Model Intercomparison Project (CMIP5). These GCM were the following: the CESM1-BGC (Lindsay *et al.* 2014) run by the National Center for Atmospheric Research (NCAR), the CMCC-CMS (Scoccimarro *et al.* 2011) run by the Centro Euro-Mediterraneo per i Cambiamenti Climatici (CMCC), the CM5A-LR (Persechino *et al.* 2013) run by the Institut Pierre-Simon Laplace (IPSL), the MIROC5 (Watanabe *et al.* 2010) run by the University of Tokyo, and the ESM-MR (Giorgetta *et al.* 2013) run by Max Planck Institute for Meteorology (MPI-M). Future projections were made for two time slices: 2041-2060 (horizon 2050) and 2061-2080 (horizon 2070). For a given horizon, projections were aggregated using the weighted average consensual approach (Marmion *et al.* 2009) using the TSS as weight.

Potential climate change impacts were assessed by comparing current and future species distribution projection using the Species Range Change (SRC, Thuiller *et al.* 2009).

Levels of protection and gap analysis

To estimate the extent to which the current terrestrial protected area network covers ecologically rare species, we carried out a gap analysis following the methodology proposed in Thuiller *et al.* (2015). We defined a conservation target for every single species, which, in terms of distribution range within the protected areas network, represents the desired level of protection we considered necessary for a species to be adequately protected. Species-specific targets were defined based on species range sizes since restricted species need more coverage than widespread ones to avoid extinction (Harnik *et al.* 2012, Thuiller *et al.* 2015). Accordingly, species-specific conservation targets were set to be inversely proportional to log-transformed species' range sizes. We hypothesized that the species with the most restricted ranges required 100% of their range to be covered, whereas the widespread species only required 10% (Rodrigues *et al.* 2004). We fitted a linear regression between these two extremes to define the target for the remaining species (see the electronic supplementary material of Thuiller *et al.* 2015). This was carried out for the two groups separately since they harbor very different distribution of range sizes. Geographic range size data (in km²) was estimated in the Behrmann projection using species' extent of occurrence

polygons from the IUCN (2016). We then extracted the proportion of range currently covered for each species from the World Database on Protected Areas (WDPA) to estimate how far species met their defined targets (species target achievement), i.e. by dividing this proportion by the defined target. Note that we restricted analyses to protected areas classified as strict protected areas, i.e. Ia, Ib and II by IUCN.

Acknowledgments

This research is a product of the FREE group funded by the synthesis center CESAB of the French Foundation for Research on Biodiversity (FRB) and EDF. WT received funding from the ERA-Net BiodivERSA - Belmont Forum, with the national funder Agence Nationale pour la Recherche (ANR-18-EBI4-0009), part of the 2018 Joint call BiodivERSA-Belmont Forum call (project 'FutureWeb'). CV was supported by the European Research Council (ERC) Starting Grant Project 'ecophysiological and biophysical constraints on domestication in crop plants' (grant ERC-StG-2014-639706-CONSTRAINTS).

Data availability

All data used in this manuscript have already been published.

Code availability

All relevant R code is available from the GitHub Repository: <https://github.com/FRBCesab/ecorar>

Discussion

1 DÉVIER DE L'OPTIMALITÉ FONCTIONNELLE

1.1 DES PROCESSUS SPÉCIFIQUES

L'optimalité fonctionnelle représente l'idée qu'il existe une seule valeur optimale de trait par communauté sous l'hypothèse d'un filtre environnemental Gaussien uni- ou multi- dimensionnel. Dans les parties 1.2, 2.1, et 2.2 nous avons montré que certains mécanismes pouvaient faire dévier des individus de l'optimalité écologique attendue dans la communauté. Je vais maintenant détailler certaines implications de ces résultats.

Le modèle *fdcoexist* présenté dans la partie 1.2 a montré qu'en fonction de la contribution des traits aux différents mécanismes d'assemblages, et à l'intensité de ceux-ci les espèces pouvaient dévier de leur environnement optimal et que les traits à l'échelle des communautés, mesurés en termes de moyennes pondérées par les abondances, pouvaient dévier de l'optimum fonctionnel. Nous avons notamment montré que la compétition hiérarchique, du fait de son asymétrie, pouvait faire dévier les communautés de l'optimum fonctionnel, malgré l'influence du filtre environnemental. En revanche, la limite à la similarité est un processus qui affecte les traits des espèces de manière symétrique (Abrams, 1983). Comme les espèces interagissent les unes sur les autres de manière symétrique, la valeur de trait moyen de la communauté ne devrait pas varier. Et si cette valeur correspond à un optimum fonctionnel, alors le trait moyen de la communauté ne sera pas différent de cet optimum. La contribution des traits aux différents processus modifie cependant nos attendus. Même si la littérature traite abondamment la question de la multi-dimensionnalité de la niche en utilisant plusieurs traits fonctionnels, la question de la contribution de chacun de ces traits aux différents processus n'est que rarement posée (Reich, 2014; Kraft, Godoy, & Levine, 2015). Ces résultats sont à relier aux résultats de la partie 2.2 où nous avons testé l'effet de l'augmentation du nombre de traits non-filtrés par l'environnement sur la relation attendue entre diversité fonctionnelle et environnement. Dans les simulations utilisées, un seul trait était filtré par l'environnement le long d'un gradient environnemental tandis que les autres traits n'étaient affectés par aucun processus le long de l'environnement. Nous avons montré que l'ajout d'un grand nombre de traits n'étant pas affectés par le filtre environnemental peut nuire à la détection du filtre environnemental en utilisant la diversité fonctionnelle. Cela correspond dans le modèle

fdcoexist à une situation où l'on ne considérerait que le filtre environnemental et beaucoup de traits n'y contribuant pas.

Dans ces deux schémas de simulations, il serait possible de tester d'autres mécanismes qui pourraient expliquer une déviation de l'attendu d'une seule valeur optimale de trait par environnement. Par exemple les filtres environnementaux disruptifs ne possèdent pas une seule valeur optimale de trait (Loranger *et al.*, 2018). Plus généralement les filtres environnementaux multi-modaux ne présentent pas non plus une seule valeur optimale de trait (Laughlin *et al.*, 2015 ; Laughlin & Joshi, 2015 ; Laughlin & Messier, 2015). Dans le cas d'un filtre environnemental à deux modes, cela pourrait correspondre à deux guildes fonctionnelles qui représentent deux stratégies possibles pour se maintenir dans la communauté (Ravel *et al.*, 2018).

La [partie 2.1](#) s'intéressait à l'effet de la saturation d'habitat sur la prédiction de richesse spécifique à l'aide de modèles de distribution d'espèces empilés (*Stacked-Species Distribution Models*, S-SDM). On entend ici la saturation d'habitat comme le fait que les espèces, individuellement, occupent l'ensemble de l'habitat qui leur est favorable. Si *a priori* cette partie ne s'intéresse pas directement à la question de l'optimalité fonctionnelle, elle traite bien de l'étude d'un processus qui pourrait amener des espèces à ne pas représenter fidèlement l'influence du filtre environnemental. Si les espèces saturent complètement leurs habitats, alors la richesse spécifique est prédite correctement par l'empilement des SDM. Si les espèces ne saturent pas toutes leurs habitats, alors la richesse prédite dévie de la richesse théorique, car les SDM supposent dans leur fonctionnement même que les espèces saturent leurs habitats. Un des points intéressants de ce modèle est qu'il ne fait aucune hypothèse sur ce qui pourrait causer la non-saturation d'habitat des espèces. Les résultats sont donc valides quels que soient les mécanismes expliquant cette non-saturation. Les processus neutres qui font varier aléatoirement l'abondance des espèces et donc leurs présences dans certaines communautés peuvent être une source de non-saturation (Hubbell, 2001). Cela rejoint aussi les études de modèles de méta-populations et méta-communautés, qui montrent qu'en fonction des probabilités de colonisation et d'extinction, certaines espèces n'occupent pas tout l'habitat qui leur est favorable (Hanski & Hanski, 1999). La dispersion limitée, c'est-à-dire le fait que les capacités de dispersion d'une espèce soient limitées dans l'espace, peut empêcher une espèce d'atteindre certains habitats favorables et causer de la non-saturation (Munguía, Peterson, & Sánchez-Cordero, 2008). Au contraire, les effets source-puits peuvent générer de la sur-saturation, avec des espèces présentes dans des habitats non-favorables, ce qui va aussi biaiser les prédictions des S-SDM (Pulliam, 1988 ; Pulliam & Danielson, 1991).

1.2 APPORTS DES APPROCHES DE MODÉLISATION

Dans le [premier chapitre](#) nous avons mis en place deux nouveaux outils de modélisation pour mieux comprendre les processus qui structurent les communautés et en particulier ceux qui favorisent une déviation de l'optimalité. Ces deux modèles sont de nature différentes tant dans leur construction que dans leurs utilisations. Le premier, *ecolottery*, est un modèle stochastique et probabiliste de l'assemblage des communautés, spatialement-implicite, basé sur les individus et utilisant la théorie du coalescent. Au contraire, le deuxième

modèle présenté, intitulé *fdcoexist* est un modèle déterministe et spatialement implicite de dynamique des populations. Ces deux modèles peuvent être utilisés comme des expériences virtuelles pour tester l'effet de processus précis sur l'assemblage des communautés ou générer des prédictions à partir des processus (Peck, 2004; Zurell *et al.*, 2010). Comme je l'ai souligné en Introduction, le fait de pouvoir manipuler indépendamment les processus d'assemblages des communautés permet de tester des effets qui seraient complexes à analyser sur le terrain. En outre, les expériences virtuelles peuvent être vues comme des « expériences de pensée » qui servent à générer de nouvelles hypothèses en modulant le nombre et le type de processus d'assemblage des communautés.

Ainsi nous avons utilisé le modèle *ecolottery* dans les deux parties du [deuxième chapitre](#) pour montrer l'influence de processus qui font dévier les communautés de l'optimalité attendue. Démontrer les effets de la non-saturation d'habitat d'une part, et des nombreux traits non affectés par le filtre environnemental d'autre part, à l'aide d'expérimentations aurait été impossible au vu de la quantité de communautés à manipuler. La démarche expérimentale est cruciale en écologie des communautés pour tester l'effet de certains processus. On observe d'ailleurs de nombreuses expériences menées sur des micro-organismes (Nemergut *et al.*, 2013). Mais le grand nombre d'individus impliqués dans les communautés rend les expériences difficiles à mettre en place pour la plupart des macro-organismes. Les modèles peuvent donc servir à générer des hypothèses dans des domaines hors de portée des expériences.

En outre le modèle *ecolottery* permet d'inférer les paramètres des processus d'assemblages des communautés. Pour ce faire, on fixe des processus qui ont lieu *a priori* dans la ou les communauté(s) d'intérêt, on simule ensuite un grand nombre de communautés avec une grande gamme de paramètres pour ces processus d'assemblage. On compare les simulations et les communautés observées avec des statistiques résumées (richesse spécifique, moments de la distribution de traits, distribution de l'abondance, etc.). Enfin on utilise l'*Approximate Bayesian Computation* (ABC, calcul approché bayésien), qui compare les statistiques résumées et permet d'explorer l'espace des paramètres utilisés dans les simulations pour obtenir une distribution *a posteriori* de la valeur des paramètres d'assemblages (Barthelemy *et al.*, in revision; Beaumont, 2010). On peut donc inférer la présence ou non de certains processus et leur intensité en utilisant ce type de modèle (Pontarp, Brännström, & Petchey, 2019). Il s'agit d'un retournement de la vision qui à un patron lie un processus sans lien mécaniste explicite. En effet, dans la mesure où plusieurs combinaisons de processus peuvent donner des patrons identiques, il est difficile d'inférer les processus simplement à partir des patrons (Mayfield & Levine, 2010; Cadotte & Tucker, 2017). En revanche, si l'on simule les processus explicitement, ce sont les combinaisons de processus dans les simulations qui génèrent les patrons observés (Pontarp, Brännström, & Petchey, 2019). En comparant les statistiques résumées des simulations et les communautés observées, on peut réaliser des tests statistiques comme la validation croisée qui permettent de juger les effets confondants et le pouvoir statistique associés aux différents processus. Il est alors possible d'inférer les combinaisons de processus qui pourraient affecter les communautés d'intérêt. Ainsi ce modèle est une contribution importante pour l'écologie des communautés qui va dans le sens d'une compréhension davantage mécaniste des communautés.

Au contraire *fdcoexist* nous sert ici d'expérience virtuelle pour réfléchir aux processus qui structurent les communautés. C'est un modèle de dynamique de populations qui modélise explicitement la croissance des populations en fonction des traits. En cela, ce modèle connecte les champs habituellement éloignés de l'écologie des populations et de l'écologie des communautés (Salguero-Gómez *et al.*, 2018). L'originalité de ce modèle réside aussi dans la formalisation explicite de deux processus de compétition (limite à la similarité et compétition hiérarchique) en fonction des traits des espèces concernées ainsi qu'à la possibilité pour les traits de contribuer de façon variable aux différents processus. *fdcoexist* permet d'explorer la dynamique temporelle des espèces, en différenciant par exemple des dynamiques précoces et tardives, même si avec l'approche « *forward* » de *ecolottery* il est aussi possible d'étudier la dynamique des espèces. Le modèle *fdcoexist* pourrait *a priori* aussi utiliser l'ABC pour inférer les paramètres des communautés observées. Seulement, le modèle *fdcoexist* est bien moins parcimonieux en paramètres que le modèle *ecolottery*. Par exemple, *fdcoexist* utilise des coefficients de compétition entre chaque paire d'espèces à partir de leur différence de valeur de trait. Pour une communauté comprenant n espèces, il faut estimer 2^n coefficients, ce qui rend le problème computationnellement difficile. Le grand nombre de paramètres impliqué peut aussi poser des problèmes d'identifiabilité, c'est-à-dire que plusieurs combinaisons différentes de valeurs de paramètres peuvent donner le même résultat (Hubbard, 2014). Les coefficients de contributions des traits aux processus qu'utilise *fdcoexist* sont des abstractions liées au modèle et ne nous sont pas accessibles dans la réalité. Si l'on voulait utiliser l'ABC avec *fdcoexist* il faudrait donc effectuer un balayage très large dans la gamme des paramètres de contribution des traits aux différents processus. Il serait peut-être possible d'utiliser l'ABC avec *fdcoexist* en faisant des hypothèses simplificatrices. Par exemple, en supposant que la compétition hiérarchique et la limite à la similarité n'ont d'effet que sur les espèces les plus proches. Cela permettrait de fixer des coefficients dans les matrices de compétition entre espèces à zéro, et donc de diminuer l'espace des paramètres à explorer.

Un intérêt supplémentaire du modèle *fdcoexist* est d'avoir considéré différentes métriques de performance locale des espèces. En écologie fonctionnelle, on utilise généralement l'abondance des espèces comme un indice de leur performance locale. Seulement, l'abondance peut ne pas être le reflet de l'adaptation des espèces aux conditions environnementales locales. La théorie des méta-communautés a par exemple montré l'importance des effets source-puits qui peuvent amener une espèce à disperser fortement dans un environnement où elle n'est pas forcément adaptée. Par la dispersion continue de cette espèce depuis une communauté « source », l'espèce peut devenir très abondante dans la communauté « puits » sans qu'elle soit adaptée à l'environnement de cette communauté. Ceci peut remettre en question l'hypothèse de l'optimalité fonctionnelle qui suppose que la distribution des abondances des espèces dépend de leur adéquation à l'environnement. Il ne suffit donc pas d'identifier des relations trait-environnement à l'aide du trait moyen des communautés pondéré par les abondances (CWM) pour identifier les filtres environnementaux.

En écologie des populations, plutôt que l'abondance locale des espèces, c'est le taux de croissance des populations qui est utilisé comme indice de performance locale (Salguero-Gómez *et al.*, 2018). Cet indicateur de performance n'est pas souvent utilisé en écologie fonctionnelle car il nécessite d'avoir des suivis temporels

des populations d'espèces pour pouvoir être calculé. *fdcoexist* étant un modèle de populations il est facile d'obtenir les taux de croissance des populations. Nous avons donc calculé trois taux de croissance différents : le taux de croissance moyen observé, qui correspond à la moyenne des taux d'accroissement sur toute la durée de simulation, le taux de croissance maximal, et le taux de croissance intrinsèque qui correspond à la partie linéaire de la courbe de croissance. Nous avons montré grâce à ce modèle que ces trois taux de croissance, ne sont pas nécessairement corrélés entre eux, ni corrélé avec l'abondance observée des espèces, en particulier si la compétition hiérarchique est forte. On ne peut pas donc recommander l'utilisation d'un seul de ces estimateurs pour définir la performance locale des espèces. L'intérêt du modèle *fdcoexist* est donc d'utiliser des outils de l'écologie des populations pour répondre à une question d'écologie fonctionnelle des communautés, ce qui permet notamment d'interroger la notion de performance locale.

ecolottery est un modèle très flexible, qui permet de tenir compte des processus neutres et non-neutres, qui permet à la fois de réaliser des expériences virtuelles et, grâce à sa performance computationnelle notamment, d'inférer les processus dans les communautés. *ecolottery* modélise les processus à l'échelle de la communauté en considérant un jeu à somme nulle avec un nombre d'individus fixes dans la communauté. À l'inverse, *fdcoexist* est un modèle déterministe, dont les processus sont définis à l'avance dans la structure du modèle, qui nous sert avant tout d'expérience virtuelle. *fdcoexist* modélise les dynamiques des populations de chaque espèce séparément avec des limites biotiques pour chaque espèce. Ainsi, le nombre d'individus dans la communauté telle que modélisée par *fdcoexist* peut fluctuer, tandis qu'il est fixe pour *ecolottery*. *fdcoexist* modélise directement les relation trait et taux de croissance dans le modèle, on est donc assuré d'avoir un trait véritablement fonctionnel, *ecolottery* ne modélise pas explicitement ce lien, mais considère que les traits influencent les probabilités d'immigration, de mort et de reproduction (et donc *in fine* la valeur sélective de l'espèce). *ecolottery* s'inscrit davantage dans une tradition de modèles de (méta-)communautés, décrivant les processus à l'échelle de la communauté (Hanski & Hanski, 1999; Leibold & Mikkelsen, 2002), tandis que *fdcoexist* est un modèle de dynamique de populations avec un formalisme plus classique (Lotka, 1927; Volterra, 1928; Beverton & Holt, 1957), plus restreint mais qui intègre des outils de l'écologie des populations et tente de lier écologie des communautés et écologie des populations (Salguero-Gómez *et al.*, 2018; Zakharova, Meyer, & Seifan, 2019; Laughlin *et al.*, n.d.). Ainsi ces deux modèles de nature distincte sont complémentaires dans les éclairages qu'ils apportent à la question de la coexistence des espèces et des processus d'assemblage des communautés.

2 L'ORIGINALITÉ FONCTIONNELLE

2.1 RARETÉ FONCTIONNELLE OU ORIGINALITÉ FONCTIONNELLE ?

En [Introduction](#) nous avons évoqué le cadre proposé par Violle et collab. (2017), inspiré de Rabinowitz (1981) qui permet de définir différentes facettes de la rareté fonctionnelle. Ce cadre repose sur deux dimensions fondamentales : d'une part, l'originalité fonctionnelle (la « rareté » en trait), qui se décompose à deux

échelles spatiales (locale et régionale); d'autre part, la rareté basée sur l'abondance locale et la fréquence régionale (Gaston, 1997). La rareté fonctionnelle correspond à une combinaison ces deux dimensions à deux échelles spatiales : dans ce cadre, il existe douze formes différentes de rareté fonctionnelle en fonction de l'originalité fonctionnelle (locale et régionale) de l'espèce et de sa rareté (locale et régionale). Par exemple, une espèce qui est localement originale et abondante mais régionalement commune et peu fréquente représente une forme de rareté fonctionnelle, même si elle n'est originale qu'à l'échelle locale. En fonction des parties de la thèse nous nous sommes concentrés soit uniquement sur l'originalité fonctionnelle (partie 3.3) soit sur la rareté fonctionnelle (partie 3.1, 4.1, et 4.2). L'intérêt initial de croiser ces deux dimensions, *a priori* indépendantes, était de permettre d'identifier des espèces susceptibles d'être menacées (du fait de leur rareté) et ayant une contribution originale en termes de traits. Les espèces fonctionnellement rares seraient donc à prioriser dans les efforts de conservation. C'est pour cela que nous nous intéressons à la rareté fonctionnelle dans les parties 4.1 et 4.2 traitant problématiques de conservation.

Dans les différentes parties traitant de l'originalité fonctionnelle, nous nous sommes concentrés sur l'originalité fonctionnelle globale des espèces (mesurée à l'échelle d'un bassin régional d'espèces). Cela permet de mettre en avant les grands gradients de différenciation phénotypique, et les combinaisons de traits qui sont originales par rapport à toutes les autres espèces considérées. Des espèces originales à l'échelle globales représenteraient donc les plus originales possibles. L'originalité fonctionnelle à cette échelle représente l'ensemble du spectre de diversification phénotypique résultant des dynamiques biogéographiques et évolutives. Utiliser l'originalité fonctionnelle à l'échelle globale est donc pertinent pour appliquer une démarche de biogéographie de la conservation (Richardson & Whittaker, 2010; Ladle & Whittaker, 2011).

Une des questions que nous avons traitée dans les différentes parties était de savoir si les dimensions de la rareté fonctionnelle étaient corrélées, c'est-à-dire de savoir si les espèces originales étaient rares et les espèces communes étaient abondantes/fréquentes. Dans la partie 3.1, sur les mammifères d'Amérique du Nord, nous n'avons pas observé de corrélation entre originalité fonctionnelle et taille de l'aire de répartition des espèces. Tandis que dans la partie 3.3 nous avons observé une légère corrélation négative entre originalité fonctionnelle et aire de répartition chez un sous-ensemble de plantes des Amériques. Dans la partie 4.1 nous avons observé une légère corrélation positive entre originalité et taille de l'aire de répartition chez les poissons de récifs coralliens du monde entier. Enfin dans la partie 4.2 nous avons observé une corrélation positive entre originalité fonctionnelle et aire de répartition chez les oiseaux et chez les mammifères à l'échelle du monde. Dans la littérature, d'autres études ont observé des relations variables entre originalité fonctionnelle et aire de répartition à des échelles diverses et sur des organismes variés (Chapman, Tunnicliffe, & Bates, 2018; Lamothe, Alofs, & Chu, 2019; Cooke, Eigenbrod, & Bates, 2020). Ainsi nos résultats ne montrent pas de relation universelle entre originalité fonctionnelle globale et aire de répartition. Les espèces ayant de petites aires de répartition sont souvent les plus menacées d'extinction, ce sont aussi les cibles prioritaires des programmes de conservation. Nous n'attendons donc pas de relation générale entre originalité fonctionnelle et statut de conservation basé sur les aires et dynamiques de distribution.

Sous l'hypothèse d'optimalité fonctionnelle certaines espèces peuvent être rare localement en abondance ou

régionalement en fréquence, sans que cela soit lié au fait de posséder des traits particuliers. Par exemple, dans le cadre d'un mécanisme source-puits, une espèce sera localement rare car dispersant d'une communauté source éloignée. De même, dans des communautés comprenant un faible nombre d'individus, des processus démographiques stochastiques peuvent réduire la taille de population de certaines espèces sans lien avec leurs traits. Les mécanismes de fréquence-dépendance négatives à l'échelle locale, c'est-à-dire où le taux de croissance d'une espèce est inversement lié à son abondance dans la communauté, peuvent aussi maintenir les espèces à des faibles tailles de populations sans forcément avoir de liens avec leurs traits (Yenni, Adler, & Ernest, 2016). À larges échelles, les espèces peuvent ne pas avoir eu le temps de (re-)coloniser tout l'habitat qui leur est favorable (cf. partie 2.1) du fait, par exemple, de changements rapides de l'environnement (Svenning *et al.*, 2015). C'est ce que l'on observe dans les dynamiques de re-colonisation des arbres en Europe depuis la dernière glaciation, beaucoup d'espèces n'occupant pas tout l'habitat qui leur est favorable du fait de leur dynamique de dispersion très lente (Svenning & Skov, 2004). Gaston (2009) propose une synthèse des facteurs qui peuvent limiter l'aire de répartition d'une espèce : immigration, reproduction, mortalité, émigration, et taille de la population. Il pointe que ces facteurs peuvent être indépendants des traits des espèces, comme l'immigration qui dépend fortement de la taille de population de la population source. La mortalité en limite d'aire de répartition en revanche, peut s'expliquer par l'absence d'adaptation des espèces aux conditions environnementales (Gaston, 2009). Les interactions entre environnement, traits des espèces, interactions biotiques, et adaptation rendent donc difficile l'interprétation des limites des aires de répartition (Holt & Keitt, 2000; Case *et al.*, 2005). Si les traits peuvent déterminer certains de ces facteurs, identifier les traits impliqués est difficile et l'originalité fonctionnelle telle que calculée dans cette thèse n'est *a priori* pas corrélée à ces facteurs. Cependant, la complexité des interactions et rétroactions en place peut expliquer la diversité des résultats obtenus entre études différentes.

2.2 QUELS TRAITS UTILISER POUR CALCULER L'ORIGINALITÉ FONCTIONNELLE ?

Pour quantifier l'originalité fonctionnelle, il faut choisir les traits sur lesquels la calculer. Dans cette thèse, nous avons choisi d'utiliser des traits qui définissent les stratégies globales des espèces (Westoby, 1998; Díaz *et al.*, 2016). Seulement, il aurait été possible de calculer ces indices sur d'autres traits. En effet le lien entre les traits et la valeur sélective des espèces n'est pas forcément évident dans tous les environnements (Shipley *et al.*, 2016). Pour que ce lien soit prouvé il faut avoir réalisé des études qui démontrent que ce trait est bien lié à la performance des espèces dans des environnements distincts. Si les traits utilisés pour calculer l'originalité fonctionnelle ne sont pas « fonctionnels » alors l'originalité ne décrira plus une position originale dans le paysage de valeur sélective multidimensionnel (Laughlin & Messier, 2015). Il sera alors plus difficile d'examiner les causes.

Une typologie des traits fonctionnels différencie les traits d'effet et les traits de réponse (Lavorel & Garnier, 2002). Les traits d'effet sont les traits associés à l'effet des espèces sur le fonctionnement des écosystèmes, comme par exemple leur capacité à recycler les nutriments (p. ex., la capacité à fixer de l'azote); alors que les

traits de réponse sont les traits associés à la réponse des espèces à des contraintes de l'environnement (p. ex., la surface spécifique foliaire, qui décrit la stratégie d'acquisition de la lumière d'une plante). Ces catégories ont été très utilisées chez les plantes et peuvent être largement corrélées (Lavorel & Garnier, 2002). Par exemple la surface spécifique foliaire est un trait de réponse car elle définit la stratégie de conservation/acquisition des ressources de l'individu selon la disponibilité des ressources; il peut s'agir aussi d'un trait d'effet car la valeur de surface spécifique foliaire dépend de la quantité de carbone fixée par la plante et donc celle fixée par l'écosystème entier. Pour d'autres taxons ces catégories peuvent être plus difficiles à distinguer (Luck *et al.*, 2012; Winemiller *et al.*, 2015; Villéger *et al.*, 2017). Calculer l'originalité fonctionnelle sur l'une ou l'autre des catégories de traits n'aura pas la même signification. L'originalité fonctionnelle calculée sur des traits de réponse aura tendance à nous indiquer les espèces qui possèdent une réponse originale à l'environnement. À l'inverse, calculer l'originalité fonctionnelle sur des traits d'effet aura plutôt tendance à nous indiquer les espèces qui ont un effet original sur les écosystèmes et donc une contribution originale au fonctionnement des écosystèmes. Dans la [partie 3.3](#) nous avons calculé l'originalité fonctionnelle des plantes des Amériques en lien avec des traits de réponse à l'environnement plutôt que d'effet. Tandis que dans les parties [4.1](#) et [4.2](#) nous avons utilisé des traits trophiques des poissons de récifs coralliens, des oiseaux et des mammifères qui peuvent être considérés comme des traits d'effets. Pour mieux comprendre le rôle des espèces originales dans le fonctionnement des écosystèmes il pourrait être pertinent de calculer l'originalité fonctionnelle sur des traits d'effet. Il serait alors possible d'estimer la contribution des espèces originales au fonctionnement des écosystèmes. En particulier, dans le cadre de la conservation des espèces, c'est peut-être davantage l'originalité fonctionnelle des traits d'effet qu'il faudrait prioriser plutôt que celle des traits de réponse.

Dans cette thèse, nous avons notamment travaillé sur des gradients biogéographiques en utilisant des traits déjà définis, étudiés, et disponibles (Díaz *et al.*, 2016). Par exemple dans la partie 3.2 nous avons étudié la distribution de l'originalité fonctionnelle des plantes à l'échelle des Amériques. Pour ce faire nous avons choisi d'utiliser les trois traits du schéma LHS [pour *Leaf*, *Height*, et *Seed Mass*; Westoby (1998)]. Ces trois traits réduisent la multi-dimensionnalité du phénotype des plantes à trois axes uniquement (Laughlin, 2014). Ces trois traits, par exemple, ne décrivent pas du tout la phénologie des plantes, c'est-à-dire les moments où ont lieu les phases du cycle de vie de ces plantes (Segrestin, Navas, & Garnier, 2020). Si la similarité fonctionnelle représente bien une similarité de niche, alors deux plantes qui possèdent exactement les mêmes traits dans le LHS ne devraient pas coexister d'après le principe d'exclusion compétitive (Gause, 1934). Mais si elles possèdent des phénologies différentes, en se développant à des moments différents dans l'année, alors elles peuvent ne pas entrer en compétition et coexister sur une année (Kraft, Godoy, & Levine, 2015). Les traits utilisés par Díaz et collab. se concentrent sur des dimensions d'économie des ressources et négligent les aspects de reproduction et d'interactions biotiques (Díaz *et al.*, 2016). Afin de mieux saisir l'importance de l'originalité fonctionnelle, il faudrait augmenter la dimensionnalité de nos analyses afin d'évaluer le plus grand nombre de facettes distinctes des niches des espèces d'intérêt (Winemiller *et al.*, 2015).

Les traits qui reflètent le mieux la performance locale des espèces dépendent du contexte environnemental dans lesquels on étudie des espèces (Shipley *et al.*, 2016). Par exemple, les plantes pyrophytes qui nécessitent

d'être exposées au feu pour se reproduire sont fonctionnellement communes dans les écosystèmes qui subissent des incendies réguliers, comme sur le pourtour méditerranéen. Pourtant, si on les compare à l'ensemble des plantes du monde, cette capacité à se reproduire après un feu est plutôt rare, et donc, ces plantes deviennent fonctionnellement originales à cette échelle-là. Or, nous ne disposons pas de la caractéristique « pyrophytique » pour l'ensemble des plantes du monde, étant donné que ce trait n'est pertinent que dans les écosystèmes avec des feux récurrents. Donc, on ne pourra pas comparer les plantes pyrophytes des non-pyrophytes sur cette caractéristique, et on ne pourra pas souligner l'originalité de la stratégie pyrophyte à l'échelle globale. Une approche consisterait à supposer les traits des espèces en fonction des environnements dans lesquelles elles sont présentes. Par exemple, si des espèces sont présentes dans un habitat régulièrement soumis au feu, on supposerait que ce sont des espèces pyrophytes. Cette façon de procéder ressemble aux valeurs indicatrices d'Ellenberg qui sont des valeurs environnementales tabulées par espèce qui indiquent ses conditions environnementales moyennes d'occurrence (Ellenberg, 1992). Ces valeurs permettent ensuite d'inférer les conditions environnementales (pH du sol, quantité de nutriments, lumière incidente, etc.) à partir de la composition floristique de la communauté. Il a fort risque de circularité dans le raisonnement, puisqu'on assigne des valeurs espèces dans un environnement donné pour s'en servir ensuite pour bio-indiquer l'environnement. Les valeurs d'Ellenberg montrent souvent une déviation par rapport aux véritables conditions environnementales des communautés étudiées (Dzwonko, 2001). Cet exemple souligne le paradoxe qu'il peut y avoir à chercher des traits qui expliquent la performance des espèces dans des environnements variés, en négligeant les traits qui expliquent la performance des espèces dans des environnements particuliers.

2.3 AUX SOURCES DE L'ORIGINALITÉ

Dans la [partie 3.3](#) nous avons exploré une origine possible des espèces originales : le fait que les espèces originales puissent l'être du fait d'adaptation à des environnements particuliers. Il y a toutefois d'autres explications possibles pour expliquer le maintien de l'originalité dans les communautés. Dans cette section je vais exposer quelques hypothèses clés sur le maintien des espèces originales.

En utilisant des traits fonctionnels *sensu* Violle et collab. (2007) il est possible de déterminer la valeur sélective des espèces en connaissant leurs valeurs de traits. Cela signifie que l'on peut expliquer les différences de valeurs sélectives entre espèces à partir de leurs différences de traits. En particulier, les performances locales des espèces comme le taux de croissance de la population de l'espèce dans l'environnement d'intérêt peuvent s'exprimer en fonction de la valeur de ses traits (Adler *et al.*, 2014; Laughlin *et al.*, n.d.). En exprimant la performance des espèces en fonction de leurs valeurs de trait on construit une courbe trait-performance. Lorsque plusieurs traits déterminent la performance d'une espèce on peut construire un paysage trait performance qui décrit, dans un environnement donné, les performances des espèces en fonction de leurs valeurs de trait (Laughlin & Messier, 2015). On appelle ce paysage multi-dimensionnel, un paysage adaptatif. Si l'on suppose une optimalité fonctionnelle sur un trait, ce paysage ressemble à une dis-

tribution Gaussienne, comme je l'ai présenté en Introduction. Mais dans le cas général on peut supposer que ce paysage est plutôt multi-modal et qu'il n'est pas nécessairement régulier (Figure 5.1). À partir de cette représentation du paysage adaptatif, je vais présenter plusieurs hypothèses qui pourraient expliquer l'originalité fonctionnelle observée localement.

Une première hypothèse est qu'une espèce originale peut se maintenir dans une communauté en occupant un pic distinct de performance, isolé du pic majoritaire de performance dans le paysage adaptatif. À l'échelle globale cela correspond à une des hypothèses testées dans la partie 3.3, c'est-à-dire que des espèces peuvent être originales suite à des processus d'adaptation à un habitat local (p. ex. l'espèce A Figure 5.1). Cela correspond au processus d'adaptation locale. C'est-à-dire que les espèces originales évoluent et leurs traits sont sélectionnés selon les conditions environnementales locales. Ce phénomène peut aussi s'envisager à une échelle locale, où une espèce s'adapte à un micro-habitat particulier.

Les dynamiques source-puits peuvent aussi expliquer l'originalité fonctionnelle locale des espèces. Ces dynamiques sont liées à la capacité d'espèces de disperser à partir de communauté « source » où elles ont des taux de croissance positifs vers des communautés « puits » où elles ont des taux de croissance négatifs [Hanski & Hanski (1999); voir Figure 5.2]. On peut donc observer des espèces dans des communautés où leur taux de croissance est négatif parce qu'elles dispersent régulièrement depuis les communautés sources (Leibold & Mikkelsen, 2002). Cela peut favoriser la présence d'espèces originales qui, par dispersion, sont présentes dans une communauté où l'optimum phénotypique est sensiblement différent. Ainsi, les effets source-puits pour permettre le maintien de l'originalité fonctionnelle locale.

Les phénomènes source-puits ne sont pas limités à la dimension spatiale des communautés, il existe des effets source-puits temporels (Johnson, 2004). Dans ce cas, les communautés sources et les communautés puits, au lieu d'être séparées dans l'espace, sont séparées dans le temps. Dans le contexte d'environnements fluctuants dans le temps, un habitat peut être favorable et favoriser l'extension d'une espèce. S'il devient défavorable, l'espèce peut voir son taux de croissance devenir négatif. Si ce changement d'environnement n'est pas trop rapide, alors la population locale de l'espèce va décroître progressivement. Cette population pourra persister alors même que l'environnement ne lui est plus favorable. Les espèces originales actuelles pourraient donc être des populations restantes mieux adaptées aux conditions environnementales passés. Elles étaient fonctionnellement communes dans les environnements passés parce qu'adaptées, mais le changement de l'environnement peut favoriser des espèces différentes fonctionnellement, ce qui fait que les espèces communes du passé peuvent être les originales du présent.

À l'échelle locale, on peut observer des espèces originales du fait des barrières de dispersion. Dans l'exemple d'une espèce A est présente dans une communauté locale. Des espèces similaires fonctionnellement peuvent exister à l'échelle régionales mais elles ne peuvent pas disperser dans la communauté locale à cause de barrières de dispersion, comme par exemple des chaînes de montagnes (Figure 5.3). L'espèce A présente localement est donc la seule qui possède ces traits, elle est donc originale fonctionnellement au niveau local. Comme elle partage des traits avec d'autres espèces au niveau régional, elle n'est pas originale

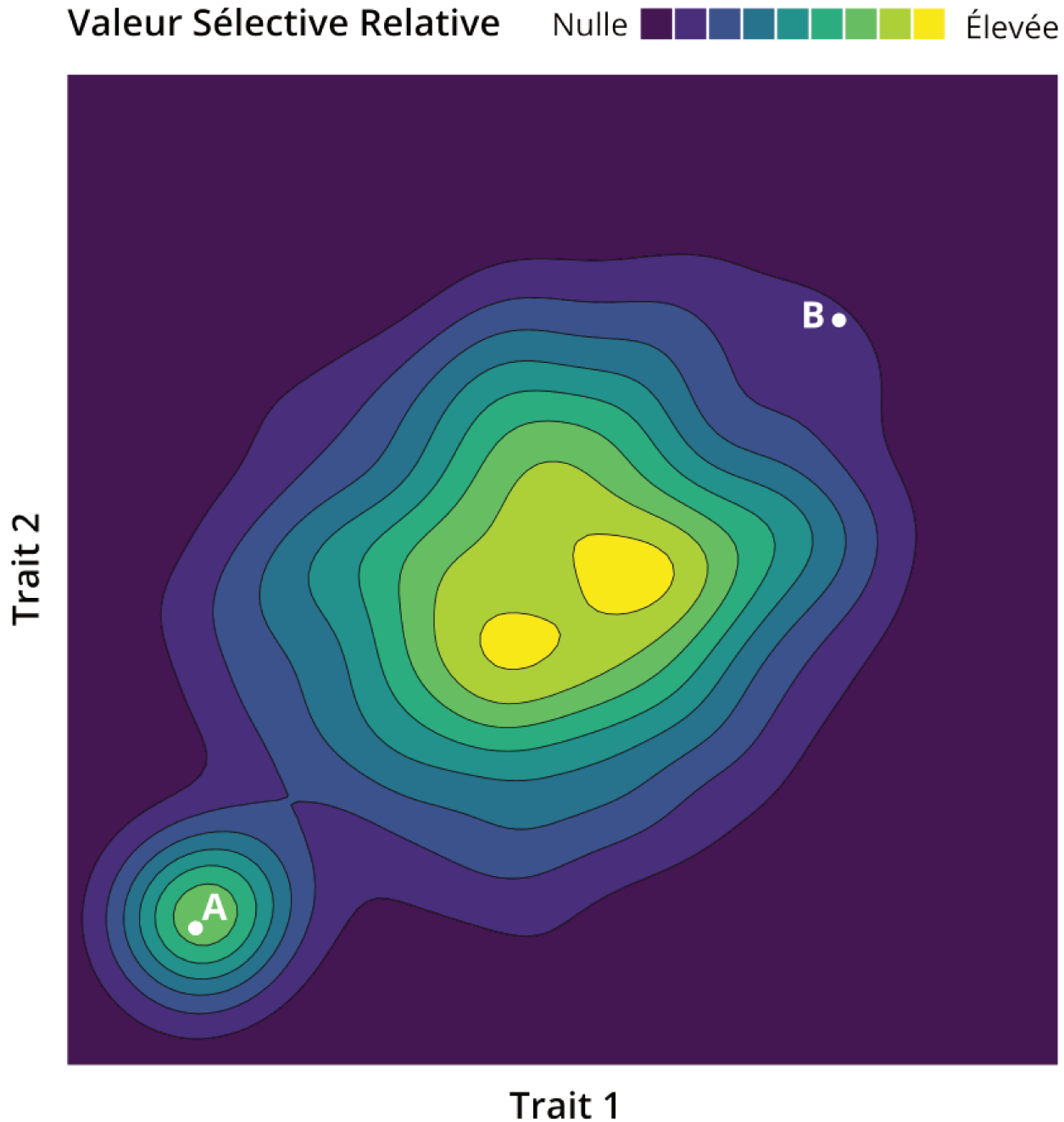


FIGURE 5.1 – Représentation schématique d'un paysage de valeur sélective en fonction de la valeur de deux traits à l'échelle régionale. Les axes des abscisses et des ordonnées représentent chacun un trait, la variation de couleur représente la valeur sélective associée à la combinaison des traits. Plus la zone est bleue, plus la valeur sélective relative de l'espèce est faible, plus la zone est jaune, plus elle est forte. L'espèce **A** occupe un pic de valeur sélective distinct du pic majoritaire, tandis que l'espèce **B** occupe la marge du pic majoritaire.

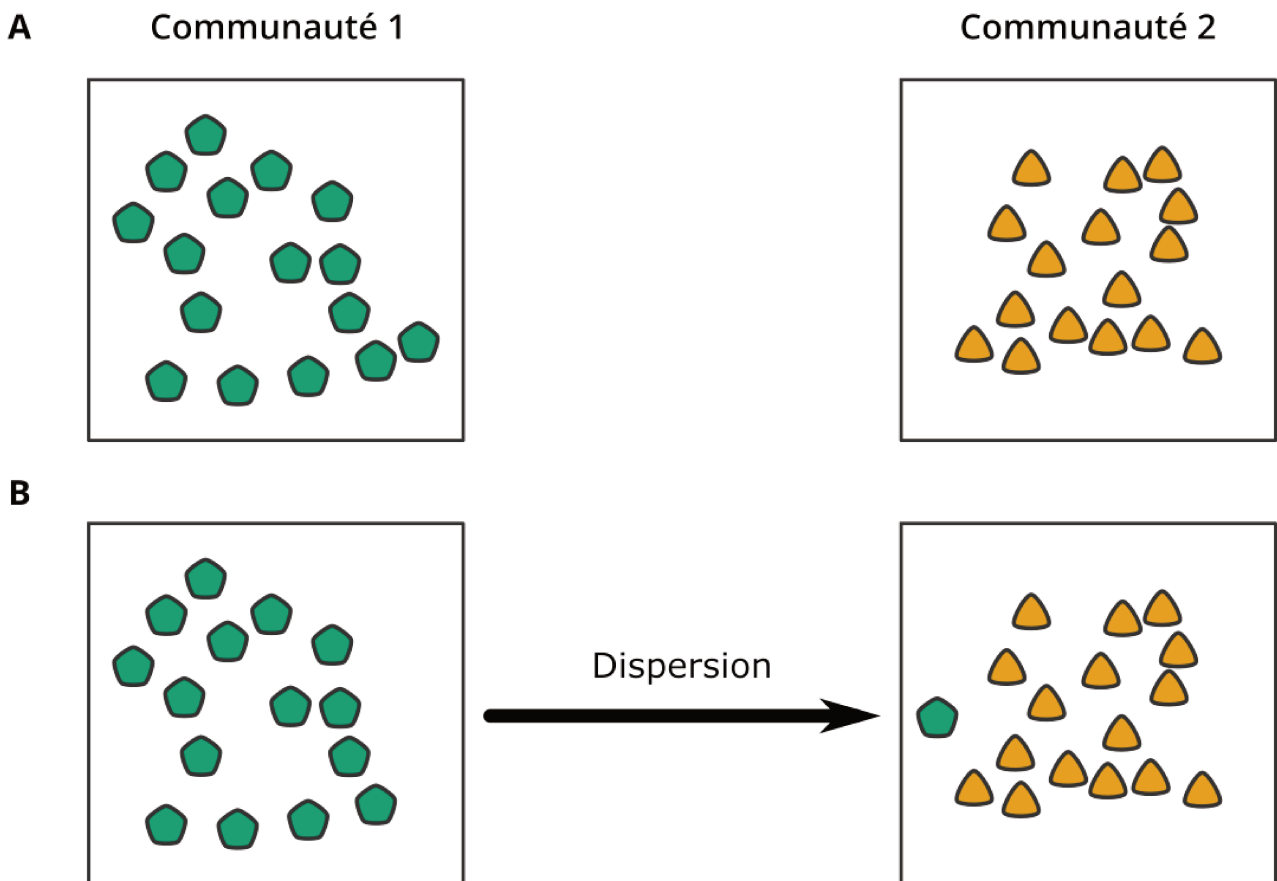


FIGURE 5.2 – Représentation schématique d’une dynamique source-puits qui pourrait contribuer au maintien de l’originalité fonctionnelle. (A) On observe deux communautés composées de deux espèces différentes : les pentagones dans la communauté 1, les triangles dans la communauté 2. Ces deux espèces diffèrent en trait, il est vert pour les pentagones, orange pour les triangles. Chaque espèce possède un trait optimal pour survivre dans son environnement initial. (B) Les pentagones sont très nombreux, un individu disperse de la communauté 1 à la communauté 2, il devient, dans cette communauté, original fonctionnellement. Étant donné que le pentagone vert n’est pas adapté à la communauté il va s’éteindre les communautés reviendront à la situation (A). Si suffisamment de pentagones dispersent en continu, alors il est possible qu’ils se maintiennent dans la communauté deux.

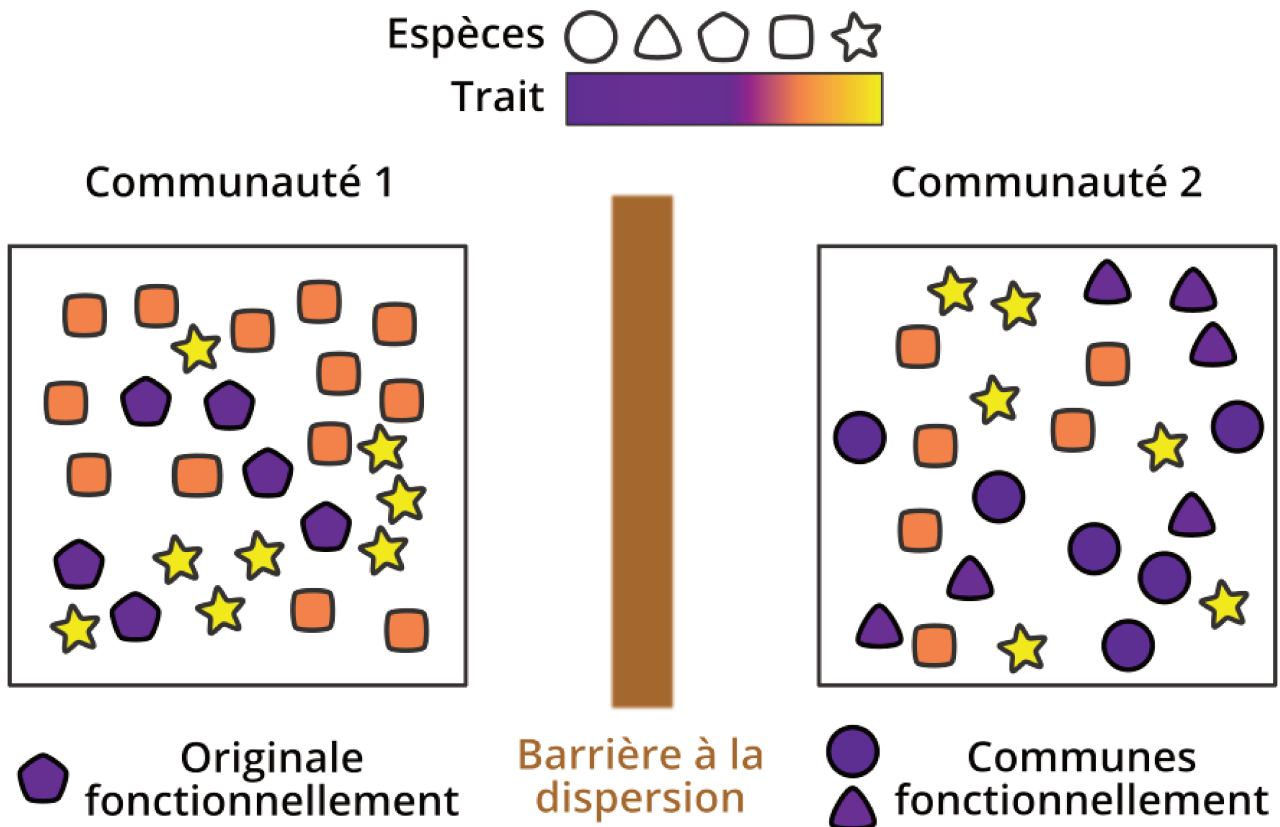


FIGURE 5.3 – Représentation schématisée des conséquences d'une barrière à la dispersion sur l'originalité fonctionnelle. Il y a deux communautés distinctes 1 et 2 séparées par une barrière à la dispersion. Les formes représentent des espèces différentes, tandis que les couleurs représentent des traits différents. Dans la communauté 1, le pentagone violet est une espèce originale fonctionnellement car c'est la seule à être violette. Dans la communauté 2, les espèces cercles et triangles sont communes fonctionnellement parce qu'elles sont proches en traits. Si l'on considère les deux communautés ensemble, l'espèce pentagone n'est pas originale fonctionnellement car fonctionnellement similaire des espèces cercle et triangle. Seulement du fait de la barrière à la dispersion, les espèces cercle et triangle ne peuvent pas disperser de la communauté 2 vers la communauté 1, ce qui rend l'espèce pentagone originale dans la communauté 1.

fonctionnellement à l'échelle régionale. Si l'on observe ce mécanisme entre populations plutôt qu'entre espèces, alors il est probable que, si les barrières de dispersion se maintiennent à long terme, la population A subisse une spéciation allopatrique qui la différencie des autres populations. On observerait donc une nouvelle espèce originale fonctionnellement au niveau local, mais avec des espèces sœurs similaires fonctionnellement au niveau régional.

Les interactions entre espèces peuvent aussi être une source importante d'originalité fonctionnelle. Par exemple une plante nurse qui modifie les conditions environnementales locales peut éviter l'extinction locale d'espèces non-adaptée à l'environnement. Ces espèces n'étant pas adaptées aux conditions locales, on peut supposer qu'elles sont plus originales fonctionnellement que les autres espèces de la communauté locale. Dans cet exemple, la plante nurse peut être originale par ses traits d'effets (c'est une espèce ingénieure

d'écosystème), alors que les espèces subordonnées sont originales par leurs traits de réponse (elles sont différentes puisqu'adaptées à des conditions environnementales différentes). Les interactions inter-spécifiques positives peuvent donc favoriser les espèces originales. Mais les processus de compétition peuvent aussi permettre le maintien des espèces originales fonctionnellement. Par compétition avec les autres espèces certaines espèces empêchent des espèces trop proches d'être présentes dans la communauté locale (Gause, 1934). Ces espèces peuvent donc être originales fonctionnellement en limitant la proximité fonctionnelle avec les autres espèces par exclusion compétitive. Sous l'hypothèse de l'optimalité fonctionnelle, une forte limite à la similarité, favorise l'émergence d'espèces originales aux valeurs de traits aux extrêmes du filtre environnemental.

À l'échelle régionale, les phénomènes évolutifs peuvent notamment permettre l'émergence d'espèces originales fonctionnellement. J'évoque plus bas certains phénomènes évolutifs dans la [section dédiée](#) aux relations entre originalités fonctionnelle et phylogénétique.

J'ai évoqué ici plusieurs mécanismes qui peuvent expliquer l'originalité fonctionnelle à l'échelle locale et, dans une moindre mesure, à l'échelle régionale. Ces processus ne sont pas mutuellement exclusifs et peuvent expliquer conjointement la présence d'espèces originales fonctionnellement dans les communautés. Pour tester ces différents processus, il serait possible d'utiliser le modèle *ecolottery* en modélisant les processus séparément et en interaction. Ainsi on pourrait disposer de prédictions théoriques sur les patrons d'originalité fonctionnelle locale et régionale attendus en fonction de l'importance relative de ces mécanismes.

2.4 COMBIEN DE TRAITS POUR CALCULER L'ORIGINALITÉ FONCTIONNELLE ?

Dans les chapitres qui ont traité de l'originalité fonctionnelle, nous avons calculé l'originalité fonctionnelle sur le maximum de traits disponibles dans les données. Faut-il cependant calculer l'originalité fonctionnelle sur des traits spécifiques ou la considérer comme une grandeur globale ? Si l'on choisit une approche multivariée, combien de traits et quels traits faut-il sélectionner ?

En calculant l'originalité fonctionnelle sur tous les traits pour toutes les espèces, on moyenne l'originalité des espèces sur chacun des traits. Or, certaines espèces peuvent être originales seulement selon certaines dimensions et peu sur d'autres. Par exemple, si la surface spécifique foliaire des plantes est extrêmement variable à travers le monde, il existe des contraintes physiologiques fortes qui l'empêchent de varier librement (Wright *et al.*, 2004). Ainsi, on ne s'attend pas nécessairement à ce que des plantes soient originales d'après leurs surfaces spécifiques foliaires. En revanche, elles peuvent l'être sur d'autres traits : configuration du réseau racinaire, couleurs des fleurs, etc. Donc il pourrait être intéressant de calculer l'originalité fonctionnelle sur quelques physiologiquement moins contraints plutôt que sur l'ensemble des traits à la fois, ce qui risque de masquer le signal de quelques traits originaux. Il est aussi possible pour une espèce d'être originale dans un espace à n -dimensions, sans l'être sur chacune des dimensions prises séparément. Cette espèce serait donc moyennement originale sur chaque dimension mais très originale en les considérant toutes à la fois. Une étude considérant les traits séparément puis pris tous ensemble pourrait séparer ces deux cas.

Nous avons montré que le nombre de traits utilisé pouvait modifier la valeur de la diversité fonctionnelle et *in fine* les patrons de diversité fonctionnelle le long de gradients environnementaux (partie 2.2). Pour l'originalité fonctionnelle on observe des patrons similaires. Plus on ajoute des traits, moins l'influence d'un seul trait va être visible. Du fait de la malédiction de la dimension (*curse of dimensionality*), plus un espace a de dimensions, plus il est vide, donc plus les points devraient être éloignés les uns des autres. Seulement, en grande dimension ($n \geq 20$) les métriques de distances ont des comportements inattendus (Aggarwal, Hinneburg, & Keim, 2001). Le rapport entre les distances d'un point au point le plus éloigné de celui-ci et de ce même point à son plus proche voisin approche un (Beyer *et al.*, 1999). Cela veut donc dire qu'avec les métriques usuelles le point d'intérêt est aussi éloigné du point le plus éloigné de lui-même que de son voisin le plus proche (Beyer *et al.*, 1999). Pour que les métriques de distances restent utilisables, on peut soit les adapter en utilisant des métriques de distance « fractionnelles » ou bien utiliser des techniques de réduction de la dimensionnalité de l'espace fonctionnel (Beyer *et al.*, 1999; Aggarwal, Hinneburg, & Keim, 2001). Étant donné que la dimensionnalité des jeux de données biologiques est relativement faible ($n = 7-8$) (Laughlin, 2014; Pianka *et al.*, 2017), si l'on applique des techniques de réduction de la dimensionnalité, on pourra continuer à utiliser la définition habituelle de l'originalité fonctionnelle.

2.5 ORIGINALITÉ, ÉCHELLE SPATIALE, ÉCHELLE BIOLOGIQUE, ET ÉCHELLE TAXONOMIQUE

Dans les chapitres 3 et 4 nous avons calculé l'originalité à l'échelle globale. Nous nous sommes concentrés sur l'échelle globale dans cette thèse dans une démarche de biogéographie fonctionnelle (Violle *et al.*, 2014). Dans le cadre proposé par Violle et collab. (2017), l'originalité fonctionnelle est calculée en croisant deux échelles spatiales : l'échelle locale et l'échelle régionale. Pour s'interroger sur les processus locaux et régionaux influençant les espèces originales, on peut calculer l'originalité fonctionnelle à ces deux échelles. Plus largement, comme rappelé dans la partie 3.2, l'originalité fonctionnelle peut se calculer à toutes les échelles spatiales, il n'y a pas d'échelle fixée *a priori*. C'est l'interprétation de l'originalité fonctionnelle qui diffèrera selon l'échelle spatiale utilisée. Dans ce mémoire, je me suis concentré à étudier l'originalité à une échelle régionale, le plus souvent à l'échelle mondiale. Seulement, comme je l'ai rappelé dans la section sur l'origine de l'originalité, la comparaison de l'originalité fonctionnelle à des multiples échelles spatiales pourrait permettre de mieux comprendre les mécanismes permettant son maintien. Il serait aussi possible d'envisager des études qui décomposent l'originalité fonctionnelle d'une espèce le long d'une échelle spatiale, comme cela a été fait pour l'étude de la diversité taxonomique (Keil & Chase, 2019). Cela permettrait notamment d'étudier les discontinuités de l'espace fonctionnel espèce par espèce. Certaines espèces pourraient par exemple, voir leur originalité réduite très rapidement à mesure que l'on agrandit l'échelle, car des espèces fonctionnellement similaires seraient présentes à proximité (Figure 5.4). Ainsi, chaque espèce pourrait avoir un profil d'originalité, qui permette de caractériser les différents types d'originalité fonctionnelle.

Tout au long de la thèse, nous nous sommes intéressés aux valeurs de traits des espèces en négligeant la variabilité intra-spécifique. Pourtant, de nombreuses études ont montré l'intérêt de la prise en compte de cette

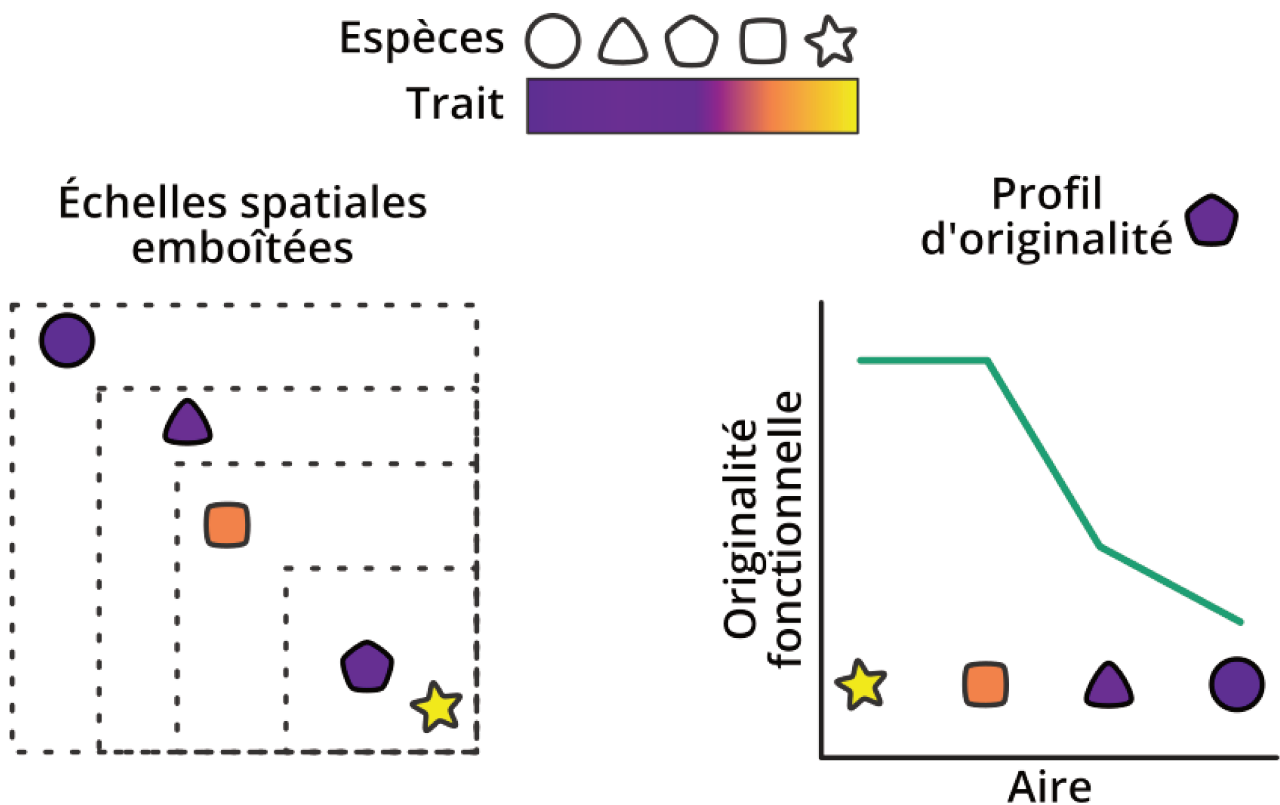


FIGURE 5.4 – Représentation schématique d'une étude sur la décomposition de l'originalité fonctionnelle d'une espèce d'intérêt en fonction de l'aire étudiée. Les formes représentent des espèces différentes, tandis que les couleurs représentent leur valeur de trait. À gauche on observe l'emboîtement successif des échelles spatiales qui permet d'accumuler de plus en plus d'espèces. À droite, il s'agit d'un graphique qui représente l'originalité de l'espèce d'intérêt, ici le pentagone, en fonction de l'aire d'étude considérée. En considérant un faible espace, le pentagone est très original fonctionnellement, puis en accumulant des espèces similaires son originalité fonctionnelle décroît.

variabilité notamment pour l'écologie des communautés (Albert *et al.*, 2010; Bolnick *et al.*, 2011; Violle *et al.*, 2012; Des Roches *et al.*, 2018). Étudier l'originalité fonctionnelle intra-spécifique pourrait nous éclairer de manière similaire sur les déterminants de l'originalité fonctionnelle. Par exemple, il serait possible de partitionner l'originalité fonctionnelle entre individus de la même espèce ou entre individus d'espèces différentes, à l'image de ce qui a été proposé pour la variance fonctionnelle (Violle *et al.*, 2012). Cela pourrait nous permettre de quantifier la part d'individus d'une espèce qui sont originaux : si en moyenne une espèce est considérée comme originale, pour quelle proportion de ses individus est-ce le cas ? Peut-on trouver des individus originaux à l'échelle de la communauté pour des espèces fonctionnellement communes et comment l'expliquer ?

Le concept d'originalité fonctionnelle n'est pas associé à une échelle biologique particulière et pourrait s'appliquer à toutes les échelles biologiques. Il serait par exemple possible de mesurer l'originalité fonctionnelle à des échelles plus grandes que celle de l'espèce, à l'échelle de populations, de communautés, ou d'écosystèmes entiers. L'idée ici n'est pas de moyenniser l'originalité fonctionnelle d'un niveau inférieur au niveau supérieur, car cela reviendrait à étudier la redondance fonctionnelle à ce niveau supérieur (Ricotta *et al.*, 2016; Kondratyeva, Grandcolas, & Pavoine, 2019). En revanche, le cadre conceptuel qui consiste à étudier l'originalité des caractéristiques d'un *item* par rapport à d'autres *items* semblables reste pertinent. On pourrait par exemple identifier les communautés dont les caractéristiques fonctionnelles (CWM, diversité fonctionnelle, etc.) diffèrent le plus des autres communautés. Cela permettrait de mettre en évidence d'éventuels effets d'échelles de l'originalité : les communautés les plus originales en traits sont-elles les plus riches en espèces originales ?

Enfin, il serait aussi possible de décomposer l'originalité fonctionnelle le long de la taxonomie. En effet, l'originalité fonctionnelle peut se calculer à l'échelle du genre, de la famille taxonomique, ou même de rangs taxonomiques supérieurs. Ainsi il serait possible d'étudier l'originalité relative d'une espèce à ces différents rangs. Par exemple pour comprendre dans quelle mesure une espèce est originale par rapport à son genre sans forcément l'être à l'échelle de sa famille. Si une espèce est originale par rapport à son genre mais pas par rapport aux autres espèces de sa famille, cela signifie que d'autres espèces possèdent des traits proches dans sa famille mais pas d'autres espèces de son genre. Des espèces d'autres genres possèdent des traits similaires à l'espèce d'intérêt, ce qui montre une convergence fonctionnelle entre ces espèces.

2.6 L'ORIGINALITÉ FONCTIONNELLE LE LONG DE L'ARBRE DU VIVANT

Dans cette thèse, nous avons travaillé sur l'originalité fonctionnelle de groupes très différents : oiseaux, végétaux (herbacées & arbres), mammifères, poissons. Le cadre utilisé peut s'appliquer de manière similaire à d'autres groupes taxonomiques dont les traits sont bien étudiés et disponibles comme les amphibiens (Oliveira *et al.*, 2017) ou les carabes (Fountain-Jones, Baker, & Jordan, 2015). Mais tous ces groupes sont des macro-organismes, c'est-à-dire des organismes de taille macroscopique. Les micro-organismes représentent pourtant le deuxième groupe de plus grande biomasse derrière les plantes (Bar-On, Phillips, & Milo, 2018).

Il peut toutefois être difficile de définir et de mesurer des traits fonctionnels sur les micro-organismes, même si des ensembles de traits sont d'ores et déjà disponibles pour des groupes spécifiques (Litchman & Klausmeier, 2008; Bestova *et al.*, 2018). Cependant, étant données la petite taille et la simplicité relative des micro-organismes, les traits mesurés sont plus facilement reliés aux composantes de la valeur sélective des espèces (Litchman & Klausmeier, 2008).

Plutôt que d'utiliser des traits fonctionnels définis de manière classique, plusieurs fronts de recherche développent des approches -omiques pour obtenir d'autres types de traits fonctionnels sur les micro-organismes. Les approches de génomique fonctionnelle qui décrivent la diversité des fonctions associées à des gènes identifiés (Brown *et al.*, 2014). La méta-génomique fonctionnelle permet de séquencer l'ADN présent dans un échantillon puis d'annoter automatiquement les fonctions des gènes identifiés dans l'échantillon (Raes *et al.*, 2011). La transcriptomique fonctionnelle permet quant à elle de s'intéresser aux gènes transcrits par l'organisme, au plus près de leur expression (Swenson & Jones, 2017). Toutes ces approches revisitent la notion de trait fonctionnel pour permettre l'utilisation de technologie génomique à haut débit tout en définissant des traits pertinents notamment pour les micro-organismes.

Chez les micro-organismes, plusieurs études montrent l'importance des espèces rares et des espèces originales pour le fonctionnement des écosystèmes (Jousset *et al.*, 2017; Banerjee, Schlaeppi, & Heijden, 2018; Jia, Dini-Andreote, & Salles, 2018). Les espèces originales pourraient donc porter ou assurer des fonctions spécifiques. La grande redondance fonctionnelle observée dans beaucoup d'écosystèmes microbiens, les espèces rares pourraient servir d'assurance écologique en cas de changement des conditions environnementales (Jousset *et al.*, 2017).

En plus d'utiliser l'originalité fonctionnelle sur des organismes différents, nous l'avons aussi comparée à l'originalité phylogénétique le long d'arbres phylogénétiques (parties 3.3, 4.1, et 4.2). Dans ces trois parties nous avons observé un signal phylogénétique positif de l'originalité fonctionnelle. Cela signifie que les espèces phylogénétiquement plus proches sont plus proches en originalité fonctionnelle qu'attendu par hasard. Ces résultats, même en considérant les critiques associées au signal phylogénétique (Münkemüller *et al.*, 2012), doivent être interprétés avec prudence dans la mesure où l'originalité fonctionnelle n'est pas un trait qui évolue le long de la phylogénie mais dépend au contraire du contexte dans laquelle elle est évaluée. Donc pour reconstruire l'originalité fonctionnelle passée des espèces il faudrait reconstruire les arbres d'évolutions des différents traits utilisés puis la calculer pour différentes profondeurs d'arbre.

Dans ces deux parties, en plus du signal phylogénétique, nous avons comparé l'originalité fonctionnelle à l'originalité phylogénétique (*evolutionary distinctiveness*, ED). L'originalité phylogénétique décrit la longueur de branches non partagée entre une espèce et toutes les autres de l'arbre (Vane-Wright, Humphries, & Williams, 1991). Plus une espèce est isolée dans une longue branche de l'arbre phylogénétique, plus elle aura une forte originalité phylogénétique (Figure 5.5). L'originalité phylogénétique a été utilisée pour prioriser les espèces sur la base de leur histoire évolutive non partagée (Vane-Wright, Humphries, & Williams, 1991; Redding & Mooers, 2006; Isaac *et al.*, 2007). L'originalité phylogénétique est le pendant spécifique de la

diversité phylogénétique, à l'image de l'originalité fonctionnelle qui est le pendant de la diversité fonctionnelle (Kondratyeva, Grandcolas, & Pavoine, 2019). La diversité phylogénétique a été très utilisée comme *proxy* de la diversité fonctionnelle notamment pour étudier l'assemblage des communautés (Webb, 2000; Webb *et al.*, 2002; Kraft *et al.*, 2007; Mouquet *et al.*, 2012). Pourtant des analyses récentes montrent qu'en fonction de l'évolution des traits, la diversité phylogénétique ne capture pas nécessairement la diversité des traits (Tucker *et al.*, 2018). De manière similaire, l'originalité phylogénétique pourrait capturer l'originalité fonctionnelle, auquel cas, il suffirait de construire des arbres phylogénétiques des espèces plutôt que de mesurer des traits pour obtenir une estimation de l'originalité fonctionnelle (Cadotte, Cardinale, & Oakley, 2008; Cadotte, Albert, & Walker, 2013). Dans les parties 3.3 et 4.1 nous avons observé une légère corrélation négative entre originalité fonctionnelle et phylogénétique, -0.12 et -0.08 respectivement; tandis que dans la partie 4.2 nous n'avons pas observé de relation entre les originalités. D'autres études n'ont pas trouvé non plus de lien entre originalité fonctionnelle et phylogénétique (Violle *et al.*, 2017; Cooke, Eigenbrod, & Bates, 2020).

L'absence de relation entre les originalités peut être liée à différents facteurs à l'image de l'absence de relations entre diversités fonctionnelle et phylogénétique (Mayfield & Levine, 2010). Des taux de spéciation et d'extinction différents entre clades peuvent faire émerger l'originalité fonctionnelle indépendamment de l'originalité phylogénétique. Les radiations évolutives peuvent générer une grande diversité de traits chez les espèces concernées, ce qui peut amener des espèces originales à émerger (Gavrilets & Losos, 2009). Les espèces ayant subi récemment une radiation évolutive sont nécessairement peu phylogénétiquement originales. Les radiations évolutives peuvent donc expliquer le découplage entre originalités fonctionnelle et phylogénétique. Les espèces spécialistes ont aussi un plus fort taux de spéciation que les espèces généralistes (Fernández & Vrba, 2005). Si les espèces originales sont plus spécialistes que généralistes, alors elles sont aussi faiblement originales phylogénétiquement. L'évolution des traits le long de la phylogénie peut aussi jouer un rôle. Pavoine et collab. (Pavoine *et al.*, 2017) ont par exemple montré avec des simulations que plus les traits utilisés pour calculer l'originalité fonctionnelle avait un signal phylogénétique fort, plus on observait une corrélation forte avec l'originalité phylogénétique. Ce n'est en revanche plus le cas si les traits utilisés ne montrent pas de signal phylogénétique (Cadotte, Albert, & Walker, 2013; Pavoine *et al.*, 2017). Pavoine et collab. ont aussi montré que plus le nombre de traits était important plus la corrélation entre originalités était forte (Pavoine *et al.*, 2017). Ces analyses pourraient être poursuivies en utilisant des modèles macro-évolutifs faisant évoluer les traits selon plusieurs modèles et en faisant varier les taux de spéciations/extinction pour tester la relation *a priori* entre originalités fonctionnelle et phylogénétique à l'échelle globale. D'autres analyses pourraient impliquer des modèles d'assemblage des communautés couplés avec des phylogénies établies pour tester la relation entre processus d'assemblage et corrélation attendue entre originalités.

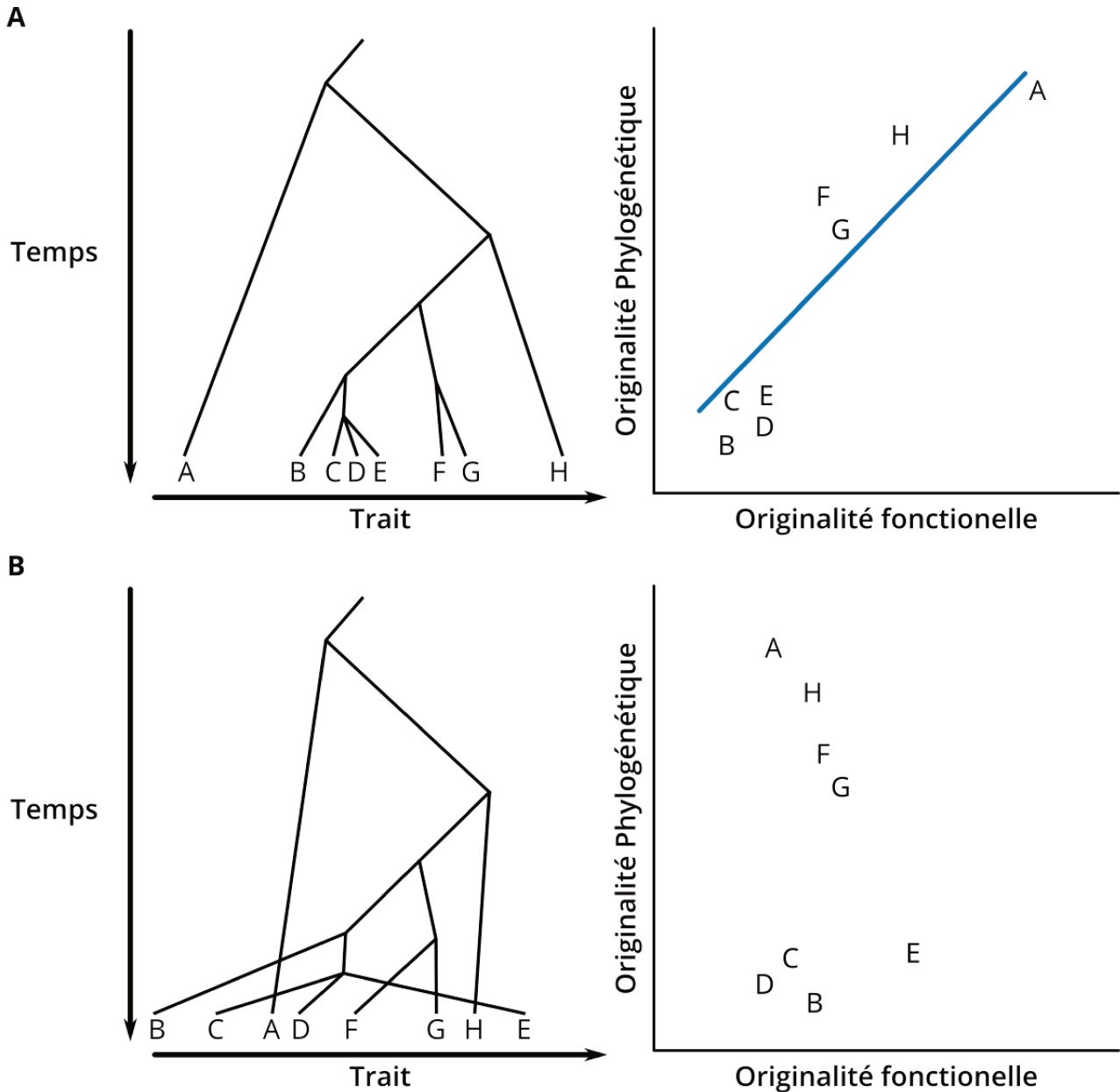


FIGURE 5.5 – Représentation schématique de la relation entre originalité phylogénétique et originalité fonctionnelle pour deux arbres distincts. La partie gauche du schéma représente des arbres phylogénétiques dont l'axe des ordonnées représente le temps depuis la divergence et l'axe des abscisses les valeurs de traits des différents nœuds. La partie droite du schéma décrit, en utilisant les arbres, les originalités fonctionnelle et phylogénétique des différentes espèces de l'arbre. (A) Cas où on observe une relation nette entre originalité fonctionnelle et originalité phylogénétique. (B) Cas où il n'y a pas de relation entre originalité fonctionnelle et originalité phylogénétique.

2.7 FAUT-IL CONSERVER SPÉCIFIQUEMENT LES ESPÈCES RARES FONCTIONNELLEMENT ?

Nous nous sommes intéressés à la conservation des espèces originales dans le [chapitre 4](#), en adoptant une démarche de biogéographie de la conservation ([Ladle & Whittaker, 2011](#)). Dans les deux parties de ce chapitre, nous avons quantifié les relations entre originalité fonctionnelle, originalité phylogénétique, et aire de répartition des espèces. Nous avons ensuite décrit les points chauds d'originalité fonctionnelle globale. En outre, dans la partie 4.2 nous avons évalué les menaces qui pesaient sur les oiseaux et mammifères fonctionnellement rares à l'échelle du monde. Ces contributions représentent un premier pas pour interroger l'intérêt du cadre de l'originalité fonctionnelle en biologie de la conservation.

Durant tout ce travail de thèse nous avons supposé que les espèces originales devaient être priorisées dans les schémas de conservation du fait de leur contribution originale au fonctionnement des écosystèmes. Même s'il existe d'autres études qui tendent à montrer que les espèces rares ont une contribution plus originale et plus grande dans le fonctionnement des écosystèmes que les espèces abondantes ([Lyons *et al.*, 2005](#); [Leitão *et al.*, 2016](#); [Jousset *et al.*, 2017](#); [Dee *et al.*, 2019](#)), nous n'avons pas démontré que c'était le cas des espèces fonctionnellement originales. Pour ce faire, comme je l'ai suggéré ci-dessus, nous pourrions évaluer l'originalité fonctionnelle sur des traits d'effet plutôt que sur des traits de réponse ([Lavorel & Garnier, 2002](#)). Pour observer l'effet des espèces originales sur le fonctionnement des écosystèmes, il faudrait pouvoir comparer des communautés, où, toute autre chose égale par ailleurs, l'espèce la plus originale a été retirée. Une expérimentation s'inscrivant dans les études de biodiversité et fonctionnement des écosystèmes pourrait manipuler l'originalité fonctionnelle des espèces comme cela a été fait pour la richesse spécifique et la diversité fonctionnelle ([Cardinale *et al.*, 2007](#); [Roscher *et al.*, 2013](#)). Seulement, les contraintes pratiques rendraient la manipulation de l'originalité fonctionnelle très difficile à réaliser. Un modèle prenant explicitement en compte les liens entre traits et fonctionnement des écosystèmes pourrait permettre de mettre en place des expériences virtuelles. La comparaison du fonctionnement de communautés comprenant des espèces plus ou moins originales permettrait d'évaluer le rôle de celles-ci dans le fonctionnement des écosystèmes. Une autre façon de procéder serait de retirer progressivement les espèces de chaque communauté et de quantifier la différence de fonctionnement associée. Des scénarios retirant les espèces selon des critères différents (rareté, originalité, sensibilité, ou aléatoire) permettraient de comparer les critères les plus déterminants pour le fonctionnement des écosystèmes ([Solan, 2004](#)).

La priorisation des espèces prend souvent en compte le risque d'extinction des espèces, afin de prioriser les espèces qui ont le plus fort risque d'extinction. Nous avons montré dans les parties 4.1 que les espèces classées comme menacées sur la liste rouge mondiale de l'UICN étaient en moyenne plus originales que les espèces non menacées. Cela signifierait que les espèces originales devraient être davantage menacées d'extinction que les espèces communes. De plus dans la partie 4.2 nous avons démontré que les espèces fonctionnellement rares comprenaient davantage d'espèces classées comme menacées sur la liste rouge mondiale de l'UICN. Nous avons aussi montré que les espèces fonctionnellement rares étaient davantage menacées par l'impact humain et le changement climatique. Pour évaluer le risque d'extinction des espèces originales,

nous devrions disposer des paramètres démographiques de ces espèces, qui sont difficilement accessibles et coûteux à acquérir (Vié *et al.*, 2009). Il est possible de prédire le risque d'extinction des espèces à partir de leurs traits en les comparant à des espèces pour lesquelles on dispose des paramètres démographiques, des menaces et des traits (Morais *et al.*, 2013; Bland & Böhm, 2016; Luiz *et al.*, 2016). Ensuite *a posteriori*, il serait possible de comparer les risques d'extinction obtenus entre espèces originales et espèces communes. Au-delà du risque de circularité lié à l'utilisation des traits à la fois pour prédire le risque d'extinction puis pour évaluer l'originalité fonctionnelle, les liens entre traits et risque d'extinction ne sont pas univoques (Chichorro, Juslén, & Cardoso, 2019). En revanche la spécialisation d'habitat et de régime alimentaire semble davantage liées au risque d'extinction (Chichorro, Juslén, & Cardoso, 2019). Il faudrait donc étudier les liens entre originalité et spécialisation, qui sont des concepts qui peuvent être rapprochés, sans être équivalents.

Par la diversité de stratégies qu'elles apportent à la communauté locale, les espèces originales peuvent assurer une forme d'assurance écologique (Yachi & Loreau, 1999). C'est-à-dire que dans des conditions environnementales qui changent, les espèces originales pourraient devenir dominantes et maintenir le fonctionnement de l'écosystème, on parle de « sauvetage écologique » (*ecological rescue*) des communautés (Loreau, Mouquet, & Gonzalez, 2003; Low-Décarie *et al.*, 2015). Les espèces originales pourraient augmenter la résilience et la résistance des communautés, en particulier dans des environnements soumis aux changements globaux.

Dans le chapitre 4, nous nous sommes concentrés sur la conservation des espèces à un niveau global. C'est pourquoi nous avons quantifié l'originalité à cette même échelle. Seulement, une espèce pourrait ne pas être originale à l'échelle globale tout en étant à une échelle plus petite. Il faudrait donc étendre nos études pour tenir compte de différentes échelles pour étudier les liens entre originalité fonctionnelle et conservation. Dans la mesure où il existe des listes rouges d'espèces menacées à différentes échelles (globale, continentale, nationale, régionale, et locale), il serait possible d'étudier les liens entre originalité fonctionnelle et statut sur ces listes à toutes les échelles spatiales.

L'originalité pourrait être aussi utilisée comme critère de priorisation dans le cadre des processus de planification systématique de la conservation (*systematic conservation planning*). La planification systématique de la conservation consiste à identifier des zones prioritaires pour la conservation en maximisant des critères prédéfinis (Arponen, 2012). Cette technique permet par exemple de maximiser le nombre d'espèces protégées en minimisant le nombre de sites à protéger. D'autres indices peuvent être utilisés comme critère à maximiser comme la diversité phylogénétique par exemple ou la diversité fonctionnelle (Arponen, 2012; Pollock, Thuiller, & Jetz, 2017). Récemment, il a d'ailleurs été démontré qu'une faible augmentation de la surface des aires protégées à l'échelle globale pouvait augmenter de manière substantielle la représentativité des diversités fonctionnelle et phylogénétique mondiales (Pollock, Thuiller, & Jetz, 2017). Dans ce cadre, les espèces originales pourraient être utilisées comme critères afin de maximiser leur représentation dans les aires protégées. Il n'est seulement pas certains que le critère d'originalité donne un résultat différent de la maximisation de la diversité fonctionnelle. En effet, pour maximiser la diversité fonctionnelle avec

un minimum de sites, les algorithmes devraient prioriser les sites qui contiennent un maximum d'espèces originales. Il n'existe toutefois pas de démonstration de ce phénomène. Les critères à maximiser pour la planification systématique de la conservation sont nombreux et il faudrait démontrer l'apport de l'originalité fonctionnelle comme critère supplémentaire (Sinclair *et al.*, 2018).

3 PERSPECTIVES

3.1 BESOINS EN DONNÉES ET DESCRIPTIONS ÉCOLOGIQUES

Les travaux menés dans le cadre de cette thèse n'ont demandé ni de nouvelles expérimentations, ni de campagne de terrain pour acquérir des données. Les deux premiers chapitres ont utilisé des approches de modélisation et de simulation, tandis que les deux chapitres suivants ont analysé des données issues de bases de données. L'objectif de cette partie est d'explorer les implications de l'utilisation des bases de données notamment dans le cadre du calcul de l'originalité fonctionnelle.

L'originalité fonctionnelle nécessite d'avoir des informations sur les traits de toutes les espèces qui composent les communautés. Contrairement au CWM d'une communauté, il ne suffit pas de disposer des traits des espèces dominantes d'une communauté pour pouvoir caractériser l'originalité fonctionnelle (Carmona *et al.*, 2015 ; Borgy *et al.*, 2017). Pour pouvoir mesurer l'originalité fonctionnelle à l'échelle locale, il faut donc avoir échantillonné exhaustivement la communauté ce qui est, en pratique, souvent impossible (Baraloto *et al.*, 2010). L'originalité fonctionnelle n'étant pas une grandeur agrégée à l'échelle de la communauté, elle est extrêmement sensible aux erreurs de mesures et de reports dans les bases de données. Si de nombreux efforts ont été faits pour consolider les bases de données de traits, et évaluer les risques d'erreurs de mesures de traits (Kattge *et al.*, 2011 ; Kattge *et al.*, 2020), nous avons dû écarter des données de nos analyses à cause d'erreurs manifestes dans les valeurs des traits reportées. En outre, si l'on veut pouvoir estimer l'originalité fonctionnelle sur de multiples traits, il faut d'autant plus intensifier son effort d'échantillonnage et mesurer plusieurs traits sur chaque individu/espèce de la communauté (Laughlin & Messier, 2015, encadré 2). C'est une des raisons pour lesquelles dans la partie 3.3 nous n'avons caractérisé les plantes qu'avec le LHS, car en plus d'être un cadre utile pour représenter les stratégies des plantes, il n'utilise que trois traits souvent mesurés et disponibles. En sélectionnant ces trois traits nous sommes passés d'un jeu de données contenant 99000 espèces à environ 2100 espèces ce qui ne représente que 2% de toutes les espèces végétales initialement présentes! Alors même que nous avons sélectionné les traits les plus couramment mesurés et disponibles dans les bases de données, ils n'étaient pas disponibles pour la majorité des espèces.

Face à la difficulté pour obtenir beaucoup de données, certains suggèrent d'utiliser des techniques statistiques pour augmenter les données et obtenir les données manquantes à partir de la structure des données (Nakagawa & Freckleton, 2011 ; Imbert & Vialaneix, 2018). La méthode la plus simple parfois encore utilisée est celle de l'imputation par le genre ou la famille. Il s'agit d'assigner à l'espèce pour laquelle la valeur du trait manquante, la valeur moyenne du trait pour les espèces du même genre ou de la même

famille. C'est une technique simple mais qui donne d'assez mauvais résultats car des espèces d'un même genre ou d'une même famille peuvent posséder des valeurs de trait très différentes (Taugourdeau *et al.*, 2014). D'autres techniques utilisent la structure multivariée du jeu de données de traits à l'image de l'imputation multiple par équations chaînées (*Multiple Imputation by Chained Equation*, MICE Buuren & Groothuis-Oudshoorn, 2011). Il s'agit d'une technique rapide, donnant de bons résultats mais n'exploitant toute l'information biologique associée aux données de traits. En effet, plusieurs approches utilisent la phylogénie, les informations environnementales, ou les matrices de co-occurrences pour augmenter la qualité de leurs imputations (Taugourdeau *et al.*, 2014; Penone *et al.*, 2014; Schrodte *et al.*, 2015; Goolsby, Bruggeman, & Ané, 2017). Toutes ces approches donnent des résultats intéressants qui pourraient être exploitables, mais qui dépendent des traits utilisés et des jeux de données (Taugourdeau *et al.*, 2014). Du fait de leur contingence au jeu de données, leur précision est difficile à estimer sur un vrai jeu de données. Une solution pourrait être d'utiliser la propagation d'incertitude pour tenir compte de la variabilité des imputations. Dans le cas précis de l'originalité fonctionnelle, étant donné que l'on cherche à prédire des espèces aux valeurs de traits particulières, il n'est pas certain que ces techniques soient applicables. Il y a, en outre, un risque de circularité dans le raisonnement s'il s'agit d'imputer les traits à partir de la phylogénie et de l'environnement, pour calculer l'originalité fonctionnelle des espèces qu'on liera ensuite à l'environnement.

Les analyses menées au cours de cette thèse ont aussi nécessité l'utilisation de base de données d'occurrence d'espèces. Même s'il existe de plus en plus de bases de données d'occurrences standardisées (Enquist *et al.*, 2016; GBIF : The Global Biodiversity Information Facility, 2020), il reste encore de nombreuses zones sous-échantillonnées (Meyer, Weigelt, & Kreft, 2016). Il faut donc continuer à consolider l'acquisition des occurrences des espèces. Pour certaines espèces comme les oiseaux ou les plantes, les sciences participatives peuvent permettre d'obtenir de nombreuses informations d'occurrences en peu de temps (Sullivan *et al.*, 2009; Joly *et al.*, 2016; Anon., 2020). Mais les incertitudes sur les données et les divers biais associés nécessitent de mettre en place des statistiques appropriées (Joly *et al.*, 2016). En particulier, pour les espèces fonctionnellement rares, étant donné qu'elles ont une faible aire de répartition ou sont présentes en faible abondance localement, il est important de disposer de données d'occurrences les plus larges possibles, avec l'effort d'échantillonnage le plus grand possible.

Les travaux présentés ici concernent des échelles spatiales importantes (partie 3.3 et chapitre 4) continentale ou globale. Or, cela peut biaiser notre vision de l'originalité fonctionnelle qui dépend de l'échelle d'étude (Estes *et al.*, 2018; McGill, 2019). Il serait possible d'évaluer l'originalité fonctionnelle à des échelles multiples et de considérer sa variation continue. De même, nous nous sommes concentrés dans nos travaux sur une seule échelle temporelle. Pour acquérir une vision générale de l'originalité fonctionnelle il faudrait aussi étudier les variations d'originalité fonctionnelle au cours du temps afin de déterminer, par exemple, si les espèces originales le restent au cours du temps. Ces difficultés d'échelles peuvent être levées en utilisant des bases de données intégrant des relevés précis à large échelle spatiale ou des relevés temporels facilement accessibles (Dornelas *et al.*, 2018; Bruelheide *et al.*, 2019).

Le manque de données a déjà été souligné à plusieurs reprises (Hortal *et al.*, 2015; Violle, Borgy, & Choler, 2015; Cornwell *et al.*, 2019). Il est donc nécessaire de continuer à mener des campagnes de terrains de mesures de traits, dans des environnements variés. En plus de mesurer des traits, il faut poursuivre les efforts existants permettant l'intégration de données dans des bases de données au format standardisé qui permettent une consolidation des données (Kattge *et al.*, 2011; Madin *et al.*, 2016; Oliveira *et al.*, 2017). Des standards ouverts bien définis peuvent faciliter l'intégration des données, plusieurs sont déjà disponibles pour pouvoir construire des bases de données de traits (McCartney & Jones, 2002; Cook *et al.*, 2018; Porter, 2018; Schneider *et al.*, 2019; Gallagher *et al.*, 2020). Il existe donc de nombreuses initiatives auxquelles participer pour permettre une consolidation des données en écologie, celles-ci peuvent être encouragées par l'implémentation de pratiques scientifiques vertueuses de citation et de partage des données.

Nous avons souligné dans cette partie les besoins en données, notamment dans le cadre de l'originalité fonctionnelle. La science ouverte (*open science*) pourrait être un cadre permettant de répondre aux besoins en données (Hampton *et al.*, 2015). Seulement, pour assurer un crédit suffisant aux personnes participant aux différentes bases de données citées dans cette section et pouvoir valoriser l'acquisition de données de « base » (Costello *et al.*, 2013). L'asymétrie de citations qui existe entre écologue de terrain et macro-écologue ré-analysant des données, peut rendre difficile la participation des premiers à la consolidation des données dans des bases de données ouvertes (Bahlai *et al.*, 2019; Benítez López & Santini, 2019). Il est donc important en plus d'exploiter les données pour répondre aux questions macro-écologiques, de renforcer la reconnaissance académique des personnes impliquées dans l'acquisition de ces données.

3.2 DE L'IMPORTANCE DES OUTILS EN ÉCOLOGIE

Cette thèse présente une forte composante méthodologique (chapitres 1 et 2, partie 3.1), les autres parties de la thèse mobilisent aussi de nombreux outils informatiques sans lesquels il aurait impossible de mener un travail identique. Dans cette section je propose d'explorer l'importance et la valorisation du travail méthodologique et de la conception d'outils informatiques.

Comme je l'ai souligné dans la section précédente, les travaux de cette thèse ont nécessité la manipulation de quantité importante de données. Si ces travaux ne font pas à strictement parler de l'écologie des mégadonnées (*big data ecology*), ils se sont basés sur d'importants outils statistiques très utilisés sur les mégadonnées (Franklin *et al.*, 2016; Wüest *et al.*, 2020). Les travaux en génomique, en télédétection, ou utilisant de nombreux capteurs embarqués génèrent en effet un débit important de données (Bush *et al.*, 2017; Farley *et al.*, 2018). Ces flux importants de données nécessitent des outils spécifiques pour les générer, y accéder, et les traiter (Wüest *et al.*, 2020). Ces outils permettent de répondre à de nouvelles questions de recherche, ils participent ainsi à l'ouverture de nouveaux fronts de recherche.

Ces dernières années plusieurs études ont mis en évidence « crise de la reproductibilité » du fait que de nombreux résultats étaient difficilement reproductibles dans différents domaines scientifiques, en psychologie notamment, mais aussi, dans une moindre mesure, en écologie (Fraser *et al.*, 16 juil. 2018; Peng, 2015;

Fidler *et al.*, 2017). Des outils ouverts et partagés permettent d'assurer une meilleure reproductibilité des analyses (Borregaard & Hart, 2016; Piccolo & Frampton, 2016). En outre, le fait de mettre en communs les outils et méthodes évite des re-développement qui pourraient s'avérer dispendieux. Le partage d'outil permet aussi la réutilisation facile des méthodes. Par exemple le package *funrar* développé en partie 3.1 a été important pour toutes les analyses traitant de l'originalité fonctionnelle, sans avoir à réimplémenter les indices d'originalité fonctionnelle (Grenié *et al.*, 2018). La publication qui en a suivi a aussi permis une réutilisation aisée par des groupes n'étant pas des collaborateurs directs (Chapman, Tunnicliffe, & Bates, 2018; Cooke, Eigenbrod, & Bates, 2020). De manière similaire, j'ai participé au développement du package *ecolottery*, en contribuant notamment à rendre le code plus robuste et la documentation plus claire (Munoz *et al.*, 2017). La publication associée au package *ecolottery* a aussi permis une réutilisation facile immédiate dans la thèse (parties 2.1 et 2.2). Cela souligne l'intérêt du développement et de la publication de logiciels scientifiques.

Malgré leur importance grandissante dans la recherche actuelle, les logiciels scientifiques ne sont pas nécessairement considérés comme des contributions scientifiques à part entière (Benureau & Rougier, 2018). En effet, ils ne sont pas nécessairement examinés avec la même rigueur que les publications en elle-même (Mislán, Heer, & White, 2016). De nombreuses revues scientifiques en écologie ne considèrent pas les logiciels scientifiques en tant que tel ou bien elles n'examinent que les articles associés sans particulièrement évaluer le logiciel associé (Mislán, Heer, & White, 2016). Il faut toutefois noter l'exception de la revue *Methods in Ecology and Evolution* qui a passé un partenariat avec *rOpenScience* pour favoriser la revue et la publication de code associé aux articles, et qui indique maintenant dans sa section dédiée à la publication des logiciels scientifiques les standards de qualité attendus (Boettiger *et al.*, 2015; Grieves, 2017; Freckleton, 2018). Il existe aussi le *Journal of Open Source Software* qui a justement été créé afin de permettre la publication de logiciels scientifiques de qualité (Katz, Niemeyer, & Smith, 2018). Mais ces deux revues font plutôt figure d'exceptions qui confirment la règle. Aucune autre revue en écologie qui publie des logiciels scientifiques n'exige une qualité logicielle minimum, ni ne propose de guide d'évaluation spécifique aux logiciels.

En outre, même si les pratiques évoluent, la citation de logiciel scientifique n'est pas systématique (Poisot, 2015). Il est souvent nécessaire de publier un article accompagnant le logiciel en question, et c'est cet article qui sera cité à son tour (Katz, Niemeyer, & Smith, 2018). Cela ne permet pas de citer les versions précises des logiciels utilisés. Ce manque de citations explicites ne permet pas de facilement déterminer qui utilisent les logiciels et dans quels buts, et donc d'améliorer les logiciels en fonction des utilisations. Enfin, ce manque de citation peut nuire à la reconnaissance pour la carrière des scientifiques ayant développé ces logiciels.

Développer ces outils nécessite des compétences domaine-spécifique afin de savoir comment s'adapter au mieux aux données, aux utilisateur-riche-s, et aux analyses. Ces compétences ne peuvent s'acquérir que part une pratique scientifique dans le domaine d'intérêt. Mais développer des outils informatiques nécessite aussi des compétences spécifiques pour suivre les meilleures pratiques possibles, documenter les logiciels, permettre la ré-utilisation, et assurer l'entretien de ces outils. Il existe en effet des recommandations pour rendre les logiciels scientifiques plus robustes, inspirés du développement informatique classique, elles per-

mettent de tendre vers des logiciels aux résultats plus sûrs (Wilson *et al.*, 2014, 2017; Poisot, 2015; Taschuk & Wilson, 2017). Seulement la mise en place de ces bonnes pratiques prend un temps bien plus grand que le développement des fonctionnalités de « base » des logiciels considérés. Ce temps supplémentaire nécessaire diminue *de facto* le temps disponible pour d'autres projets de recherche. Si les logiciels ne sont pas considérés comme une contribution scientifique spécifique, cela diminue donc la motivation à suivre les bonnes pratiques de développement et rend les écosystèmes logiciels scientifiques plus fragiles.

La formation au développement logiciel est aussi manquante dans les carrières scientifiques classiques. Peu d'universités proposent des cours de développement logiciels dans les cursus d'écologie et peu de formations destinées aux scientifiques existent (Touchon & McCoy, 2016; Farrell & Carey, 2018). On peut toutefois citer les ateliers proposés par *The Carpentries* qui visent justement à donner les bases de pratiques « assez bonnes » pour renforcer la robustesse de la programmation scientifique (Wilson, 2016). L'utilisation des bonnes pratiques de développement logiciel dépend donc de l'auto-formation et des connaissances préalables des scientifiques sur le sujet. Le recours à des cadres informels d'échanges de pratiques du type *coding club* peut aussi permettre un échange horizontal de ces compétences.

Dans le cadre de l'évaluation quantitative des carrières scientifiques basées sur les publications, la publication de logiciels scientifiques est moins « compétitive » que les publications classiques étant donné le temps supplémentaire qu'elle suppose. En outre, si le développement de logiciels nouveaux est un acte de recherche en soi, leur entretien ne l'est plus considéré comme faisant partie d'une démarche de recherche et ne peut donc plus être valorisé sous forme de publications (Poisot, 2015). Il est toutefois possible de publier un deuxième article sur un même logiciel si les changements sont conséquents, c'est le cas par exemple de *ecolottery* qui sera bientôt mis à jour (Barthelemy *et al.*, in revision). Il n'y a donc pas de motivation extrinsèque à suivre des bonnes pratiques pour entretenir des logiciels scientifiques (Katz *et al.*, 2018). Les personnes qui malgré tout suivent les bonnes pratiques, disposent de l'expertise domaine-spécifique, ainsi que de l'expertise pour développer des logiciels scientifiques de qualité, ne sont donc pas compétitives par rapport à d'autres personnes qui auraient des contributions scientifiques plus classiques. Il n'existe pas en France de poste de *Research Software Engineer*, contrairement au Royaume-Uni ou à l'Allemagne, qui pourrait correspondre à une carrière scientifique davantage tournée vers les logiciels scientifiques (Cohen *et al.*, 2020). Au CNRS par exemple, les Ingénieur-e-s de Recherche (IR) pourraient avoir des postes qui correspondent à ces missions, mais la diminution des recrutements d'IR à l'échelle nationale fait qu'il est difficile d'envisager d'obtenir un poste sur ces sujets (Clavey, 2019). Il existe donc un risque de perte de compétences associées à des pratiques vertueuses du développement logiciel, ce qui nuit à la robustesse du logiciel scientifique.

Ainsi, les logiciels scientifiques doivent être considérés comme des contributions scientifiques à part entière, nécessitant des compétences spécifiques en plus d'une expertise scientifique du domaine. Afin de pérenniser le maintien de ces compétences pour augmenter la robustesse des logiciels scientifiques, il faut davantage de reconnaissance entre pairs et institutionnelle. Il faut aussi renforcer la formation autour du développement logiciel pour participer à la diffusion de bonnes pratiques. Et ainsi, la recherche en écologie pourra aller vers

un véritable commun des logiciels scientifiques.

3.3 L'ORIGINALITÉ FONCTIONNELLE : UN NOUVEAU CONCEPT OU UN INDICE DE PLUS ?

Dans cette thèse nous avons exploré le cadre conceptuel de l'originalité fonctionnelle proposé par Violle et collab. (Violle *et al.*, 2017). Or, il existe déjà un grand nombre d'indices de diversité fonctionnelle qui peuvent être mobilisés pour identifier des patrons au sein des communautés (Schleuter *et al.*, 2010; Mason *et al.*, 2013). Ces indices sont d'ores et déjà répandus et utilisés, ce qui est bien moins le cas pour l'originalité fonctionnelle. Il est donc légitime de se demander quels sont les apports de l'originalité fonctionnelle par rapport aux indices de diversité fonctionnelle plus classiques.

Contrairement à la diversité fonctionnelle, l'originalité fonctionnelle se calcule à l'échelle spécifique (voire individuelle). C'est un véritable changement de perspectives vis-à-vis de l'écologie des communautés fonctionnelle classique. Plutôt que de s'intéresser à des statistiques résumant la dispersion des traits fonctionnels des espèces composant une communauté, il s'agit de mettre en avant des espèces particulières. On passe d'une métrique qui décrit les propriétés d'un ensemble (la communauté), à une métrique qui décrit individuellement les propriétés des composants de cet ensemble (Pavoine, Ollier, & Dufour, 2005). L'originalité fonctionnelle s'intéresse donc à des espèces en particulier et met en avant les espèces les plus originales. D'ailleurs, dans les travaux menés dans cette thèse nous n'avons pas calculé d'originalité à l'échelle de la communauté. Cela reviendrait en effet à calculer une mesure de diversité fonctionnelle, la redondance fonctionnelle, ou plutôt l'inverse de la redondance fonctionnelle, puisqu'en mesurant l'originalité moyenne des espèces on mesure leur originalité moyenne (Ricotta *et al.*, 2016; Kondratyeva, Grandcolas, & Pavoine, 2019). Or, puisque précisément l'objectif de l'originalité fonctionnelle est s'intéresser aux composants de la communauté, ce cadre n'est pas utile pour calculer une moyenne à l'échelle de la communauté par rapport au cadre plus général de la diversité fonctionnelle.

L'originalité fonctionnelle est une tentative nouvelle de généralisation pour mieux comprendre les communautés. Si l'écologie des communautés est un « fouillis » (« *community ecology is a mess* », Lawton, 1999), l'étude de la diversité fonctionnelle a permis de mettre en évidence des principes généraux qui structurent les communautés (Keddy, 1992; Weiher & Keddy, 1995; McGill *et al.*, 2006; Malaterre *et al.*, 2019). La diversité fonctionnelle, puisque nombreux des indices qui sont calculés pour la mesure intègre l'abondance des espèces, est le reflet de mécanismes qui affectent plutôt la majorité de la communauté, la plus abondante ou la plus commune en trait (Mouchet *et al.*, 2010). L'originalité fonctionnelle est une tentative de généraliser des mécanismes nouveaux avec la proposition suivante : les espèces ayant des traits originaux subissent certainement des mécanismes qui diffèrent de ceux affectant les espèces communes car elles sont différentes de la communauté. Si l'on imagine, l'effet du seul filtre environnemental dans les communautés, alors après un temps très long on s'attendrait à ce que dans un environnement donné, toutes les espèces possèdent des valeurs de traits proches, et donc qu'il n'y ait plus d'espèces originales. Pourtant le fait que l'on observe des espèces originales dans les communautés montre qu'il existe des mécanismes qui permettent

leur émergence et leur maintien.

L'objectif de l'étude de l'originalité fonctionnelle est aussi la conservation des écosystèmes (Pavoine, Ollier, & Dufour, 2005; Violle *et al.*, 2017; Kondratyeva, Grandcolas, & Pavoine, 2019). Comme nous l'avons rappelé dans la section spécifique plus haut, l'originalité fonctionnelle permet d'évaluer l'« irremplaçabilité » fonctionnelle des espèces, et ainsi de les prioriser. L'idée sous-jacente étant que les espèces originales, du fait de leurs traits fonctionnels particuliers, ont une contribution particulière au fonctionnement des écosystèmes.

L'originalité fonctionnelle propose un nouveau programme de recherches complémentaire aux questions posées par l'étude de la diversité fonctionnelle. Durant mes travaux de thèse, nous n'avons pu qu'effleurer la surface des questions posées par l'étude de l'originalité fonctionnelle. Nous nous sommes peu intéressés par exemple à la question des mécanismes évolutifs d'émergence de l'originalité fonctionnelle : dans quelles conditions peut-elle émerger et persister ? La question de la relation entre originalité fonctionnelle et fonctionnement des écosystèmes devrait aussi être davantage examinée. D'un point de vue mathématique, il est aussi nécessaire de mieux caractériser les indices utilisés pour mesurer l'originalité fonctionnelle afin de mieux les intégrer dans leurs relations aux indices de diversité fonctionnelle. L'originalité fonctionnelle est donc une dimension supplémentaire de la biodiversité, complémentaire à la diversité fonctionnelle, qui permet de renouveler les questions que l'on se pose en écologie.

A

Annexes à l'article 1.1

Appendix 1: basic functions and examples

Munoz, F., Grenié, M., Denelle, P., Taudière, A., Laroche, F., Tucker, C. & Violle, C. (2017) *ecolottery*: simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution*

Contents

1	Neutral dynamics in species pool and communities	2
2	User-defined species pool	4
3	Trait distribution in communities	5
4	Environmental filtering	6
5	Filtering of multiple traits	9
6	Investigating phylogenetic structure in communities	10
7	Forward-in-time simulation	12
	References	14

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

coalesc is the key function of the *ecolottery* package for coalescent-based simulation of local communities. The user can define parameters of community size, migration rate, and a custom function specifying environmental filtering, according to the situation he wants to model. He can also provide relative abundances and species trait values in the reference species pool.

1 Neutral dynamics in species pool and communities

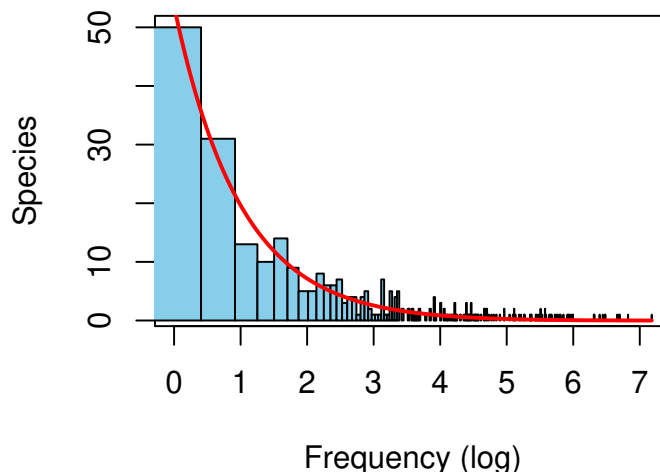
Before simulating communities, we need first to define a reference pool from which immigrants are drawn. In the neutral theory (Hubbell 2001), speciation and extinction events occur randomly irrespective of species properties and drive the composition and species abundances in the reference pool (neutral speciation-drift dynamics). We can define such dynamics by setting the input argument *theta* in *coalesc* to a non-null value, and the composition of the pool can be simulated by setting $m = 1$. The simulated species abundances are then expected to follow a logseries distribution (Hubbell 2001).

The fit of simulated pool composition to the logseries distribution can be estimated using the *fisherfit* function of package *vegan*.

```
set.seed(1)

Jpool <- 25000
logser <- coalesc(Jpool, m = 1, theta = 50)
abund <- abund(logser)

# The expected distribution of abundances in the reference pool is log-series
require(vegan)
fit <- fisherfit(abund$pool$ab)
freq <- as.numeric(names(fit$fisher))
plot(log(freq), fit$fisher, xlab = "Frequency (log)", ylab = "Species", type="n")
rect(log(freq-0.5),0,log(freq+0.5),fit$fisher, col="skyblue")
alpha <- fit$estimate; k <- fit$nuisance
curve(alpha*k^exp(x)/exp(x), log(0.5), max(log(freq)), col="red", lwd=2, add=TRUE)
```



ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

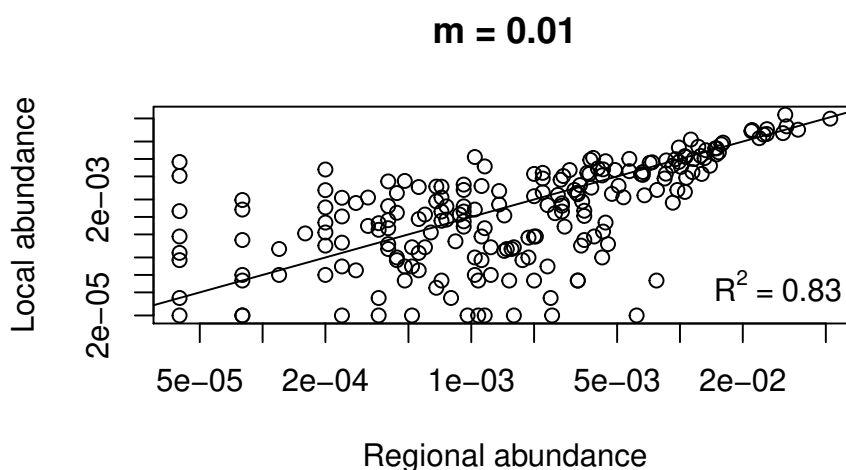
We now simulate neutral communities of size $J = 500$ in which immigrants from the pool establish at rate either $m = 0.01$ or $m = 0.95$. In this case, all the species available in the reference pool have the same prospect of immigrating, persisting and reproducing in the local community. We can define such dynamics by setting the input argument *filt* = *NULL* in *coalesc* (default value).

The smaller is m the more local species abundances fluctuate due to stochastic demographic variations. Where m is close to 1, most of the dead individuals are replaced by immigrants, and local species abundances are then more closely related to regional abundances.

```
# Local abundances are averaged over 100 replicate communities
J <- 500
m <- 0.01
comm1a <- data.frame()
for (i in 1:100) {
  comm1a <- rbind(comm1a, coalesc(J, m, pool = logser$pool)$com)
}
comm1a <- list(pool = logser$pool, com = comm1a)

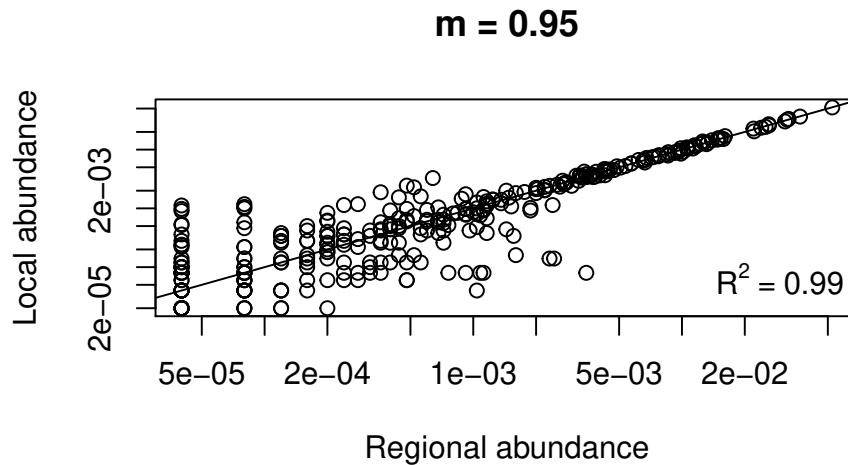
m <- 0.95
comm1b <- data.frame()
for (i in 1:100) {
  comm1b <- rbind(comm1b, coalesc(J, m, pool = logser$pool)$com)
}
comm1b <- list(pool = logser$pool, com = comm1b)

plot_comm(comm1a, type = "abund", main = "m = 0.01")
r_sqa <- summary(lm(abund(comm1a)$pool[rownames(abund(comm1a)$com), "relab"] ~
  abund(comm1a)$com$relab))
r_sqa <- signif(r_sqa$r.squared, 2)
legend("bottomright", legend = bquote(R2 ~ "=" ~. (r_sqa)), bty = "n")
```



```
plot_comm(comm1b, type = "abund", main = "m = 0.95")
r_sqb <- summary(lm(abund(comm1b)$pool[rownames(abund(comm1b)$com), "relab"] ~
  abund(comm1b)$com$relab))
r_sqb <- signif(r_sqb$r.squared, 2)
legend("bottomright", legend = bquote(R2 ~ "=" ~. (r_sqb)), bty = "n")
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



The first figure shows the relationship between local and regional species abundances in neutral communities with low immigration rate, and the second the relationship with a high immigration rate. Each point is averaged over 100 communities.

2 User-defined species pool

In the previous case, the composition of the reference pool was simulated depending on the parameter *theta* of neutral speciation-drift dynamics.

In the following example, the user provides a custom species pool including 500 species with equal abundances.

```
Jpool <- 50*J
pool <- cbind(1:Jpool, rep(1:500, Jpool/500))
# Generate a neutral community drawn from the pool
comm2 <- coalesc(J, m, pool=pool)
abund2 <- abund(comm2)
summary(abund2$pool$relab)
```

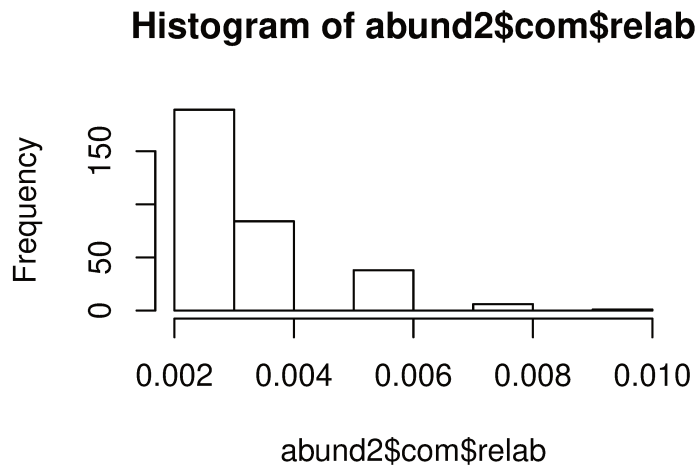
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.002  0.002  0.002  0.002  0.002  0.002
```

```
summary(abund2$com$relab)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.002000 0.002000 0.002000 0.003145 0.004000 0.010000
```

```
hist(abund2$com$relab)
```


ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

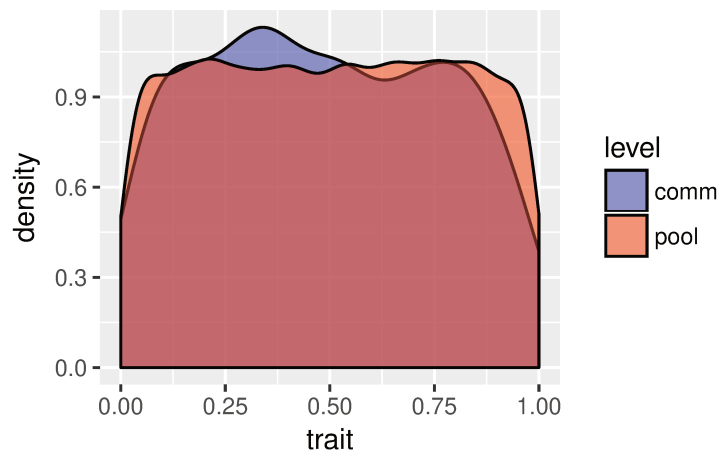


While species relative abundances are set equal in the reference pool, local species abundances fluctuate due to migration-drift dynamics.

3 Trait distribution in communities

We can also examine the trait composition of simulated communities. If the user does not provide trait values in the reference pool, the values of a unique trait are randomly assigned to the species of the reference pool following a uniform distribution between 0 and 1. Alternatively, the user can include in *pool* the values of one or several traits for each individual of the species pool, or provide a separate *traits* data frame including the values of one or several traits for each species.

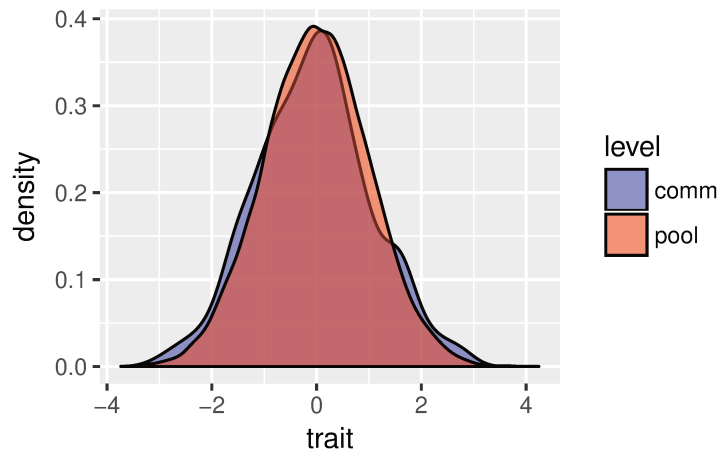
```
# With uniform trait values in the species pool
pool <- cbind(1:Jpool, rep(1:500, Jpool/500), runif(Jpool))
comm3a <- coalesc(J, m, pool = pool)
plot_comm(comm3a, type = "trait")
```



```
# With Gaussian trait values in the species pool
pool <- cbind(1:Jpool, rep(1:500, Jpool/500), rnorm(Jpool))
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

```
comm3b <- coalesc(J, m, pool = pool)
plot_comm(comm3b, type="trait")
```



`plot_comm` with `type = trait` here displays the trait distributions in species pool (red) and local community (blue). With high immigration rate $m = 0.95$, the local and regional distributions are quite similar. The user may also simulate a distribution of mean species trait values, and consider additional intraspecific variation of these values around the mean.

```
pool <- cbind(1:Jpool, rep(1:500, Jpool/500), NA)
# Distribution of the mean species trait values
t.sp <- runif(500)
# Gaussian intraspecific variation with standard deviation = 0.01
for(i in 1:500) pool[pool[,2] == i, 3] <- rnorm(sum(pool[,2]==i),
                                             mean = t.sp[i], sd = 0.01)
comm3c <- coalesc(J, m, pool = pool)
```

4 Environmental filtering

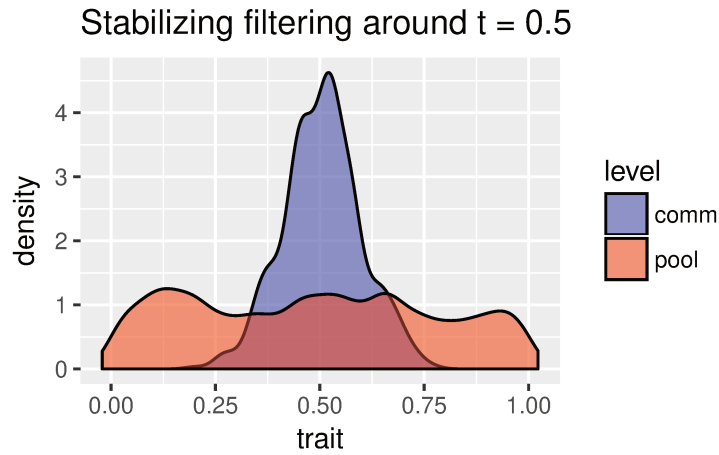
The user can provide an environmental filtering function weighting the probability that individuals from the reference pool successfully immigrate in the community, depending on their trait value(s). In the following example, the filtering function is Gaussian with mean t and standard deviation 0.1 (stabilizing filtering, Shipley 2013).

```
sigma <- 0.1
filt_gaussian <- function(t,x) exp(-(x-t)^2/(2*sigma^2))
```

We simulate a community undergoing stabilizing environmental filtering around $t = 0.5$.

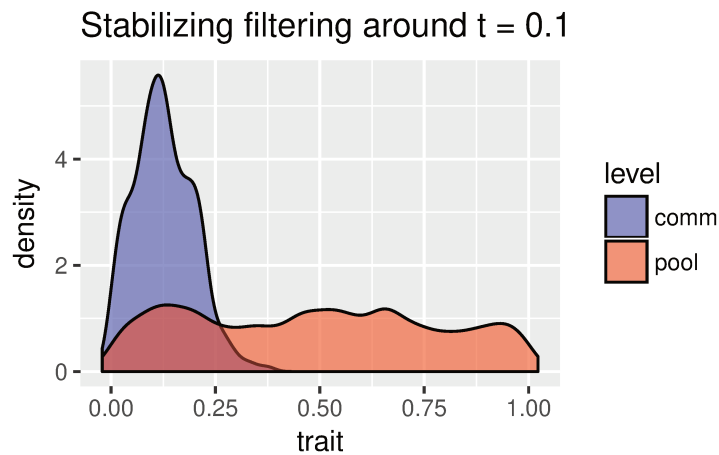
```
J <- 500; m <- 0.5;
comm4a <- coalesc(J, m, filt = function(x) filt_gaussian(0.5, x), pool = pool)
plot_comm(comm4a, main = "Stabilizing filtering around t = 0.5")
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



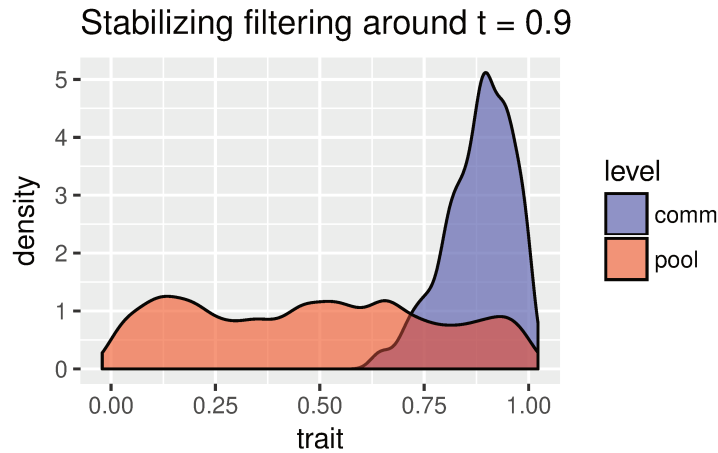
We can also simulate stabilizing environmental filtering around $t = 0.1$ and $t = 0.9$.

```
J <- 500; m <- 0.5;
comm4b <- coalesc(J, m, filt = function(x) filt_gaussian(0.1, x), pool = pool)
plot_comm(comm4b, main = "Stabilizing filtering around t = 0.1")
```



```
comm4c <- coalesc(J, m, filt = function(x) filt_gaussian(0.9, x), pool = pool)
plot_comm(comm4c, main = "Stabilizing filtering around t = 0.9")
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



When stabilizing filtering operates around different optimal values among communities, we expect corresponding changes in the local mean trait values (Cornwell & Ackerly 2009).

```
mean(comm4b$com[, 3])
```

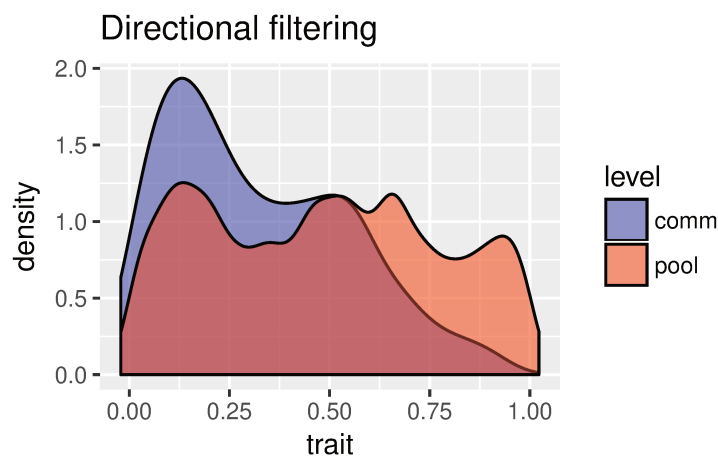
```
## [1] 0.1265049
```

```
mean(comm4c$com[, 3])
```

```
## [1] 0.8809555
```

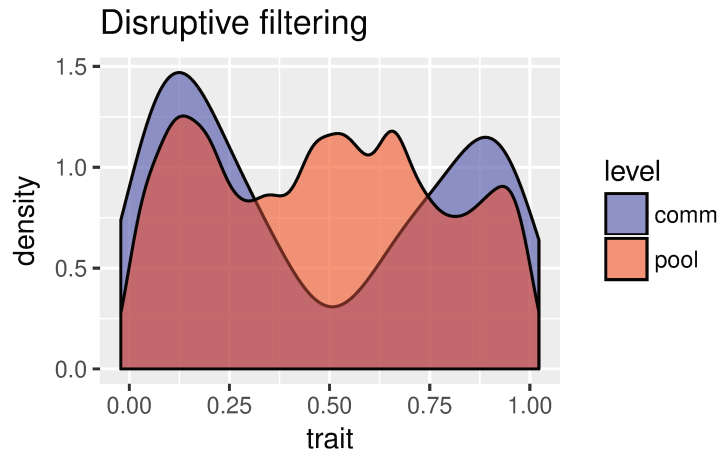
Different filtering functions can be designed to represent different types of environmental filtering. By analogy with selection regimes in evolutionary theory (Shipley 2013), we can define the outcome of directional and disruptive filtering functions.

```
# Directional environmental filtering toward  $t = 0$ 
comm4d <- coalesc(J, m, filt = function(x) 1 - min(x,1), pool = pool)
plot_comm(comm4d, main = "Directional filtering")
```



```
# Disruptive environmental filtering around  $t = 0.5$ 
comm4e <- coalesc(J, m, filt = function(x) abs(0.5 - x), pool = pool)
plot_comm(comm4e, main = "Disruptive filtering")
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



Disruptive filtering here represents the greater success of species with trait values away from $t = 0.5$. It corresponds to a separation of ecological groups in the community, and represents a form of niche differentiation (Vergnon *et al.* 2009).

5 Filtering of multiple traits

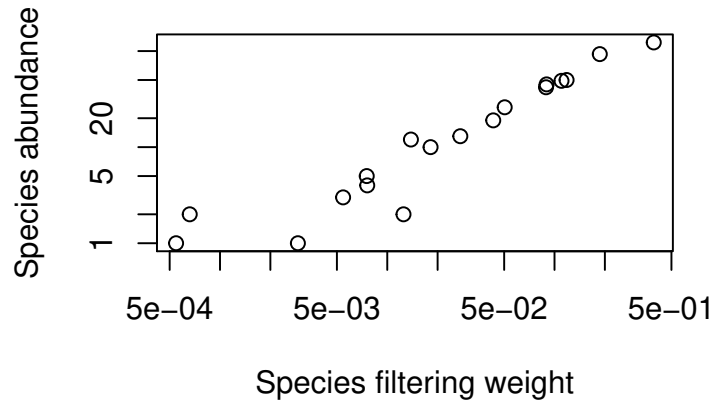
Previous examples represented environmental filtering depending on the values of a single trait. It is also possible to define environmental filtering operating on multiple traits. In the following example, three traits have values uniformly distributed between 0 and 1 in the reference pool, and undergo stabilizing filtering in the local community with distinct optimal values, 0.5, 0.25 and 0.75.

```
# An example with 3 traits
traits <- cbind(runif(Jpool), runif(Jpool), runif(Jpool))
filt <- function(x)
  filt_gaussian(0.5, x[1])*filt_gaussian(0.25, x[2])*filt_gaussian(0.75, x[3])
comm5 <- coalesc(J, m, pool=cbind(1:10000, rep(1:100, 100)), filt = filt, traits = traits)
```

The filtering function determines the success of immigrants depending on the combination of their trait values. It entails a correlation between the local species abundances and the species weights given by the filtering function (Shipley *et al.* 2006).

```
# Relationship between species weight in environmental filtering and local abundance
par(mfrow = c(1, 1))
plot(tapply(comm5$com[,3], comm5$com[,2], length)~tapply(apply(comm5$com[,3:5],
  1, function(x)
    filt(x)), comm5$com[,2], mean), xlab = "Species filtering weight",
  ylab = "Species abundance", log = "xy")
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



6 Investigating phylogenetic structure in communities

Species trait values are the legacy of evolutionary history. The processes that affect the distribution of trait values in the community can also affect the phylogenetic composition of the community depending, e.g., on the conservatism of traits among close relatives (Mouquet *et al.* 2012).

We can simulate a phylogenetic tree of the species of the reference pool, and the distribution of species trait values under an assumption of niche conservatism (Wiens & Graham 2005). In this case, descendant species can retain characteristics of their ancestors, and the trait values of more closely related species are then more similar than trait values of distantly related species (phylogenetic signal, Blomberg & Garland 2002).

```
require(ape)
```

```
## Loading required package: ape
```

```
## Warning: package 'ape' was built under R version 3.3.3
```

```
require(phyloilm)
```

```
## Loading required package: phyloilm
```

```
require(picante)
```

```
## Loading required package: picante
```

```
## Loading required package: nlme
```

```
## Warning: package 'nlme' was built under R version 3.3.3
```

```
tre <- rcoal(200)
```

```
Jpool <- 10000
```

```
J <- 500
```

```
pool <- data.frame(ind=1:Jpool, sp=rep(tre$tip.label,Jpool/50),
                  tra=rep(NA,Jpool), stringsAsFactors=F)
```

```
# Brownian model of trait evolution
```

```
t.sp = rTraitCont(n = 1, phy = tre, model = "BM", sigma = 0.2, root.value = 0.5)
```

```
pool$tra <- t.sp[pool$sp]
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

phylosignal measures the phylogenetic signal of the simulated trait values in the phylogenetic tree, using the Blomberg's K statistic (Blomberg & Garland 2002). The p-value is calculated based on the null distribution of K when shuffling species trait values in the phylogeny. A small p-value indicates that trait variation among close relatives is smaller than expected by chance.

```
phylosignal(t.sp[tre$tip.label], phy = tre)
```

K	PIC.variance.obs	PIC.variance.rnd.mean	PIC.variance.P	PIC.variance.Z
1.702134	0.0460415	23.57097	0.001	-1.876859

We then simulate two communities related to this reference pool, one undergoing neutral dynamics, the other undergoing stabilizing environmental filtering. In the first case, any species of the reference pool can perform as well in the local community, and then the phylogenetic structure of the community should not be different from the structure of a random sample with same size from the reference pool. In the second case, environmental filtering limits the range of trait values in the community around an optimal value. Because more closely related species are more likely to have similar trait values in the reference pool, we then expect that the Mean Nearest Taxon Distance (MNTD) among coexisting species of the community is smaller than the average distance in the reference pool (Webb *et al.* 2002).

```
m <- 1
tab <- array(0, c(2, 200)); colnames(tab) <- tre$tip.label
# First simulation of a neutral community
com <- abund(coalesc(J, m, pool = pool, filt = NULL))$pool
tab[1, com$sp] <- com$ab
# Second simulation of a community undergoing stabilizing environmental filtering
topt <- quantile(t.sp, 0.25)
sigma <- 0.01
com <- abund(coalesc(J, m, pool = pool, filt = function(x)
  exp(-(x-topt)^2/(2*sigma^2))))$com
tab[2, com$sp] <- com$ab
```

These patterns can be tested against a null model of random sampling in the phylogeny. The standard effect size (SES) quantifies the departure of observed phylogenetic structure from the null distribution.

```
ses.mntd(tab[, tre$tip.label], cophenetic(tre), null.model = "taxa.labels")
```

Table 2: Standard Effect Size of Mean Nearest Taxon Distance

ntaxa	mntd.obs	mntd.rand.mean	mntd.rand.sd	mntd.obs.rank	mntd.obs.z	mntd.obs.p	runs
186	0.0243106	0.0250491	0.0018181	222	-0.4061801	0.222	999
14	0.0862791	0.2573882	0.1184873	9	-1.4441125	0.009	999

The first community does not depart from random phylogenetic structure, while the second shows phylogenetic structuring, i.e., smaller Mean Nearest Taxon Distance than expected by chance ($mntd.obs.p < 0.01$).

Alternatively, in the absence of niche conservatism in the phylogeny, environmental filtering in community assembly would not result in phylogenetic clustering. To illustrate this case, we shuffle the trait values of species in the reference pool, and simulate a new local community with stabilizing environmental filtering.

```
names(t.sp) <- sample(names(t.sp))
pool$tra <- t.sp[pool$sp]
# Again simulating stabilizing environmental filtering
com <- abund(coalesc(J, m, pool = pool, filt = function(x)
  exp(-(x-topt)^2/(2*sigma^2))))$com
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

```
tab[2,com$sp] <- com$ab
ses.mntd(tab[,tre$tip.label], cophenetic(tre), null.model = "taxa.labels")
```

Table 3: Standard Effect Size of Mean Nearest Taxon Distance

ntaxa	mntd.obs	mntd.rand.mean	mntd.rand.sd	mntd.obs.rank	mntd.obs.z	mntd.obs.p	runs
186	0.0243106	0.0251842	0.0015468	187	-0.5647577	0.187	999
28	0.2312387	0.1558493	0.0632384	808	1.1921457	0.808	999

In this case, neither the neutral community nor the community with environmental filtering shows phylogenetic clustering. Therefore, when the condition of niche conservatism is not met, the absence of phylogenetic clustering does not mean that environmental filtering does not play (Mouquet *et al.* 2012).

7 Forward-in-time simulation

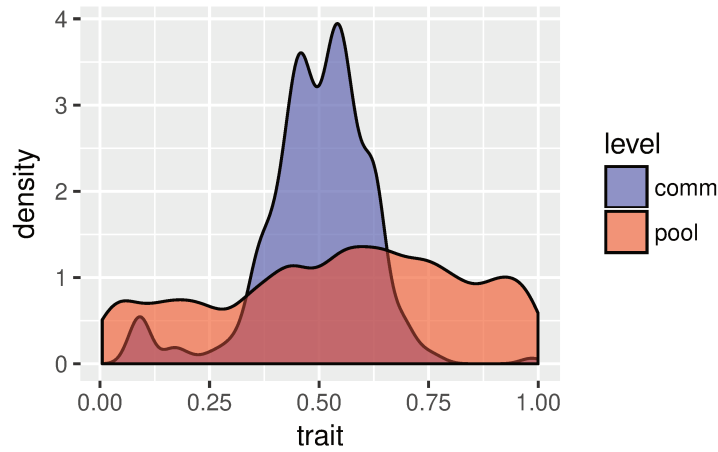
The *ecolottery* package also includes a function *forward* to perform simulation of community dynamics from an initial composition. As for *coalesc*, the user can define the composition of the species pool from which immigrants are drawn, and can specify environmental filtering based on one or several traits.

We can for instance simulate stabilizing environmental filtering in a way analogous to the previous example with *coalesc*. $d = 10$ represents the number of individuals that die in the community at each time step, $prob = 0.5$ represents the immigration rate at each time step. The number of simulated time steps is $gens = 500$.

```
pool <- data.frame(ind=1:Jpool, sp=rep(as.character(1:500),Jpool/500),
                  tra=rep(NA,Jpool), stringsAsFactors=F)
t.sp <- data.frame(sp=as.character(1:500), tra = runif(500))
pool$tra <- t.sp[pool$sp,]$tra
# Initial community composed of 500 individuals
initial <- pool[sample(1:Jpool,500),]
# Forward-in-time simulation
sigma <- 0.1
filt_gaussian <- function(t,x) exp(-(x-t)^2/(2*sigma^2))
final.envfilt <- forward(initial = initial, prob = 0.25, d = 10, gens = 500,
                       pool = pool, filt = function(x) filt_gaussian(0.5,x))

plot_comm(final.envfilt)
```


ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

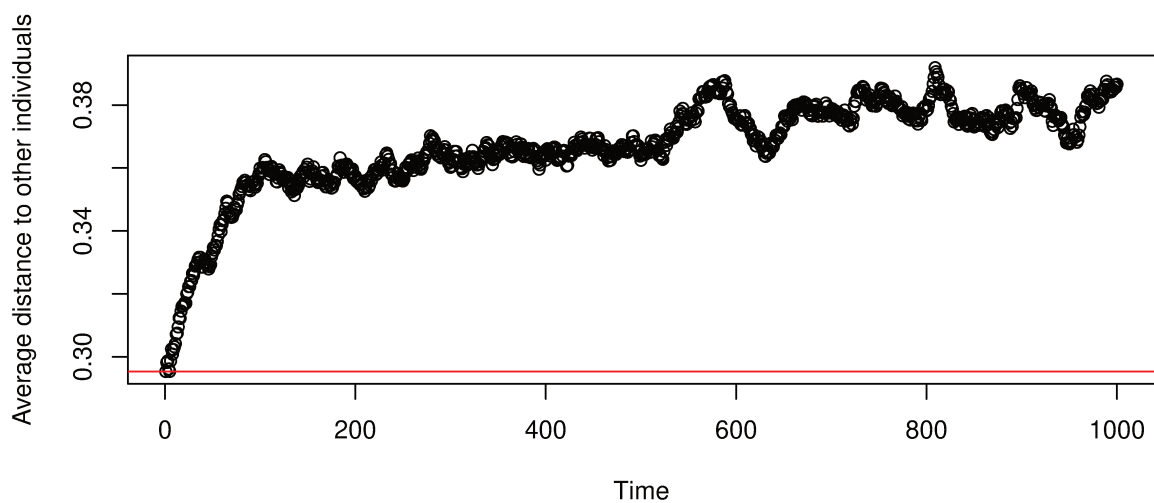


A specific advantage of *forward* is to simulate the sequence of community assembly events over time. If the birth, death and immigration probabilities depend on community composition at a specific time, the coalescent-based approach may not be appropriate.

In the following example, community dynamics with limiting similarity are simulated. In this case, the probability of individual death depends on the similarity of trait values of each individual to the other individuals of the community.

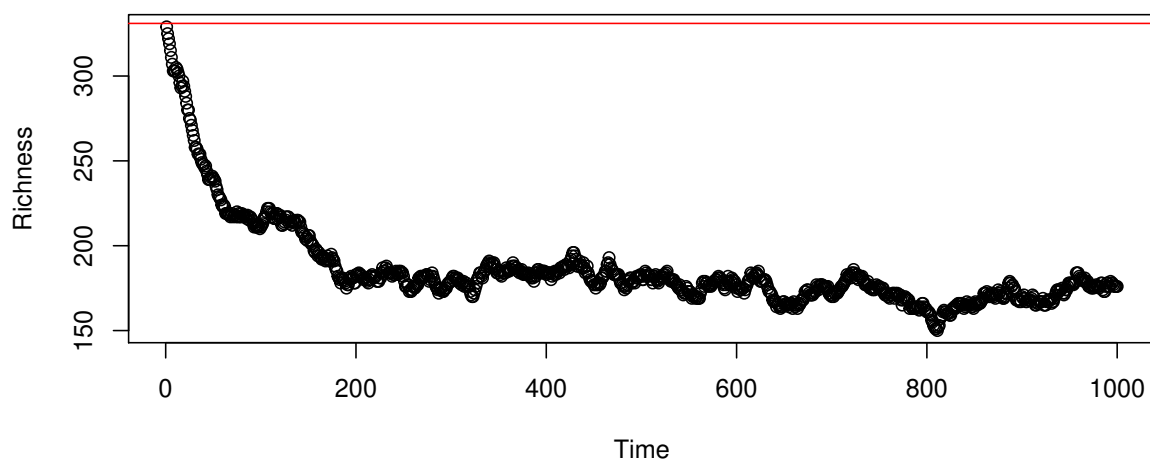
```
final.limsim <- forward(initial = initial, prob = 0.25, d = 10, gens = 1000,
  pool = pool, keep = F, limit.sim = T, coeff.lim.sim = 1)
```

```
plot(final.limsim$dist.t, xlab = "Time", ylab = "Average distance to other individuals")
init.dist <- matrix(dist(initial$tra))
diag(init.dist) <- NA
abline(mean(init.dist, na.rm=T), 0, col="red")
```



```
plot(final.limsim$sp_t, xlab = "Time", ylab="Richness")
abline(length(unique(initial$sp)), 0, col="red")
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



The first figure represents the temporal trajectory of the average distance of each individual to the other individuals of the community. The distance fluctuates and increases over time due to the influence of limiting similarity. The second figure shows the variation of richness, basically decreasing due to limiting similarity, until reaching stationarity.

References

- Blomberg, P. & Garland, T. (2002). Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology*, **15**, 899–910.
- Cornwell, W.K. & Ackerly, D.D. (2009). Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, **79**, 109–126.
- Hubbell, S. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton and Oxford.
- Mouquet, N., Devictor, V., Meynard, C.N., Munoz, F., Bersier, L.-F., Chave, J., Couteron, P., Dalecky, A., Fontaine, C., Gravel, D., Hardy, O.J., Jabot, F., Lavergne, S., Leibold, M., Mouillot, D., Münkemüller, T., Pavoine, S., Prinzing, A., Rodrigues, A.S.L., Rohr, R.P., Thebault, E. & Thuiller, W. (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, **87**, 769–785.
- Shipley, B. (2013). *From plant traits to vegetation structure: Chance and selection in the assembly of ecological communities*. Cambridge University Press.
- Shipley, B., Vile, D. & Garnier, E. (2006). From plant traits to plant communities: A statistical mechanistic approach to biodiversity. *Science*, **314**, 812–814.
- Vergnon, R., Dulvy, N.K. & Freckleton, R.P. (2009). Niches versus neutrality: Uncovering the drivers of diversity in a species-rich community. *Ecology Letters*, **12**, 1079–1090.
- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.
- Wiens, J.J. & Graham, C.H. (2005). Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology Evolution and Systematics*, **36**, 519–539.

Appendix 2: comparison of computation time for coalescent-based and forward-in-time methods

Munoz, F., Grenié, M., Denelle, P., Taudière, A., Laroche, F., Tucker, C. & Violle, C. (2017) *ecolottery*: simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution*

We first generated with *coalesc* a species pool of size $J_{pool} = 10000$ undergoing neutral speciation-extinction dynamics (with $\theta = 50$).

```
# Generate a regional pool/metacommunity at speciation-drift equilibrium
Jpool = 10000
pool = coalesc(Jpool, theta = 50, m = 1)$pool

nb_sp_pool = length(unique(pool[, 2]))
```

We then simulated communities receiving immigrants from this pool with immigration rate $m = 0.5$ and stabilizing environmental filtering $filt(x) = \exp(-(x - 0.5)^2 / (2 * \sigma^2))$. We set community size as 10, 50, 100, 250, 500, 1000 or 2000 individuals.

```
sigma <- 0.1
filt_gaussian <- function(x) exp(-(x-0.5)^2/(2*sigma^2))

# Vector of community sizes
com_size = c(10, 50, 100, 250, 500, 1000, 2000)
```

We alternatively used *coalesc* and *forward* for community simulation.

We examined the number of generations needed to reach stationary richness with *forward*, using the *get_number_of_gens* function of *ecolottery* package.

This function simulates replicate communities with *forward* from a random initial sample of the regional pool. The time to reach stationarity, t_{stat} , is estimated using least-square adjustment of a segmented function varying exponentially below t_{stat} , and being constant above t_{stat} .

The maximum value of t_{stat} over replicate runs of *forward* provides a reference number of generations needed to reach stationarity for a given community size.

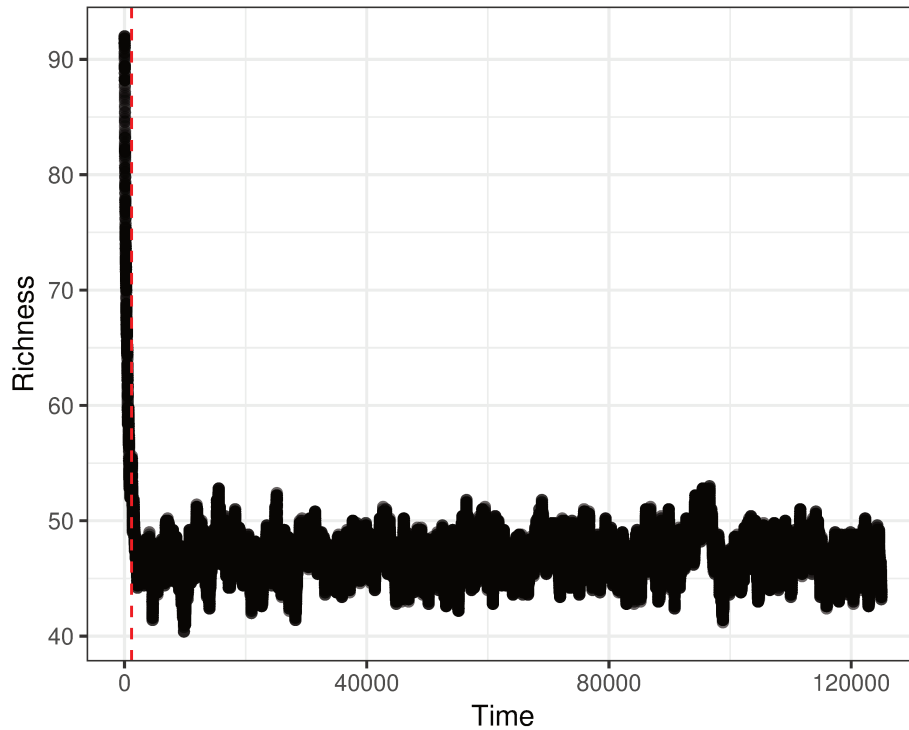
```
set.seed(285)

convergent_gens = c()
for(i in 1:length(com_size))
  print(system.time(convergent_gens[[i]] <- get_number_of_gens(com_size[[i]],
    pool, prob = 0.5, filt = filt_gaussian)))

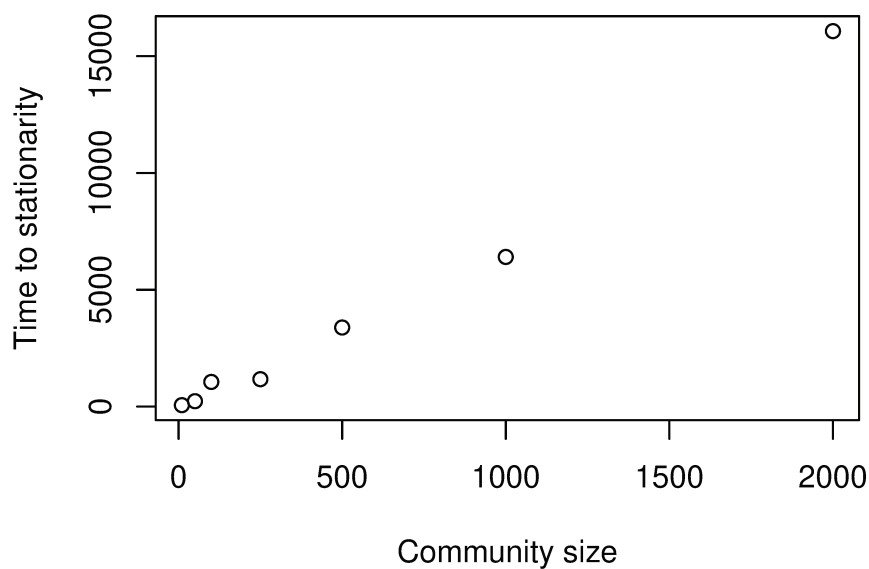
names(convergent_gens) <- as.character(com_size)

# Show the richness variation over time for a given community size, along with
# the number of generations needed to reach stationarity (vertical red dashed line)
convergent_gens[["250"]]$plot + labs(x = "Time", y = "Richness")
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



```
# Variation of time to stationarity with forward for increasing community size
plot(com_size, unlist(lapply(convergent_gens, function(x) x$n)),
     xlab = "Community size", ylab = "Time to stationarity")
```



We then simulated 100 communities with *forward* for each community size, using the number of generations

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

recommended by *get_number_of_gens*.

We also simulated 100 communities with *coalesc* using the same parameterization, except the number of generations since it is not used for coalescent-based modelling.

In both cases, we calculated the simulation time for each simulation.

```
coal.time <- c()
forw.time <- c()

for(size in com_size)
{
  for(rep in 1:100)
    coal.time <- rbind(coal.time, c(size, system.time(coalesc(J = size,
      filt = filt_gaussian, m = 0.5, pool = pool))[[3]]))
  colnames(coal.time) <- c("size", "time")

  initial <- pool[sample(1:nrow(pool), size),]
  for(rep in 1:100)
    forw.time <- rbind(forw.time, c(size, system.time(forward(initial = initial,
      prob = 0.5, gens = convergent_gens[[as.character(size)]]$n,
      filt = filt_gaussian, pool = pool))[[3]]))
  colnames(forw.time) <- c("size", "time")

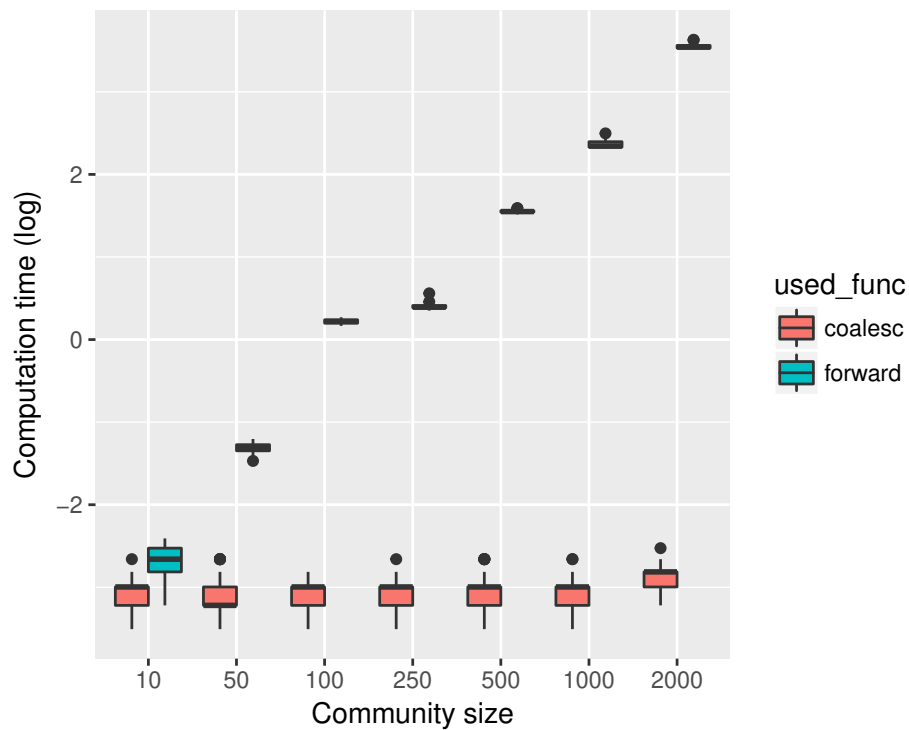
  print(size)
}
```

The following figure compares the simulation time of *coalesc* and *forward* for varying community size.

```
simul.time <- data.frame(rbind(coal.time, forw.time))
simul.time$used_func <- c(rep("coalesc", nrow(coal.time)), rep("forward", nrow(forw.time)))

require(ggplot2)
ggplot(data = simul.time, aes(x = as.factor(size), y = log(time))) +
  labs(x = "Community size", y = "Computation time (log)") +
  geom_boxplot(aes(fill = used_func))
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



While simulation time of *coalesc* increases slowly with community size, the increase with *forward* is dramatically greater.

Both algorithms have rather comparable speed for communities of 10 individuals (0.05s for *coalesc* and 0.07s for *forward*). For communities with 2000 individuals, *coalesc* provides result within 0.056 s on average, while *forward* needs 34.77s. Despite possible variations of these calculation times depending on the computer configuration, the prominent picture here is the substantial benefit of using *coalesc* for large communities.

Appendix 3: ABC estimation of neutral parameters in Barro Colorado Island rainforest, using coalescent-based simulation

Munoz, F., Grenié, M., Denelle, P., Taudière, A., Laroche, F., Tucker, C. & Violle, C. (2017) *ecolottery*: simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution*

We illustrate here how to use coalescent-based simulation to estimate neutral parameters of regional biodiversity dynamics, θ , and local migration rate, m , from observed patterns of taxonomic diversity within and between communities.

Neutral theory proved successful to predict patterns of species diversity in tropical forests (Hubbell 2001), while other studies also emphasized the influence of niche-based processes (Baraloto *et al.* 2012). With coalescent-based modelling, it is possible to address the relative ability of purely neutral and environmental filtering models to explain patterns of biodiversity, using the Approximate Bayesian Computation (ABC) approach implemented in *ecolottery*.

Barro Colorado Island is a 50ha lowland rainforest plot established in Panama. It has become a flagship case study to test competing theories of community assembly in tropical forests. The dataset available in *vegan* package includes a census of tree species above 10 cm DBH in 1ha subplots.

```
require(vegan)

## Warning: package 'vegan' was built under R version 3.3.3
## Warning: package 'lattice' was built under R version 3.3.3

data(BCI)

# Size (= number of individual trees) of subplots
comm.size <- rowSums(BCI)
# Minimum subplot size
comm.size.min <- min(comm.size)
```

For sake of simplicity in joint analyses of diversity patterns across subplots, we first subsample the subplots to have the same minimum subplot size.

```
# Rarefy to minimum sample size
bci.res <- rrarefy(BCI, sample = comm.size.min)
```

We consider a set of summary statistics representing for each subplot the local richness, local Shannon diversity, and the average Bray-Curtis beta diversity with other subplots.

```
library(betapart)

# Compute diversity indices
rich.obs <- apply(bci.res, 1, function(x) sum(x!=0))
shan.obs <- apply(bci.res, 1, function(x) diversity(x, index = "shannon"))
beta.obs <- lapply(beta.pair.abund(bci.res), function(x) {
  X = rowMeans(as.matrix(x), na.rm=T)
})$beta.bray

stats.obs <- c(rich.obs, shan.obs, beta.obs)
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

```
names(stats.obs) <- paste0(rep(c("rich", "shan", "beta"), each = 50), 1:50)
```

To estimate the θ and m parameters of neutral dynamics, we consider a vector of $2 * 10^5$ values drawn from a prior uniform distribution.

```
m.samp <- runif(2*10^5, min = 0, max = 1)
theta.samp <- runif(2*10^5, min = 0, max = 100)
```

Parallel computing can be used to perform the simulations, by using multiple cores in desktops, or multiple clusters. The *parallel* package allows handling parallel computing.

```
library(parallel)

# Start up a parallel cluster
parallelCluster <- makeCluster(parallel::detectCores())
print(parallelCluster)

# Function to perform simulations
mkWorker <- function(m.samp, theta.samp, J)
{
  require(ecolottery)
  require(untb)

  force(J)
  force(m.samp)
  force(theta.samp)

  summCalc <- function(j, m.samp, theta.samp, J)
  {
    pool.samp <- ecolottery::coalesc(100*J, theta = theta.samp[j])$pool
    meta.samp <- array(0, c(50, length(unique(pool.samp$sp))))
    colnames(meta.samp) <- unique(pool.samp$sp)

    for(i in 1:50)
    {
      comm.samp <- ecolottery::coalesc(J, m.samp[j], pool = pool.samp);
      tab <- table(comm.samp$com[,2])
      meta.samp[i, names(tab)] <- tab
    }

    rich.samp <- apply(meta.samp, 1, function(x) sum(x != 0))
    shan.samp <- apply(meta.samp, 1, function(x) vegan::diversity(x, index = "shannon"))
    beta.samp <- lapply(betapart::beta.pair.abund(meta.samp),
                       function(x) rowMeans(as.matrix(x), na.rm=T)
                       )$beta.bray

    return(list(sum.stats = c(rich.samp, shan.samp, beta.samp),
                  param = c(m.samp[j], theta.samp[j])))
  }
  worker <- function(j) {
    summCalc(j, m.samp, theta.samp, J)
  }
  return(worker)
}
```


ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

The function *mkWorker* will be used in parallel instances of R to perform the simulations. For each values of *m* and θ the summary statistics are calculated and returned.

The overall set of statistics and corresponding parameter values are stored in a list.

```
modelbci <- parLapply(parallelCluster, 2:10^5, mkWorker(m.samp, theta.samp, comm.size.min))

# IMPORTANT
# Shutdown cluster after calculation
if(!is.null(parallelCluster)) {
  stopCluster(parallelCluster)
  parallelCluster <- c()
}

# Summary statistics and parameter values are extracted
# and stored in matrices
stats <- t(sapply(modelbci, function(x) x$sum.stats))
stats.sd <- apply(stats, 2, sd)
stats.mean <- apply(stats, 2, mean)

stats <- t(apply(stats, 1, function(x) (x - stats.mean)/stats.sd))
colnames(stats) <- paste0(rep(c("rich", "shan", "beta"), each = 50), 1:50)

stats.obs <- (stats.obs-stats.mean)/stats.sd

param <- t(sapply(modelbci, function(x) x$param))
colnames(param) <- c("m", "theta")
```

Then we use the *abc* function from package *abc* to estimate the parameters in observed rainforest subplots.

```
require(abc)
bci.abc <- abc(target = stats.obs, param = param, sumstat = stats, tol = 0.01, method = "neuralnet")
```

The function *coalesc_abc* encompasses the steps of simulation and ABC analysis.

The previous calculations can thus be performed using the following command.

```
# Define the function providing the summary statistics
f.sumstats <- function(tab)
{
  rich <- apply(tab, 1, function(x) sum(x!=0))
  shan <- apply(tab, 1, function(x) vegan::diversity(x, index="shannon"))
  beta <- lapply(betapart::beta.pair.abund(tab),
                function(x) rowMeans(as.matrix(x), na.rm=T))$beta.bray

  stats <- c(rich, shan, beta)
  names(stats) <- paste0(rep(c("rich", "shan", "beta"), each = 50), 1:50)

  return(stats)
}

# Perform the simulations and the ABC analysis
bci.abc <- coalesc_abc(bci.res, multi = T, traits = NULL,
                     f.sumstats = f.sumstats, params = NULL,
                     theta.max = 100, nb.samp = 2*10^5, tol = 0.01,
                     pkg = c("vegan", "betapart"), method = "neuralnet",
                     parallel = T)
```

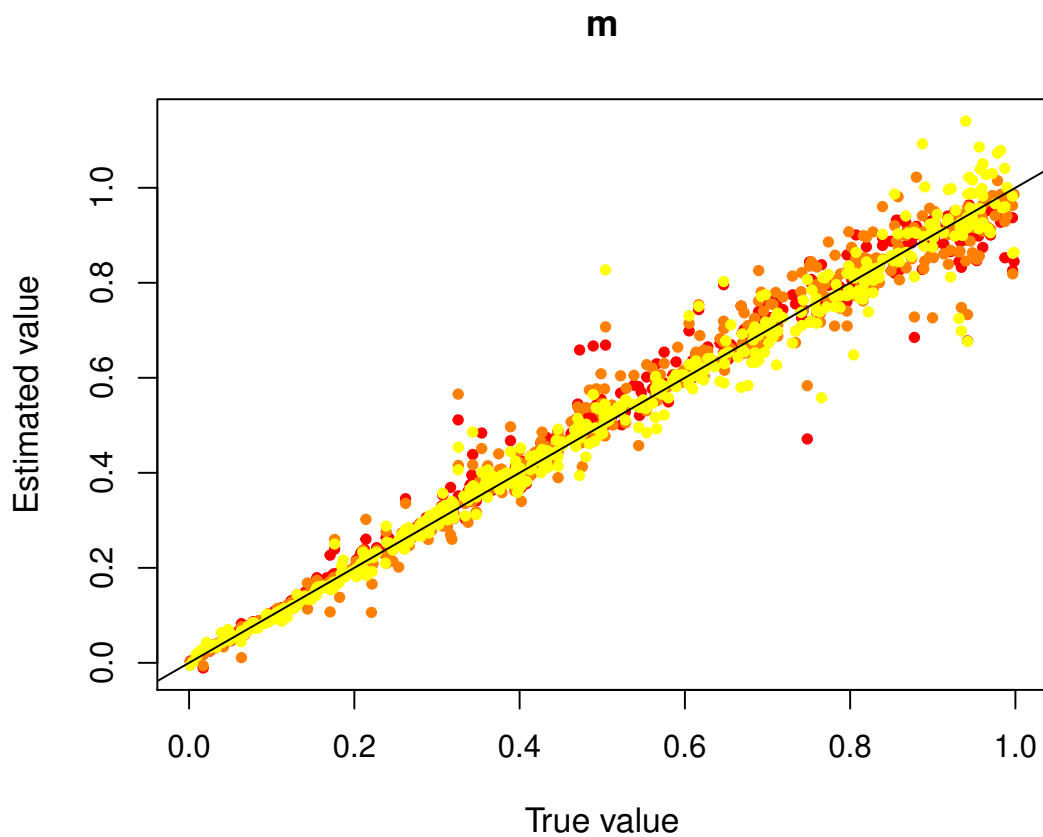
ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

bci.abc then includes the matrix of parameter values (*par* output), the matrix of simulated statistics values (*ss* output) and an *abc* object including the results of ABC analysis.

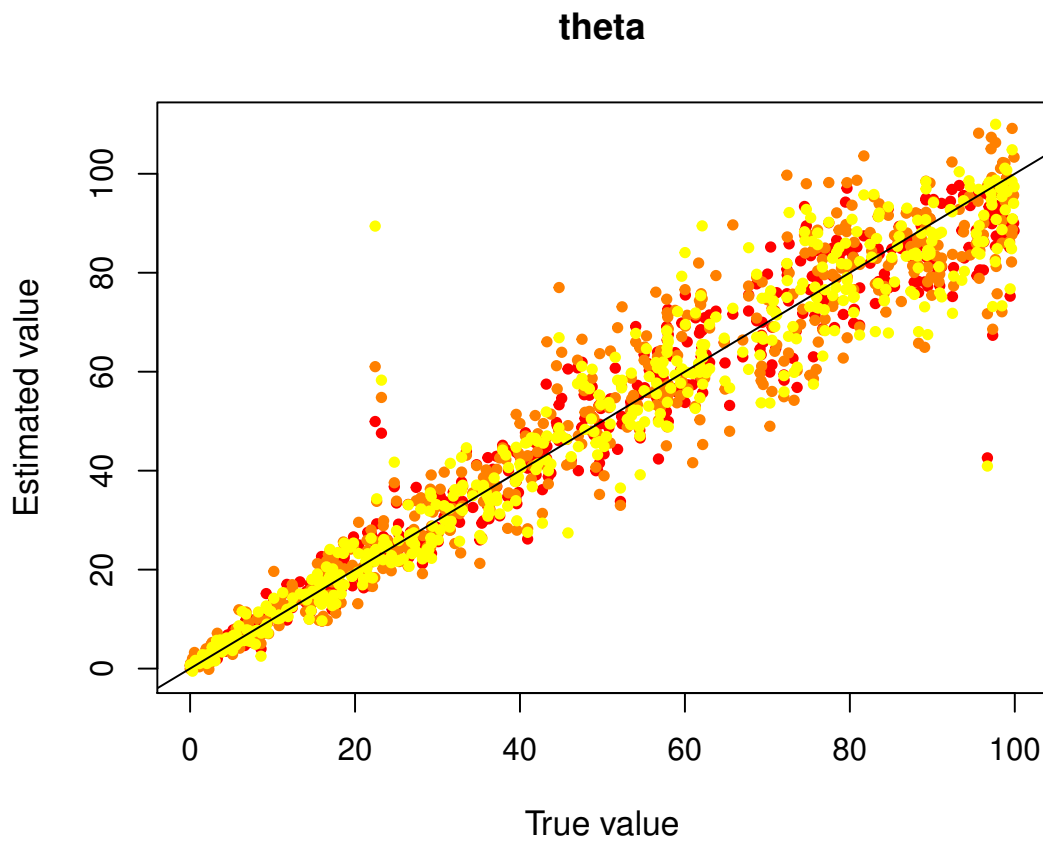
To test if we can correctly infer m and θ from the chosen set of summary statistics, we can perform cross validation.

```
cv <- cv4abc(param = bci.abc$par, sumstat = bci.abc$ss,  
            nval = 500, tols = c(10^-2, 10^-1, 1), method="neuralnet")
```

```
plot(cv)
```



ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

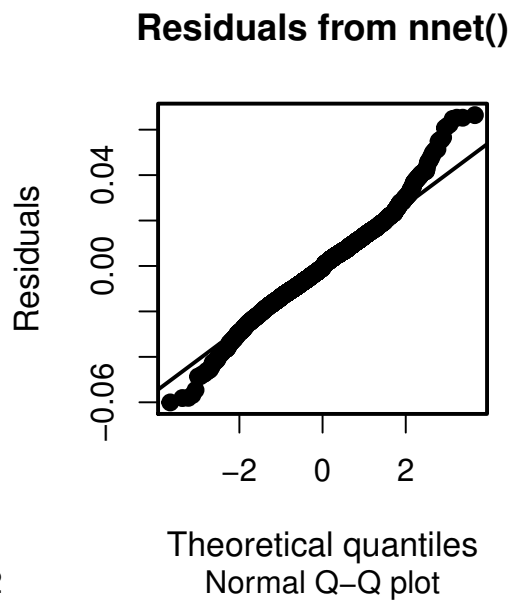
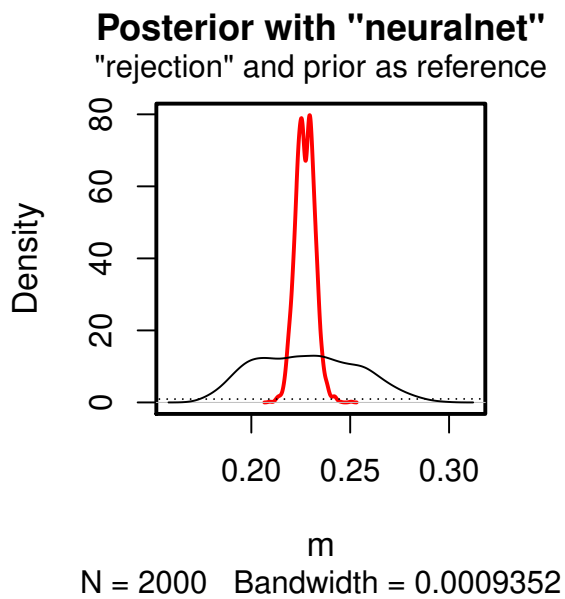
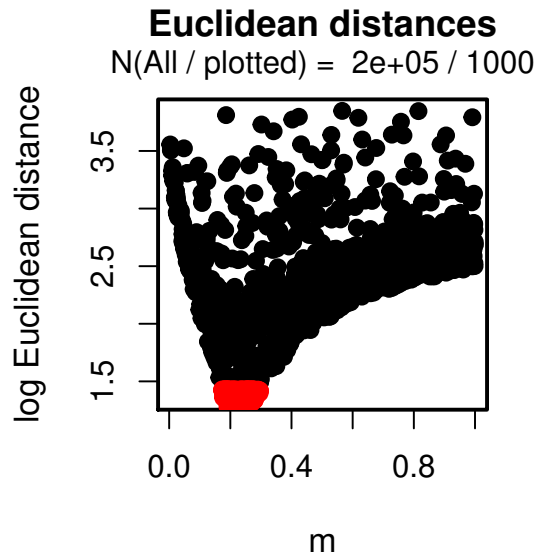
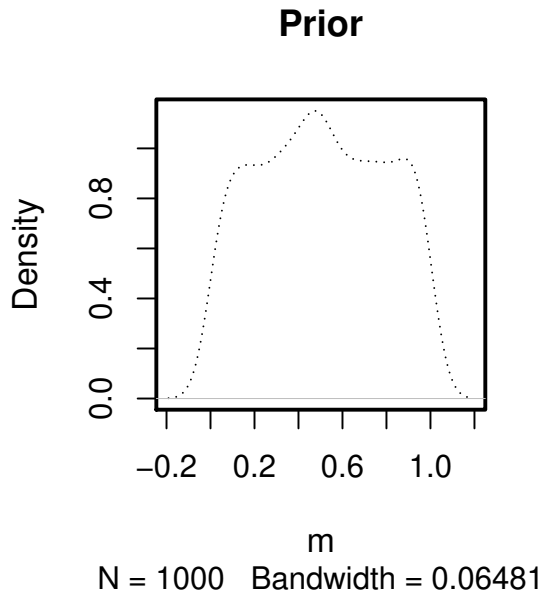


Three colors, yellow, orange and red, represent results of cross-validation for different tolerance levels in ABC analysis, 1, 10^{-1} and 10^{-2} , respectively.

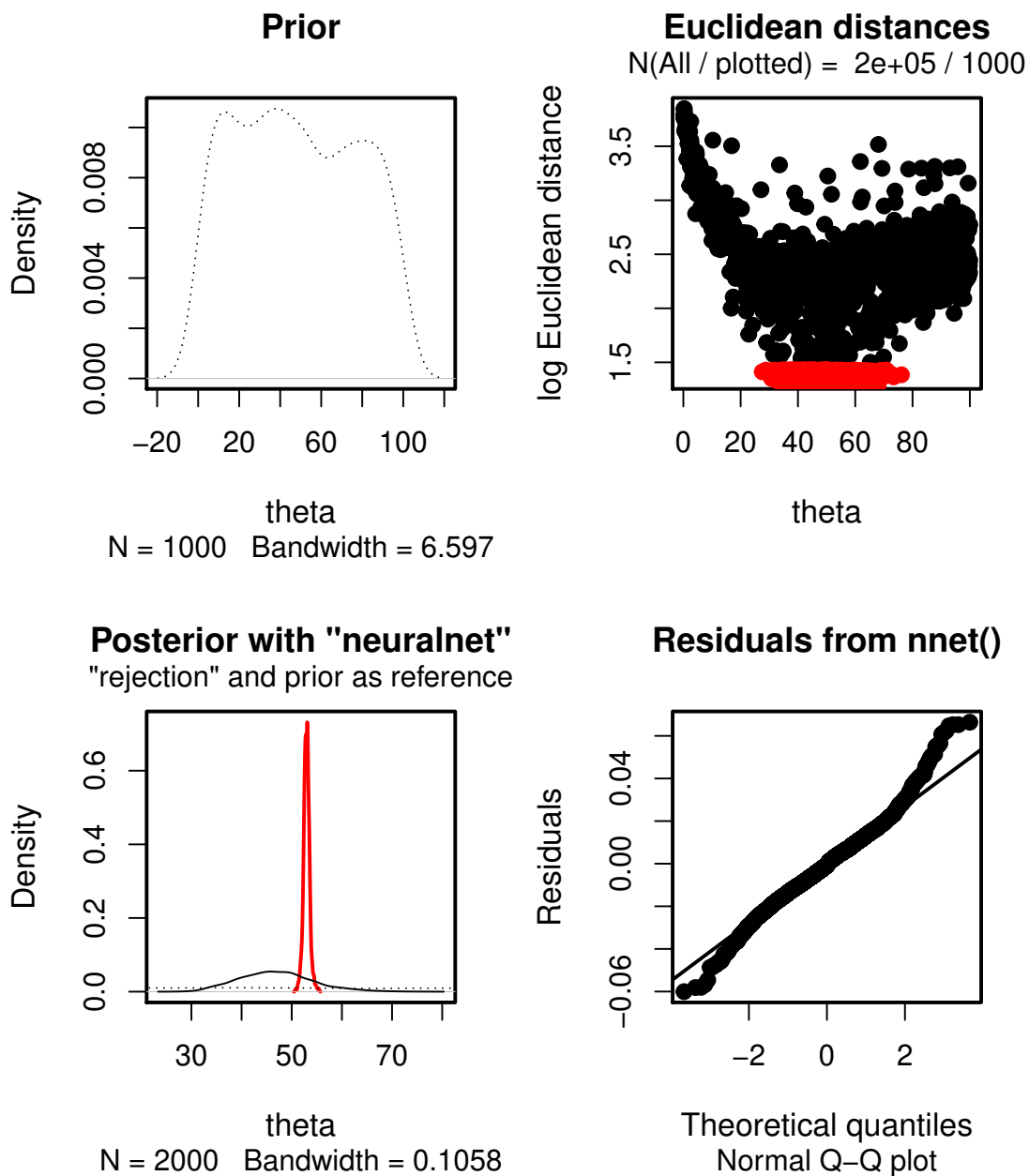
The plot indicates that good estimation of the parameters can be obtained with the selected tolerance levels. We can then reliably estimate the parameters from observed diversity patterns at Barro Colorado Island.

```
plot(bci.abc$abc, param=bci.abc$par)
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics



```
summary(bci.abc$abc)
```

```
## Call:
## abc(target = stats.obs, param = param, sumstat = stats, tol = 0.01,
##      method = "neuralnet")
## Data:
## abc.out$adj.values (2000 posterior samples)
## Weights:
## abc.out$weights
##
## m theta
```

ecolottery - simulating and assessing community assembly with environmental filtering and neutral dynamics

```
## Min.:                0.2094 50.7249
## Weighted 2.5 % Perc.: 0.2192 52.0346
## Weighted Median:    0.2271 52.9654
## Weighted Mean:      0.2273 52.9632
## Weighted Mode:      0.2250 53.1335
## Weighted 97.5 % Perc.: 0.2356 53.9297
## Max.:                0.2505 55.4273
```

We found 95% confidence interval of [52; 54] for θ , compared to $\theta = 50$ in (Hubbell 2001). We found an interval of [0.22; 0.24] for m . While it is somewhat larger than the estimated $m = 0.1$ in (Hubbell 2001), it is close to the estimation of (Etienne & Olf 2004), $m = 0.2$, and still lower than the estimation of (Condit *et al.* 2012), $m = 0.38$, which acknowledged the census of smaller trees. While previous studies mostly addressed migration-drift in the whole 50-ha plot compared to a regional background, our summary statistics here quantifies species turnover and variation of local diversity among the 1-ha subplots. A larger estimate of m can then reflect the fact that β diversity is lower than expected with $m = 0.1$, possibly under the influence of greater replacement dynamics within the plot than between the plot and its regional background.

A next step of analysis could be to incorporate in simulations the influence of environmental variation across subplots on community assembly, depending on the species' ecological properties. For this purpose, the previous ABC analysis can be reproduced with additional environmental filtering and variation of the linkage of environmental filtering to functional traits, as shown in the example in main text. Comparison of the results of the two ABC analyses, one with purely neutral dynamics and one incorporating the influence of environmental filtering, can then allow testing the influence of niche-based dynamics in the dynamics of the rainforest.

References

- Baraloto, C., Hardy, O.J., Paine, C.E.T., Dexter, K.G., Cruaud, C., Dunning, L.T., Gonzalez, M.-A., Molino, J.-F., Sabatier, D., Savolainen, V. & Chave, J. (2012). Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. *Journal of Ecology*, **100**, 690–701.
- Condit, R., Chisholm, R.A. & Hubbell, S.P. (2012). Thirty years of forest census at barro Colorado and the importance of immigration in maintaining diversity. *PLoS ONE*, **7**, e49826.
- Etienne, R.S. & Olf, H. (2004). A novel genealogical approach to neutral biodiversity theory. *Ecology Letters*, **7**, 170–175.
- Hubbell, S. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton and Oxford.

B

Annexes à l'article 4.1

Supplementary Material

- Table A1: Correlation table between traits and Principal Components of PCoA.
- Table A2: Correlation table between number of top 10% species per cell using t-test modified version (Dutilleul et al., 1993).
- Figure A1: Data selection flowchart from Kulbicki et al. 2013.
- Figure A2: Relationship between the species richness at the cell-level between the full dataset and our subset of species.
- Figure A3: Number and proportion of the species Red List status among our subset of species compared to all the coral reef fishes.
- Figure A4: Number and proportion of the species Red List status between the Atlantic and the Indo-Pacific Realms.
- Figure A5: Distinctiveness estimated using all species worldwide in function of Realm-specific worldwide.
- Figure A6: Relationship between functional distinctiveness computed on all traits with functional distinctiveness computed on all but one trait.
- Figure A7: Species Restrictedness estimated using the full dataset of 331 cells including coral and tropical reefs vs. Restrictedness estimated only using the 259 coral reef cells.
- Figure A8: Principal Coordinates Analysis of coral reef fishes functional trait space including the 6316 species available from the full dataset.
- Figure A9: Maps of top 5%, 10% and 15% species of different functional rarity indices.
- Figure A10: Comparison of number of top 10% functionally distinct species between full dataset and subset of species.
- Figure A11: Comparison of cell Phylogenetic Endemism and number of top 10% Evolutionary Rare species

Table A1: Correlation table between variables and Principal Components of PCoA. Coefficients are Spearman's correlation coefficients. †: the reported value is the eta value from a Kruskal-Wallis test because the variable is nominal ***: p-value < 0.001

Variable	PC1 correlation	PC2 correlation
Activity Period	0.62***	0.50***
Diet	0.86***†	0.57***†
Mobility	0.26***	0.58***
Vertical Position	-0.54***	0.61***
Gregariousness	-0.76***	0.36***
Size Class	0.57***	0.54***

Table A2: Correlation table between number of top 10% species per cell using t-test modified version (Dutilleul et al., 1993). p-values significance codes: *** < 0.001, ** < 0.01, * < 0.05, . < 0.01, ns > 0.01.

First metric	Second metric	Correlation coefficient	Corrected degrees of freedom	F-statistic
Functional distinctiveness	Restrictedness	-0.29*	72.52	6.55
Functional distinctiveness	Functional rarity	0.15ns	44.33	1.02
Functional distinctiveness	Evolutionary distinctiveness	0.89***	7.30	27.16
Restrictedness	Functional rarity	0.31***	100.10	10.88
Restrictedness	Evolutionary distinctiveness	-0.23*	88.56	5.11
Functional rarity	Evolutionary distinctiveness	0.19ns	32.39	1.18

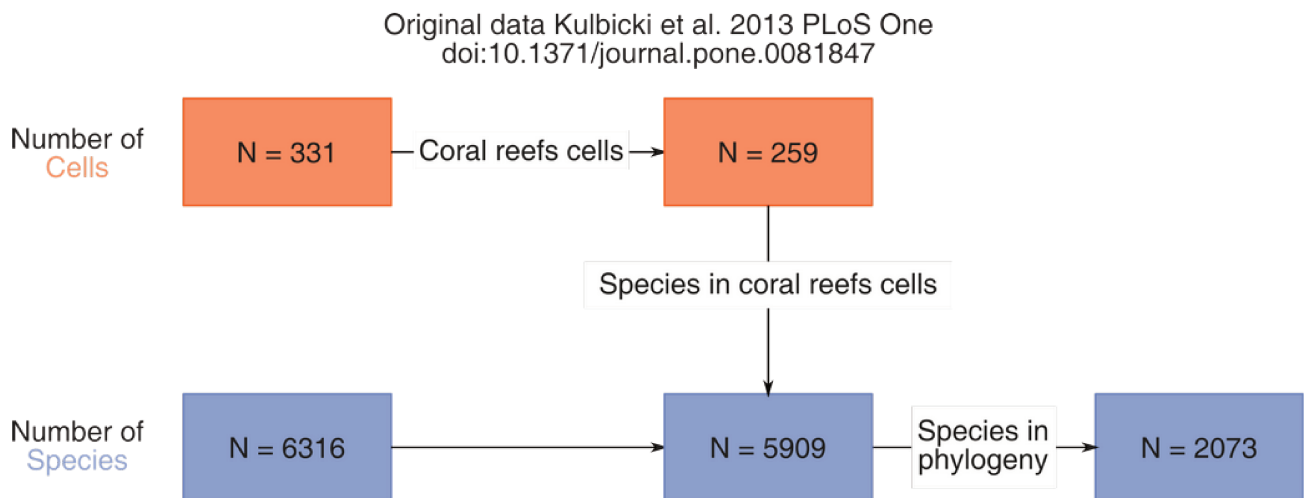


Figure A1: Data flow chart from the data by Kulbicki et al. 2013, showing our sub-selection. The top row, in orange, shows the number of grid cells in our analysis. The bottom row, in blue, show the number of species selected. We started with 331 cells and retained the only 259 which contained coral reefs. Similarly the initial dataset contained 6316 species, of which 5909 occurred in the 259 selected cells. We kept only 2073 species for which we had phylogenetic information.

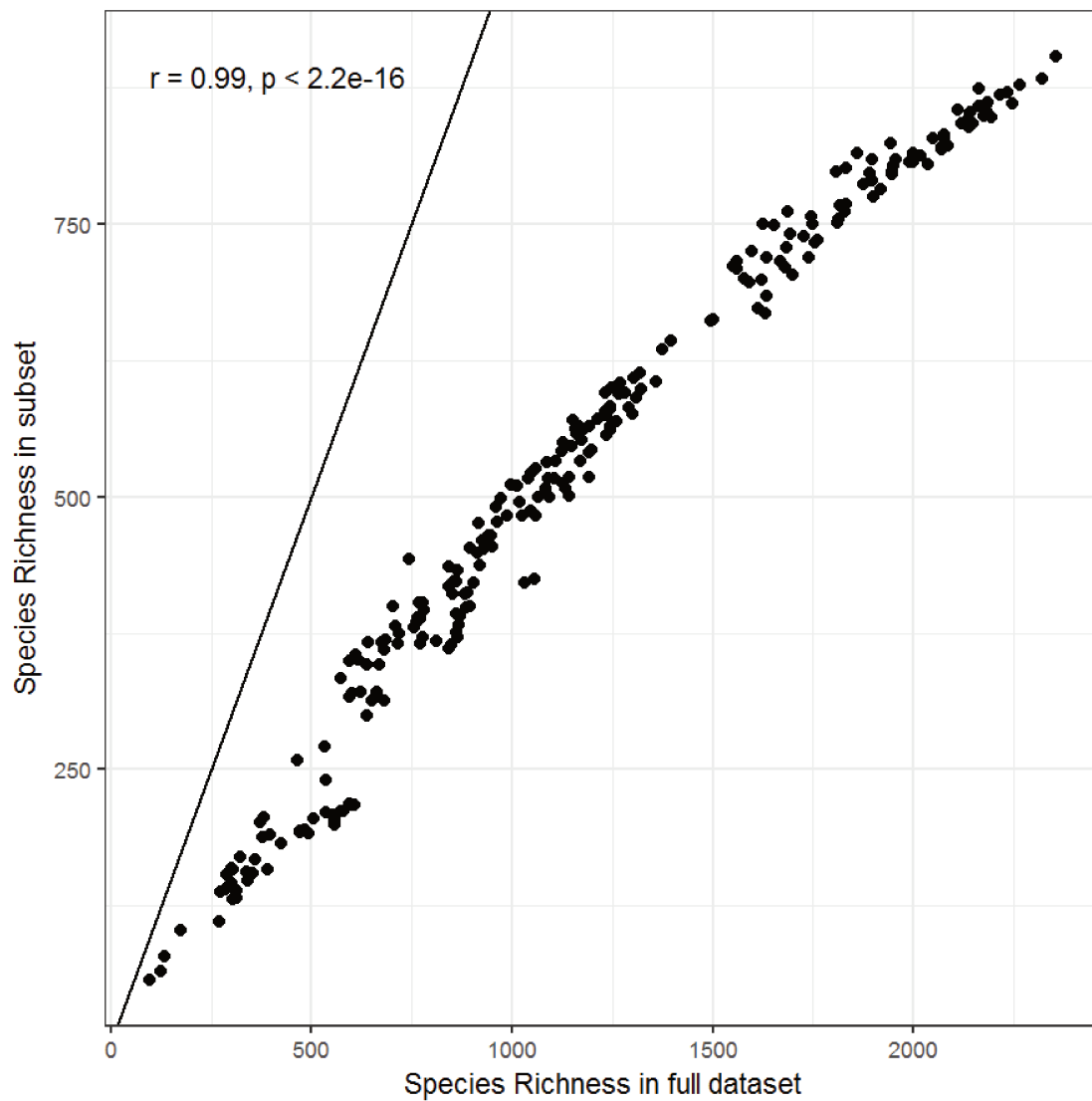


Figure A2: Relationship between the species richness at the cell-level between the full dataset (5909 species) and our subset of species (2073 species), refer to Figure A1 for more details on data selection. Each points represent one of the 259 cell included in the analysis. The solid line is the identity line. In the top left corner is Spearman's rho coefficient with the corresponding p-value indicated.

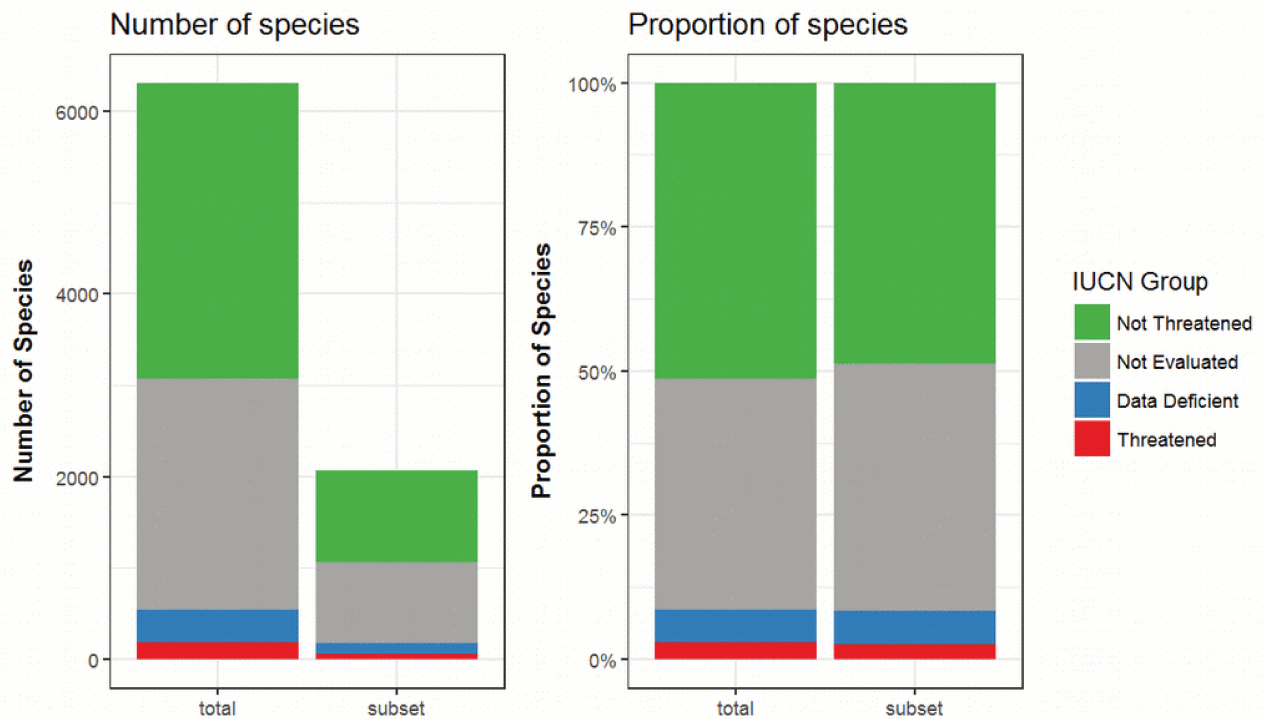


Figure A3: Number and proportion of species for each IUCN Red List status among our subset of 2073 species among the global set (6316 species) of coral reef fishes (please refer to Figure A1 for details on data selection). The left-hand graph show the absolute number of species in each IUCN threat group, while the right-hand graph shows the proportion of species in each threat group between the full dataset and our subset.

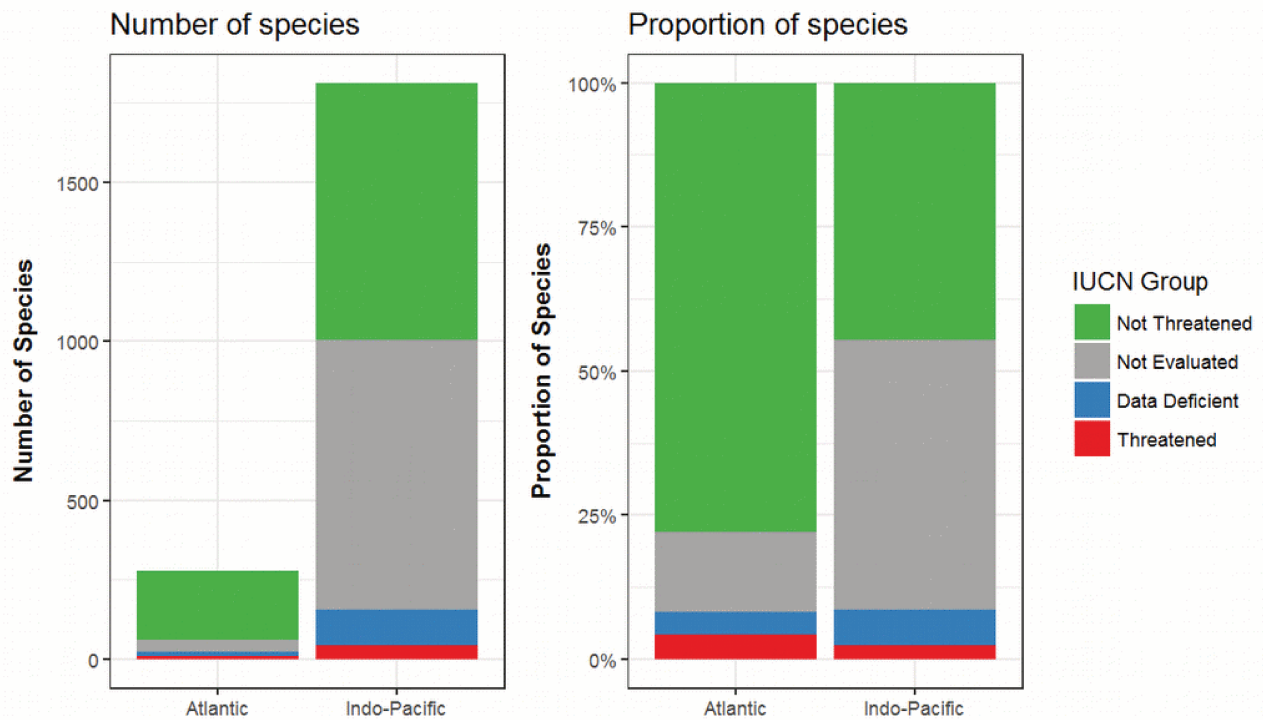


Figure A4: Number and proportion of species for each IUCN Red List status among the two Realms Atlantic and Indo-Pacific. The left-hand graph shows the absolute number of species in each IUCN threat group, while the right-hand graph shows the proportion of species in each threat group.

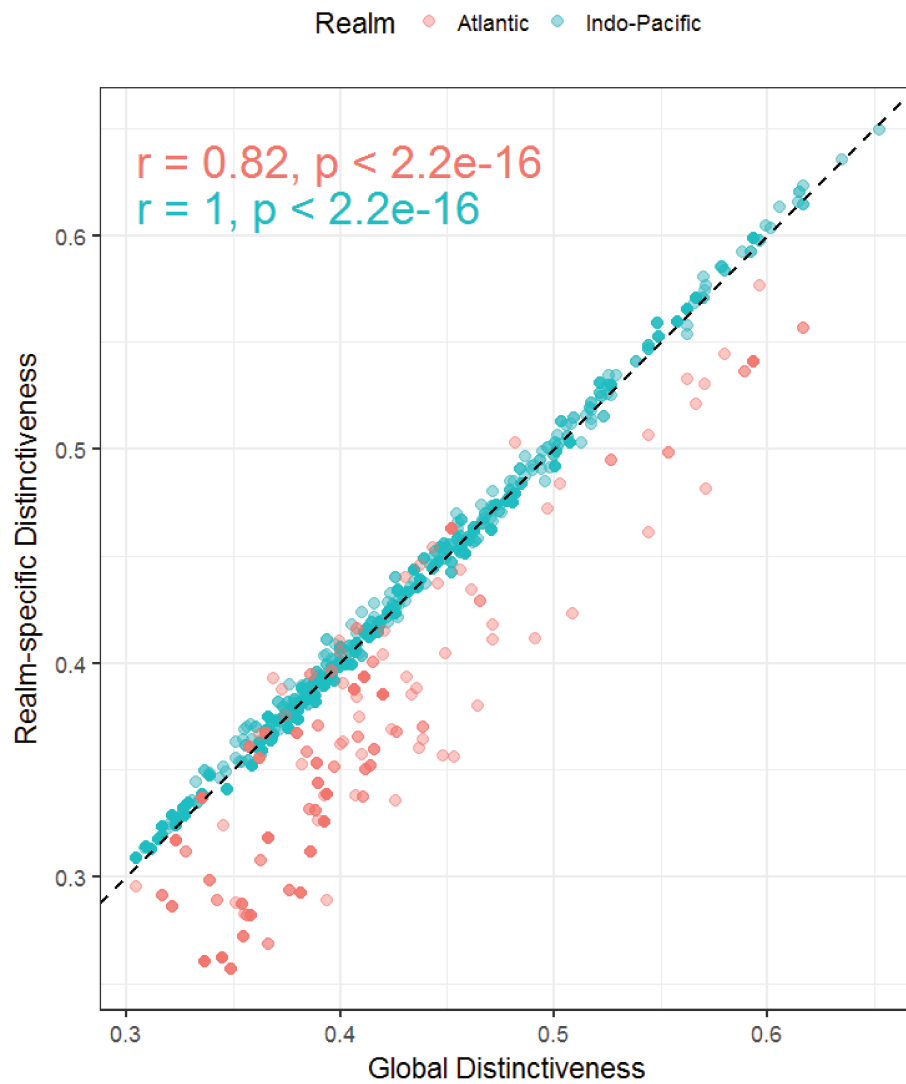


Figure A5: Distinctiveness estimated using all species worldwide in function of Realm-specific worldwide. A point represents a species present in one of the Realm. The dashed line is the identity line. The global dataset is split in two realms: the Atlantic and Indo-Pacific following (Mouillot et al., 2013). The numbers on the top left corner of the plot show Spearman's rho value associated to the corresponding p-value.

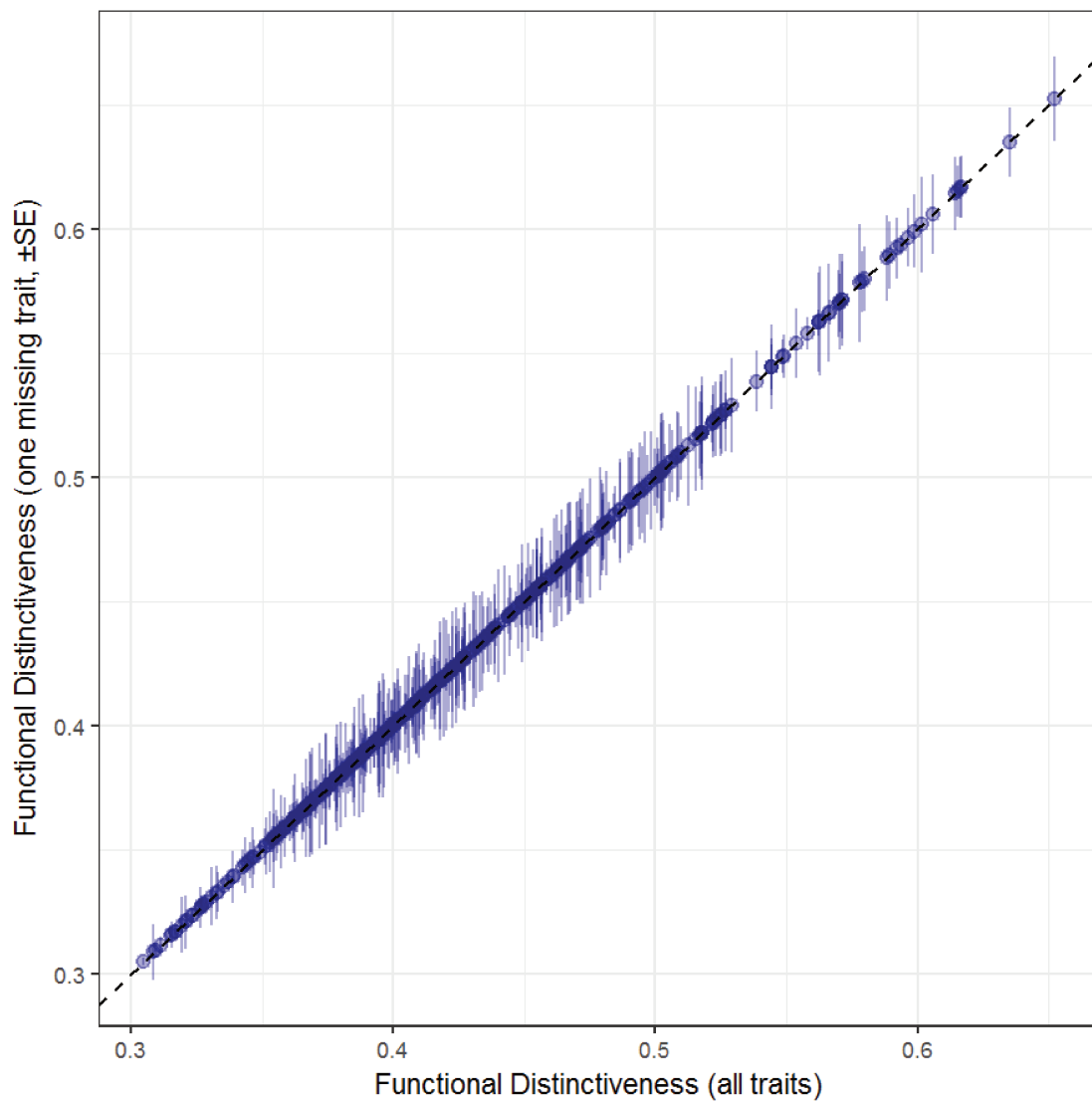


Figure A6: Relationship between functional distinctiveness computed on all traits with functional distinctiveness computed on all but one trait. One point represents the value of distinctiveness of a single species, the y-axis interval is a standard error interval. The dashed line is the identity line. From the full trait matrix we removed each trait, one at a time, then we computed functional distinctiveness values. The standard error interval represents the standard error of each of the six different distinctiveness values we obtained with one missing trait.

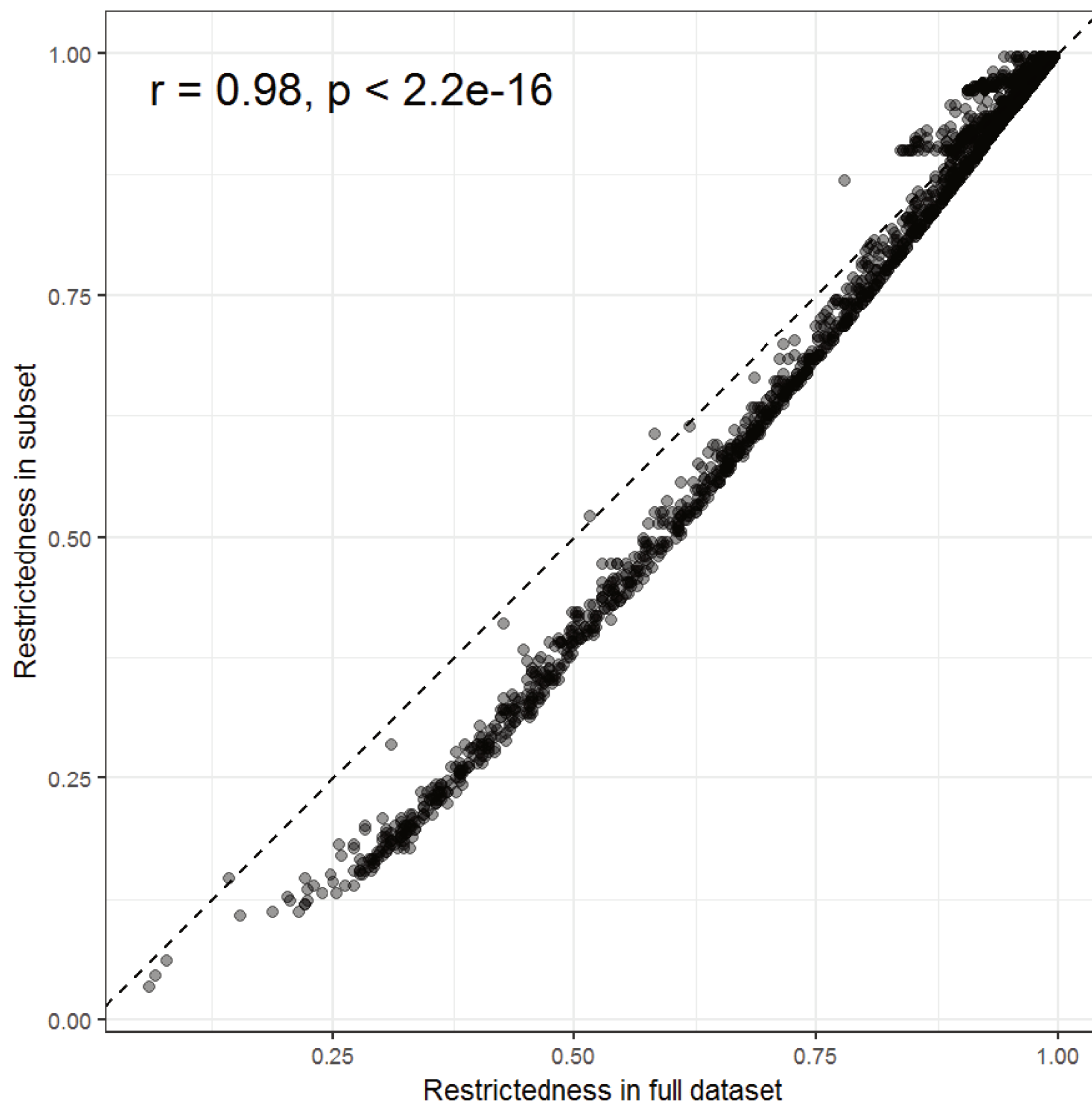


Figure A7: Species Restrictedness estimated using the full dataset of 331 cells including coral and tropical reefs vs. Restrictedness estimated only using the 259 coral reef cells (please refer to Figure A1 for more details on data selection). The correlation coefficient indicates Spearman's rho and the corresponding p-value.



Figure A8: Principal Coordinates Analysis (PCoA) computed on functional traits values on a total of 6315 coral reef fishes. left panel illustrates the 2073 species included in our functional rarity analyses, and the right panel the remaining 4242 species that were not included (see Figure A1 for details on data selection). Please refer to appendix B for the meaning and the interpretation of each functional trait.

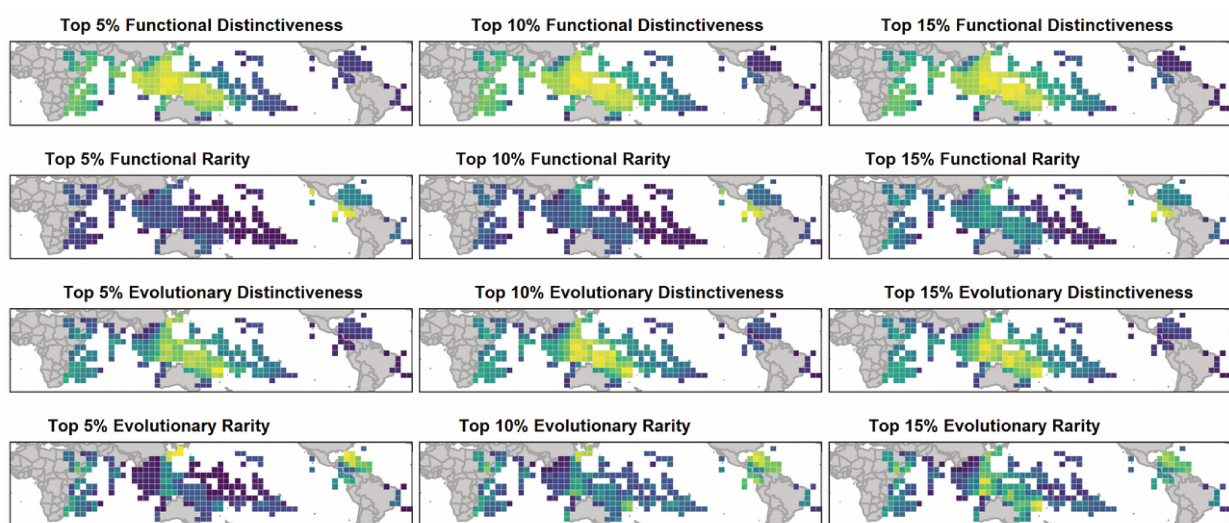


Figure A9: Maps of top 5%, 10% and 15% species of different rarity indices. A darker cell indicates a lower value while a brighter cell indicates a higher number of species pertaining to the top percentile. The color scales linearly with the values, thus a yellow cell represents always the same relative value.

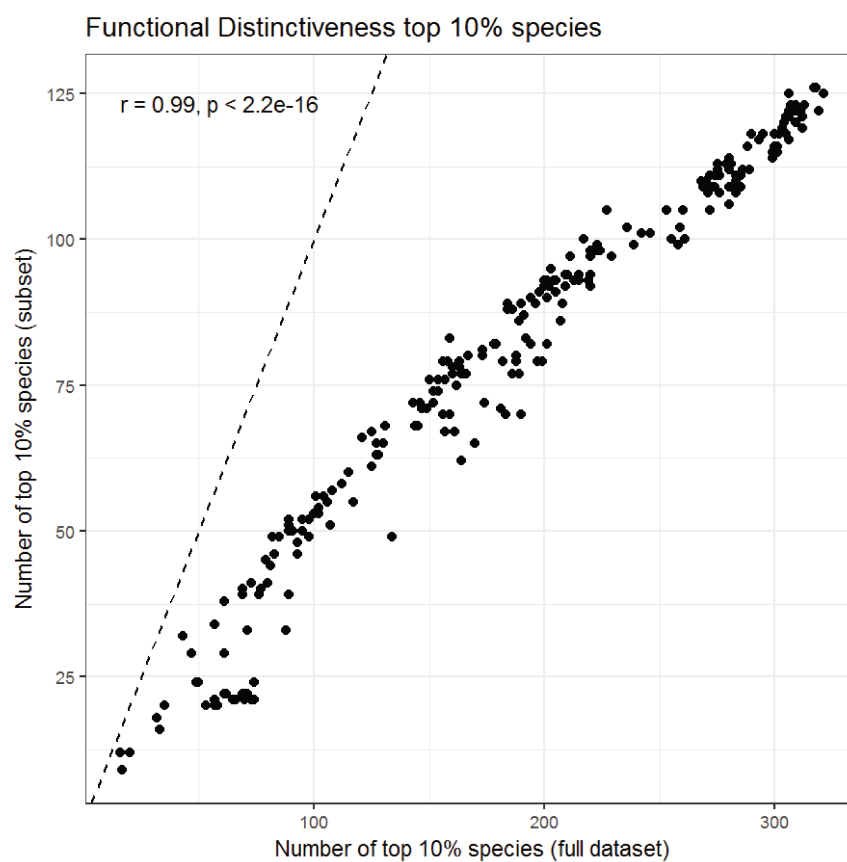
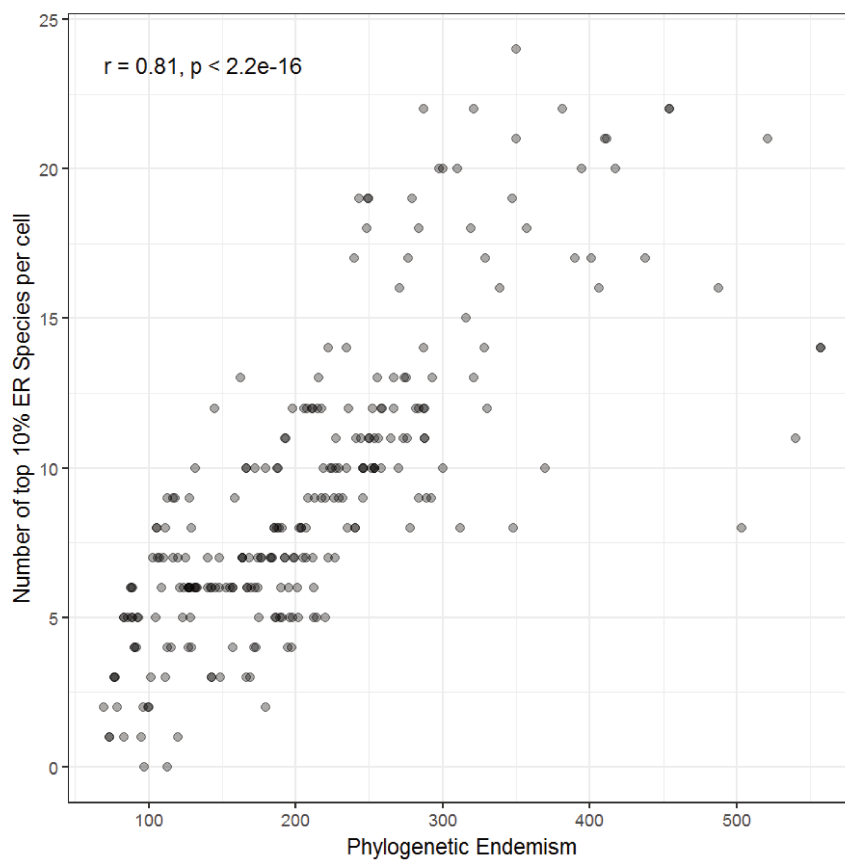


Figure A10: Comparison of number of top 10% functionally distinct species between full dataset and subset of species. We estimated the number of top 10% functionally distinct species in cells. The x-axis is the number estimated on the dataset comprising 5909 species (see Figure A1 for details on data subset), while the y-axis is the number estimated from our subset of 2073 species. The correlation coefficient in the top left corner is Spearman's rho with the associated p-value.

Figure A11: Comparison of cell Phylogenetic Endemism and number of top 10% Evolutionary Rare species per cell. The x-axis is the Phylogenetic Endemism per cell as defined by Rosauer et al. 2009 Mol. Ecol.; the y-axis is the number of top 10% Evolutionary Rare species per cell. The correlation coefficient in the top left corner is Spearman's rho with the associated p-value.



Appendix B: Ecological relevance of fish functional traits

Fish plays important roles in aquatic ecosystems mainly through regulation of food webs and nutrient cycling (Holmlund and Hammer, 1999; Villéger et al., 2017). The ability of each species to impact these ecosystems processes depends on several biological traits linked to food acquisition and locomotion (Villéger et al., 2010; Winemiller, 1991). For instance, the trophic impact of a species depends on its foraging activity, i.e. which prey items it targets, when and how many? For instance, an ambush solitary benthic predator (e.g. scorpionfish) will not have the same trophic impact than a mobile pelagic gregarious predator (e.g. barracuda) on pelagic fishes. Therefore, describing fish functional niche requires considering a set of complementary functional traits. Here, we selected 6 traits which describe the main facets of fish ecology (Belmaker et al., 2013; Buisson et al., 2008; Guillemot et al., 2011; Olden et al., 2008) and which are available for a wide range of reef species (Mouillot et al., 2014).

BODY SIZE has a primary role in defining fish ecological niche (Fisher et al., 2010; Wilson, 1975). More specifically, size determines energy needs through the amount of energy required per unit of body mass (L. Munday and Jones, 1998), and constrain prey-predator relationships since mouth gap scales with body size (Costa, 2009). Size also influences growth rate with small fishes growing faster than larger ones (Brown et al., 2004). Mortality rate tends to be higher for smaller fishes (L. Munday and Jones, 1998) while temperature tolerance is at least partly related to body size in reef fishes (Ospina and Mora, 2004).

DIET, as size, is an essential component of reef fish ecological niche as indicated in general reviews (Bellwood et al., 2006; Burkepile and Hay, 2008; Hobson, 1975; Sale, 1977). In particular diet determines fish impact on ecosystem functioning through trophic interactions with other food web components (Green and Bellwood, 2009; Mumby et al., 2006) and, consequently, on nutrient cycling (Burkepile et al., 2013; Vanni, 2002).

MOBILITY determines energy needs with mobile species requiring a lot of energy by mass unit compared to sedentary species (Norman and Jones, 1984). Mobility also affects the spatial extent at which fishes control their resources and transfer nutrients, especially between habitats around reefs (Meyer and Schultz, 1985; Nagelkerken et al., 2000).

PERIOD OF THE DAY AT WHICH FISHES ARE ACTIVE has implication on the trophic role a species plays in the food web through both bottom-up, i.e. the set of resources it can target (Fox and Bellwood, 2011), and top-down controls, i.e. the susceptibility it has to be predated. For instance, most of nocturnal species escape predation from active predators during the day and vice-versa (S. Helfman, 1986).

VERTICAL POSITION IN THE WATER COLUMN occupied by fish is critical for determining fish ecological niche as it influences the set of potential prey available (Bellwood et al., 2006) and fish impacts on nutrient transfer between vertical strata (Schaus and Vanni, 2000).

GREGARIOUSNESS of fish is an important component of fish behavior that determines the ability of (i) escaping from predation (J. Hoare et al., 2005; Stier et al., 2013) and (ii) impacting local ecological

processes with schooling species inducing potentially massive nutrient cycling and resource depletion (Foster, 1985; Meyer and Schultz, 1985; Robertson et al., 1976).

All these traits are also impacting sensibility of fish to fishing (e.g. a small sedentary benthic fish will less likely be caught by a net than a pelagic large fish).

REFERENCES

- Bellwood, D.R., Wainwright, P.C., Fulton, C.J., Hoey, A.S., 2006. Functional versatility supports coral reef biodiversity. *Proc. R. Soc. Lond. B Biol. Sci.* 273, 101–107. <https://doi.org/10.1098/rspb.2005.3276>
- Belmaker, J., Parravicini, V., Kulbicki, M., 2013. Ecological traits and environmental affinity explain Red Sea fish introduction into the Mediterranean. *Glob. Change Biol.* 19, 1373–1382. <https://doi.org/10.1111/gcb.12132>
- Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M., West, G.B., 2004. Toward a Metabolic Theory of Ecology. *Ecology* 85, 1771–1789. <https://doi.org/10.1890/03-9000>
- Buisson, L., Blanc, L., Grenouillet, G., 2008. Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. *Ecol. Freshw. Fish* 17, 244–257. <https://doi.org/10.1111/j.1600-0633.2007.00276.x>
- Burkepile, D.E., Allgeier, J.E., Shantz, A.A., Pritchard, C.E., Lemoine, N.P., Bhatti, L.H., Layman, C.A., 2013. Nutrient supply from fishes facilitates macroalgae and suppresses corals in a Caribbean coral reef ecosystem. *Sci. Rep.* 3, 1493. <https://doi.org/10.1038/srep01493>
- Burkepile, D.E., Hay, M.E., 2008. Herbivore species richness and feeding complementarity affect community structure and function on a coral reef. *Proc. Natl. Acad. Sci.* 105, 16201–16206. <https://doi.org/10.1073/pnas.0801946105>
- Costa, G.C., 2009. Predator size, prey size, and dietary niche breadth relationships in marine predators. *Ecology* 90, 2014–2019. <https://doi.org/10.1890/08-1150.1>
- Fisher, J.A.D., Frank, K.T., Leggett, W.C., 2010. Global variation in marine fish body size and its role in biodiversity–ecosystem functioning. *Mar. Ecol. Prog. Ser.* 405, 1–13. <https://doi.org/10.3354/meps08601>
- Foster, S.A., 1985. Group foraging by a coral reef fish: a mechanism for gaining access to defended resources. *Anim. Behav.* 33, 782–792. [https://doi.org/10.1016/S0003-3472\(85\)80011-7](https://doi.org/10.1016/S0003-3472(85)80011-7)
- Fox, R.J., Bellwood, D.R., 2011. Unconstrained by the clock? Plasticity of diel activity rhythm in a tropical reef fish, *Siganus lineatus*. *Funct. Ecol.* 25, 1096–1105. <https://doi.org/10.1111/j.1365-2435.2011.01874.x>
- Green, A., Bellwood, D., 2009. Monitoring Functional Groups of Herbivorous Reef Fishes as Indicators of Coral Reef Resilience A practical guide for coral reef managers in the Asia Pacific Region.
- Guillemot, N., Kulbicki, M., Chabanet, P., Vigliola, L., 2011. Functional Redundancy Patterns Reveal Non-Random Assembly Rules in a Species-Rich Marine Assemblage. *PLOS ONE* 6, e26735. <https://doi.org/10.1371/journal.pone.0026735>
- Hobson, E.S., 1975. Feeding patterns among tropical reef fishes. *Am Sci* 63, 382–392.
- Holmlund, C.M., Hammer, M., 1999. Ecosystem services generated by fish populations. *Ecol. Econ.* 29, 253–268. [https://doi.org/10.1016/S0921-8009\(99\)00015-4](https://doi.org/10.1016/S0921-8009(99)00015-4)
- J. Hoare, D., Krause, J., Peuhkuri, N., Godin, J.-G., 2005. Body size and shoaling in fish. *J. Fish Biol.* 57, 1351–1366. <https://doi.org/10.1111/j.1095-8649.2000.tb02217.x>

- L. Munday, P., Jones, G., 1998. The ecological implications of small body size among coral-reef fishes. *Oceanogr. Mar. Biol.* 36, 373–411.
- Meyer, J.L., Schultz, E.T., 1985. Migrating haemulid fishes as a source of nutrients and organic matter on coral reefs. *Limnol. Oceanogr.* 30, 146–156. <https://doi.org/10.4319/lo.1985.30.1.0146>
- Mouillot, D., Villéger, S., Parravicini, V., Kulbicki, M., Arias-González, J.E., Bender, M., Chabanet, P., Floeter, S.R., Friedlander, A., Vigliola, L., Bellwood, D.R., 2014. Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. *Proc. Natl. Acad. Sci.* 111, 13757–13762. <https://doi.org/10.1073/pnas.1317625111>
- Mumby, P.J., Dahlgren, C.P., Harborne, A.R., Kappel, C.V., Micheli, F., Brumbaugh, D.R., Holmes, K.E., Mendes, J.M., Broad, K., Sanchirico, J.N., Buch, K., Box, S., Stoffle, R.W., Gill, A.B., 2006. Fishing, Trophic Cascades, and the Process of Grazing on Coral Reefs. *Science* 311, 98–101. <https://doi.org/10.1126/science.1121129>
- Nagelkerken, I., Dorenbosch, M., Verberk, W., De La Morinière, E.C., Van Der Velde, G., 2000. Day-night shifts of fishes between shallow-water biotopes of a Caribbean bay, with emphasis on the nocturnal feeding of Haemulidae and Lutjanidae. *Mar. Ecol. Prog. Ser.* 55–64.
- Norman, M.D., Jones, G.P., 1984. Determinants of territory size in the pomacentrid reef fish, *Parma victoriae*. *Oecologia* 61, 60–69. <https://doi.org/10.1007/BF00379090>
- Olden, J.D., Poff, N.L., Bestgen, K.R., 2008. Trait synergisms and the rarity, extirpation, and extinction risk of desert fishes. *Ecology* 89, 847–856.
- Ospina, A.F., Mora, C., 2004. Effect of Body Size on Reef Fish Tolerance to Extreme Low and High Temperatures. *Environ. Biol. Fishes* 70, 339–343. <https://doi.org/10.1023/B:EBFI.0000035429.39129.34>
- Robertson, D.R., P.A. Sweatman, H., A. Fletcher, E., G. Cleland, M., 1976. Schooling as a Mechanism for Circumventing the Territoriality of Competitors. *Ecology* 57, 1208. <https://doi.org/10.2307/1935045>
- S. Helfman, G., 1986. The Behaviour of Teleost Fish, in: *The Behaviour of Teleost Fishes*. pp. 366–387. https://doi.org/10.1007/978-1-4684-8261-4_14
- Sale, P.F., 1977. Maintenance of High Diversity in Coral Reef Fish Communities. *Am. Nat.* 111, 337–359. <https://doi.org/10.1086/283164>
- Schaus, M.H., Vanni, M.J., 2000. Effects of Gizzard Shad on Phytoplankton and Nutrient Dynamics: Role of Sediment Feeding and Fish Size. *Ecology* 81, 1701–1719. [https://doi.org/10.1890/0012-9658\(2000\)081\[1701:EOGSOP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[1701:EOGSOP]2.0.CO;2)
- Stier, A.C., Geange, S.W., Bolker, B.M., 2013. Predator density and competition modify the benefits of group formation in a shoaling reef fish. *Oikos* 122, 171–178. <https://doi.org/10.1111/j.1600-0706.2012.20726.x>
- Vanni, M.J., 2002. Nutrient Cycling by Animals in Freshwater Ecosystems. *Annu. Rev. Ecol. Syst.* 33, 341–370. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150519>
- Villéger, S., Brosse, S., Mouchet, M., Mouillot, D., Vanni, M.J., 2017. Functional ecology of fish: current approaches and future challenges. *Aquat. Sci.* 79, 783–801. <https://doi.org/10.1007/s00027-017-0546-z>
- Villéger, S., Miranda, J.R., Hernández, D.F., Mouillot, D., 2010. Contrasting changes in taxonomic vs. functional diversity of tropical fish communities after habitat degradation. *Ecol. Appl.* 20, 1512–1522. <https://doi.org/10.1890/09-1310.1>

- Wilson, D.S., 1975. The Adequacy of Body Size as a Niche Difference. *Am. Nat.* 109, 769–784. <https://doi.org/10.1086/283042>
- Winemiller, K., 1991. Ecomorphological Diversification in Lowland Freshwater Fish Assemblages from Five Biotic Regions. *Ecol. Monogr. - ECOL MONOGR* 61, 343–365. <https://doi.org/10.2307/2937046>

C

Annexes à l'article 4.2

List of Supplementary Materials

This PDF file includes:

Supplementary tables 1 to 3.

Supplementary figures 1 to 11.

Figures

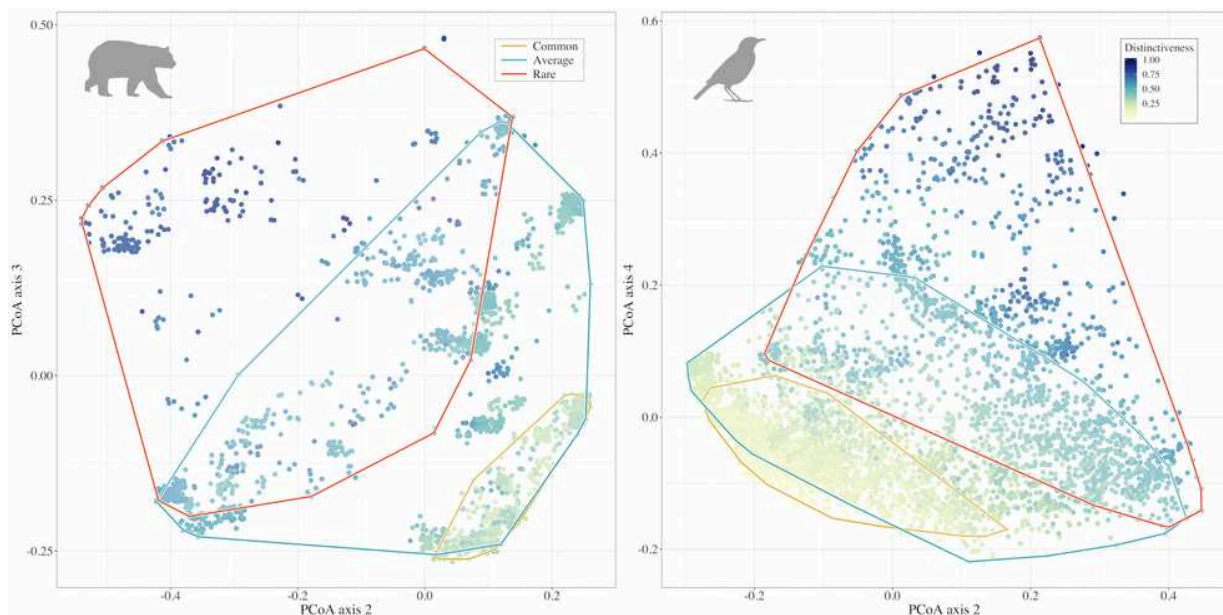


Figure 1. Functional spaces of mammals and birds. Principal Coordinates Analysis (PCoA) representing the functional space of mammals and birds based on their functional traits. The two axes (PC2 and PC3 for mammals, PC2 and PC4 for birds,) explain respectively 35.3% and 26.4% of the variance in the distance matrix (Supplementary table S1 and S2). Species are colored by functional distinctiveness values. Functional space of ecologically rare species is in red, common and average species are respectively in orange and blue.

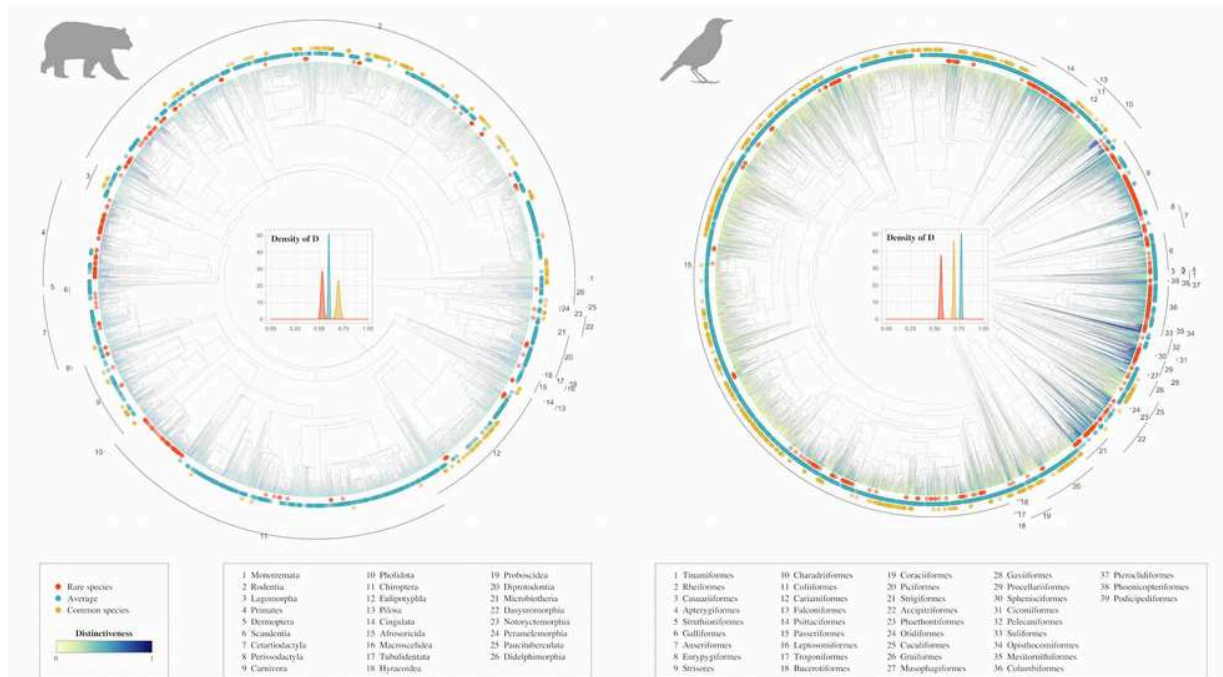


Figure 2. Phylogenies of terrestrial mammals and birds. Functional distinctiveness values are represented on terminal branches (color gradient). Ecologically rare species are highlighted by red dots, common by orange and average by blue. Distribution of the values of phylogenetic signal of ecological rarities (index D) are plotted in the center of the tree. Orders names are indicated on the outside arcs. The figure represents a single phylogeny from the 100 phylogenies generated (see methods).

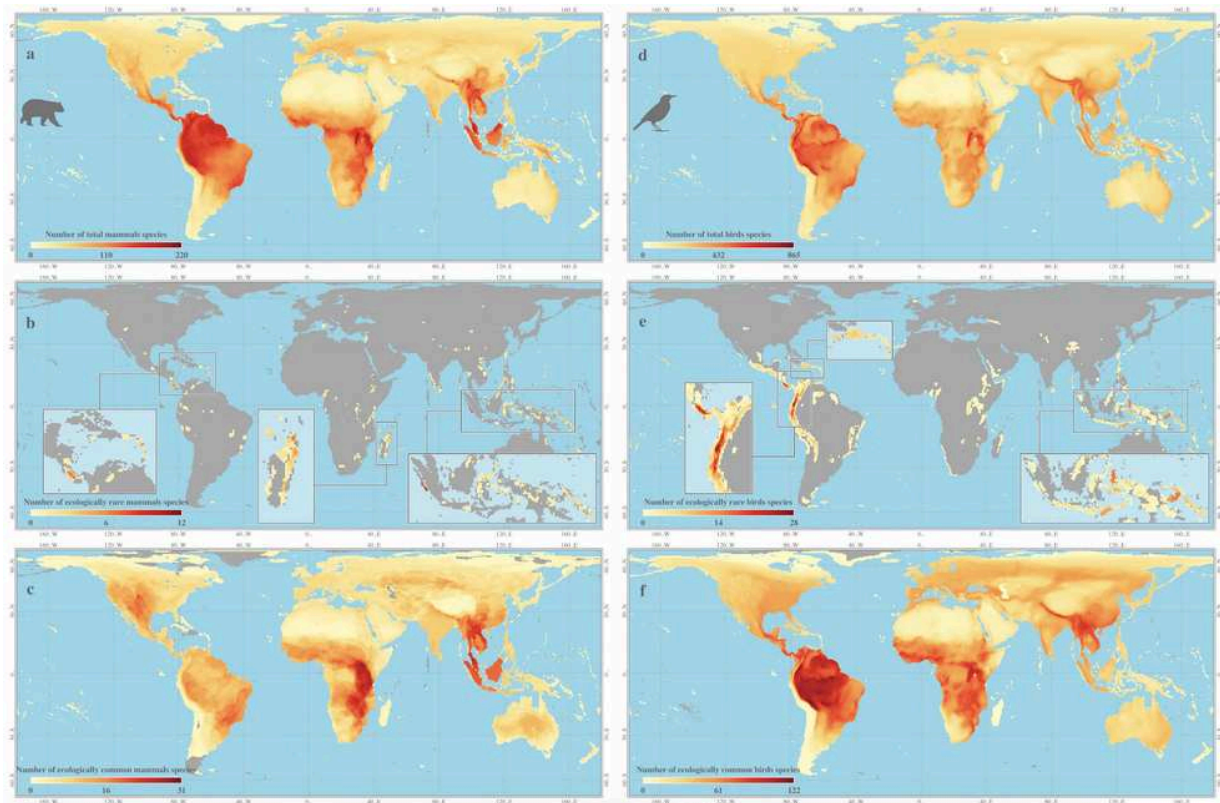


Figure 3. Global distribution considering only unglaciated areas of total species richness (a, d), number of ecologically rare (b, e) and common (c, f) mammals (left) and birds (right) species.

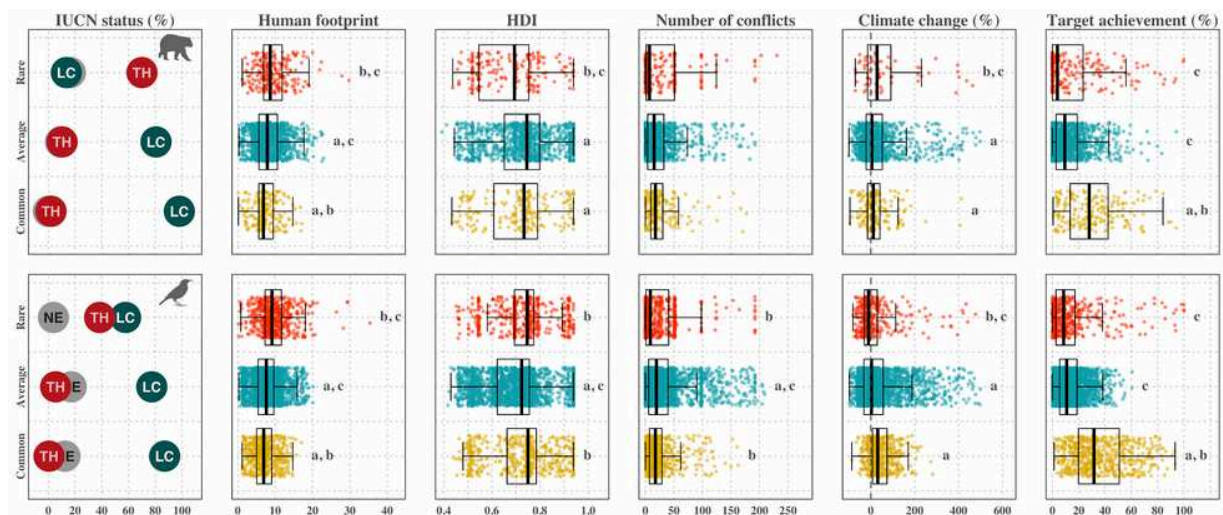


Figure 4. From left to right: percentage of species per IUCN status according to criteria by the IUCN Red List, Threatened (TH red: Vulnerable, Endangered, Critically Endangered), Least concern (LC green: Least Concern and Near Threatened) and Not Evaluated (NE grey: Not Evaluated or inadequate information). Main threats on ecologically rare (red), average (blue) and common (orange) species; Human Footprint that measure the cumulative impact of direct human pressures; Human development index (HDI); Change in distribution range based on climate change projections (scenario RCP 8.5, Horizon 2041 - 2060, see Figure S8 supplementary material for Horizon 2061 - 2080) and Target achievement (extent to which species are represented within PAs regarding their restrictiveness). Letters indicate significant difference between group ($P < 0.05$, table S3) via ANOVA and Tukey's post-hoc tests.

SUPPLEMENTARY MATERIALS**Functional rarity in the Anthropocene****This PDF file includes:**

Supplementary Tables 1 to 3

Supplementary Figures 1 to 11

Supplementary Tables

Supplementary Table 1: Correlations between functional traits and the first four axes of the PcoA of mammals (explaining respectively 29.3%, 21.8%, 13.5%, 8.2% of the total variance). For numerical traits we used Spearman correlations and for nominal traits we used Goodman and Kruskal's gamma correlations. The first axis of functional differentiation integrated both a body mass gradient ($R^2 = -0.58$) and diet (herbivores $R^2 = -0.80$ vs. invertebrate-eaters $R^2 = 0.68$). The second axis integrated both, body mass gradient (body mass, $R^2 = -0.45$) and diet (frugivores $R^2 = -0.61$ vs. invertebrate-eaters $R^2 = 0.53$). The third axis integrated mostly diet (granivores $R^2 = -0.26$ vs. invertebrate-eaters $R^2 = 0.40$). The fourth axis integrated mostly diet (scavenger $R^2 = -0.41$ vs. herbivores $R^2 = 0.44$).

Functions	Traits	Correlation	PC1	PC2	PC3	PC4
Diet	Invertebrate	Spearman	0.68	0.53	0.40	-0.07
Diet	Mammals/birds	Spearman	0.01	0.06	0.01	-0.18
Diet	Reptiles	Spearman	0.11	0.28	-0.005	-0.32
Diet	Fish	Spearman	-0.02	0.10	0.06	-0.10
Diet	Vertebrates	Spearman	0.01	-0.08	0.08	-0.15
Diet	Scavenge	Spearman	0.11	0.41	0.02	-0.41
Diet	Fruit	Spearman	0.01	-0.62	-0.22	-0.36
Diet	Nectar	Spearman	0.17	-0.23	-0.05	-0.10
Diet	Seed	Spearman	-0.24	-0.06	-0.26	-0.15
Diet	Plant	Spearman	-0.80	-0.23	-0.23	0.44
Foraging stratum	Foraging stratum	Kruskal	0.34	0.38	0.1	0.24
Activity	Nocturnal	Kruskal	0.15	0.13	0.25	0.02
Activity	Crepuscular	Kruskal	0.21	0.01	0.11	0.02
Activity	Diurnal	Kruskal	0.26	0.03	0.33	0.01
BodyMass	BodyMass	Spearman	-0.60	-0.45	-0.09	0.01

Supplementary Table 2: Correlations between functional traits and the first four axes of the PcoA of birds (explaining respectively 26.4%, 17.8%, 11.4%, 8.6% of the total variance). For numerical traits we used Spearman correlations and for nominal traits we used Goodman and Kruskal's gamma correlations. The first axis of functional differentiation integrated mostly foraging strata position (ground $R^2 = -0.88$ vs. middle high $R^2 = 0.82$). The second axis integrated mostly diet (invertebrate-eaters $R^2 = -0.90$ vs. frugivores $R^2 = 0.50$). The third axis integrated mostly foraging strata position (forest understory $R^2 = -0.84$ vs. canopy $R^2 = 0.39$). The fourth axis integrated foraging strata position (water surface $R^2 = 0.37$ vs. ground $R^2 = -0.29$) and diet (piscivores $R^2 = 0.36$ vs. frugivores $R^2 = -0.62$).

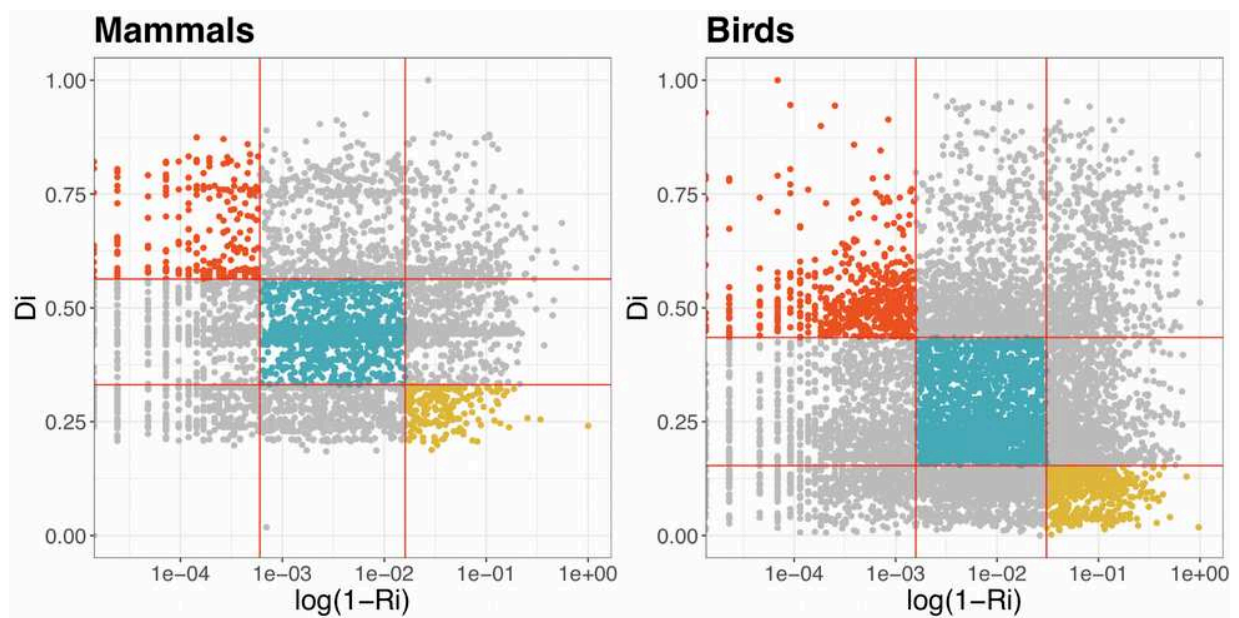
Functions	Traits	Correlation	PC1	PC2	PC3	PC4
Diet	Invertebrates	Spearman	-0.12	-0.91	0.21	-0.03
Diet	Mammals/birds	Spearman	-0.15	0.14	0.10	0.18
Diet	Reptiles	Spearman	-0.17	0.05	0.13	0.11
Diet	Fish	Spearman	-0.14	0.11	0.30	0.36
Diet	Vertebrates	Spearman	-0.05	0.02	0.04	0.03
Diet	Scavenge	Spearman	-0.13	0.11	0.08	0.10
Diet	Fruit	Spearman	0.40	0.51	-0.003	-0.62
Diet	Nectar	Spearman	0.36	0.11	-0.22	0.23
Diet	Seed	Spearman	-0.34	0.42	-0.30	-0.17
Diet	Plant	Spearman	-0.15	0.33	-0.10	0.03
Foraging stratum	below the water surfaces	Spearman	-0.07	0.09	0.23	0.32
Foraging stratum	<5 inches below water surface	Spearman	-0.20	0.09	0.30	0.38
Foraging stratum	ground	Spearman	-0.88	0.16	-0.03	-0.29
Foraging stratum	understory	Spearman	0.21	-0.40	-0.84	0.15
Foraging stratum	Middle high tree	Spearman	0.82	-0.03	0.02	-0.16
Foraging stratum	above tree canopy	Spearman	0.61	0.21	0.39	-0.16
Foraging stratum	ForStrat-aerial	Spearman	0.08	-0.04	0.1	0.19
Diet	predominantly pelagic	Kruskal	0.002	0.01	0.05	0.07

Nocturnal	foraging activity at night	Kruskal	0.004	0.0001	0.004	0.04
Bodymass	body mass	Spearman	-0.28	0.45	0.34	0.02

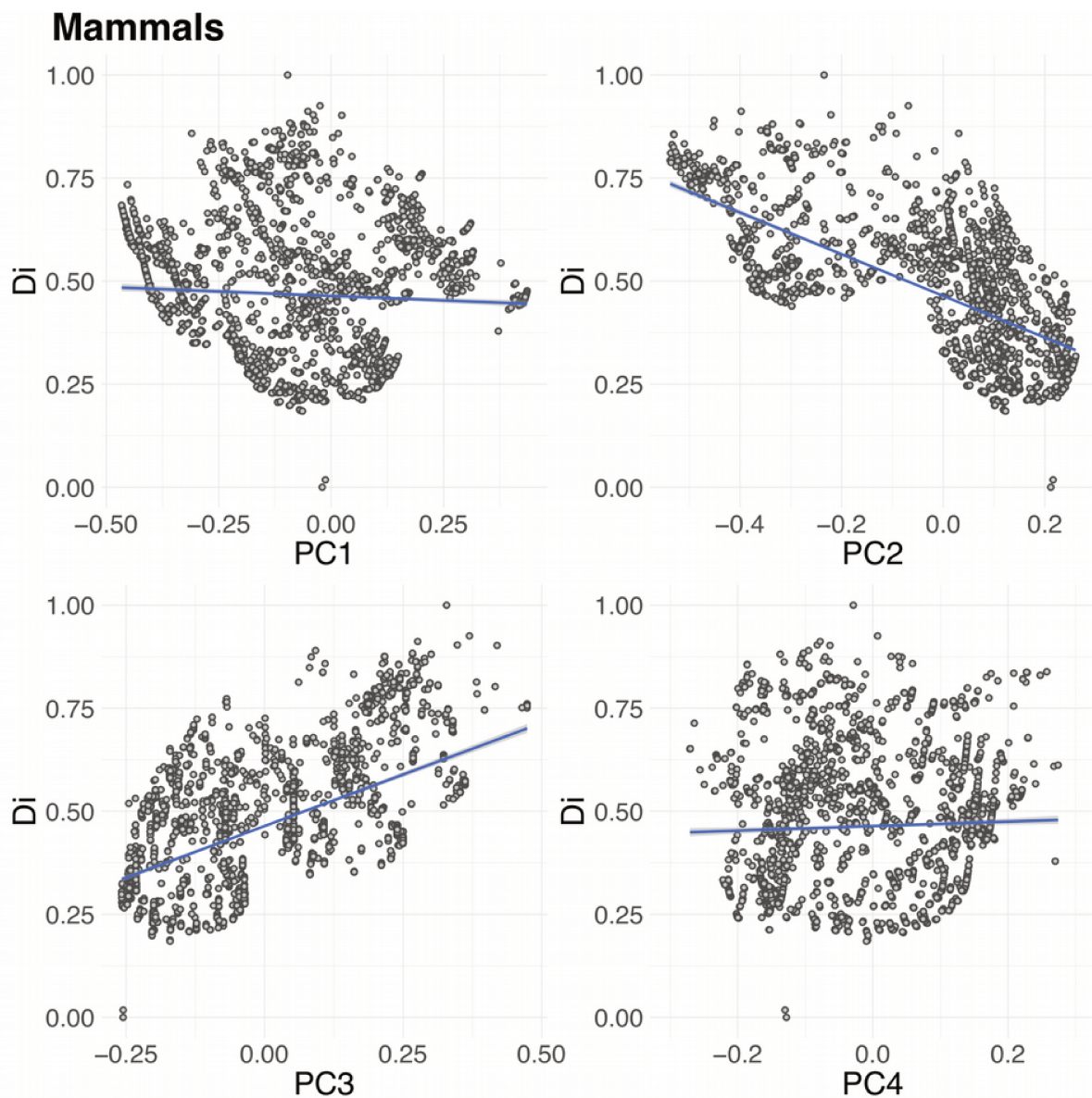
Supplementary Table 3: P-values indicate if there is or not a significant difference between ecologically rare, common, and average species via ANOVA and Tukey's post-hoc tests. From left to right: Human Footprint that measure the cumulative impact of direct human pressures; Human development index (HDI); Change in distribution range based on climate change projections (scenario RCP 8.5, Horizon 2041-2060 and 2061-2080) and Target achievement (extent to which species are represented within PAs regarding their restrictiveness). P-value in bold are significant ($P < 0.05$).

Pairs	Human Footprint	HDI	Conflicts	Climate Change 2041-2060	Climate Change 2061-2080	Target Achievement
<i>Mammals</i>						
Common - Average	6.9e-03	0.25	0.88	0.32	0.37	< 1.e-16
Rare - Average	8.3e-05	4.2e-09	0.09	0.01	0.09	0.27
Rare - Common	9.7e-08	3.2e-03	0.47	2.4e-03	0.02	< 1.e-16
<i>Birds</i>						
Common - Average	3.4e-04	2.7e-06	7.2e-04	0.07	0.07	9.2e-11
Rare - Average	3.5e-09	1.7e-08	3.3e-05	0.01	6.2e-03	0.06
Rare - Common	3.5e-09	0.72	0.8	3.5e-04	1.2e-04	9.2e-11

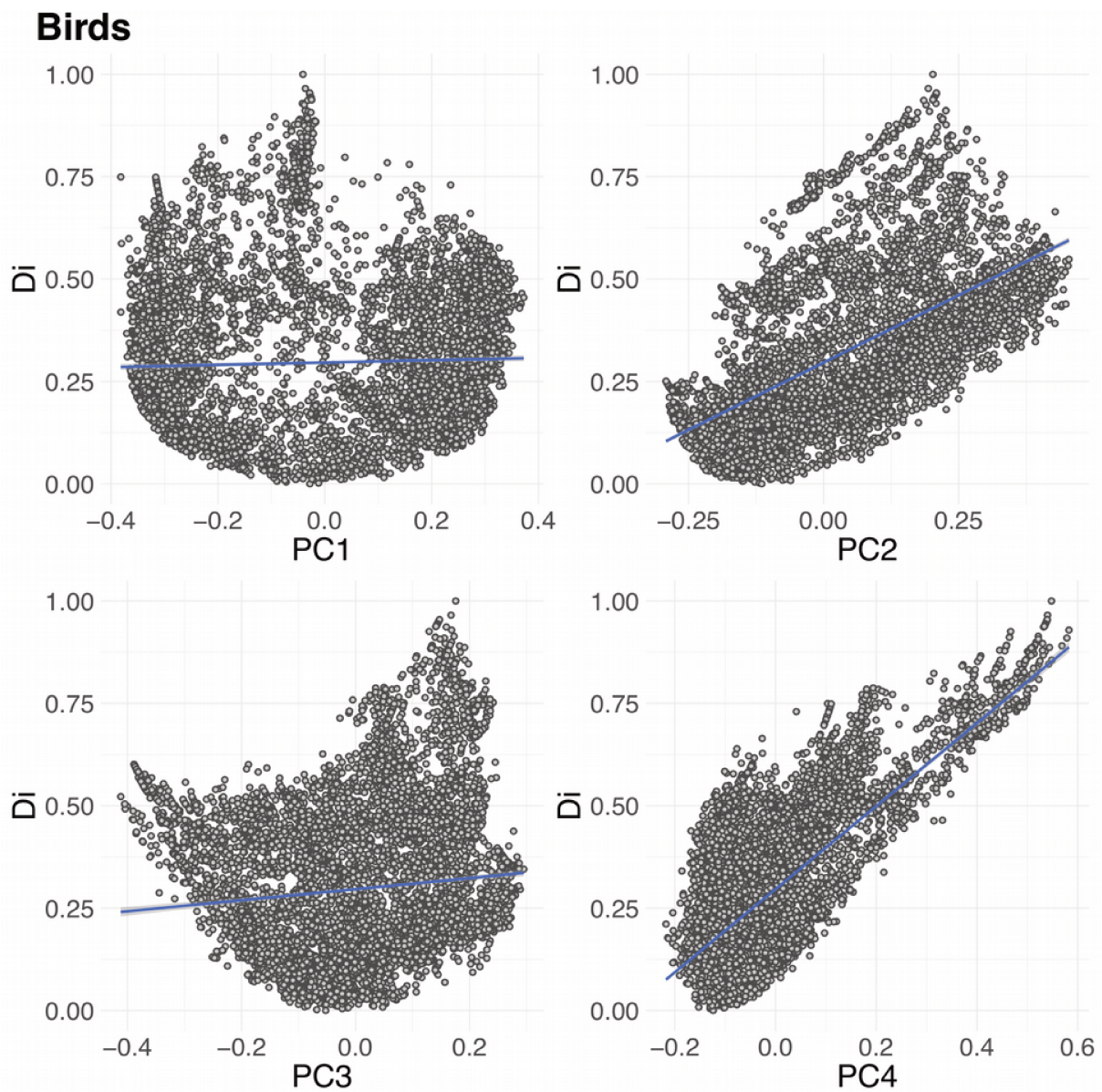
Supplementary Figures



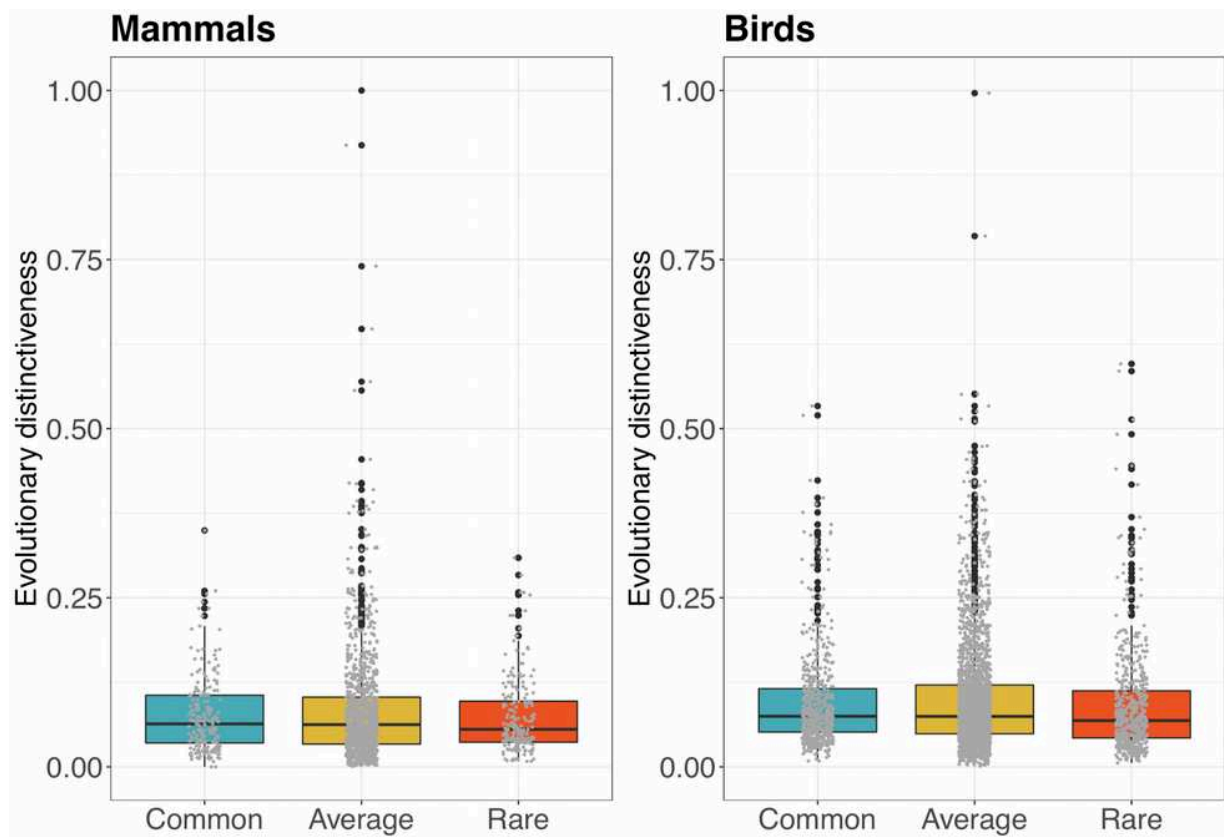
Supplementary Figure 1: Relation between functional distinctiveness (D_i) and geographical restrictiveness (R_i , we use $\log(1-R_i)$ for clarity as most of the R_i values are close to 0 for mammals and birds). Ecologically rare species are in red, average in blue, and common in orange.



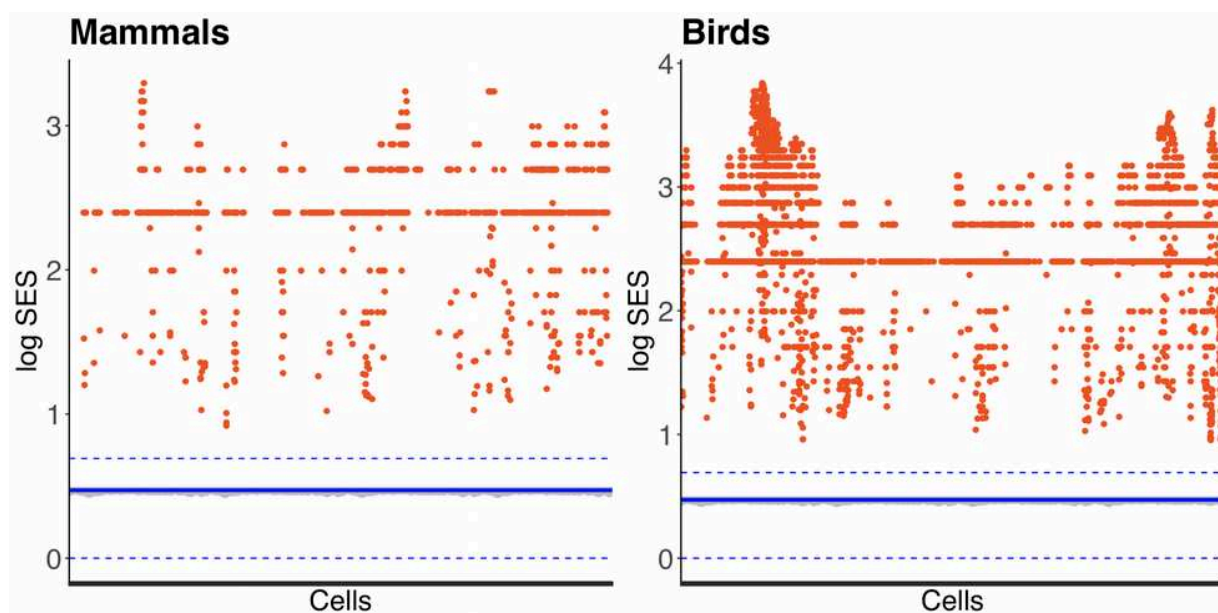
Supplementary Figure 2: Relation between functional distinctiveness (\bar{D}_i) and the four axes of the functional traits PcoA for mammals. Blue line represents the linear model between \bar{D}_i and PcoA axes. Distinctiveness was mainly correlated with axes 2 and 3 of the PcoA ($R^2 = 0.51$, $R^2 = 0.31$, p -values < 0.001).



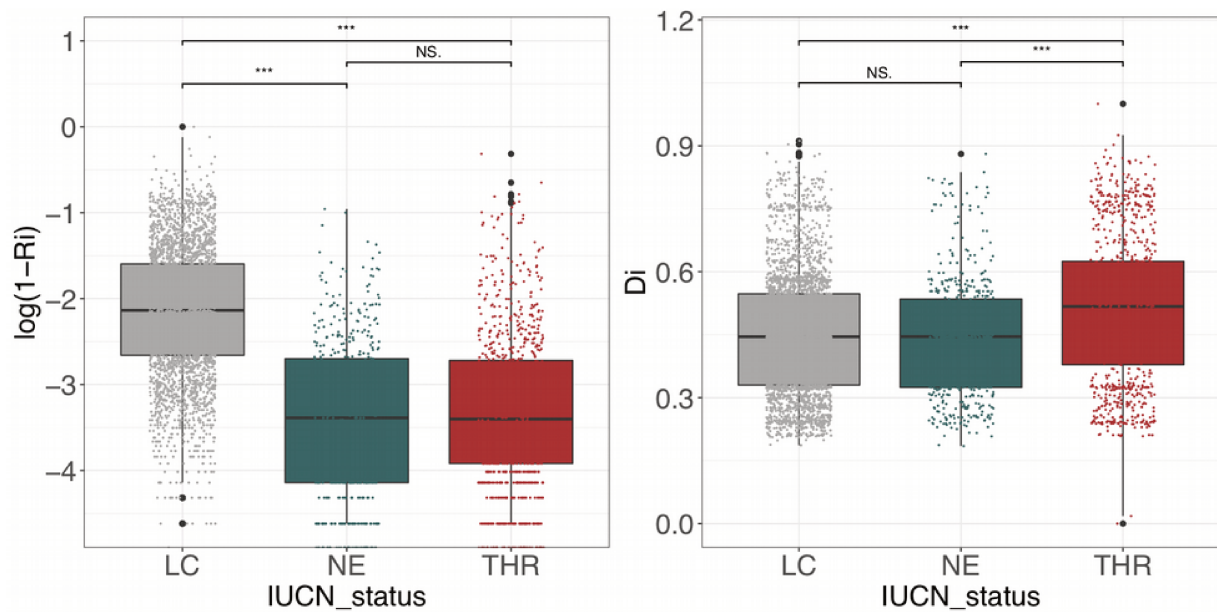
Supplementary Figure 3: Relation between functional distinctiveness (D_i) and the four axes of the functional traits PcoA for birds. Blue line represents the linear model between D_i and PcoA axes. Distinctiveness was mainly correlated with axes 2 and 4 of the PcoA ($R^2 = 0.42$, $R^2 = 0.48$, p -values < 0.001).



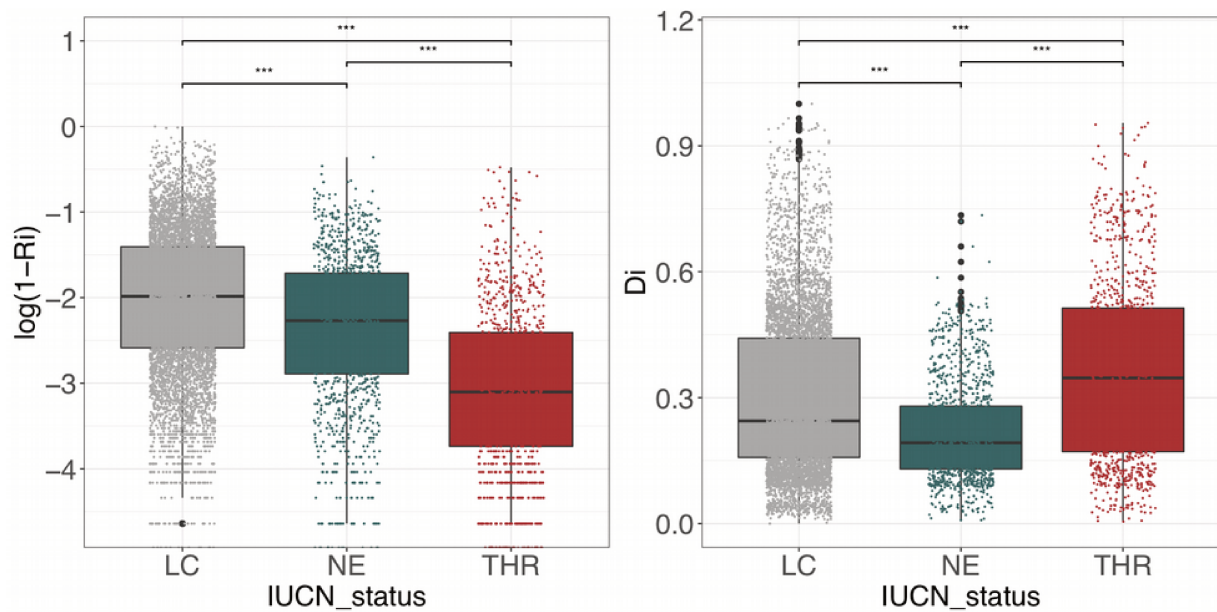
Supplementary Figure 4: Evolutionary distinctiveness of ecologically common (blue), average (orange) and rare (red). The Evolutionary Distinctiveness of species i , is high when the species has a long unshared branch length with all the other species. The more “isolated” a species is in a phylogenetic tree, the higher its evolutionary distinctiveness.



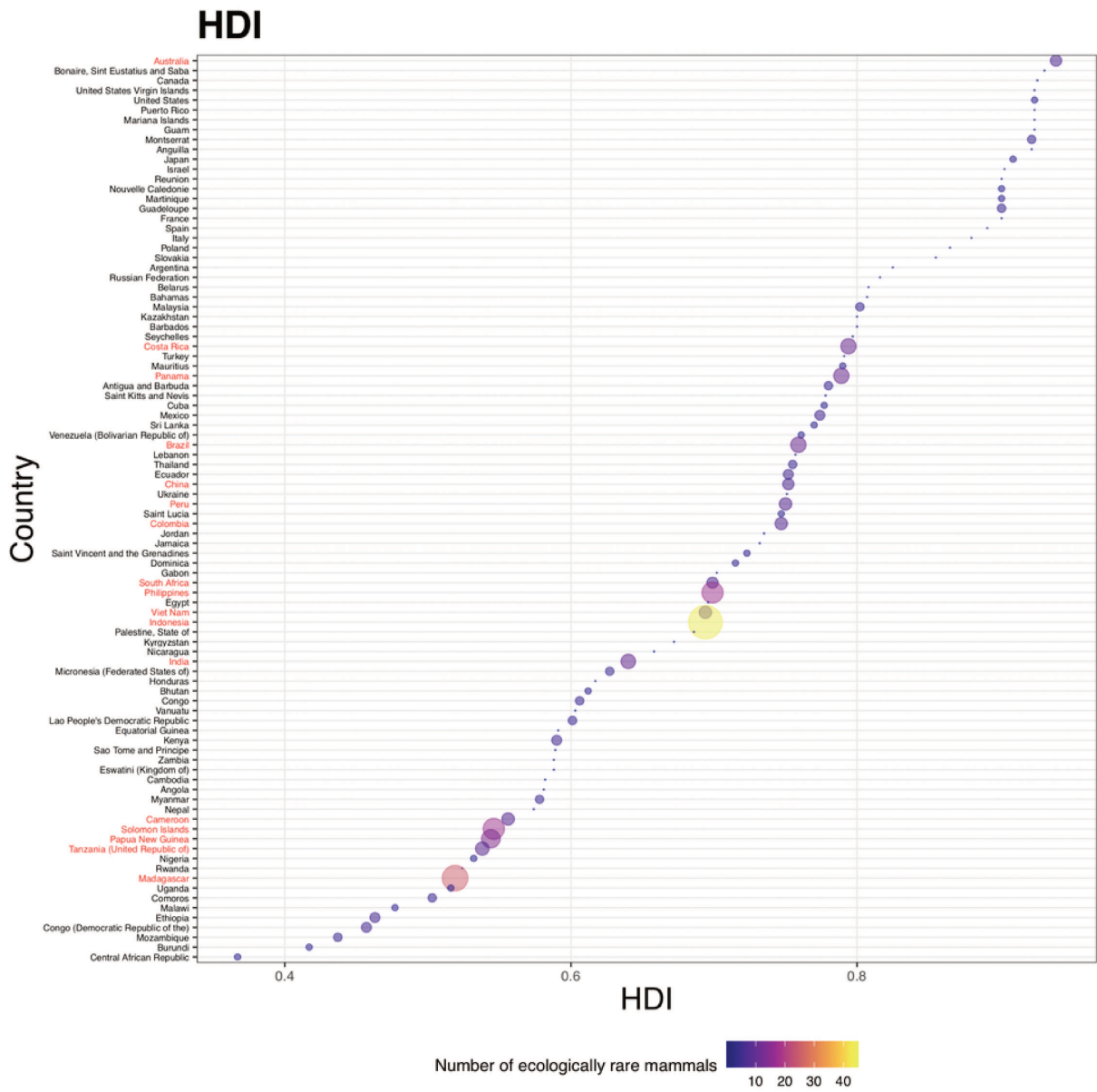
Supplementary Figure 5: Standardized effect size (SES) of the number of species on the number of ecologically rare species per cells. For clarity we plotted SES on a log scale $\log(\text{SES} + \text{abs}(\min(\text{SES}))+1)$. Solide blue line indicate $\text{SES} = 0$, for random values. Bluedotted lines indicate the $\alpha = 0.05$ threshold of $\text{SES} = 1.96$ for significantly non-random values. Each data point represents a cell. Values above the null expectation indicate that the cell host more ecologically rare species than expected given the number of species in the cell. In red, cells hosting at least on ecologically rare species, in grey, cells that do not host any ecologically rare species.

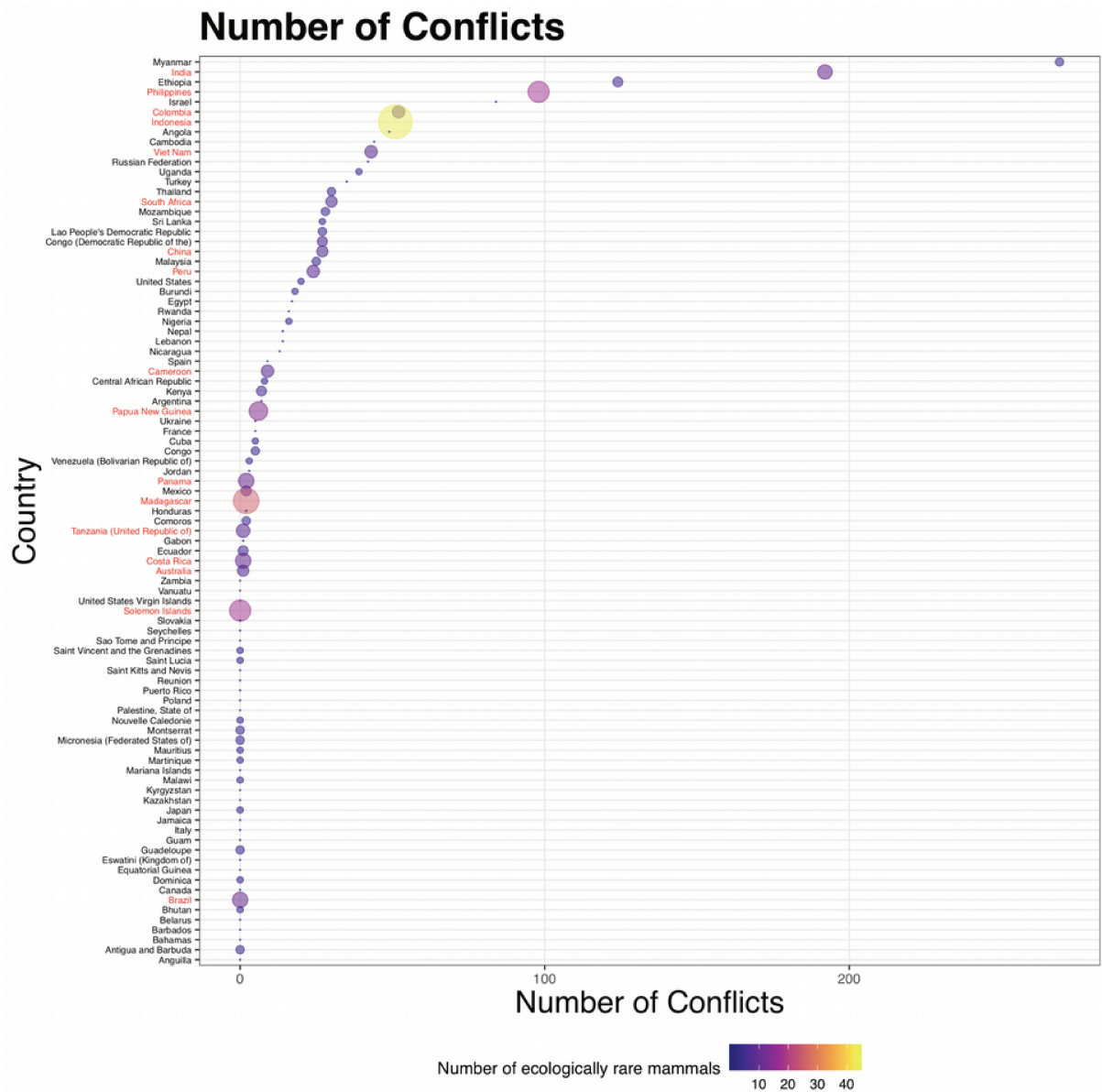


Supplementary Figure 6: Mammals distinctiveness (D_i) and geographical restrictiveness (R_i , we use $\log(1-R_i)$ for clarity as most of the R_i values are close to 0 for mammals and birds) in relation to IUCN status according to criteria by the IUCN Red List, Threatened (TH red: Vulnerable, Endangered, Critically Endangered), Least concern (LC blue: Least Concern and Near Threatened) and None Evaluated (NE grey: Not Evaluated or inadequate information). The difference between IUCN Status was tested with Wilcox test and is shown along with the significance of this difference: NS for $p > 0.05$, * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

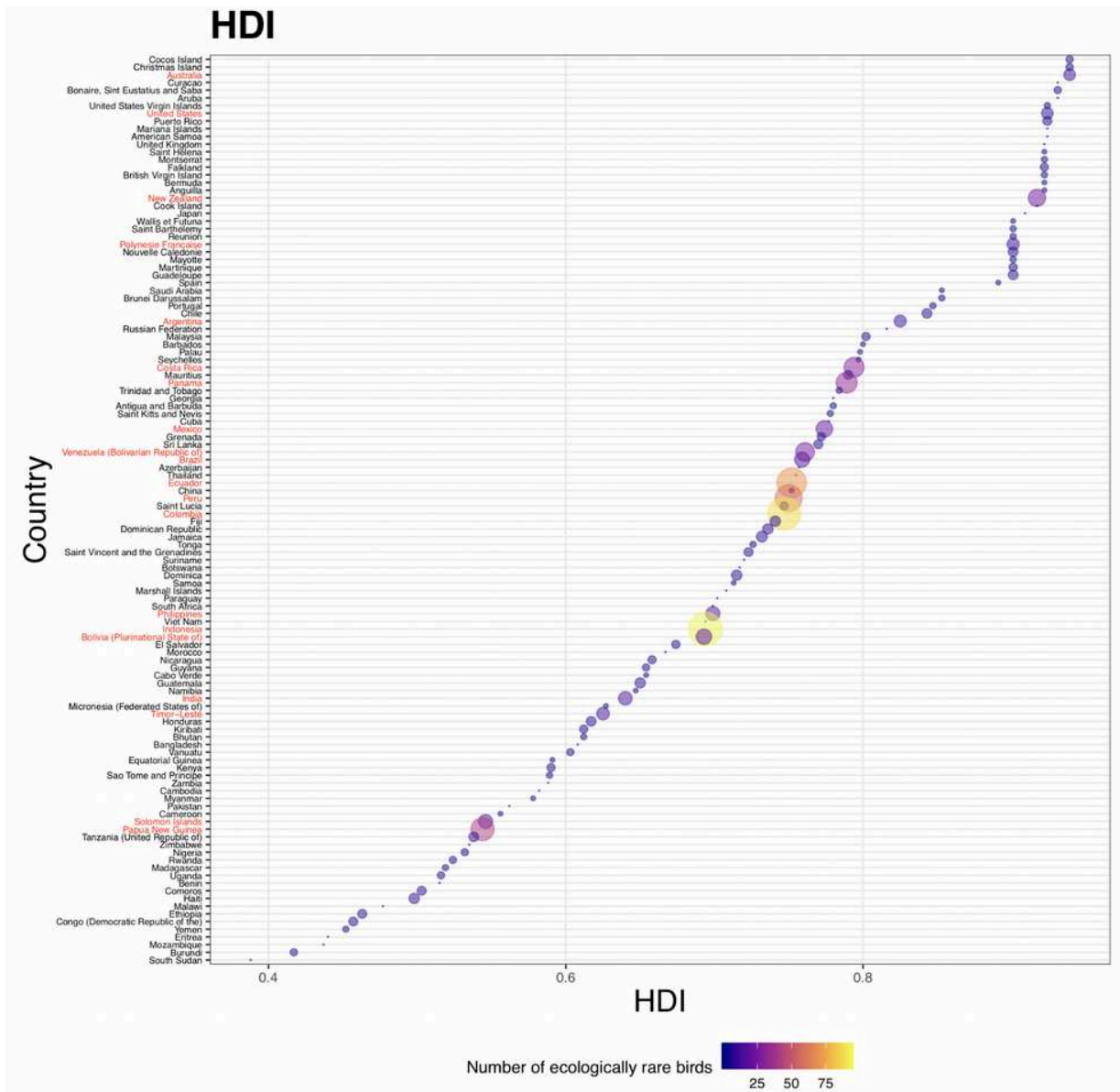


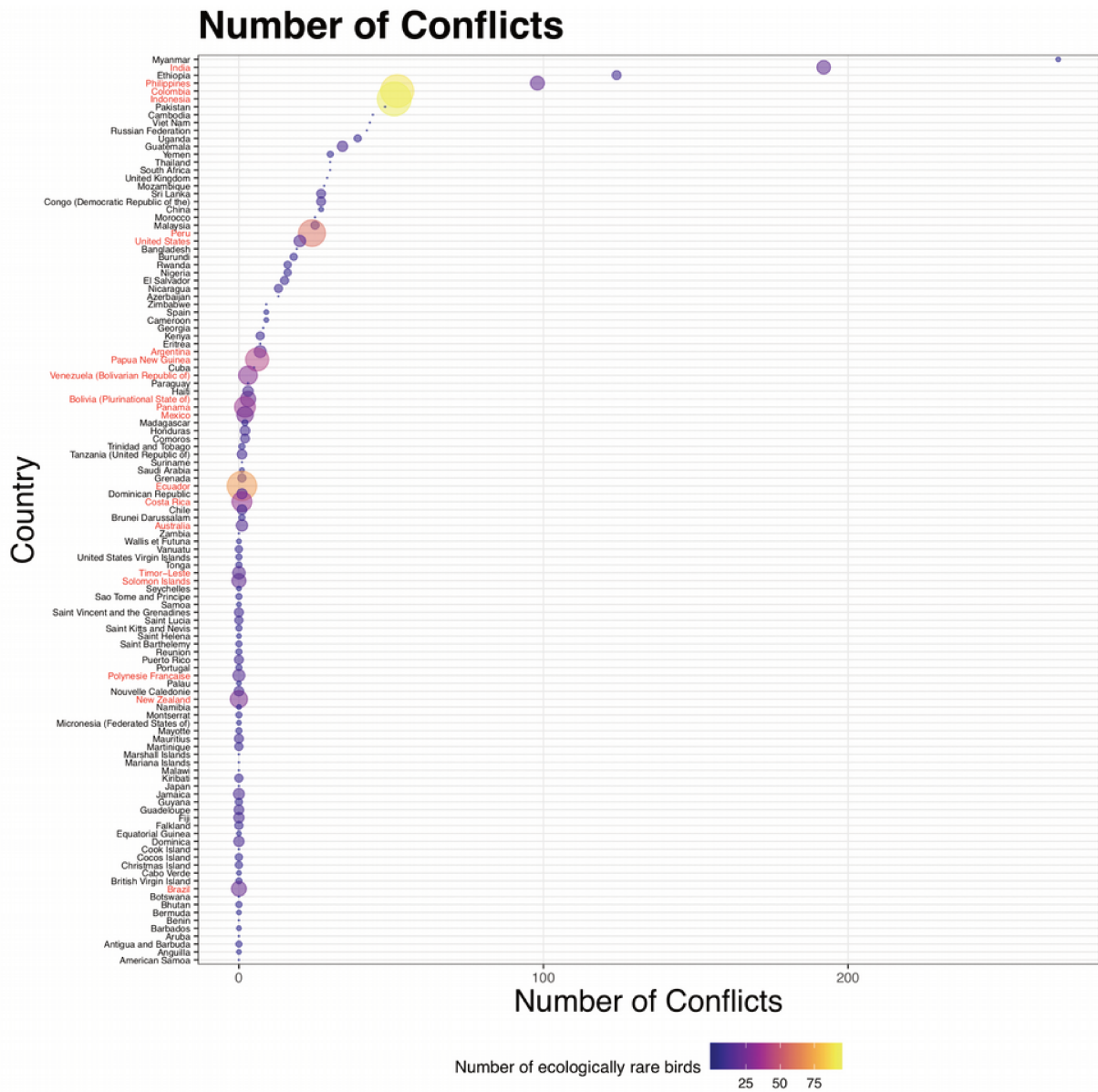
Supplementary Figure 7: Birds distinctiveness (D_i) and geographical restrictiveness (R_i , we use $\log(1-R_i)$ for clarity as most of the R_i values are close to 0 for mammals and birds) in relation to IUCN status according to criteria by the IUCN Red List, Threatened (TH red: Vulnerable, Endangered, Critically Endangered), Least concern (LC blue: Least Concern and Near Threatened) and None Evaluated (NE grey: Not Evaluated or inadequate information). The difference between IUCN Status was tested with Wilcox test and is shown along with the significance of this difference: NS for $p > 0.05$, * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.



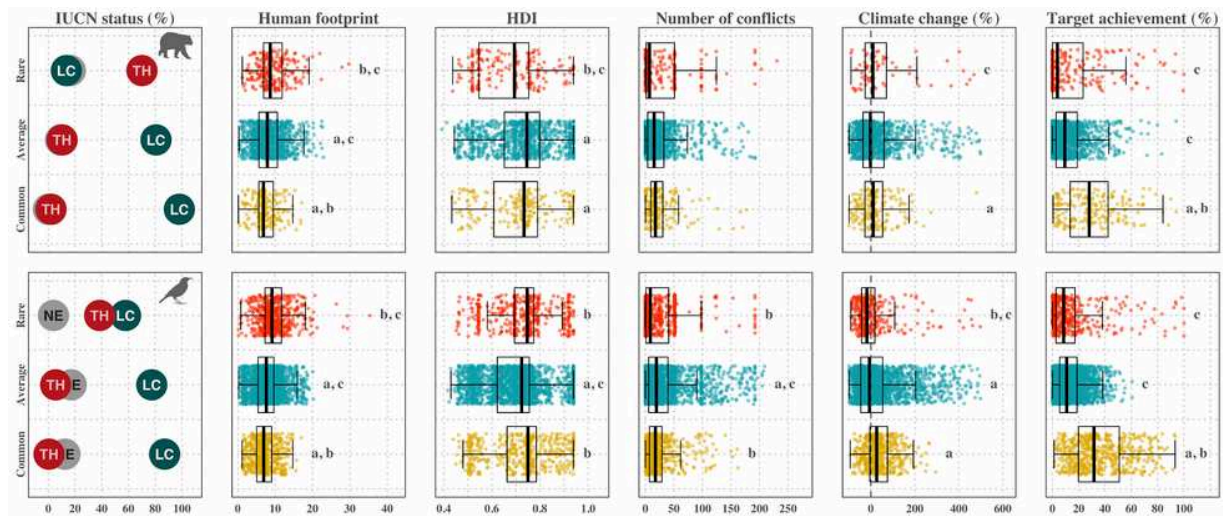


Supplementary Figure 8: HDI (top) and conflicts (bottom) in countries hosting ecologically rare mammals (92 countries). In red we highlight countries with more than 5 ecologically rare species.

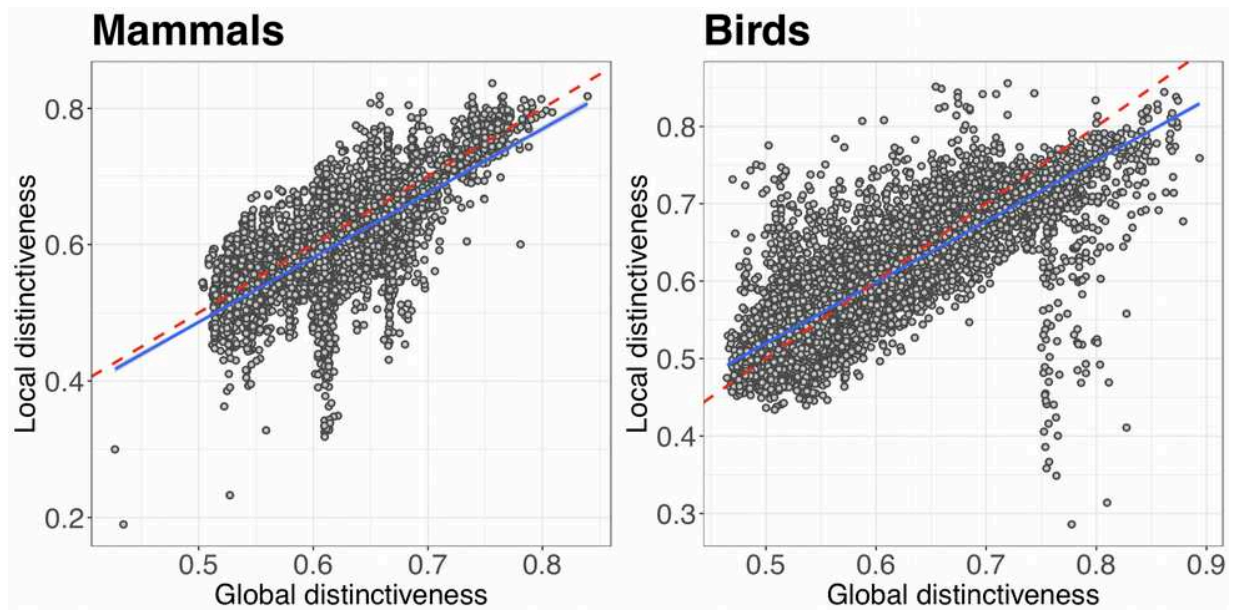




Supplementary Figure 9: HDI (top) and conflicts (bottom) in countries hosting ecologically rare birds (118 countries). In red we highlight countries with more than 10 ecologically rare species.



Supplementary Figure 10. Figure 4 from the main manuscript with scenario RCP 8.5, Horizon 2061 - 2080 for influence of Climate Change. From left to right: percentage of species per IUCN status according to criteria by the IUCN Red List, Threatened (TH red: Vulnerable, Endangered, Critically Endangered), Least concern (LC green: Least Concern and Near Threatened) and Not Evaluated (NE grey: Not Evaluated or inadequate information). Main threats on ecologically rare (red), average (blue) and common (orange) species; Human Footprint that measure the cumulative impact of direct human pressures; Human development index (HDI); Change in distribution range based on climate change projections (scenario RCP 8.5, Horizon 2061 - 2080) and Target achievement (extent to which species are represented within PAs regarding their restrictiveness). Letters indicate significant difference between group ($P < 0.05$, table S3) via ANOVA and Tukey's post-hoc tests.



Supplementary Figure 11 : Relation between functional distinctiveness computed at global scale and computed at local scale (within cells) for mammals and birds. In red: linear model between local and global distinctiveness ($R^2 = 0.50$ for mammals, $R^2 = 0.64$ for birds and p -values < 0.001). Blue lines represent the hypothetical case where global functional distinctiveness is equal to local functional distinctiveness (i.e. first bisector).

D

Barnagaud et. al (2019) *Global Ecology and
Biogeography*

Received: 1 August 2018 | Revised: 27 November 2018 | Accepted: 14 December 2018

DOI: 10.1111/geb.12910

RESEARCH PAPER

WILEY

Global Ecology
and BiogeographyA Journal of
Macroecology

Functional biogeography of dietary strategies in birds

Jean-Yves Barnagaud¹  | Nathan Mazet¹ | François Munoz² | Matthias Grenié¹ | Pierre Denelle¹  | Mar Sobral¹ | W. Daniel Kissling³ | Çağan H. Şekercioğlu^{4,5} | Cyrille Violle¹ 

¹Centre d'Ecologie Fonctionnelle et Evolutive—UMR 5175 CEFE, CNRS, PSL Research University, EPHE, University of Montpellier, SupAgro, IND, INRA, Montpellier, France

²Laboratoire d'Écologie Alpine, University of Grenoble Alpes, Grenoble, France

³Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, The Netherlands

⁴Department of Biology, University of Utah, Salt Lake City, Utah

⁵College of Sciences, Koç University, Istanbul, Turkey

Correspondence

Jean-Yves Barnagaud, Centre d'Ecologie Fonctionnelle et Evolutive—UMR 5175 CEFE, CNRS, PSL Research University, EPHE, University of Montpellier, SupAgro, IND, INRA, 1919 route de Mende, Montpellier, France
Email: jean-yves.barnagaud@cefe.cnrs.fr

Funding information

European Research Council Starting Grant, Grant/Award Number: ERC-StG-2014-639706-CONSTRAINTS; European Research Council

Editor: Adriana Ruggiero

Abstract

Aim: Diet is key to understanding resource use by species, their relationships with their environment and biotic interactions. We aimed to identify the major strategies that shape the diet space of birds and to investigate their spatial distributions in association with biogeographical, bioclimatic and anthropogenic drivers.

Location: Global.

Time period: Current.

Major taxa studied: Birds.

Methods: We analysed score-based assessments of eight diet categories for 8,937 out of 10,964 extant bird species. We constructed a multivariate diet space by ordinating these data in a principal coordinates analysis and assessed its dimensionality as a balance between the representation of original diet scores and parsimony. We averaged the positions of species along each dimension for 12,705 species assemblages and used quantile regressions to infer the relative contributions of species richness, climate, primary productivity, topography and human footprint to the spatial distribution of the diet space at a global scale.

Results: The diet space of birds was structured by four dimensions ordinating species along continua ranging from insectivory to plant-based strategies, granivory to frugivory, common to rare diets, and nectarivory to carnivory and piscivory. Although orthogonal at the species level, these dimensions were correlated among species assemblages, with regional variation consistent with past climatic and tectonic events. Human footprint packed bird assemblages in the diet space, whereas warm climate, high productivity and high topographic variability were associated with high variability in the prevalence of dietary strategies among assemblages.

Main conclusions: The tremendous variability in bird diets can be explained by a few basic ecological continua sustained by morphological and ecophysiological differences among species. Strong biogeographical legacies on top of bioclimatic drivers distribute this diet space in species assemblages through environmental filtering and niche packing. However, these patterns are altered at macroecological scales by human-mediated functional homogenization, which might, in turn, affect the global distribution of bird functions and services.

KEYWORDS

biotic homogenization, birds, dietary strategies, ecological traits, functional biogeography, human footprint, multivariate analysis

1 | INTRODUCTION

How the variety of life-forms is organized in the geographical space and along environmental gradients is a long-standing question in biogeography, at the crossroads of evolutionary and functional ecology (Grinnell, 1917; MacArthur, 1972; Soberon, 2007; Violle, Reich, Pacala, Enquist, & Kattge, 2014). Trait-based strategies for growth, survival and reproduction represent a key connection between this diversity and the environment (Garnier, Navas, & Grigulis, 2016; McGill, Enquist, Weiher, & Westoby, 2006), by providing a functional explanation of species' geographical distributions and assemblages (McGill et al., 2006; Violle et al., 2014). Understanding how and why trait strategies vary in space and time should, in particular, illuminate the causes of species distributional shifts related to global changes and their vulnerability to anthropogenization (Jiguet, Gadot, Julliard, Newson, & Couvet, 2007; Purvis et al., 2000). However, inferences on trait–environment relationships have proved challenging because of the multidimensionality of the underlying niche space (Hutchinson, 1957; Laughlin, 2014), unequal contributions of traits to niche dimensions (Tucker, Davies, Cadotte, & Pearse, 2018) and data-deficient trait descriptions (Hortal et al., 2015). As a result, the macroecological distribution of trait strategies remains unknown for the vast majority of taxonomic groups (Garnier et al., 2016).

The > 10,000 avian species in the world exhibit a tremendous variability in ecological traits, which has allowed them to colonize virtually all emerged habitats on Earth (Lovette & Fitzpatrick, 2016). Owing to this diversity, birds sustain numerous ecosystem functions, including pollination, seed dispersal, pest regulation and creation of microhabitats, making them a key element of ecosystem-based conservation strategies (Şekercioğlu, Wenny, & Whelan, 2016). Diet is central to these functions and to various aspects of bird life cycle, ecosystem position and evolution (Burin, Kissling, Guimarães, Şekercioğlu, & Quental, 2016; Duffy, 2002; Lovette & Fitzpatrick, 2016). It thus forms one of the most crucial dimensions of bird life history (Hutchinson, 1959), because it determines the main lines of a species' energetic investment, survival, reproduction and, in turn, fitness (Costa, Vitt, Pianka, Mesquita, & Colli, 2008; MacArthur & Pianka, 1966; Sibly et al., 2012). Illustrating the pivotal role of diet in animal ecophysiology, herbivorous bird species exhibit lower basal metabolic rates, higher field metabolic rates and higher body temperatures than carnivores (Anderson & Jetz, 2005; Speakman, 1999), whereas energetic and water turnovers reach extreme values in nectarivorous birds (Nicolson & Fleming, 2003). As a corollary of these physiological constraints, the size and energetic value of prey are related to essential functional traits, such as body size or reproductive output, and to other ecological characteristics, such as habitat use (Janes, 1994; Sam, Koane, Jeppy, Sykorova, & Novotny, 2017; Sibly et al., 2012). Diet is under strong selection pressure because

of its tight link with individual fitness, notably through competitive interactions and co-evolutionary dynamics associated with extreme morphological or ecophysiological adaptations (Grant & Grant, 2006; Nicolson & Fleming, 2003). Hence, the multiple convergences and trade-offs that have structured the evolution of traits in birds are likely to be reflected in dietary strategies that are related to biogeographical patterns of bird–environment relationships.

The highly uneven geographical distribution of bird diets could be explained by a combination of climate, primary productivity, distribution of trophic resources, and phylogenetic niche conservatism (Barnagaud et al., 2014; Kissling, Şekercioğlu, & Jetz, 2012; Lein, 1972). Dietary guild richness is strongly correlated with species richness on top of these environmental gradients, consistent with the hypothesis that long-term environmental variations have shaped dietary diversification within avian assemblages (Belmaker, Şekercioğlu, & Jetz, 2012; Burin et al., 2016; Kissling et al., 2012). High flowering and fruit plant diversities are typically sustained by climate stability and high primary productivity in the tropics, favouring the diversification of frugivores and nectarivores (Dalsgaard et al., 2011, 2017; Jetz et al., 2009; Kissling, Rahbek, & Böhning-Gaese, 2007). Inversely, at northern latitudes, carnivore and scavenger bird assemblages are composed of phylogenetically clustered species, as a consequence of past diversification events related to the physiological tolerance of the species and climatic constraints that affect prey distributions (Kissling et al., 2012; Vejrikova et al., 2016).

Previous studies on spatial patterns in bird diets have examined guild-level species richness (Burin et al., 2016; Kissling et al., 2012) or were geographically restricted (Pigot et al., 2016). One strong limitation of guild-based approaches is that they do not account for covariations among diets resulting from the evolutionary constraints of species. For instance, it is unclear whether the global congruence between high insectivore and frugivore species richness is attributable to species preying simultaneously on insects and fruits or to convergences in the distributions of species specialized on each diet (Kissling et al., 2012). Furthermore, guild approaches rely on a priori categorizations that hinder the probabilistic nature of dietary choices, more adequately quantified by continua reflecting the observation that most species rely on multiple diets in varying proportions. Covariation in the diets of birds is thus better represented within a multivariate space structured by a few continuous, orthogonal dimensions (MacArthur, 1972; Pianka, Vitt, Pelegrin, Fitzgerald, & Winemiller, 2017; Winemiller, Fitzgerald, Bower, Pianka, & Arita, 2015). The shape of this diet space is assumed to have emerged as an outcome of niche convergence and niche exclusion processes (Morowitz, 1980; Winemiller et al., 2015), providing an interface between species' diversification processes and their functions (Grime & Pierce, 2012; Wüest, Münkemüller, Lavergne, Pollock, & Thuiller, 2018), which remains,

however, poorly quantified. In particular, investigating how the diet space is spread among bird assemblages at a macroecological scale might help to uncover the relative contributions of historical legacies, environmental filtering and niche convergence to the distribution of bird diets and associated functions.

Geographical variation in the diet space of a taxonomic group can be mapped by averaging the position of co-occurring species along each of its dimensions to provide a community-level attribute comparable to the widely used community-weighted mean traits approach (Ricotta & Moretti, 2011; Wüest et al., 2018). This approach resolves the double limitation of guild-based studies: first, it incorporates species-level covariations among diets in a quantitative assessment of community composition; and second, scores computed in an ordination space permit a continuous representation of gradients from assemblages dominated by opposed dietary strategies. Two patterns are likely to be identified when correlating the average positions of bird assemblages in the diet space. A systematic variation in the average diet positions of species assemblages along a well-identified environmental gradient would indicate diet-mediated environmental filtering, which is typically expected with gradual changes in climate or dominant resources along latitudinal or elevational gradients (Cadotte & Tucker, 2017). Alternatively, an increase or decrease in the range of average diet positions would reveal a differentiation or homogenization of assemblages, respectively, as a consequence of changes in resources, typically in association with productivity or high environmental heterogeneity in transition areas or mountains (Dehling et al., 2014; Pellissier, Barnagaud, Kissling, Şekercioğlu, & Svenning, 2018; Pigot et al., 2016). A packing of species assemblages in the diet space is also expected under the effect of man-mediated functional homogenization, as an outcome of the reduction in environmental turnover and simplification of the ecosystem network in highly anthropogenized areas (Baiser & Lockwood, 2011).

The aim of this study was to uncover the main dimensions of the diet space of birds and their geographical and environmental spread at a global scale. We constructed a diet space for 83% of extant bird species, mapped it and studied its distribution along climatic, topographic, productivity and anthropogenization gradients, accounting for variations in species richness. Following the prediction by MacArthur (1972), we hypothesized that no more than two to four dimensions would summarize parsimoniously the continua that structure the diets of the birds of the world. We also expected that the average positions of species assemblages along these continua would be patterned strongly in the geographical space under functional homogenization and differentiation mediated by environmental gradients and human impact.

2 | MATERIALS AND METHODS

2.1 | Data

2.1.1 | Bird diets

We compiled diet information for 9,098 out of the 10,694 extant bird species currently recognized by the International Ornithological

Union, covering all 40 orders (Gill & Donsker, 2017). Among the 1,596 missing species, 1,583 were data deficient and 13 had unclassifiable diets. We also removed 161 species for which we did not have distribution or phylogenetic data, resulting in a final dataset of 8,937 species. We assessed scores of use from zero (non-use) to 10 (exclusive use) for each species and eight diet categories (fishes, vertebrates, carrion, insects, nectar, fruits, seeds and other plant material) based on the most comprehensive available information to date (Del Hoyo, Elliott, & Sargatal, 2013; Kissling et al., 2012; Şekercioğlu, Daily, & Ehrlich, 2004). Scores summed to 10 per species over all categories, thus providing a standardized, probabilistic assessment of species' diets.

2.1.2 | Species assemblages

We defined species assemblages by overlaying species' breeding extent-of-occurrence maps provided by BirdLife International (BirdLife International & Nature Serve, 2012) on a 110 km × 110 km resolution grid in cylindrical equal area projection. Extent-of-occurrence maps are not exempt from criticism, because they were built from expert opinion and incomplete observational records that are subject to error, especially at species range limits (Herkt, Skidmore, & Fahr, 2017). In spite of these limitations, these maps remain the best available description of bird distributions at the world scale. We only exploited breeding-period distributions, a limitation imposed by data deficiency on the wintering ranges of most migratory species. Diet assessments of migratory species are, however, usually conducted during the nestling period or as birds arrive on/leave breeding grounds (Del Hoyo et al., 2013), ensuring consistency between the diet and distribution datasets.

The spatial dataset encompassed 12,802 bird assemblages (grid cells), after exclusion of aquatic (> 10% of open waters in the cell) and Antarctic cells. We also removed cells with species richness < 10 species ($n = 97$ cells) because they would inflate statistical uncertainty. Furthermore, exploring species lists in these cells revealed that their extremely low species richness resulted, at least in part, from deficient assessments of species distributions in poorly surveyed areas, such as parts of North Africa or boreal Siberia. Hence, all our analyses relied on a sample size of 12,705 species assemblages (species richness from 10 to 849 species, mean \pm SD = 179.6 ± 121.3 species).

2.1.3 | Environmental gradients

We retrieved rasters of mean annual temperatures (in degrees Celsius) from 1970 to 2000 at a 30" native resolution (www.worldclim.org; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), elevational range (in metres) at a 1 km native resolution (global land cover characterization data base, <https://lta.cr.usgs.gov/GLCC>) and net primary productivity (NPP; in kilograms of carbon per square metre) at a 1 km native resolution (<https://modis.gsfc.nasa.gov/data/dataproduct/mod17.php>; Zhao, Heinsch, Nemani, & Running, 2005). We also retrieved the human influence index (HII),

a score-based synthetic measure of human impacts on ecosystems at a 1 km native resolution (Sanderson et al., 2002; updated for the 1995–2004 period, <http://sedac.ciesin.columbia.edu/data/set/wildareas-v2-human-influence-index-geographic>). We discarded precipitation, another influential predictor of bird distributions, owing to its high correlation with NPP ($R^2 = 0.6$); the other four variables were weakly correlated (maximum $R^2 = 0.21$, between NPP and temperature). We averaged all variables in each cell of the 110 km \times 110 km grid for match with bird assemblage data. Correlations between bird species richness and the four bioclimatic variables peaked at $R^2 = 0.20$ (with temperature) and $R^2 = 0.24$ (with NPP).

2.2 | Building the diet space of birds

We characterized the main dimensions of the diet space of birds by the most parsimonious set of principal components of a principal coordinates analysis (PCoA; Legendre & Legendre, 2012). Principal coordinates analysis summarizes a species \times species dissimilarity matrix into orthogonal components (or dimensions), and thus emphasizes relative differences among species rather than correlations among traits as a principal components analysis would do. In our analysis, these dimensions represented synthetic indices of the variation of dietary strategies among birds, clustering species with an average diet at the centre of the space and relegating the most distinct diets at its margins. Unlike principal components analysis, PCoA is robust to extreme values and double-zeros, which occurred in our data sets as a consequence of rare diets. Furthermore, PCoA is adequate to summarize non-continuous variables, such as pseudo-quantitative fuzzy scores, as in our data set.

To build the PCoA, we first calculated a species \times species dissimilarity matrix from the original species \times diet matrix using a modified version of Gower's distance that accommodates non-independent fuzzy variables with appropriate weighting (Pavoine, Vallet, Dufour, Gachet, & Daniel, 2009). We optimized the dimensions of the diet space by selecting a parsimonious set of PCoA components, which reflected well the pairwise distances defined in the Gower's dissimilarity matrix. For this, we calculated the mean squared deviation (mSD) between the Gower's matrix and an Euclidean matrix based on the scores of species along the PCoA dimensions, and selected the minimum dimensionality reaching an mSD of 0.01 (Maire, Grenouillet, Brosse, & Villéger, 2015). We validated the biological interpretability of the dimensions retained by projecting the original diet categories as supplementary variables on the Euclidean diet space formed by the PCoA components (Legendre & Legendre, 2012).

We computed Spearman rank correlation coefficients as a quantitative measure of the individual contribution of each diet category to each dimension of the diet space (Appendix S1). We also tested for a phylogenetic signal to measure the imprint of niche conservatism on each dimension of the diet space (Additional methods and results in Appendix S2).

2.3 | Spatial variations in the diet space of birds

We mapped the geographical spread of the diet space of birds by averaging species scores on each of its dimensions across all species occurring within each 110 km \times 110 km grid cells. Similar to community-weighted mean trait values in functional ecology, this average score reflects the centroid of a species assemblage in the diet space (Newbold, Butchart, Şekercioglu, Purves, & Scharlemann, 2012). We then regressed these assemblage-level scores on each dimension of the diet space against species richness, temperature, NPP, topographic heterogeneity and HII with quantile regressions, testing the 0.1, 0.25, 0.5, 0.75 and 0.9 quantiles (Davino, Furno, & Vistocco, 2013; Koenker, 2016). In these models, parallel regression slopes across all quantiles would indicate environmental filtering (a systematic shift from a position to another in the diet space; Cadotte & Tucker, 2017), whereas converging or diverging slopes would indicate diet homogenization or differentiation among assemblages along a bioclimatic gradient. Note that we did not account for residual spatial autocorrelation because quantile regressions do not permit the incorporation of spatial errors easily, while alternative methods do not accommodate highly heteroscedastic data.

We assessed goodness-of-fit of quantile regressions through an approximate correlation coefficient (ρ) based on a comparison of the sum of weighted deviations of each model (ρ_1) with the sum of weighted deviations of a regression without any covariate (ρ_0): $\rho = 1 - (\rho_1/\rho_0)$ (Koenker & Machado, 1999). We performed all computations and analyses in the R environment (R Core Team, 2016) with packages *ade4* (Dray & Dufour, 2007) and *quantreg* (Koenker, 2016).

3 | RESULTS

3.1 | The structure of the diet space of birds

Four PCoA components summarized well the dissimilarity of the 8,937 bird species diets (mSD = 0.01; mSD = 0.06, 0.025 and 0.005 for two-, three- and five-dimensional spaces, respectively) and explained 42.5, 20.4, 13.4 and 9.6%, respectively, of the total variance of the Gower's dissimilarity matrix, giving rise to a strongly supported four-dimensional diet space. The ordination of species in this multidimensional space reflects dietary dissimilarities between species without relying on a priori categorical classifications: the dietary strategies of birds spread continuously from the most undifferentiated at the centroid to the most distinct at the margins of the diet space.

The highest proportion of total variance in the diet space was explained by a first dimension separating insectivory from frugivory and granivory [Figure 1; principal component 1 (PC1)]. This structure implies that species feeding on fruits and/or seeds shared similar dietary regimes, whereas species feeding on insects had distinct diet compositions. Accordingly, the negative side of PC1 was dominated by insectivore specialists (Galbuliformes, Caprimulgiformes and Cuculiformes), whereas highly positive values included a mixture of typical frugivores (Psittaciformes) and typical granivores (Pteroclitiformes; Appendix S2).

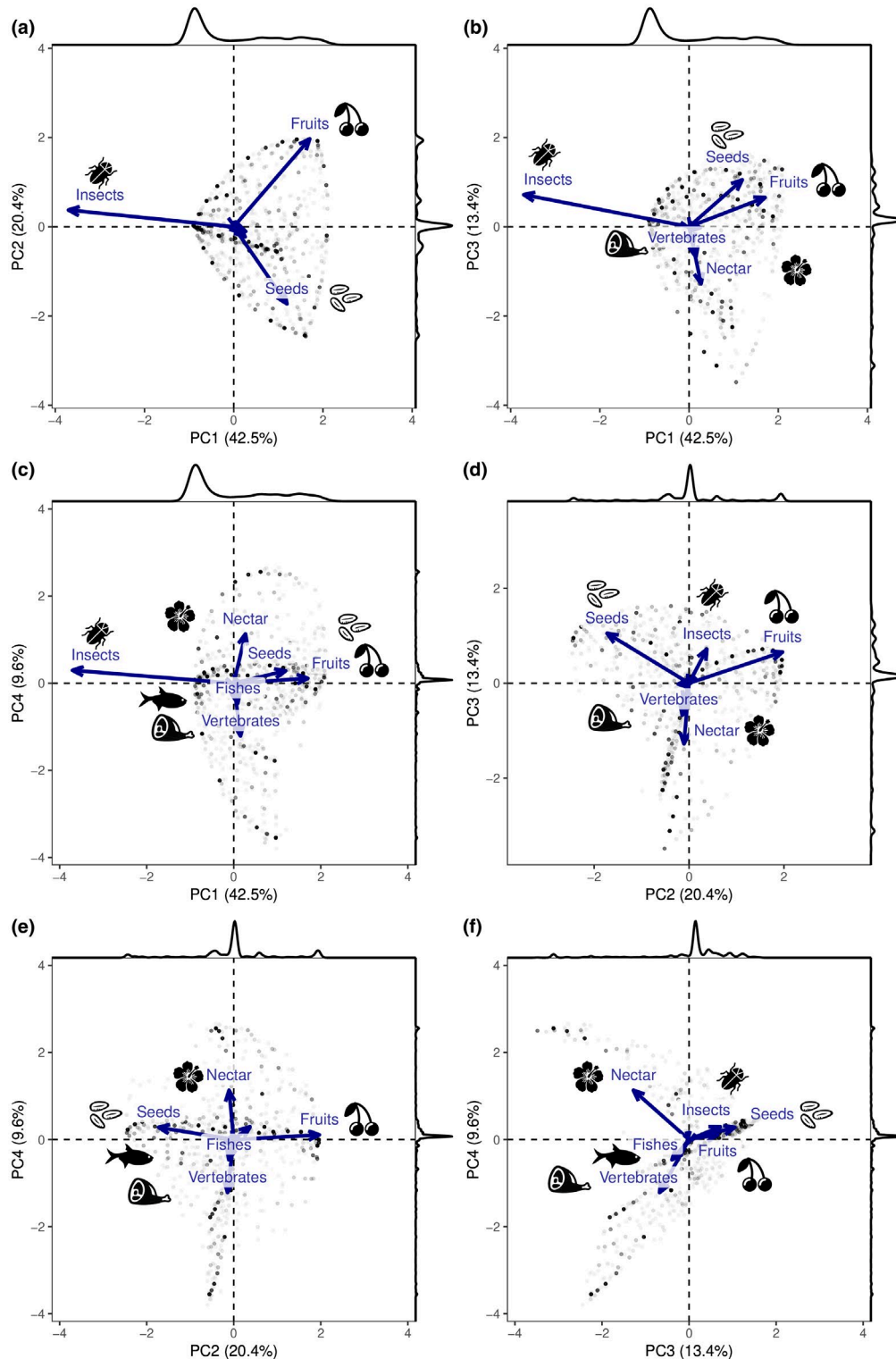


FIGURE 1 Bivariate planes representing the four dimensions of the diet space of birds and projections of the original diet categories (blue arrows). The dot cloud is shaded according to point density (one point per species, $n = 8,937$ species; darker shades imply that several species are superimposed). Marginal densities of species distributions along each dimension are shown on the top and the right sides of each graph [Colour figure can be viewed at wileyonlinelibrary.com]

The second dimension refined this first segregation, by opposing granivory and frugivory [Figure 1: principal component 2 (PC2)], typically separating Pterocliiformes (sandgrouse) from Musophagiformes (turacos) and Bucerotiformes (hornbills; Appendix

S2). Diets based on other plant materials had a negligible contribution to the diet space, being exploited by only 768 species, among which they formed a primary diet for only 181 species. Plotting these two first dimensions on a plane (Figure 1a) revealed that species spread

evenly along continua from primary consumers to insectivores (PC1) and from frugivory to granivory (PC2), with representatives of all intermediate strategies along these two dimensions. The score-based structure of the original diet data resulted in species projections that form a sharp conical cloud in the first plane of the diet space (Figure 1a).

This coarse overview of dissimilarities among the diets of world bird species hinders variability associated with finer specializations that are not captured in our data. Nevertheless, it reveals that whereas insectivorous species may use seeds and fruits simultaneously as alternative prey (negative PC1 scores associated with null scores on PC2), non-insectivorous species tend to feed either on fruits or seeds, but not simultaneously on these two diets (few species associate positive PC1 scores and null scores on PC2).

The third dimension isolated nectarivory, carnivory and piscivory from all other diets [Figure 1; principal component 3 (PC3)]. Orders as different as Apodiformes, Accipitriformes and Pelecaniformes were thus grouped on its negative side (Appendix S2), suggesting high specialization levels (i.e., high scores in one or a few dietary categories). This pattern is in line with the prediction that nectarivory, carnivory and piscivory, three comparatively rare diets, involve highly energetic foods that require metabolic and morphological adaptations and restrict the use of alternative prey. Although they were grouped in negative values of PC3, no species, expectedly, combined nectarivory with carnivory or piscivory. As a result, the fourth dimension [Figure 1; principal component 4 (PC4)] ordinated species along a continuum from piscivory and carnivory (Accipitriformes and Pelecaniformes; Appendix S2) to nectarivory (Apodiformes; Appendix S2). Plotting PC4 against PC1 and PC2

(Figure 1e,f) revealed that many species used combinations of nectar and fruits or insects (e.g., Meliphagid honeyeaters), but none associated nectar and seeds, probably owing to morphological constraints. Conversely, species using carrion or fish-based diets (negative values on PC4) spread on both sides of PC1 (Figure 1c), but few expanded on positive values of PC2 (Figure 1e); these patterns reveal that strategies mixing varying levels of carnivory with reliance on insects or seeds are more common than strategies associating carnivory and frugivory.

3.2 | Spatial variations of dietary strategies of birds

Once averaged over species assemblages, all dimensions of the diet space exhibited clear and relatively congruent geographical structures matching biome boundaries or physical barriers, such as the Himalayas and Andes (Figure 2). Bird assemblages of the Western Palearctic, Eastern Asia down to the Indo-Malayan archipelago, the Guineo-Congolian region, Madagascar and South Australia exhibited strongly negative scores along PC1, indicative of assemblages dominated by insectivory, whereas granivory and frugivory (positive scores on PC1) were more widespread, with less delineated regional patterns. Highly positive scores along PC1 in boreal regions and deserts are likely to reflect assemblages that are largely dominated by seed eaters. Supporting this interpretation, negative scores on PC1 were correlated with positive scores on PC2 in Old World bird assemblages (Figure 3a; Pearson's $R^2 = 0.42$ between the two axes, 7,880 assemblages). This pattern indicates that frugivory and insectivory were associated in species assemblages in the Old World, whereas granivory tended to be relegated to desertic and high-elevation

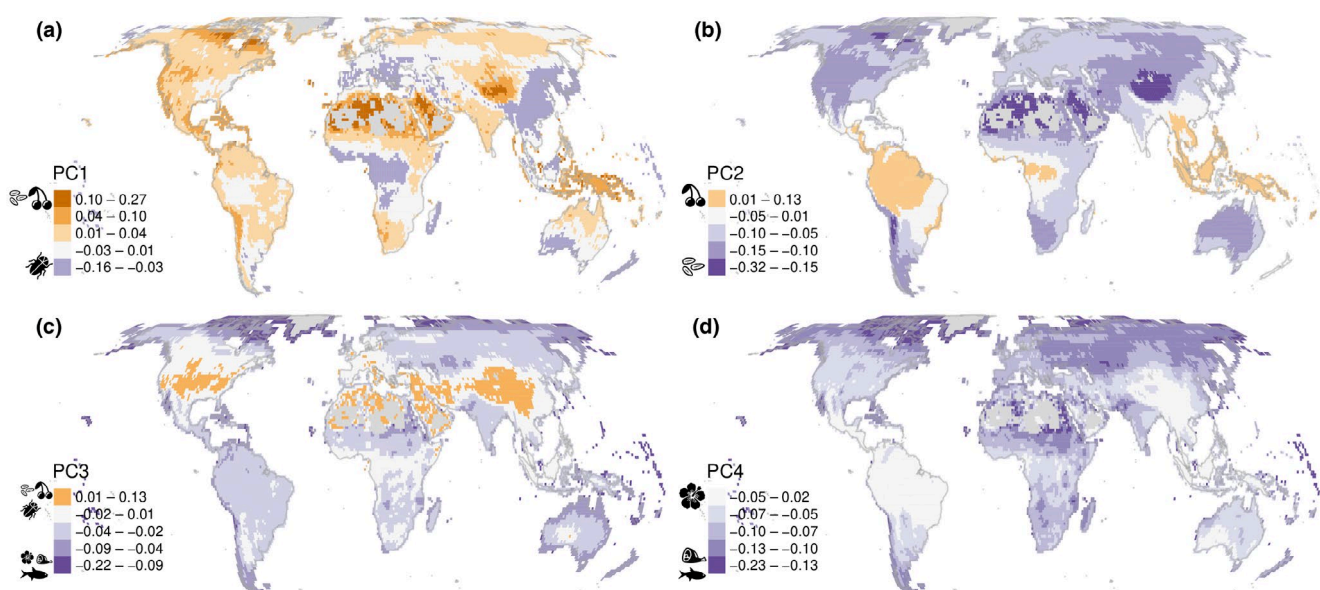


FIGURE 2 The geographical distribution of the four main dimensions (PC1–PC4) of the diet space of birds. These dimensions correspond to four continua between opposing dietary strategies, as assessed from trait covariations in 8,937 bird species. The colour gradient represents variations in the dominant dietary strategy as assessed from the average score of species along each dimension within species assemblages found in 100 km × 100 km square pixels ($n = 12,705$ assemblages). Eckert IV projection [Colour figure can be viewed at wileyonlinelibrary.com]

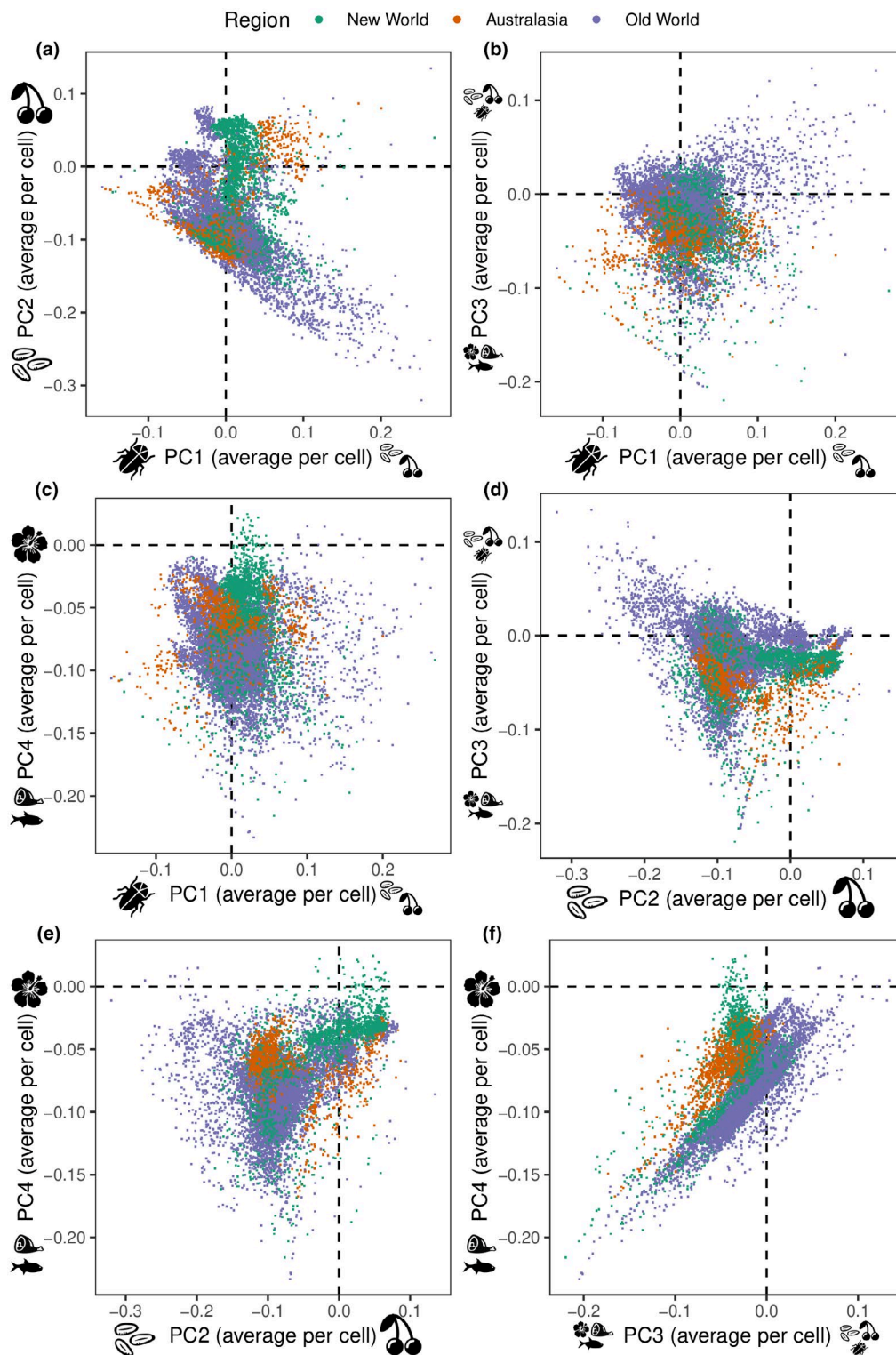


FIGURE 3 Spatial correlations among the dimensions of the diet space of birds. Each point corresponds to the average position of species found in one of 12,705 bird assemblages along each of the four dimensions of the diet space. Colours separate the correlations observed in the New World (both American continents and the Panamean region), the Old World (Eurasia, the Malay archipelago and Africa) and Australasia (encompasses Australia and New Zealand) [Colour figure can be viewed at wileyonlinelibrary.com]

regions, such as the Sahara or the Tibetan plateau, and some savannah regions (Figure 2a,b). The association between PC1 and PC2 did not hold in the New World ($R^2 = 0.03$, 3,711 assemblages) and was

reversed in Australasia ($R^2 = 0.18$, 1,114 assemblages), suggesting a key role of regional biogeography in separating dietary strategies across bird assemblages.

Another striking association was the positive correlation between the average scores of bird assemblages along PC3 and PC4 in all regions (Figure 3f; $R^2 = 0.44$, 12,705 assemblages). This pattern reflected the geographical dominance of carnivory and piscivory (negative scores on PC3 and PC4) in desertic, coastal and boreal areas (Figure 2d), whereas nectarivory (positive scores on PC4) was found only in association with frugivory, granivory and insectivory (positive scores on PC3; Figure 3f), especially in southern America and southeast Asia (Figure 2c,d).

3.3 | Diet space of birds along environmental gradients

The approximate correlation coefficients of quantile regressions (used to assess the fit of the relationship between each diet space dimension and the set of environmental variables considered) ranged from .06 (PC1, .5 quantile) to 0.54 (PC2, .1 quantile), averaging 0.24 ± 0.14 (all uncertainties in standard deviation units). All variance inflation factors but two were below a threshold of five, and 87 out of 120 were below three, suggesting that regression coefficients were not affected by collinearity (Zuur, 2011). Quantile regression lines crossed in PC1, PC3 and PC4 (Figure 4a,c,d), in spite of our attempts to constrain these models at exploratory stages; correcting this artefact would lead to unnecessary complexity given the clarity of the observed patterns.

Quantile lines converged as species richness increased along all dimensions of the diet space except PC2, suggesting greater similarity of richer assemblages in terms of average diet (Figure 4a,k,p). Conversely, quantile lines diverged as the average diet composition of species assemblages became increasingly variable towards warm, productive and topographically heterogeneous environments (e.g., Figure 4b,h,d). These contradicting patterns hint that the dissimilarity of species assemblages in terms of diet increases towards tropical climates. Conversely, human influence was associated with a reduction in the range of dietary strategies along all dimensions of the diet space (Figure 4e,j,o,t), conforming to a macroecological-level signature of functional homogenization associated with anthropogenic activity.

Several quantile regressions showed a systematic shift towards undifferentiated dietary strategies along environmental gradients (assemblages with null values on a given dimension of the diet space: e.g., Figure 4,b,j,m,p), consistent with environmental filtering. However, we could not draw any general pattern from these trends because they were non-congruent across the four dimensions and the four bioclimatic variables. Although no strong evidence of filtering emerged on PC1 (Figure 4a–e), increasing NPP was associated with a marked shift from assemblages dominated by granivory to frugivory along PC2 (Figure 4h), and to a lesser extent towards rare diets along PC3 (Figure 4m). The increase in PC3 scores (lower prevalence of rare diets) towards warmer climates seemed to contradict this picture (Figure 4l), but could be driven by the negative response of carnivory and piscivory to temperature, which was more obvious along PC4 (Figure 4q). Topographic barriers imposed another strong

environmental constraint, in which frugivory and carnivory were gradually replaced by granivory and nectarivory towards mountainous areas (Figure 4d–s).

4 | DISCUSSION

Our study revealed that bird species are spread inside a four-dimensional diet space, consistent with the prediction by MacArthur (1972) that the high diversity of life-forms can be summarized parsimoniously by a few basic continua bearing signs of the environmental constraints that preside to diversification. The biogeography of dietary strategies has previously been explored at regional scales (Pigot et al., 2016) or with a guild-richness approach, in which species were classified categorically according to their most prevalent diet (e.g., Kissling et al., 2012). In contrast, our study uniquely investigates the global and environmental distributions of dietary strategies in light of covariations that arise from niche convergence and exclusion (Díaz et al., 2016; Pianka et al., 2017).

The four-dimensional structure of the diet space of birds emerged from covariations among diets that species exploit either jointly or exclusively. The absence of any major discontinuities in this space shown in Figure 1 reflects the continuous, probabilistic nature of bird diets, exemplified by the tendency of passerines to feed on mixtures of fruits, seeds and insects (Herrera, Hobson, Carlos, & Méndez, 2006; Morton, 1973; Nazaro & Blendinger, 2017). In spite of this form of diet complementation, morphological differences (e.g., in bill size and shape), distinct foraging modes and digestive metabolism associated with insect or fruit-based regimes are consistent with the separation of insectivory, frugivory and granivory along the first two dimensions of the diet space (Figure 1; Hanken & Hall, 1993; Levey & Karasov, 1989; Snow & Snow, 1971). The next two dimensions (PC3 and PC4; Figure 1) corresponded to rarer diets with peculiar ecophysiological and morphological adaptations. Nectarivory implies a rapid metabolism efficient in assimilating high sugar concentrations at the cost of energetic storage, restricting this diet to the world's smallest bodied birds, hummingbirds (Trochilidae) and sunbirds (Nectariniidae), although it may also complement fruit- or insect-based diets in larger tropical species, such as honeyeaters (Meliphagidae) and *Loriculus* hanging parrots (Nicolson & Fleming, 2003; Wooller & Richardson, 2008). Nectarivory was thus logically opposed to carnivory and piscivory along the fourth dimension of the diet space (Figure 1c,e,f), with the latter two requiring large body sizes and morphological adaptations to cope with the hunting and capture of large prey (Barton & Houston, 2009; Hilton, Houston, Barton, Furness, & Ruxton, 1999). The diet space of birds is therefore likely to be sustained by a combination of behaviour, digestive physiology and morphology, which gives rise to well-identified diet-based convergences and trade-offs at the family level (see also Appendix S2; Fitzpatrick, 1985). Adaptive constraints are thus likely to provide a causal basis to our diet space, similar to the interpretation of constraints shaping plant and reptile functional spaces (Díaz et al., 2016; Pianka et al., 2017).

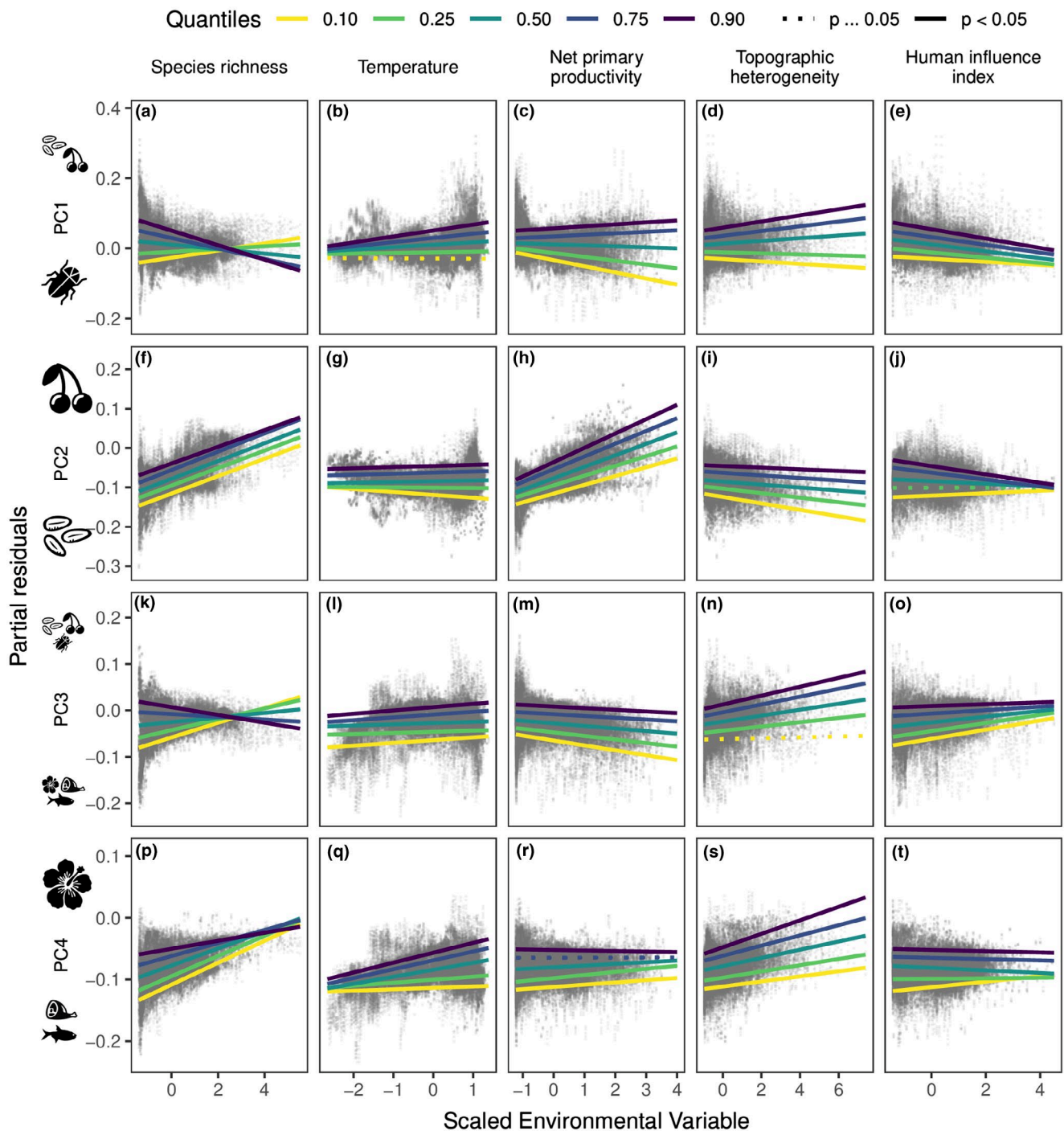


FIGURE 4 Quantile regressions of the four dimensions of the diet space of birds (PC1–PC4) against species richness, net primary productivity, temperature, topographic heterogeneity and human influence gradients. Regression lines correspond to the 0.1, 0.25, 0.5, 0.75 and 0.90 quantiles, and grey dots depict partial residuals [Colour figure can be viewed at wileyonlinelibrary.com]

We built this diet space with the most comprehensive diet assessment currently available for the birds of the world. These categories and scores form a representation of real bird diets that ignores specialization over specific prey but that are accurate at segregating major dietary groups. Given the sparsity of fine-grained data, we deliberately traded ecological resolution for an exhaustive comparative analysis unrestricted taxonomically or geographically, which

comes at the cost of some coarseness in our representation of the diet space. Although diet data are likely to be more exhaustive than in most other traits (Hortal et al., 2015), they are not standardized across studies and often rely on gut contents of one to fewer than 10 individuals captured in specific locations and seasons (Del Hoyo et al., 2013). Effort on supplementary data collection would warrant a quantification of intraspecific variability in the shape of the diet

space, a necessary perspective to enforce its macroevolutionary underpinnings (Violle et al., 2014). Furthermore, in addition to spatial variations across populations (Terraube & Arroyo, 2011), species may use different diets depending on age, sex, environmental constraints or resource availability (Davies, Hipfner, Hobson, & Ydenberg, 2009; León, Podos, Gardezi, Herrel, & Hendry, 2014; Murphy, 1994). In particular, numerous migratory species typically shift from insect-dominated diets to fruits at the onset of migration, whereas the diets of sedentary species follow food availability across seasons (Lovette & Fitzpatrick, 2016; Newton, 2010). This seasonality is usually aggregated or ignored in sources of species-level data, which probably leads to under-representation of the dispersion of species inside the diet space and to blurring of assemblage-level patterns, especially in temperate regions, where migratory species are most numerous. Although these limitations are unlikely to explain the whole structure of the diet space, they probably account for some (lack of) patterns, calling for increased effort into trait acquisition in poorly surveyed areas and at a finer resolution (Hortal et al., 2015).

The geographical mapping of the diet space through assemblage-averaged scores (Figure 2) supported the hypothesis that biogeographical legacies determine the distribution of bird dietary strategies on top of climatic influences and resource distributions (Belmaker et al., 2012; Ericson, 2012; Kissling et al., 2012). Increasing metabolic constraints in favour of high-order consumers promote assemblages essentially formed of carnivores and piscivores at high latitudes (Figure 2d), which accounts for the strong relationship between assemblage-averaged scores on PC3 and PC4 visible in Figure 3f (Clarke & O'Connor, 2014; Vejrikova et al., 2016; Zimmerman & Tracy, 1989). Likewise, the dominance of seed-feeding species in deserts and topographically complex regions (Figure 2b) can be related to seasonal climates and short growing seasons that restrict the availability of alternative foods for most of the year (Brown, Reichman, & Davidson, 1979; Kissling et al., 2012). Conversely, long-term climatic stability and niche conservatism confined the diversification of fruit trees and their consumers to the inter-tropical band (Jetz et al., 2009; Kissling et al., 2007, 2012), leading to a geographical association of insectivory and frugivory in the Old World (Figure 3a), although they were opposed at the species level along the first dimension of the diet space. Interestingly, however, this association was reversed in Australasia and was absent in the New World. These differences suggest that long-term climatic constraints were overcome by other processes, such as the admixture that resulted from the Great American Interchange or ancient diversification events that followed the colonization of cone-billed oscine clades from the Old-World since the late Miocene (Newton, 2003; Webb, 2006). Hence, the diet space of birds has probably been shaped largely by varying diversification rates across clades exploiting distinct dietary niches (e.g., hummingbirds, ducks and geese and various specialized passerine groups), associated with successive colonization events and recent radiations after climatic oscillations or modifications of physical barriers (Jetz, Thomas, Joy, Hartmann, & Mooers, 2012). Still consistent with a role of biogeographical

legacies associated with long-term climatic variability and plate tectonics (Ficetola, Mazel, & Thuiller, 2017), assemblage-averaged scores were more variable in topographically heterogeneous regions (i.e., near mountain chains) along the four dimensions of the diet space, and clear-cut spatial structures were associated with transition areas, such as the Panamanian isthmus, the Guineo-Congolian region, the Indo-Malayan and Papua-Melanesian archipelagos. These regional patterns support the hypothesis that admixture and diversification events that have resulted from orographic or tectonic processes imprint the global distribution of the diet space and call for further studies on its diversification history using phylogenetic reconstruction methods (Burin et al., 2016; Jetz et al., 2012; Newton, 2003).

Bird assemblages tended to become more concentrated in the diet space as species richness increased, as shown by the tendency of the quantile regression line to converge towards high species richness along all dimensions but PC2 (Figure 4). This result suggests that species-rich assemblages imply the coexistence of most dietary strategies, which could be promoted by niche packing in highly productive environments (Pellissier et al., 2018; Pigot et al., 2016). In line with the interpretation that productivity enhances the coexistence of multiple dietary strategies, the variance in the positions of assemblages in the diet space increased towards warm climates, high NPP and high topographic heterogeneity (Figure 4). These patterns concur with the hypothesis that warm and productive environments in addition to endemism associated with physical barriers promote diversity in the composition of species assemblages (Ficetola et al., 2017; Hawkins et al., 2003; Jetz et al., 2009; Ruggiero & Hawkins, 2008). At the low-productive side of these gradients, environmental filters could impose constraints on the diets of species-poor assemblages facing harsh climates or low energetic input (Hawkins et al., 2003). Examples of such filtering processes typically lie in pelagic bird colonies formed almost exclusively by piscivores at boreal latitudes (Figure 4q), or in desert assemblages mainly formed by seed-eaters (Figure 4h). In line with this energetic interpretation, the 23 poorest assemblages (species richness = 10) were distinctly segregated along the first dimension of the diet space, from desert communities of the Middle East, where specialized seed-eaters such as larks (Alaudidae) and sandgrouse (Pteroclididae) account for half of the species, to coastal assemblages of eastern Russia consisting of a mixture of insectivorous or omnivorous passerines and pelagic piscivores. Inversely, the 23 richest assemblages (700–848 species), all concentrated along the Andean chain, were clustered in the diet space and held specialists of the eight dietary categories (score of 10 in a single diet). Hence, the distribution of assemblages in the diet space is consistent with a combination of regional history and environmental constraints towards the pole and deserts.

A major outcome of our study was the packing of bird assemblages close to the centroid of the diet space when the human footprint increased, in agreement with the increased risk of extinction that accompanies ecological specialization by species (Şekercioğlu, 2011). This result suggests a trophic basis to the decrease in bird trait turnover along global gradients of human influence (Barnagaud

et al., 2017) and conveys a macroecological-level signal of functional homogenization in association with the loss of marginal dietary strategies and a rise in species using average diets (Baiser & Lockwood, 2011; Clavel, Julliard, & Devictor, 2011; Devictor et al., 2008). Our results thus add strong evidence that the human footprint is sufficiently pervasive to affect macroecological patterns (Šizling et al., 2016). Diets influence demographic and distributional responses of species to modifications in land use through their determinant role on offspring productivity and other aspects of energetic demands (Newbold et al., 2013; Sibly et al., 2012). The outcomes of a dietary homogenization of bird assemblages associated with the human imprint would thus deserve a more mechanistic assessment, especially with respect to the persistence and resilience of critical bird functions, such as pollination, endozoochorous dispersal and pest control (Ko, Schmitz, Barbet-Massin, & Jetz, 2013; Şekercioğlu, 2006, 2011).

4.1 | Conclusion

Our work illustrates the potential of multivariate approaches in functional biogeography analyses (Violle et al., 2014). It permits, in particular, an explicit mapping of functional strategies along continuous axes rather than using guild categorizations, and the explicit integration of covariations among traits or trait modalities. A formal test of the role of diversification processes on the diet space of birds was beyond the aims of this study, but will have to be tackled through phylogenetically informed null models (Díaz et al., 2016) or trait evolution modelling (Burin et al., 2016) for both basic and conservation-oriented aims. Our results should also draw interest on the potential alterations of macroecological-level trophic networks by anthropogenization, and their consequences for the distribution of ecosystem functions and services on earth. We hope that these outcomes, adjoined to studies on other taxa and aspects of the structure of functional spaces (Díaz et al., 2016; Pianka et al., 2017), will stimulate efforts towards bridging macroevolution and functional biogeography through the empirical exploration of global trait datasets.

ACKNOWLEDGMENTS

We are grateful to dozens of volunteers and students, especially Monte Neate-Clegg, Joshua Horns, Evan Buechley, Jason Socci, Sherron Bullens, Debbie Fisher, David Hayes, Beth Karpas and Kathleen McMullen, for their dedicated help with the world bird ecology database. We also thank Roger Koenker and Sandrine Pavoine for their advice regarding quantile regression and multivariate analyses, the editor and two anonymous referees for their constructive comments on a previous version of this manuscript. P.D. and C.V. were supported by the European Research Council (ERC) Starting Grant Project 'Ecophysiological and biophysical constraints on domestication of crop plants' (grant ERC-StG-2014-639706-CONSTRAINTS).

DATA ACCESSIBILITY

Data for environmental covariates and species assemblages at a 110 km × 110 km resolution, and species scores on the four dimensions of the diet space, are available through Dryad (<https://doi.org/10.5061/dryad.24c3s89>). Original diet data are not released, by request of their owner, but can be requested from the corresponding author.

ORCID

Jean-Yves Barnagaud  <https://orcid.org/0000-0002-8438-7809>

Pierre Denelle  <https://orcid.org/0000-0001-5037-2281>

Cyrille Violle  <https://orcid.org/0000-0002-2471-9226>

REFERENCES

- Anderson, K. J., & Jetz, W. (2005). The broad-scale ecology of energy expenditure of endotherms. *Ecology Letters*, 8, 310–318. <https://doi.org/10.1111/j.1461-0248.2005.00723.x>
- Baiser, B., & Lockwood, J. L. (2011). The relationship between functional and taxonomic homogenization. *Global Ecology and Biogeography*, 134–144. <https://doi.org/10.1111/j.1466-8238.2010.00583.x>
- Barnagaud, J.-Y., Kissling, W. D., Sandel, B., Eiserhardt, W. L., Şekercioğlu, Ç. H., Enquist, B. J., ... Svenning, J.-C. (2014). Ecological traits influence the phylogenetic structure of bird species co-occurrences worldwide. *Ecology Letters*, 17, 811–820. <https://doi.org/10.1111/ele.12285>
- Barnagaud, J. Y., Kissling, W. D., Tsirogiannis, C., Fisikopoulos, V., Sébastien, V., Şekercioğlu, C. H., & Svenning, J. C. (2017). Biogeographical, environmental and anthropogenic determinants of global patterns in bird taxonomic and trait turnover. *Global Ecology and Biogeography*, 26, 1190–1200. <https://doi.org/10.1111/geb.12629>
- Barton, N. W. H., & Houston, D. C. (2009). Morphological adaptation of the digestive tract in relation to feeding ecology of raptors. *Journal of Zoology*, 232, 133–150. <https://doi.org/10.1111/j.1469-7998.1994.tb01564.x>
- Belmaker, J., Şekercioğlu, C. H., & Jetz, W. (2012). Global patterns of specialization and coexistence in bird assemblages. *Journal of Biogeography*, 39, 193–203. <https://doi.org/10.1111/j.1365-2699.2011.02591.x>
- BirdLife International & Nature Serve. (2012). *Bird species distribution maps of the world*. Cambridge, UK; Arlington, TX: BirdLife International, NatureServe.
- Brown, J. H., Reichman, O. J., & Davidson, D. W. (1979). Granivory in desert ecosystems. *Annual Review of Ecology and Systematics*, 10, 201–227. <https://doi.org/10.1146/annurev.es.10.110179.001221>
- Burin, G., Kissling, W. D., Guimarães, P. R., Şekercioğlu, Ç. H., & Quental, T. B. (2016). Omnivory in birds is a macroevolutionary sink. *Nature Communications*, 7, 11250. <https://doi.org/10.1038/ncomms11250>
- Cadotte, M. W., & Tucker, C. M. (2017). Should environmental filtering be abandoned? *Trends in Ecology and Evolution*, 32, 429–437. <https://doi.org/10.1016/j.tree.2017.03.004>
- Clarke, A., & O'Connor, M. I. (2014). Diet and body temperature in mammals and birds. *Global Ecology and Biogeography*, 23, 1000–1008. <https://doi.org/10.1111/geb.12185>
- Clavel, J., Julliard, R., & Devictor, V. (2011). Worldwide decline of specialist species: Towards a global functional homogenization? *Frontiers in Ecology and the Environment*, 9, 222–238. <https://doi.org/10.1890/080216>

- Costa, G. C., Vitt, L. J., Pianka, E. R., Mesquita, D. O., & Colli, G. R. (2008). Optimal foraging constrains macroecological patterns: Body size and dietary niche breadth in lizards. *Global Ecology and Biogeography*, 17, 670–677. <https://doi.org/10.1111/j.1466-8238.2008.00405.x>
- Dalsgaard, B., Magård, E., Fjeldså, J., González, A. M. M., Rahbek, C., Olesen, J. M., ... Svenning, J.-C. (2011). Specialization in plant-hummingbird networks is associated with species richness, contemporary precipitation and quaternary climate-change velocity. *PLoS ONE*, 6, e25891. <https://doi.org/10.1371/journal.pone.0025891>
- Dalsgaard, B., Schleuning, M., Maruyama, P. K., Dehling, D. M., Sonne, J., Vizin-Bugoni, J., ... Rahbek, C. (2017). Opposed latitudinal patterns of network-derived and dietary specialization in avian plant-frugivore interaction systems. *Ecography*, 40, 1395–1401. <https://doi.org/10.1111/ecog.02604>
- Davies, W. E., Hipfner, J. M., Hobson, K. A., & Ydenberg, R. C. (2009). Seabird seasonal trophodynamics: Isotopic patterns in a community of Pacific alcids. *Marine Ecology Progress Series*, 382, 211–219. <https://doi.org/10.3354/meps07997>
- Davino, C., Furno, M., & Vistocco, D. (2013). *Quantile regression: Theory and applications* (1st ed.). Chichester, UK: Wiley.
- Dehling, D. M., Fritz, S. A., Töpfer, T., Päckert, M., Estler, P., Böhning-Gaese, K., & Schleuning, M. (2014). Functional and phylogenetic diversity and assemblage structure of frugivorous birds along an elevational gradient in the tropical Andes. *Ecography*, 37, 1047–1055. <https://doi.org/10.1111/ecog.00623>
- Del Hoyo, J., Elliott, A., & Sargatal, J. (2013). *Handbook of the birds of the world*. Barcelona, Spain: Lynx Edicions.
- Devictor, V., Julliard, R., Clavel, J., Jiguet, F., Lee, A., & Couvet, D. (2008). Functional biotic homogenization of bird communities in disturbed landscapes. *Global Ecology and Biogeography*, 17, 252–261. <https://doi.org/10.1111/j.1466-8238.2007.00364.x>
- Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., ... Gorné, L. D. (2016). The global spectrum of plant form and function. *Nature*, 529, 167–171. <https://doi.org/10.1038/nature16489>
- Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22, 1–20. <https://doi.org/10.18637/jss.v022.i04>
- Duffy, J. E. (2002). Biodiversity and ecosystem function: The consumer connection. *Oikos*, 99, 201–219. <https://doi.org/10.1034/j.1600-0706.2002.990201.x>
- Ericson, P. G. P. (2012). Evolution of terrestrial birds in three continents: Biogeography and parallel radiations. *Journal of Biogeography*, 39, 813–824. <https://doi.org/10.1111/j.1365-2699.2011.02650.x>
- Ficetola, G. F., Mazel, F., & Thuiller, W. (2017). Global determinants of zoogeographical boundaries. *Nature Ecology & Evolution*, 1, 0089. <https://doi.org/10.1038/s41559-017-0089>
- Fitzpatrick, J. W. (1985). Form, foraging behavior, and adaptive radiation in the Tyrannidae. *Ornithological Monographs*, 36, 447–470. <https://doi.org/10.2307/40168298>
- Garnier, E., Navas, M.-L., & Grigulis, K. (2016). *Plant functional diversity: Organism traits, community structure, and ecosystem properties* (1st ed.). Oxford, UK: Oxford University Press.
- Gill, F., & Donsker, D. (2017). IOC world bird list v.7.3. <https://doi.org/10.14344/IOC.ML.7.3>
- Grant, P. R., & Grant, B. R. (2006). Evolution of character displacement in Darwin's finches. *Science*, 313, 224–226. <https://doi.org/10.1126/science.1128374>
- Grime, J. P., & Pierce, S. (2012). *The evolutionary strategies that shape ecosystems* (1st ed.). Chichester, UK; Hoboken, NJ: Wiley-Blackwell.
- Grinnell, J. (1917). Field tests of theories concerning distributional control. *The American Naturalist*, 51, 115–128. <https://doi.org/10.1086/279591>
- Hanken, J., & Hall, B. K. (1993). *The skull, volume 3: Functional and evolutionary mechanisms*. Chicago: University of Chicago Press.
- Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guegan, J. F., Kaufman, D. M., ... Turner, J. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84, 3105–3117. <https://doi.org/10.1890/03-8006>
- Herk, K. M. B., Skidmore, A. K., & Fahr, J. (2017). Macroecological conclusions based on IUCN expert maps: A call for caution. *Global Ecology and Biogeography*, 26, 930–941. <https://doi.org/10.1111/geb.12601>
- Herrera, M. L. G., Hobson, K. A., Carlos, M. J., & Méndez, C. G. (2006). Tracing the origin of dietary protein in tropical dry forest birds. *Biotropica*, 38, 735–742. <https://doi.org/10.1111/j.1744-7429.2006.00201.x>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. <https://doi.org/10.1002/joc.1276>
- Hilton, G. M., Houston, D. C., Barton, N. W. H., Furness, R. W., & Ruxton, G. D. (1999). Ecological constraints on digestive physiology in carnivorous and piscivorous birds. *Journal of Experimental Zoology*, 283, 365–376. [https://doi.org/10.1002/\(SICI\)1097-010X\(19990301/01\)283:4/5<365:AID-JEZ6>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-010X(19990301/01)283:4/5<365:AID-JEZ6>3.0.CO;2-9)
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. In D. J. Futuyma (Ed.), *Annual review of ecology, evolution, and systematics* (Vol. 46, pp. 523–549). Palo Alto, CA: Annual Reviews. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415–427. <http://dx.doi.org/10.1101/SQB.1957.022.01.039>
- Hutchinson, G. E. (1959). Homage to Santa Rosalia or why are there so many kinds of animals? *The American Naturalist*, 93, 145–159. <https://doi.org/10.1086/282070>
- Janes, S. W. (1994). Variation in the species composition and mean body size of an avian foliage-gleaning guild along an elevational gradient: Correlation with arthropod body size. *Oecologia*, 98, 369–378. <https://doi.org/10.1007/BF00324226>
- Jetz, W., Kreft, H., Ceballos, G., & Mutke, J. (2009). Global associations between terrestrial producer and vertebrate consumer diversity. *Proceedings of the Royal Society B: Biological Sciences*, 276, 269–278. <https://doi.org/10.1098/rspb.2008.1005>
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491, 444–448. <https://doi.org/10.1038/nature11631>
- Jiguet, F., Gadot, A. S., Julliard, R., Newson, S. E., & Couvet, D. (2007). Climate envelope, life history traits and the resilience of birds facing global change. *Global Change Biology*, 13, 1672–1684. <https://doi.org/10.1111/j.1365-2486.2007.01386.x>
- Kissling, W. D., Rahbek, C., & Böhning-Gaese, K. (2007). Food plant diversity as broad-scale determinant of avian frugivore richness. *Proceedings of the Royal Society B: Biological Sciences*, 274, 799–808. <https://doi.org/10.1098/rspb.2006.0311>
- Kissling, W. D., Şekercioğlu, C. H., & Jetz, W. (2012). Bird dietary guild richness across latitudes, environments and biogeographic regions. *Global Ecology and Biogeography*, 21, 328–340. <https://doi.org/10.1111/j.1466-8238.2011.00679.x>
- Ko, C.-Y., Schmitz, O. J., Barbet-Massin, M., & Jetz, W. (2013). Dietary guild composition and disaggregation of avian assemblages under climate change. *Global Change Biology*, 20, 790–802. <https://doi.org/10.1111/gcb.12419>
- Koenker, R. (2016). quantreg: Quantile Regression. R Package Version 5.29. Retrieved from <https://CRAN.R-project.org/package=quantreg>
- Koenker, R., & Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94, 1296–1310. <https://doi.org/10.1080/01621459.1999.10473882>

- Laughlin, D. C. (2014). The intrinsic dimensionality of plant traits and its relevance to community assembly. *Journal of Ecology*, 102, 186–193. <https://doi.org/10.1111/1365-2745.12187>
- Legendre, P., & Legendre, L. F. J. (2012). *Numerical ecology* (3rd ed., Vol. 24). Amsterdam, The Netherlands: Elsevier.
- Lein, M. R. (1972). A trophic comparison of avifaunas. *Systematic Zoology*, 21, 135–150. <https://doi.org/10.2307/2412285>
- De León, L. F., Podos, J., Gardezi, T., Herrel, A., & Hendry, A. P. (2014). Darwin's finches and their diet niches: The sympatric coexistence of imperfect generalists. *Journal of Evolutionary Biology*, 27, 1093–1104. <https://doi.org/10.1111/jeb.12383>
- Levey, D. J., & Karasov, W. H. (1989). Digestive responses of temperate birds switched to fruit or insect diets. *The Auk*, 106, 675–686. <https://doi.org/10.1093/auk/106.4.675>
- Lovette, I. J., & Fitzpatrick, J. W. (2016). *Handbook of bird biology* (3rd ed.). Chichester, UK: John Wiley & Sons Inc.
- MacArthur, R. H. (1972). Coexistence of species. In J. Behnke (Ed.), *Challenging biological problems* (pp. 253–259). Oxford: Oxford University Press.
- MacArthur, R. H., & Pianka, E. R. (1966). On optimal use of a patchy environment. *The American Naturalist*, 100, 603–609. <https://doi.org/10.1086/282454>
- Maire, E., Grenouillet, G., Brosse, S., & Villéger, S. (2015). How many dimensions are needed to accurately assess functional diversity? A pragmatic approach for assessing the quality of functional spaces. *Global Ecology and Biogeography*, 24, 728–740. <https://doi.org/10.1111/geb.12299>
- McGill, B., Enquist, B., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution*, 21, 178–185. <https://doi.org/10.1016/j.tree.2006.02.002>
- Morowitz, H. J. (1980). The dimensionality of niche space. *Journal of Theoretical Biology*, 86, 259–263. [https://doi.org/10.1016/0022-5193\(80\)90006-5](https://doi.org/10.1016/0022-5193(80)90006-5)
- Morton, E. (1973). On the evolutionary advantages and disadvantages of fruit eating in tropical birds. *The American Naturalist*, 107, 8–22. <https://doi.org/10.1086/282813>
- Murphy, M. (1994). Dietary complementation by wild birds—Considerations for field studies. *Journal of Biosciences*, 19, 355–368. <https://doi.org/10.1007/BF02703173>
- Nazaro, M. G., & Blendinger, P. G. (2017). How important are arthropods in the diet of fruit-eating birds? *Wilson Journal of Ornithology*, 129, 520–527. <https://doi.org/10.1676/16-083.1>
- Newbold, T., Butchart, S. H. M., Şekercioğlu, Ç. H., Purves, D. W., & Scharlemann, J. P. W. (2012). Mapping functional traits: Comparing abundance and presence-absence estimates at large spatial scales. *PLoS ONE*, 7, e44019. <https://doi.org/10.1371/journal.pone.0044019>
- Newbold, T., Scharlemann, J. P. W., Butchart, S. H. M., Şekercioğlu, Ç. H., Alkemade, R., Booth, H., & Purves, D. W. (2013). Ecological traits affect the response of tropical forest bird species to land-use intensity. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20122131. <https://doi.org/10.1098/rspb.2012.2131>
- Newton, I. (2003). *The speciation and biogeography of birds*. Oxford, UK: Academic Press.
- Newton, I. (2010). *The migration ecology of birds* (1st ed.). Oxford: Academic Press.
- Nicolson, S. W., & Fleming, P. A. (2003). Nectar as food for birds: The physiological consequences of drinking dilute sugar solutions. *Plant Systematics and Evolution*, 238, 139–153. <https://doi.org/10.1007/s00606-003-0276-7>
- Pavoine, S., Vallet, J., Dufour, A.-B., Gachet, S., & Daniel, H. (2009). On the challenge of treating various types of variables: Application for improving the measurement of functional diversity. *Oikos*, 118, 391–402. <https://doi.org/10.1111/j.1600-0706.2008.16668.x>
- Pellissier, V., Barnagaud, J.-Y., Kissling, W. D., Şekercioğlu, Ç., & Svenning, J.-C. (2018). Niche packing and expansion account for species richness–productivity relationships in global bird assemblages. *Global Ecology and Biogeography*, 27, 604–615. <https://doi.org/10.1111/geb.12723>
- Pianka, E. R., Vitt, L. J., Pelegrin, N., Fitzgerald, D. B., & Winemiller, K. O. (2017). Toward a periodic table of niches, or exploring the lizard niche hypervolume. *The American Naturalist*, 190, 601–616. <https://doi.org/10.1086/693781>
- Pigot, A. L., Trisos, C. H., & Tobias, J. A. (2016). Functional traits reveal the expansion and packing of ecological niche space underlying an elevational diversity gradient in passerine birds. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20152013. <https://doi.org/10.1098/rspb.2015.2013>
- Purvis, A., Gittleman, J. L., Cowlshaw, G., & Mace, G. M. (2000). Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences*, 267, 1947–1952. <https://doi.org/10.1098/rspb.2000.1234>
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ricotta, C., & Moretti, M. (2011). CWM and Rao's quadratic diversity: A unified framework for functional ecology. *Oecologia*, 167, 181–188. <https://doi.org/10.1007/s00442-011-1965-5>
- Ruggiero, A., & Hawkins, B. A. (2008). Why do mountains support so many species of birds? *Ecography*, 31, 306–315. <https://doi.org/10.1111/j.0906-7590.2008.05333.x>
- Sam, K., Koane, B., Jeppy, S., Sykora, J., & Novotny, V. (2017). Diet of land birds along an elevational gradient in Papua New Guinea. *Scientific Reports*, 7, 44018. <https://doi.org/10.1038/srep44018>
- Sanderson, E. W., Jaiteh, M., Levy, M. A., Redford, K. H., Wannebo, A. V., & Woolmer, G. (2002). The human footprint and the last of the wild. *BioScience*, 52, 891–904. [https://doi.org/10.1641/0006-3568\(2002\)052\[0891:THFATL\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052[0891:THFATL]2.0.CO;2)
- Şekercioğlu, C. H. (2006). Increasing awareness of avian ecological function. *Trends in Ecology & Evolution*, 21, 464–471. <https://doi.org/10.1016/j.tree.2006.05.007>
- Şekercioğlu, C. H. (2011). Functional extinctions of bird pollinators cause plant declines. *Science*, 331, 1019–1020. <https://doi.org/10.1126/science.1202389>
- Şekercioğlu, Ç. H., Daily, G. C., & Ehrlich, P. R. (2004). Ecosystem consequences of bird declines. *Proceedings of the National Academy of Sciences USA*, 101, 18042–18047. <https://doi.org/10.1073/pnas.0408049101>
- Şekercioğlu, Ç. H., Wenny, D. G., & Whelan, C. J. (2016). *Why birds matter: Avian ecological function and ecosystem services*. Chicago, IL: University of Chicago Press.
- Sibly, R. M., Witt, C. C., Wright, N. A., Venditti, C., Jetz, W., & Brown, J. H. (2012). Energetics, lifestyle, and reproduction in birds. *Proceedings of the National Academy of Sciences USA*, 109, 10937–10941. <https://doi.org/10.1073/pnas.1206512109>
- Šizling, A. L., Pokorný, P., Juříčková, L., Horáčková, J., Abraham, V., Šizlingová, E., ... Kerr, J. (2016). Can people change the ecological rules that appear general across space? *Global Ecology and Biogeography*, 25, 1072–1084. <https://doi.org/10.1111/geb.12467>
- Snow, B. K., & Snow, D. W. (1971). The feeding ecology of tanagers and honeycreepers in Trinidad. *The Auk*, 88, 291–322. <https://doi.org/10.2307/4083882>
- Soberon, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10, 1115–1123. <https://doi.org/10.1111/j.1461-0248.2007.01107.x>
- Speakman, J. R. (1999). The cost of living: Field metabolic rates of small mammals. In A. H. Fitter & D. G. Raffaelli (Eds.), *Advances in ecological research* (pp. 177–297). [https://doi.org/10.1016/S0065-2504\(08\)60019-7](https://doi.org/10.1016/S0065-2504(08)60019-7)
- Terraube, J., & Arroyo, B. (2011). Factors influencing diet variation in a generalist predator across its range distribution. *Biodiversity and Conservation*, 20, 2111–2131. <https://doi.org/10.1007/s10531-011-0077-1>

- Tucker, C. M., Davies, T. J., Cadotte, M. W., & Pearse, W. D. (2018). On the relationship between phylogenetic diversity and trait diversity. *Ecology*, *99*, 1473–1479. <https://doi.org/10.1002/ecy.2349>
- Vejrikova, I., Vejrik, L., Syvaranta, J., Kiljunen, M., Cech, M., Blabolil, P., ... Peterka, J. (2016). Distribution of herbivorous fish is frozen by low temperature. *Scientific Reports*, *6*, 39600. <https://doi.org/10.1038/srep39600>
- Violle, C., Reich, P. B., Pacala, S. W., Enquist, B. J., & Kattge, J. (2014). The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences USA*, *111*, 13690–13696. <https://doi.org/10.1073/pnas.1415442111>
- Webb, S. D. (2006). The great American biotic interchange: Patterns and processes. *Annals of the Missouri Botanical Garden*, *93*, 245–257. [https://doi.org/10.3417/0026-6493\(2006\)93\[245:TGABIP\]2.0.CO;2](https://doi.org/10.3417/0026-6493(2006)93[245:TGABIP]2.0.CO;2)
- Winemiller, K. O., Fitzgerald, D. B., Bower, L. M., Pianka, E. R., & Arita, H. (2015). Functional traits, convergent evolution, and periodic tables of niches. *Ecology Letters*, *18*, 737–751. <https://doi.org/10.1111/ele.12462>
- Wooller, R. D., & Richardson, K. C. (2008). Morphological relationships of passerine birds from Australia and New Guinea in relation to their diets. *Zoological Journal of the Linnean Society*, *94*, 193–201. <https://doi.org/10.1111/j.1096-3642.1988.tb00106.x>
- Wüest, R. O., Münkemüller, T., Lavergne, S., Pollock, L. J., & Thuiller, W. (2018). Integrating correlation between traits improves spatial predictions of plant functional composition. *Oikos*, *127*, 472–481. <https://doi.org/10.1111/oik.04420>
- Zhao, M., Heinsch, F. A., Nemani, R. R., & Running, S. W. (2005). Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, *95*, 164–176. <https://doi.org/10.1016/j.rse.2004.12.011>
- Zimmerman, L., & Tracy, C. (1989). Interactions between the environment and ectothermy and herbivory in reptiles. *Physiological Zoology*, *62*, 374–409. <https://doi.org/10.1086/physzool.62.2.30156176>
- Zuur, A. (2011). *Mixed effects models and extensions in ecology with R, softcover reprint of hardcover* (1st ed., 2009 edition). New York, NY: Springer.

BIOSKETCH

The author team is broadly interested in the functional and evolutionary implications of trait distributions across species assemblages at a macroecological scale. The authors have specific expertise in functional ecology, conservation biology, biogeography and macroevolution, focusing on observational and ecoinformatics approaches.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Barnagaud J-Y, Mazet N, Munoz F, et al. Functional biogeography of dietary strategies in birds. *Global Ecol Biogeogr.* 2019;28:1004–1017. <https://doi.org/10.1111/geb.12910>

Bibliographie

- 10 ABRAMS, P. (1983) The Theory of Limiting Similarity. *Annual Review of Ecology and Systematics* **14**, 359–376.
- ACKERLY, D. (2009) Conservatism and diversification of plant functional traits : Evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences* **106**, 19699–19706.
- ADLER, P.B., HILLERISLAMBERS, J. & LEVINE, J.M. (2007) A niche for neutrality. *Ecology Letters* **10**, 95–104.
- ADLER, P.B., SALGUERO-GÓMEZ, R., COMPAGNONI, A., HSU, J.S., RAY-MUKHERJEE, J., MBEAUACHE, C. & FRANCO, M. (2014) Functional traits explain variation in plant life history strategies. *Proceedings of the National Academy of Sciences* **111**, 740–745. National Academy of Sciences.
- AGGARWAL, C.C., HINNEBURG, A. & KEIM, D.A. (2001) On the Surprising Behavior of Distance Metrics in High Dimensional Space. In *Database Theory ICDT 2001* (eds G. GOOS, J. HARTMANIS, J. VAN LEEUWEN, J. VAN DEN BUSSCHE & V. VIANU), pp. 420–434. Springer Berlin Heidelberg, Berlin, Heidelberg.
- ALBERT, C.H., THUILLER, W., YOCCOZ, N.G., SOUDANT, A., BOUCHER, F., SACCONI, P. & LAVOREL, S. (2010) Intraspecific functional variability : Extent, structure and sources of variation. *Journal of Ecology* **98**, 604–613.
- ANON. (2020) iNaturalist. *iNaturalist*.
- ARPONEN, A. (2012) Prioritizing species for conservation planning. *Biodiversity and Conservation* **21**, 875–893.
- BAHLAI, C., BARTLETT, L.J., BURGIO, K.R., FOURNIER, A.M., KEISER, C.N., POISOT, T. & WHITNEY, K.S. (2019) Open science isn't always open to all scientists. *American Scientist* **107**, 78–82.
- BANERJEE, S., SCHLAEPI, K. & HEIJDEN, M.G.A. VAN DER (2018) Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology* **16**, 567–576. Nature Publishing Group.

- BARALOTO, C., TIMOTHY PAINE, C.E., PATIÑO, S., BONAL, D., HÉRAULT, B. & CHAVE, J. (2010) Functional trait variation and sampling strategies in species-rich plant communities. *Functional Ecology* **24**, 208–216.
- BAR-ON, Y.M., PHILLIPS, R. & MILO, R. (2018) The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* **115**, 6506–6511.
- BARTHELEMY, E., DENELLE, P., BLANCHARD, G., VIOLLE, C. & MUNOZ, F. (in revision) An ecologist's guide to infer mechanisms of community assembly using ABC. *Methods in Ecology and Evolution*.
- BEAUMONT, M.A. (2010) Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics* **41**, 379–406. Annual Reviews.
- BENÍTEZ LÓPEZ, A. & SANTINI, L. (2019) Game of Tenure : The role of 'hidden' citations on researchers' ranking in Ecology. *Frontiers of Biogeography* **0**.
- BENUREAU, F.C.Y. & ROUGIER, N.P. (2018) Re-run, Repeat, Reproduce, Reuse, Replicate : Transforming Code into Scientific Contributions. *Frontiers in Neuroinformatics* **11**.
- BESTOVA, H., MUNOZ, F., SVOBODA, P., SKALLOUD, P. & VIOLLE, C. (2018) Ecological and biogeographical drivers of freshwater green algae biodiversity : From local communities to large-scale species pools of desmids. *Oecologia* **186**, 1017–1030. Springer, New York.
- BEVERTON, R.J. & HOLT, S.J. (1957) On the dynamics of exploited fish populations. Fishery Investigations, Series II, 19. *London : Her Majesty's Stationary Office*.
- BEYER, K., GOLDSTEIN, J., RAMAKRISHNAN, R. & SHAFT, U. (1999) When Is 'Nearest Neighbor' Meaningful? In *Database Theory ICDT'99* (eds G. GOOS, J. HARTMANIS, J. VAN LEEUWEN, C. BEERI & P. BUNEMAN), pp. 217–235. Springer Berlin Heidelberg, Berlin, Heidelberg.
- BLAND, L.M. & BÖHM, M. (2016) Overcoming data deficiency in reptiles. *Biological Conservation* **204**, 16–22.
- BLONDER, B., KAPAS, R.E., DALTON, R.M., GRAAE, B.J., HEILING, J.M. & OPEDAL, Ø.H. (2018) Microenvironment and functional-trait context dependence predict alpine plant community dynamics. *Journal of Ecology* **106**, 1323–1337.
- BOETTIGER, C., CHAMBERLAIN, S., HART, E. & RAM, K. (2015) Building Software, Building Community : Lessons from the rOpenSci Project. *Journal of Open Research Software* **3**.
- BOLNICK, D.I., AMARASEKARE, P., ARAÚJO, M.S., BÜRGER, R., LEVINE, J.M., NOVAK, M., RUDOLF, V.H.W., SCHREIBER, S.J., URBAN, M.C. & VASSEUR, D.A. (2011) Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution* **26**, 183–192.

- BORGY, B., VIOLLE, C., CHOLER, P., GARNIER, E., KATTGE, J., LORANGER, J., AMIAUD, B., CELLIER, P., DEBARROS, G., DENELLE, P., DIQUELOU, S., GACHET, S., JOLIVET, C., LAVOREL, S., LEMAUVIEL-LAVENANT, S., et al. (2017) Sensitivity of community-level trait-environment relationships to data representativeness : A test for functional biogeography. *Global Ecology and Biogeography* **26**, 729–739.
- BORREGAARD, M.K. & HART, E.M. (2016) Towards a more reproducible ecology. *Ecography* **39**, 349–353.
- BOTTA-DUKÁT, Z. & CZÚCZ, B. (2015) Testing the ability of functional diversity indices to detect trait convergence and divergence using individual-based simulation. *Methods in Ecology and Evolution*, n/a–n/a.
- BROWN, M.V., OSTROWSKI, M., GRZYMSKI, J.J. & LAURO, F.M. (2014) A trait based perspective on the biogeography of common and abundant marine bacterioplankton clades. *Marine Genomics* **15**, 17–28.
- BRUELHEIDE, H., DENGLER, J., JIMÉNEZ-ALFARO, B., PURSCHKE, O., HENNEKENS, S.M., CHYTRÝ, M., PILLAR, V.D., JANSEN, F., KATTGE, J., SANDEL, B., AUBIN, I., BIURRUN, I., FIELD, R., HAIDER, S., JANDT, U., et al. (2019) sPlot A new tool for global vegetation analyses. *Journal of Vegetation Science* **30**, 161–186.
- BUSH, A., SOLLMANN, R., WILTING, A., BOHMANN, K., COLE, B., BALZTER, H., MARTIUS, C., ZLINSZKY, A., CALVIGNAC-SPENCER, S., COBBOLD, C.A., DAWSON, T.P., EMERSON, B.C., FERRIER, S., GILBERT, M.T.P., HEROLD, M., et al. (2017) Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology & Evolution* **1**, 0176.
- BUUREN, S. & GROOTHUIS-OUDSHOORN, K. (2011) MICE : Multivariate imputation by chained equations in R. *Journal of statistical software* **45**.
- CABRAL, J.S., VALENTE, L. & HARTIG, F. (2016) Mechanistic simulation models in macroecology and biogeography : State-of-art and prospects. *Ecography*, n/a–n/a.
- CADOTTE, M., ALBERT, C.H. & WALKER, S.C. (2013) The ecology of differences : Assessing community assembly with trait and evolutionary distances. *Ecology Letters* **16**, 1234–1244.
- CADOTTE, M.W., CARDINALE, B.J. & OAKLEY, T.H. (2008) Evolutionary history and the effect of biodiversity on plant productivity. *Proceedings of the National Academy of Sciences* **105**, 17012–17017.
- CADOTTE, M.W. & TUCKER, C.M. (2017) Should Environmental Filtering be Abandoned? *Trends in Ecology & Evolution* **32**, 429–437.

- CARDINALE, B.J., WRIGHT, J.P., CADOTTE, M.W., CARROLL, I.T., HECTOR, A., SRIVASTAVA, D.S., LOREAU, M. & WEIS, J.J. (2007) Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proceedings of the National Academy of Sciences* **104**, 18123–18128. National Academy of Sciences.
- CARMONA, C.P., ROTA, C., AZCÁRATE, F.M. & PECO, B. (2015) More for less : Sampling strategies of plant functional traits across local environmental gradients. *Functional Ecology* **29**, 579–588. John Wiley & Sons, Ltd.
- CASE, T.J., HOLT, R.D., MCPEEK, M.A. & KEITT, T.H. (2005) The community context of species' borders : Ecological and evolutionary perspectives. *Oikos* **108**, 28–46. Wiley Online Library.
- CHAPMAN, A.S.A., TUNNICLIFFE, V. & BATES, A.E. (2018) Both rare and common species make unique contributions to functional diversity in an ecosystem unaffected by human activities. *Diversity and Distributions* **24**, 568–578.
- CHICHORRO, F., JUSLÉN, A. & CARDOSO, P. (2019) A review of the relation between species traits and extinction risk. *Biological Conservation* **237**, 220–229.
- CLAVEY, M. (2019) Évolution du recrutement des ingénieurs de recherche au CNRS de 2007 à 2018. *The Sound Of Science*.
- COHEN, J., KATZ, D.S., BARKER, M., CHUE HONG, N.P., HAINES, R. & JAY, C. (2020) The Four Pillars of Research Software Engineering. *IEEE Software*, 0–0.
- CONNOR, E.F. & SIMBERLOFF, D. (1979) The assembly of species communities : Chance or competition? *Ecology* **60**, 1132–1140. Wiley Online Library.
- COOK, R.B., WEI, Y., HOOK, L.A., VANNAN, S.K.S. & MCNELIS, J.J. (2018) Preserve : Protecting Data for Long-Term Use. In *Ecological Informatics* (eds F. RECKNAGEL & W.K. MICHENER), pp. 89–113. Springer International Publishing, Cham.
- COOKE, R.S.C., EIGENBROD, F. & BATES, A.E. (2020) Ecological distinctiveness of birds and mammals at the global scale. *Global Ecology and Conservation*, e00970.
- CORNWELL, W.K. & ACKERLY, D.D. (2009) Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs* **79**, 109–126.
- CORNWELL, W.K., PEARSE, W.D., DALRYMPLE, R.L. & ZANNE, A.E. (2019) What we (don't) know about global plant diversity. *Ecography* **42**, 1819–1831. John Wiley & Sons, Ltd.
- COSTELLO, M.J., MICHENER, W.K., GAHEGAN, M., ZHANG, Z.-Q. & BOURNE, P.E. (2013) Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution* **28**, 454–461.

- DANET, A., ANTHELME, F., GROSS, N. & KÉFI, S. (2018) Effects of indirect facilitation on functional diversity, dominance and niche differentiation in tropical alpine communities. *Journal of Vegetation Science* **29**, 835–846.
- DEE, L.E., COWLES, J., ISBELL, F., PAU, S., GAINES, S.D. & REICH, P.B. (2019) When Do Ecosystem Services Depend on Rare Species? *Trends in Ecology & Evolution* **34**, 746–758.
- DES ROCHES, S., POST, D.M., TURLEY, N.E., BAILEY, J.K., HENDRY, A.P., KINNISON, M.T., SCHWEITZER, J.A. & PALKOVACS, E.P. (2018) The ecological importance of intraspecific variation. *Nature Ecology & Evolution* **2**, 57–64.
- DIAMOND, J.M. (1975) Assembly of species communities. *Ecology and evolution of communities* **342**, 444.
- DIAMOND, J.M. (1989) The present, past and future of human-caused extinctions. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **325**, 469–477. The Royal Society London.
- DÍAZ, S., KATTGE, J., CORNELISSEN, J.H.C., WRIGHT, I.J., LAVOREL, S., DRAY, S., REU, B., KLEYER, M., WIRTH, C., PRENTICE, I.C., GARNIER, E., BÖNISCH, G., WESTOBY, M., POORTER, H., REICH, P.B., et al. (2016) The global spectrum of plant form and function. *Nature* **529**, 167–171.
- DORNELAS, M., ANTÃO, L.H., MOYES, F., BATES, A.E., MAGURRAN, A.E., ADAM, D., AKHMETZHANOVA, A.A., APPELTANS, W., ARCOS, J.M., ARNOLD, H., AYYAPPAN, N., BADIHI, G., BAIRD, A.H., BARBOSA, M., BARRETO, T.E., et al. (2018) BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography* **27**, 760–786.
- DZWONKO, Z. (2001) Assessment of light and soil conditions in ancient and recent woodlands by Ellenberg indicator values. *Journal of Applied Ecology* **38**, 942–951. John Wiley & Sons, Ltd.
- ELLENBERG, H. (1992) Zeigerwerte der Gefäßpflanzen (ohne Rubus). *Scripta Geobotanica* **18**, 9–166.
- ENQUIST, B.J., CONDIT, R., PEET, R.K., SCHILDHAUER, M. & THIERS, B.M. (2016) Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. PeerJ Inc.
- ENQUIST, B.J., NORBERG, J., BONSER, S.P., VIOLLE, C., WEBB, C.T., HENDERSON, A., SLOAT, L.L. & SAVAGE, V.M. (2015) Scaling from Traits to Ecosystems. *Advances in Ecological Research* **52**, 249–318.
- ESPELAND, E.K. & EMAM, T.M. (2011) The value of structuring rarity: The seven types and links to reproductive ecology. *Biodiversity and Conservation* **20**, 963–985.

- ESTES, L., ELSÉN, P.R., TREUER, T., AHMED, L., CAYLOR, K., CHANG, J., CHOI, J.J. & ELLIS, E.C. (2018) The spatial and temporal domains of modern ecology. *Nature Ecology & Evolution* **2**, 819.
- FARLEY, S.S., DAWSON, A., GORING, S.J. & WILLIAMS, J.W. (2018) Situating Ecology as a Big-Data Science : Current Advances, Challenges, and Solutions. *BioScience* **68**, 563–576. Oxford Academic.
- FARRELL, K.J. & CAREY, C.C. (2018) Power, pitfalls, and potential for integrating computational literacy into undergraduate ecology courses. *Ecology and Evolution* **8**, 7744–7751. John Wiley & Sons, Ltd.
- FERNÁNDEZ, M.H. & VRBA, E.S. (2005) Macroevolutionary Processes and Biomic Specialization : Testing the Resource-use Hypothesis. *Evolutionary Ecology* **19**, 199–219. Kluwer Academic Publishers.
- FIDLER, F., CHEE, Y.E., WINTLE, B.C., BURGMAN, M.A., MCCARTHY, M.A. & GORDON, A. (2017) Metaresearch for Evaluating Reproducibility in Ecology and Evolution. *BioScience* **67**, 282–289. Oxford Academic.
- FOREST, F., GRENYER, R., ROUGET, M., DAVIES, T.J., COWLING, R.M., FAITH, D.P., BALMFORD, A., MANNING, J.C., PROCHEŞ, Ş., BANK, M. VAN DER, REEVES, G., HEDDERSON, T.A.J. & SAVOLAINEN, V. (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**, 757–760.
- FOUNTAIN-JONES, N.M., BAKER, S.C. & JORDAN, G.J. (2015) Moving beyond the guild concept : Developing a practical functional trait framework for terrestrial beetles. *Ecological Entomology* **40**, 1–13. John Wiley & Sons, Ltd.
- FOURNIER, B., GILLET, F., LE BAYON, R.-C., MITCHELL, E.A.D. & MORETTI, M. (2015) Functional responses of multitaxa communities to disturbance and stress gradients in a restored floodplain. *Journal of Applied Ecology* **52**, 1364–1373.
- FRANKLIN, J., SERRA-DIAZ, J.M., SYPHARD, A.D. & REGAN, H.M. (2016) Big data for forecasting the impacts of global change on plant communities. *Global Ecology and Biogeography*, n/a–n/a.
- FRASER, H., PARKER, T., NAKAGAWA, S., BARNETT, A. & FIDLER, F. (16 juil. 2018) Questionable research practices in ecology and evolution. *PLOS ONE* **13**, e0200303. Public Library of Science.
- FRECKLETON, R.P. (2018) Accessibility, reusability, reliability : Improving the standards for publishing code in Methods in Ecology and Evolution. *Methods in Ecology and Evolution* **9**, 4–6. John Wiley & Sons, Ltd.
- GALLAGHER, R.V., FALSTER, D.S., MAITNER, B.S., SALGUERO-GÓMEZ, R., VANDVIK, V., PEARSE, W.D., SCHNEIDER, F.D., KATTGE, J., POELEN, J.H., MADIN, J.S., ANKENBRAND, M.J., PENONE, C., FENG, X., ADAMS, V.M., ALROY, J., et al. (2020) Open Science principles for accelerating trait-

- based science across the Tree of Life. *Nature Ecology & Evolution*, 1–10. Nature Publishing Group.
- GARNIER, E., CORTEZ, J., BILLÈS, G., NAVAS, M.-L., ROUMET, C., DEBUSSCHE, M., LAURENT, G., BLANCHARD, A., AUBRY, D., BELLMANN, A., NEILL, C. & TOUSSAINT, J.-P. (2004) Plant Functional Markers Capture Ecosystem Properties During Secondary Succession. *Ecology* **85**, 2630–2637.
- GASTON, K.J. (1997) What is rarity? In *The Biology of Rarity* (eds W.E. KUNIN & K.J. GASTON), pp. 30–47. Springer Netherlands.
- GASTON, K.J. (1998) Rarity as double jeopardy. *Nature* **394**, 229–230. Nature Publishing Group.
- GASTON, K.J. (2009) Geographic range limits : Achieving synthesis. *Proceedings of the Royal Society B : Biological Sciences* **276**, 1395–1406. Royal Society.
- GAUSE, G.F. (1934) Experimental analysis of Vito Volterra's mathematical theory of the struggle for existence. *Science* **79**, 16–17.
- GAVRILETS, S. & LOSOS, J.B. (2009) Adaptive Radiation : Contrasting Theory with Data. *Science* **323**, 732–737. American Association for the Advancement of Science.
- GBIF : THE GLOBAL BIODIVERSITY INFORMATION FACILITY (2020) What is GBIF? *GBIF*.
- GIAM, X. & OLDEN, J.D. (2017) Drivers and interrelationships among multiple dimensions of rarity for freshwater fishes. *Ecography*, n/a–n/a.
- GIMENEZ, O., BUCKLAND, S.T., MORGAN, B.J.T., BEZ, N., BERTRAND, S., CHOQUET, R., DRAY, S., ETIENNE, M.-P., FEWSTER, R., GOSSELIN, F., MÉRIGOT, B., MONESTIEZ, P., MORALES, J.M., MORTIER, F., MUNOZ, F., et al. (2014) Statistical ecology comes of age. *Biology Letters* **10**, 20140698. Royal Society.
- GOLDBERG, D.E. (1990) Components of resource competition in plant communities. *Perspectives on plant competition*, 27–49.
- GOOLSBY, E.W., BRUGGEMAN, J. & ANÉ, C. (2017) Rphylopars : Fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution* **8**, 22–27. John Wiley & Sons, Ltd.
- GOTELLI, N.J. & GRAVES, G.R. (1996) *Null models in ecology*. Smithsonian Institution Press, Washington.
- GRENIÉ, M., MOUILLOT, D., VILLÉGER, S., DENELLE, P., TUCKER, C.M., MUNOZ, F. & VIOLLE, C. (2018) Functional rarity of coral reef fishes at the global scale : Hotspots and challenges for conservation. *Biological Conservation* **226**, 288–299.

- GRIEVES, C. (2017) Software Review Collaboration with rOpenSci. *methods.blog*.
- GRIME, J.P. (1998) Benefits of plant diversity to ecosystems : Immediate, filter and founder effects. *Journal of Ecology* **86**, 902–910.
- HAEGEMAN, B. & LOREAU, M. (2011) A mathematical synthesis of niche and neutral theories in community ecology. *Journal of Theoretical Biology* **269**, 150–165.
- HAMPTON, S.E., ANDERSON, S.S., BAGBY, S.C., GRIES, C., HAN, X., HART, E.M., JONES, M.B., LENHARDT, W.C., MACDONALD, A., MICHENER, W.K., MUDGE, J., POURMOKHTARIAN, A., SCHILDHAUER, M.P., WOO, K.H. & ZIMMERMAN, N. (2015) The Tao of open science for ecology. *Ecosphere* **6**, art120. John Wiley & Sons, Ltd.
- HANSKI, I. & HANSKI, P. IN THE D. OF E. AND S.I. (1999) *Metapopulation Ecology*. OUP Oxford.
- HOLT, R.D. & KEITT, T.H. (2000) Alternative causes for range limits : A metapopulation perspective. *Ecology letters* **3**, 41–47. Wiley Online Library.
- HORTAL, J., DE BELLO, F., DINIZ-FILHO, J.A.F., LEWINSOHN, T.M., LOBO, J.M. & LADLE, R.J. (2015) Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics* **46**, 523–549.
- HUBBARD, B.A. (2014) Parameter Redundancy with Applications in Statistical Ecology. PhD thesis, University of Kent,.
- HUBBELL, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton.
- HUGHES, T.P., BELLWOOD, D.R., CONNOLLY, S.R., CORNELL, H.V. & KARLSON, R.H. (2014) Double jeopardy and global extinction risk in corals and reef fishes. *Current Biology* **24**, 2946–2951. Elsevier.
- HUTCHINSON, G.E. (1957) Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology* **22**, 415–427.
- IMBERT, A. & VIALANEIX, N. (2018) Décrire, prendre en compte, imputer et évaluer les valeurs manquantes dans les études statistiques : une revue des approches existantes **159**, 55.
- ISAAC, N.J.B., TURVEY, S.T., COLLEN, B., WATERMAN, C. & BAILLIE, J.E.M. (2007) Mammals on the EDGE : Conservation Priorities Based on Threat and Phylogeny. *PLOS ONE* **2**, e296.
- IUCN (2017) *The IUCN Red List of Threatened Species. Version 2017-1*.
- JAIN, M., FLYNN, D.F.B., PRAGER, C.M., HART, G.M., DEVAN, C.M., AHRESTANI, F.S., PALMER, M.I., BUNKER, D.E., KNOPS, J.M.H., JOUSEAU, C.F. & NAEEM, S. (2014) The importance of rare

- species : A trait-based assessment of rare species contributions to functional diversity and possible ecosystem function in tall-grass prairies. *Ecology and Evolution* **4**, 104–112.
- JETZ, W., THOMAS, G.H., JOY, J.B., REDDING, D.W., HARTMANN, K. & MOOERS, A.O. (2014) Global Distribution and Conservation of Evolutionary Distinctness in Birds. *Current Biology* **24**, 919–930.
- JIA, X., DINI-ANDREOTE, F. & SALLES, J.F. (2018) Community Assembly Processes of the Microbial Rare Biosphere. *Trends in Microbiology* **26**, 738–747. Elsevier.
- JOHNSON, D.M. (2004) Sourcesink dynamics in a temporally heterogeneous environment. *Ecology* **85**, 2037–2045. John Wiley & Sons, Ltd.
- JOLY, A., BONNET, P., GOËAU, H., BARBE, J., SELMI, S., CHAMP, J., DUFOUR-KOWALSKI, S., AFFOUARD, A., CARRÉ, J., MOLINO, J.-F., BOUJEMAA, N. & BARTHÉLÉMY, D. (2016) A look inside the Pl@ntNet experience. *Multimedia Systems* **22**, 751–766. Springer Berlin Heidelberg.
- JOUSSET, A., BIENHOLD, C., CHATZINOTAS, A., GALLIEN, L., GOBET, A., KURM, V., KÜSEL, K., RILLIG, M.C., RIVETT, D.W., SALLES, J.F., VAN DER HEIJDEN, M.G.A., YOUSSEF, N.H., ZHANG, X., WEI, Z. & HOL, W.H.G. (2017) Where less may be more : How the rare biosphere pulls ecosystems strings. *The ISME Journal* **11**, 853–862.
- KATTGE, J., BÖNISCH, G., DÍAZ, S., LAVOREL, S., PRENTICE, I.C., LEADLEY, P., TAUTENHAHN, S., WERNER, G.D.A., AAKALA, T., ABEDI, M., ACOSTA, A.T.R., ADAMIDIS, G.C., ADAMSON, K., AIBA, M., ALBERT, C.H., et al. (2020) TRY plant trait database enhanced coverage and open access. *Global Change Biology* **26**, 119–188. John Wiley & Sons, Ltd.
- KATTGE, J., DÍAZ, S., LAVOREL, S., PRENTICE, I.C., LEADLEY, P., BÖNISCH, G., GARNIER, E., WESTOBY, M., REICH, P.B., WRIGHT, I.J., CORNELISSEN, J.H.C., VIOLLE, C., HARRISON, S.P., VAN BODEGOM, P.M., REICHSTEIN, M., et al. (2011) TRY a global database of plant traits. *Global Change Biology* **17**, 2905–2935.
- KATZ, D.S., MCINNES, L.C., BERNHOLDT, D.E., MAYES, A.C., HONG, N.P.C., DUCKLES, J., GESING, S., HEROUX, M.A., HETTRICK, S., JIMENEZ, R.C., PIERCE, M., WEAVER, B. & WILKINS-DIEHR, N. (2018) Community Organizations : Changing the Culture in Which Research Software Is Developed and Sustained. *Computing in Science & Engineering*, 1–1.
- KATZ, D.S., NIEMEYER, K.E. & SMITH, A.M. (2018) Publish your software : Introducing the Journal of Open Source Software (JOSS). *Computing in Science Engineering* **20**, 84–88.
- KEDDY, P.A. (1992) Assembly and response rules : Two goals for predictive community ecology. *Journal of Vegetation Science* **3**, 157–164.

- KEIL, P. & CHASE, J.M. (2019) Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution* **3**, 390–399.
- KNEITEL, J.M. & CHASE, J.M. (2004) Trade-offs in community ecology : Linking spatial scales and species coexistence. *Ecology Letters* **7**, 69–80.
- KONDRATYEVA, A., GRANDCOLAS, P. & PAVOINE, S. (2019) Reconciling the concepts and measures of diversity, rarity and originality in ecology and evolution. *Biological Reviews* **94**, 1317–1337.
- KRAFT, N.J.B., ADLER, P.B., GODOY, O., JAMES, E.C., FULLER, S. & LEVINE, J.M. (2015) Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology* **29**, 592–599.
- KRAFT, N.J.B., CORNWELL, W.K., WEBB, C.O. & ACKERLY, D.D. (2007) Trait Evolution, Community Assembly, and the Phylogenetic Structure of Ecological Communities. *The American Naturalist* **170**, 271–283.
- KRAFT, N.J.B., GODOY, O. & LEVINE, J.M. (2015) Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences* **112**, 797–802.
- LADLE, R.J. & WHITTAKER, R.J. (2011) *Conservation biogeography*. John Wiley & Sons.
- LAMOTHE, K.A., ALOFS, K.M. & CHU, C. (2019) Evaluating functional diversity conservation for freshwater fishes resulting from terrestrial protected areas. *Freshwater Biology* **64**, 2057–2070. John Wiley & Sons, Ltd.
- LAUGHLIN, D.C. (2014) The intrinsic dimensionality of plant traits and its relevance to community assembly. *Journal of Ecology* **102**, 186–193.
- LAUGHLIN, D.C. & JOSHI, C. (2015) Theoretical consequences of trait-based environmental filtering for the breadth and shape of the niche : New testable hypotheses generated by the TraitSpace model. *Ecological Modelling* **307**, 10–21.
- LAUGHLIN, D.C., JOSHI, C., RICHARDSON, S.J., PELTZER, D.A., MASON, N.W.H. & WARDLE, D.A. (2015) Quantifying multimodal trait distributions improves trait-based predictions of species abundances and functional diversity. *Journal of Vegetation Science* **26**, 46–57. John Wiley & Sons, Ltd.
- LAUGHLIN, D.C. & MESSIER, J. (2015) Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends in Ecology & Evolution* **30**, 487–496.
- LAUGHLIN, D.C., STRAHAN, R.T., ADLER, P.B. & MOORE, M.M. (undated) Survival rates indicate that correlations between community-weighted mean traits and environments can be unreliable estimates of the adaptive value of traits. *Ecology Letters* **21**, 411–421.

- LAVOREL, S. & GARNIER, E. (2002) Predicting changes in community composition and ecosystem functioning from plant traits : Revisiting the Holy Grail. *Functional Ecology* **16**, 545–556.
- LAWTON, J.H. (1999) Are There General Laws in Ecology? *Oikos* **84**, 177.
- LE BAGOUSSE-PINGUET, Y., GROSS, N., MAESTRE, F.T., MAIRE, V., DE BELLO, F., FONSECA, C.R., KATTGE, J., VALENCIA, E., LEPS, J. & LIANCOURT, P. (2017) Testing the environmental filtering concept in global drylands. *Journal of Ecology* **105**, 1058–1069. John Wiley & Sons, Ltd.
- LEIBOLD, M.A. & MCPEEK, M.A. (2006) Coexistence of the Niche and Neutral Perspectives in Community Ecology. *Ecology* **87**, 1399–1410.
- LEIBOLD, M.A. & MIKKELSON, G.M. (2002) Coherence, species turnover, and boundary clumping : Elements of meta-community structure. *Oikos* **97**, 237–250.
- LEITÃO, R.P., ZUANON, J., VILLÉGER, S., WILLIAMS, S.E., BARALOTO, C., FORTUNEL, C., MENDONÇA, F.P. & MOUILLOT, D. (2016) Rare species contribute disproportionately to the functional structure of species assemblages. *Proceedings of the Royal Society B : Biological Sciences* **283**, 20160084.
- LEVINS, R. (1969) Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the Entomological Society of America* **15**, 237–240.
- LHOTSKY, B., KOVÁCS, B., ÓNODI, G., CSECSEKITS, A., RÉDEI, T., LENGYEL, A., KERTÉSZ, M. & BOTTA-DUKÁT, Z. (2016) Changes in assembly rules along a stress gradient from open dry grasslands to wetlands. *Journal of Ecology* **104**, 507–517.
- LITCHMAN, E. & KLAUSMEIER, C.A. (2008) Trait-Based Community Ecology of Phytoplankton. *Annual Review of Ecology, Evolution, and Systematics* **39**, 615–639.
- LORANGER, J., MUNOZ, F., SHIPLEY, B. & VIOLLE, C. (2018) What makes traitabundance relationships when both environmental filtering and stochastic neutral dynamics are at play? *Oikos* **127**, 1735–1745. John Wiley & Sons, Ltd.
- LOREAU, M., MOUQUET, N. & GONZALEZ, A. (2003) Biodiversity as spatial insurance in heterogeneous landscapes. *Proceedings of the National Academy of Sciences* **100**, 12765–12770. National Academy of Sciences.
- LORTIE, C.J., BROOKER, R.W., CHOLER, P., KIKVIDZE, Z., MICHALET, R., PUGNAIRE, F.I. & CALLAWAY, R.M. (2004) Rethinking plant community theory. *Oikos* **107**, 433–438.
- LOTKA, A.J. (1927) Fluctuations in the abundance of a species considered mathematically. *Nature* **119**, 12–12. Nature Publishing Group.

- LOW-DÉCARIE, E., KOLBER, M., HOMME, P., LOFANO, A., DUMBRELL, A., GONZALEZ, A. & BELL, G. (2015) Community rescue in experimental metacommunities. *Proceedings of the National Academy of Sciences* **112**, 14307–14312.
- LUCK, G.W., LAVOREL, S., MCINTYRE, S. & LUMB, K. (2012) Improving the application of vertebrate trait-based frameworks to the study of ecosystem services. *Journal of Animal Ecology* **81**, 1065–1076. John Wiley & Sons, Ltd.
- LUIZ, O.J., WOODS, R.M., MADIN, E.M.P. & MADIN, J.S. (2016) Predicting IUCN Extinction Risk Categories for the World's Data Deficient Groupers (Teleostei : Epinephelidae). *Conservation Letters* **9**, 342–350.
- LYONS, K.G., BRIGHAM, C.A., TRAUT, B.H. & SCHWARTZ, M.W. (2005) Rare Species and Ecosystem Functioning. *Conservation Biology* **19**, 1019–1024.
- MACARTHUR, R. & LEVINS, R. (1967) The Limiting Similarity, Convergence, and Divergence of Co-existing Species. *The American Naturalist* **101**, 377–385. The University of Chicago Press.
- MACE, G.M., GITTLEMAN, J.L. & PURVIS, A. (2003) Preserving the Tree of Life. *Science* **300**, 1707–1709. American Association for the Advancement of Science.
- MADIN, J.S., ANDERSON, K.D., ANDREASEN, M.H., BRIDGE, T.C.L., CAIRNS, S.D., CONNOLLY, S.R., DARLING, E.S., DIAZ, M., FALSTER, D.S., FRANKLIN, E.C., GATES, R.D., HOOGENBOOM, M.O., HUANG, D., KEITH, S.A., KOSNIK, M.A., et al. (2016) The Coral Trait Database, a curated database of trait information for coral species from the global oceans. *Scientific Data* **3**, sdata201617.
- MALATERRE, C., DUSSAULT, A.C., ROUSSEAU-MERMANS, S., BARKER, G., BEISNER, B.E., BOUCHARD, F., DESJARDINS, E., HANDA, I.T., KEMBEL, S.W., LAJOIE, G., MARIS, V., MUNSON, A.D., ODENBAUGH, J., POISOT, T., SHAPIRO, B.J., et al. (2019) Functional Diversity : An Epistemic Roadmap. *BioScience* **69**, 800–811. Oxford Academic.
- MARCHESE, C. (2015) Biodiversity hotspots : A shortcut for a more complicated concept. *Global Ecology and Conservation* **3**, 297–309.
- MARGULES, C.R. & PRESSEY, R.L. (2000) Systematic conservation planning. *Nature* **405**, 243–253. Nature Publishing Group.
- MARKS, C.O. & LECHOWICZ, M.J. (2006) Alternative Designs and the Evolution of Functional Diversity. *The American Naturalist* **167**, 55–66.
- MASON, N.W.H., DE BELLO, F., MOUILLOT, D., PAVOINE, S. & DRAY, S. (2013) A guide for using functional diversity indices to reveal changes in assembly processes along ecological gradients. *Journal of Vegetation Science* **24**, 794–806.

- MAYFIELD, M.M. & LEVINE, J.M. (2010) Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters* **13**, 1085–1093.
- MAZEL, F., GUILHAUMON, F., MOUQUET, N., DEVICTOR, V., GRAVEL, D., RENAUD, J., CIANCIA-RUSO, M.V., LOYOLA, R., DINIZ-FILHO, J.A.F., MOUILLOT, D. & THUILLER, W. (2014) Multifaceted diversity-area relationships reveal global hotspots of mammalian species, trait and lineage diversity. *Global Ecology and Biogeography* **23**, 836–847.
- MCCARTNEY, P. & JONES, M. (2002) Using XML-encoded metadata as a basis for advanced information systems for ecological research. In *Proc. 6th World Multiconference Systemics, Cybernetics and Informatics* pp. 379–384.
- MCDONALD, R.I. (2009) The promise and pitfalls of systematic conservation planning. *Proceedings of the National Academy of Sciences* **106**, 15101–15102.
- MCGILL, B.J. (2019) The what, how and why of doing macroecology. *Global Ecology and Biogeography* **28**, 6–17.
- MCGILL, B.J., ENQUIST, B.J., WEIHER, E. & WESTOBY, M. (2006) Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution* **21**, 178–185.
- MEYER, C., WEIGELT, P. & KREFT, H. (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, n/a–n/a.
- MISLAN, K.A.S., HEER, J.M. & WHITE, E.P. (2016) Elevating The Status of Code in Ecology. *Trends in Ecology & Evolution* **31**, 4–7.
- MITTELBACH, G.G. & SCHEMSKE, D.W. (2015) Ecological and evolutionary perspectives on community assembly. *Trends in Ecology & Evolution* **30**, 241–247.
- MORAIS, A.R., SIQUEIRA, M.N., LEMES, P., MACIEL, N.M., DE MARCO, P. & BRITO, D. (2013) Unraveling the conservation status of Data Deficient species. *Biological Conservation* **166**, 98–102.
- MOUCHET, M.A., VILLÉGER, S., MASON, N.W.H. & MOUILLOT, D. (2010) Functional diversity measures : An overview of their redundancy and their ability to discriminate community assembly rules : Functional diversity measures. *Functional Ecology* **24**, 867–876.
- MOUILLOT, D., BELLWOOD, D.R., BARALOTO, C., CHAVE, J., GALZIN, R., HARMELIN-VIVIEN, M., KULBICKI, M., LAVERGNE, S., LAVOREL, S., MOUQUET, N., PAINE, C.E.T., RENAUD, J. & THUILLER, W. (2013) Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. *PLOS Biology* **11**, e1001569.
- MOUQUET, N., DEVICTOR, V., MEYNARD, C.N., MUNOZ, F., BERSIER, L.-F., CHAVE, J., COUTERON, P., DALECKY, A., FONTAINE, C., GRAVEL, D., HARDY, O.J., JABOT, F., LAVERGNE, S., LEI-

- BOLD, M., MOUILLOT, D., et al. (2012) Ecophylogenetics : Advances and perspectives. *Biological Reviews* **87**, 769–785.
- MUNGUÍA, M., PETERSON, A.T. & SÁNCHEZ-CORDERO, V. (2008) Dispersal limitation and geographical distributions of mammal species. *Journal of Biogeography* **35**, 1879–1887.
- MUNOZ, F., GRENIÉ, M., DENELLE, P., TAUDIERE, A., LAROCHE, F., TUCKER, C. & VIOLLE, C. (2017) Ecolottery : Simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution* **in press**. John Wiley & Sons, Ltd.
- MÜNKEMÜLLER, T. & GALLIEN, L. (2015) VirtualCom : A simulation model for eco-evolutionary community assembly and invasion. *Methods in Ecology and Evolution* **6**, 735–743.
- MÜNKEMÜLLER, T., LAVERGNE, S., BZEZNIK, B., DRAY, S., JOMBART, T., SCHIFFERS, K. & THUILLER, W. (2012) How to measure and test phylogenetic signal : How to measure and test phylogenetic signal. *Methods in Ecology and Evolution* **3**, 743–756.
- MYERS, N., MITTERMEIER, R.A., MITTERMEIER, C.G., DA FONSECA, G.A.B. & KENT, J. (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858.
- NAKAGAWA, S. & FRECKLETON, R.P. (2011) Model averaging, missing data and multiple imputation : A case study for behavioural ecology. *Behavioral Ecology and Sociobiology* **65**, 103–116. Springer-Verlag.
- NEMERGUT, D.R., SCHMIDT, S.K., FUKAMI, T., O'NEILL, S.P., BILINSKI, T.M., STANISH, L.F., KNELMAN, J.E., DARCY, J.L., LYNCH, R.C., WICKEY, P. & FERRENBURG, S. (2013) Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews* **77**, 342–356.
- OLIVEIRA, B.F., SÃO-PEDRO, V.A., SANTOS-BARRERA, G., PENONE, C. & COSTA, G.C. (2017) AmphiBIO, a global database for amphibian ecological traits. *Scientific Data* **4**, 170123.
- PARSONS, P.A. (1990) The metabolic cost of multiple environmental stresses : Implications for climatic change and conservation. *Trends in Ecology & Evolution* **5**, 315–317.
- PAVOINE, S., BONSALE, M.B., DUPAIX, A., JACOB, U. & RICOTTA, C. (2017) From phylogenetic to functional originality : Guide through indices and new developments. *Ecological Indicators* **82**, 196–205.
- PAVOINE, S., OLLIER, S. & DUFOUR, A.-B. (2005) Is the originality of a species measurable? *Ecology Letters* **8**, 579–586.
- PECK, S.L. (2004) Simulation as experiment : A philosophical reassessment for biological modeling. *Trends in Ecology & Evolution* **19**, 530–534.

- PENG, R. (2015) The reproducibility crisis in science : A statistical counterattack. *Significance* **12**, 30–32.
- PENONE, C., DAVIDSON, A.D., SHOEMAKER, K.T., DI MARCO, M., RONDININI, C., BROOKS, T.M., YOUNG, B.E., GRAHAM, C.H. & COSTA, G.C. (2014) Imputation of missing data in life-history trait datasets : Which approach performs the best? *Methods in Ecology and Evolution* **5**, 961–970. John Wiley & Sons, Ltd.
- PIANKA, E.R., VITT, L.J., PELEGRIN, N., FITZGERALD, D.B. & WINEMILLER, K.O. (2017) Toward a Periodic Table of Niches, or Exploring the Lizard Niche Hypervolume. *The American Naturalist* **190**, 601–616.
- PICCOLO, S.R. & FRAMPTON, M.B. (2016) Tools and techniques for computational reproducibility. *GigaScience* **5**. Oxford Academic.
- PISTÓN CABALLERO, N., BELLO, F. DE, DIAS, A.T.C., GÖTZENBERGER, L., ROSADO, B.H.P., MATOS, E.A. DE, SALGUERO-GÓMEZ, R. & CARMONA, C.P. (2019) Multidimensional ecological analyses demonstrate how interactions between functional traits shape fitness and life history strategies. John Wiley & Sons.
- POISOT, T. (2015) Best publishing practices to improve user confidence in scientific software. *Ideas in Ecology and Evolution* **8**.
- POLLOCK, L.J., THUILLER, W. & JETZ, W. (2017) Large conservation gains possible for global biodiversity facets. *Nature* **546**, 141–144.
- PONTARP, M., BRÄNNSTRÖM, Å. & PETCHEY, O.L. (2019) Inferring community assembly processes from macroscopic patterns using dynamic eco-evolutionary models and Approximate Bayesian Computation (ABC). *Methods in Ecology and Evolution* **10**, 450–460. John Wiley & Sons, Ltd.
- PORTER, J.H. (2018) Scientific Databases for Environmental Research. In *Ecological Informatics* (eds F. RECKNAGEL & W.K. MICHENER), pp. 27–53. Springer International Publishing, Cham.
- PRENDERGAST, J.R., QUINN, R.M., LAWTON, J.H., EVERS HAM, B.C. & GIBBONS, D.W. (1993) Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* **365**, 335–337.
- PRESTON, F.W. (1948) The Commonness, And Rarity, of Species. *Ecology* **29**, 254.
- PRESTON, F.W. (1962) The Canonical Distribution of Commonness and Rarity : Part I. *Ecology* **43**, 185–215. John Wiley & Sons, Ltd.
- PULLIAM, H.R. (1988) Sources, Sinks, and Population Regulation. *The American Naturalist* **132**, 652–661.

- PULLIAM, H.R. & DANIELSON, B.J. (1991) Sources, Sinks, and Habitat Selection : A Landscape Perspective on Population Dynamics. *The American Naturalist* **137**, S50–S66.
- RABINOWITZ, D. (1981) Seven forms of rarity. *Biological aspects of rare plant conservation*.
- RAES, J., LETUNIC, I., YAMADA, T., JENSEN, L.J. & BORK, P. (2011) Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Molecular Systems Biology* **7**, 473. John Wiley & Sons, Ltd.
- RAEVEL, V., ANTHELME, F., MENESES, R.I. & MUNOZ, F. (2018) Cushion-plant protection determines guild-dependent plant strategies in high-elevation peatlands of the Cordillera Real, Bolivian Andes. *Perspectives in Plant Ecology, Evolution and Systematics* **30**, 103–114.
- RAPACCIUOLO, G. & BLOIS, J.L. (2019) Understanding ecological change across large spatial, temporal and taxonomic scales : Integrating data and methods in light of theory. *Ecography* **42**, 1247–1266.
- REDDING, D.W., DEWOLFF, C. & MOOERS, A.Ø. (2010) Evolutionary distinctiveness, threat status, and ecological oddity in primates. *Conservation Biology* **24**, 1052–1058.
- REDDING, D.W. & MOOERS, A.Ø. (2006) Incorporating Evolutionary Measures into Conservation Prioritization. *Conservation Biology* **20**, 1670–1678.
- REICH, P.B. (2014) The world-wide “fastslow” plant economics spectrum : A traits manifesto. *Journal of Ecology* **102**, 275–301.
- RELYEA, R., RICKLEFS, R.E. & LEMPÉRIÈRE, G. (2019) *Écologie : l'économie de la nature*.
- RICHARDSON, D.M. & WHITTAKER, R.J. (2010) Conservation biogeography foundations, concepts and challenges. *Diversity and Distributions* **16**, 313–320. John Wiley & Sons, Ltd.
- RICOTTA, C., BELLO, F. DE, MORETTI, M., CACCIANIGA, M., CERABOLINI, B.E.L. & PAVOINE, S. (2016) Measuring the functional redundancy of biological communities : A quantitative guide. *Methods in Ecology and Evolution* **7**, 1386–1395.
- RODRIGUES, A.S. & GASTON, K.J. (2002) Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biological Conservation* **105**, 103–111. Elsevier.
- ROLHAUSER, A.G. & PUCHETA, E. (2017) Directional, stabilizing, and disruptive trait selection as alternative mechanisms for plant community assembly. *Ecology* **98**, 668–677.
- ROSCHER, C., SCHUMACHER, J., LIPOWSKY, A., GUBSCH, M., WEIGELT, A., POMPE, S., KOLLE, O., BUCHMANN, N., SCHMID, B. & SCHULZE, E.-D. (2013) A functional trait-based approach to understand community assembly and diversity-productivity relationships over 7 years in experimental grasslands. *Perspectives in Plant Ecology, Evolution and Systematics* **15**, 139–149.

- ROSINDELL, J., HUBBELL, S.P. & ETIENNE, R.S. (2011) The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. *Trends in Ecology & Evolution* **26**, 340–348.
- SALGUERO-GÓMEZ, R., VIOLLE, C., GIMENEZ, O. & CHILDS, D. (2018) Delivering the promises of trait-based approaches to the needs of demographic approaches, and vice versa. *Functional Ecology* **32**, 1424–1435. John Wiley & Sons, Ltd.
- SCHLEUTER, D., DAUFRESNE, M., MASSOL, F. & ARGILLIER, C. (2010) A user's guide to functional diversity indices. *Ecological Monographs* **80**, 469–484.
- SCHNEIDER, F.D., FICHTMUELLER, D., GOSSNER, M.M., GÜNTSCH, A., JOCHUM, M., KÖNIG-RIES, B., LE PROVOST, G., MANNING, P., OSTROWSKI, A., PENONE, C. & SIMONS, N.K. (2019) Towards an ecological trait-data standard. *Methods in Ecology and Evolution* **10**, 2006–2019. John Wiley & Sons, Ltd.
- SCHRODT, F., KATTGE, J., SHAN, H., FAZAYELI, F., JOSWIG, J., BANERJEE, A., REICHSTEIN, M., BÖNISCH, G., DÍAZ, S., DICKIE, J., GILLISON, A., KARPATNE, A., LAVOREL, S., LEADLEY, P., WIRTH, C.B., et al. (2015) BHPMF a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecology and Biogeography* **24**, 1510–1521.
- SEGRESTIN, J., NAVAS, M.-L. & GARNIER, E. (2020) Reproductive phenology as a dimension of the phenotypic space in 139 plant species from the Mediterranean. *New Phytologist* **225**, 740–753.
- SHIPLEY, B. (2010) *From plant traits to vegetation structure : Chance and selection in the assembly of ecological communities*. Cambridge University Press.
- SHIPLEY, B., BELLO, F.D., CORNELISSEN, J.H.C., LALIBERTÉ, E., LAUGHLIN, D.C. & REICH, P.B. (2016) Reinforcing loose foundation stones in trait-based plant ecology. *Oecologia*, 1–9.
- SHIPLEY, B., VILE, D. & GARNIER, É. (2006) From Plant Traits to Plant Communities : A Statistical Mechanistic Approach to Biodiversity. *Science* **314**, 812–814. American Association for the Advancement of Science.
- SINCLAIR, S.P., MILNER-GULLAND, E.J., SMITH, R.J., MCINTOSH, E.J., POSSINGHAM, H.P., VERCAMMEN, A. & KNIGHT, A.T. (2018) The use, and usefulness, of spatial conservation prioritizations. *Conservation Letters* **11**, e12459. John Wiley & Sons, Ltd.
- SOLAN, M. (2004) Extinction and Ecosystem Function in the Marine Benthos. *Science* **306**, 1177–1180.
- SOUTHWOOD, T.R.E. (1987) The concept and nature of the community. *Organization of Communities Past and Present*, 3–28. Blackwell Sci. Publ.

- STUART-SMITH, R.D., BATES, A.E., LEFCHECK, J.S., DUFFY, J.E., BAKER, S.C., THOMSON, R.J., STUART-SMITH, J.F., HILL, N.A., KININMONTH, S.J., AIROLDI, L., BECERRO, M.A., CAMPBELL, S.J., DAWSON, T.P., NAVARRETE, S.A., SOLER, G.A., et al. (2013) Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* **501**, 539–542.
- SULLIVAN, B.L., WOOD, C.L., ILIFF, M.J., BONNEY, R.E., FINK, D. & KELLING, S. (2009) eBird : A citizen-based bird observation network in the biological sciences. *Biological Conservation* **142**, 2282–2292.
- SVENNING, J.-C., EISERHARDT, W.L., NORMAND, S., ORDONEZ, A. & SANDEL, B. (2015) The Influence of Paleoclimate on Present-Day Patterns in Biodiversity and Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* **46**, 551–572.
- SVENNING, J.-C. & SKOV, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters* **7**, 565–573.
- SWENSON, N.G., ENQUIST, B.J., PITHER, J., KERKHOFF, A.J., BOYLE, B., WEISER, M.D., ELSE, J.J., FAGAN, W.F., FORERO-MONTAÑA, J., FYLLAS, N., KRAFT, N.J.B., LAKE, J.K., MOLES, A.T., PARTIÑO, S., PHILLIPS, O.L., et al. (2012) The biogeography and filtering of woody plant functional diversity in North and South America. *Global Ecology and Biogeography* **21**, 798–808.
- SWENSON, N.G. & JONES, F.A. (2017) Community transcriptomics, genomics and the problem of species co-occurrence. *Journal of Ecology* **105**, 563–568. John Wiley & Sons, Ltd.
- TASCHUK, M. & WILSON, G. (2017) Ten simple rules for making research software more robust. *PLOS Computational Biology* **13**, e1005412.
- TAUGOURDEAU, S., VILLERD, J., PLANTUREUX, S., HUGUENIN-ELIE, O. & AMIAUD, B. (2014) Filling the gap in functional trait databases : Use of ecological hypotheses to replace missing data. *Ecology and Evolution* **4**, 944–958. John Wiley & Sons, Ltd.
- THOMSON, F.J., MOLES, A.T., AULD, T.D. & KINGSFORD, R.T. (2011) Seed dispersal distance is more strongly correlated with plant height than with seed mass. *Journal of Ecology* **99**, 1299–1307.
- TOUCHON, J.C. & MCCOY, M.W. (2016) The mismatch between current statistical practice and doctoral training in ecology. *Ecosphere* **7**, n/a–n/a. John Wiley & Sons, Ltd.
- TUCKER, C.M., AZE, T., CADOTTE, M.W., CANTALAPIEDRA, J.L., CHISHOLM, C., DÍAZ, S., GRENYER, R., HUANG, D., MAZEL, F., PEARSE, W.D., PENNELL, M.W., WINTER, M. & MOOERS, A.O. (2019) Assessing the utility of conserving evolutionary history. *Biological Reviews* **94**, 1740–1760.
- TUCKER, C.M., DAVIES, T.J., CADOTTE, M.W. & PEARSE, W.D. (2018) On the relationship between phylogenetic diversity and trait diversity. *Ecology* **99**, 1473–1479.

- VAN DER PLAS, F., JANZEN, T., ORDONEZ, A., FOKKEMA, W., REINDERS, J., ETIENNE, R.S. & OLFF, H. (2015) A new modeling approach estimates the relative importance of different community assembly processes. *Ecology* **96**, 1502–1515.
- VANE-WRIGHT, R.I., HUMPHRIES, C.J. & WILLIAMS, P.H. (1991) What to protect? Systematics and the agony of choice. *Biological Conservation* **55**, 235–254.
- VIÉ, J.-C., HILTON-TAYLOR, C., POLLOCK, C., RAGLE, J., SMART, J., STUART, S.N. & TONG, R. (2009) The IUCN Red List : A key conservation tool. *Wildlife in a changing world An analysis of the 2008 IUCN Red List of Threatened Species*, 1.
- VILLÉGER, S., BROSSE, S., MOUCHET, M., MOUILLOT, D. & VANNI, M.J. (2017) Functional ecology of fish : Current approaches and future challenges. *Aquatic Sciences* **79**, 783–801.
- VIOLLE, C., BORGY, B. & CHOLER, P. (2015) Trait databases : Misuses and precautions. *Journal of Vegetation Science* **26**, 826–827.
- VIOLLE, C., ENQUIST, B.J., MCGILL, B.J., JIANG, L., ALBERT, C.H., HULSHOF, C., JUNG, V. & MESSIER, J. (2012) The return of the variance : Intraspecific variability in community ecology. *Trends in Ecology & Evolution* **27**, 244–252.
- VIOLLE, C., NAVAS, M.-L., VILE, D., KAZAKOU, E., FORTUNEL, C., HUMMEL, I. & GARNIER, E. (2007) Let the concept of trait be functional! *Oikos* **116**, 882–892.
- VIOLLE, C., REICH, P.B., PACALA, S.W., ENQUIST, B.J. & KATTGE, J. (2014) The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences* **111**, 13690–13696.
- VIOLLE, C., THUILLER, W., MOUQUET, N., MUNOZ, F., KRAFT, N.J.B., CADOTTE, M.W., LIVINGSTONE, S.W. & MOUILLOT, D. (2017) Functional Rarity : The Ecology of Outliers. *Trends in Ecology & Evolution* **32**, 356–367.
- VOLTERRA, V. (1928) Variations and Fluctuations of the Number of Individuals in Animal Species living together. *ICES Journal of Marine Science* **3**, 3–51.
- VON HUMBOLDT, A. (1877) *Personal narrative of travels to the equinoctial regions of the new continent during the years 1799-1804*. G. Bell.
- WEBB, C.O., ACKERLY, D.D., MCPEEK, M.A. & DONOGHUE, M.J. (2002) Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics* **33**, 475–505.
- WEBB, NULL (2000) Exploring the Phylogenetic Structure of Ecological Communities : An Example for Rain Forest Trees. *The American Naturalist* **156**, 145–155.

- WEIHER, E. & KEDDY, P.A. (1995) Assembly Rules, Null Models, and Trait Dispersion : New Questions from Old Patterns. *Oikos* **74**, 159.
- WESTOBY, M. (1998) A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil* **199**, 213–227.
- WILLIAMS, N.M., CRONE, E.E., ROULSTON, T.H., MINCKLEY, R.L., PACKER, L. & POTTS, S.G. (2010) Ecological and life-history traits predict bee species responses to environmental disturbances. *Biological Conservation* **143**, 2280–2291.
- WILSON, G. (2016) Software Carpentry : Lessons learned. *1000Research*.
- WILSON, G., ARULIAH, D.A., BROWN, C.T., CHUE HONG, N.P., DAVIS, M., GUY, R.T., HADDOCK, S.H.D., HUFF, K.D., MITCHELL, I.M., PLUMBLEY, M.D., WAUGH, B., WHITE, E.P. & WILSON, P. (2014) Best Practices for Scientific Computing. *PLoS Biol* **12**, e1001745.
- WILSON, G., BRYAN, J., CRANSTON, K., KITZES, J., NEDERBRAGT, L. & TEAL, T.K. (2017) Good enough practices in scientific computing. *PLOS Computational Biology* **13**, e1005510. Public Library of Science.
- WINEMILLER, K.O., FITZGERALD, D.B., BOWER, L.M. & PIANKA, E.R. (2015) Functional traits, convergent evolution, and periodic tables of niches. *Ecology Letters* **18**, 737–751.
- WINTER, M., DEVICTOR, V. & SCHWEIGER, O. (2013) Phylogenetic diversity and nature conservation : Where are we? *Trends in Ecology & Evolution* **28**, 199–204.
- WRIGHT, I.J., REICH, P.B., WESTOBY, M., ACKERLY, D.D., BARUCH, Z., BONGERS, F., CAVENDER-BARES, J., CHAPIN, T., CORNELISSEN, J.H.C., DIEMER, M., FLEXAS, J., GARNIER, E., GROOM, P.K., GULIAS, J., HIKOSAKA, K., et al. (2004) The worldwide leaf economics spectrum. *Nature* **428**, 821–827.
- WÜEST, R.O., ZIMMERMANN, N.E., ZURELL, D., ALEXANDER, J.M., FRITZ, S.A., HOF, C., KREFT, H., NORMAND, S., CABRAL, J.S., SZEKELY, E., THUILLER, W., WIKELSKI, M. & KARGER, D.N. (2020) Macroecology in the age of Big Data Where to go from here? *Journal of Biogeography* **47**, 1–12.
- YACHI, S. & LOREAU, M. (1999) Biodiversity and ecosystem productivity in a fluctuating environment : The insurance hypothesis. *Proceedings of the National Academy of Sciences* **96**, 1463–1468.
- YENNI, G., ADLER, P. & ERNEST, M. (2016) Do persistent rare species experience stronger negative frequency dependence than common species? *bioRxiv*.
- ZAKHAROVA, L., MEYER, K.M. & SEIFAN, M. (2019) Trait-based modelling in ecology : A review of two decades of research. *Ecological Modelling* **407**, 108703.

- ZANNE, A.E., PEARSE, W.D., CORNWELL, W.K., MCGLINN, D.J., WRIGHT, I.J. & UYEDA, J.C. (2018) Functional biogeography of angiosperms : Life at the extremes. *New Phytologist* **218**, 1697–1709.
- ZURELL, D., BERGER, U., CABRAL, J.S., JELTSCH, F., MEYNARD, C.N., MÜNKEMÜLLER, T., NEHRBASS, N., PAGEL, J., REINEKING, B., SCHRÖDER, B. & GRIMM, V. (2010) The virtual ecologist approach : Simulating data and observers. *Oikos* **119**, 622–635.

Résumés

Une hypothèse centrale en écologie fonctionnelle est que certaines combinaisons de traits procurent un taux de croissance maximal dans des conditions environnementales données, et représentent ainsi un optimum fonctionnel. Les valeurs de traits abondantes sont ainsi supposées refléter la meilleure performance des espèces proches de l'optimum. Cependant certains processus, comme les processus neutres, peuvent affecter la présence d'un optimum fonctionnel. Certaines espèces possèdent des combinaisons de traits originales différentes de l'optimum fonctionnel, localement dans les communautés ainsi qu'au sein des bassins d'espèces régionaux. Les objectifs généraux de cette thèse sont (i) de proposer des modèles d'assemblage des communautés prenant en compte les différents processus qui peuvent affecter la présence d'un optimum fonctionnel, (ii) d'identifier des processus précis qui affectent l'optimum fonctionnel à l'aide de ces modèles, (iii) d'identifier les espèces originales fonctionnellement à l'échelle globale et d'étudier leurs déterminants, (iv) d'évaluer la prise en compte de l'originalité fonctionnelle dans la conservation. Dans le premier chapitre nous proposons deux modèles d'assemblages des communautés, l'un couplant processus locaux et régionaux, l'autre s'inspirant des modèles de dynamiques de populations. Dans le deuxième chapitre, nous mettons en avant deux processus qui affectent l'optimum fonctionnel : la saturation d'habitat des espèces et le présence de traits n'étant pas affecté par l'environnement. Dans le troisième chapitre nous proposons un outil pour quantifier l'originalité fonctionnelle, et nous évaluons les relations qui existent entre originalité fonctionnelle régionale et déterminants climatiques. Enfin dans le quatrième chapitre, nous montrons que les espèces rares fonctionnellement sont davantage menacées d'extinction que les autres espèces et que davantage de menaces pèsent sur elles comparées aux espèces communes fonctionnellement. La thèse montre l'importance de la prise en compte des différentes échelles spatiales, temporelles et taxonomiques pour caractériser l'originalité fonctionnelle. La diversité des modèles biologiques utilisés (herbacées, arbres, oiseaux, mammifères et poissons de récifs coralliens) au cours de cette thèse souligne l'intérêt conceptuel de l'originalité fonctionnelle dans des contextes écologiquement variés.

mots-clés : biogéographie, assemblage des communautés, échelles de diversité, traits fonctionnels

A central hypothesis in functional ecology is that certain trait combinations let individuals have a maximum growth rate in certain environmental conditions, they represent as such a functional optimum. The abundant trait values are thus supposed to reflect the best performance of species close to the optimum. However, some processes, such as neutral processes, can affect the presence of a functional optimum. Some species also harbor original trait combination compared to the functional optimum, locally among communities as well as in regional species pools. The general goals of this thesis are (i) to propose community assembly models to test the influence of different processes that can affect the presence of a functional optimum, (ii) to identify specific processes that affect functional optimum by using these models, (iii) to identify globally original species and study their determinants, (iv) to evaluate the importance of originality in a conservation context. In the first chapter we propose to community assembly models, the first couples regional and local processes, while the second is more inspired by population dynamics model. In the second chapter, we identify two processes that affect the presence of a functional optimum: species habitat saturation and traits not affected by the environment. In the third chapter, we propose a statistical tool to quantify functional originality and we evaluate the relationships between regional functional originality and climatic factors. In the fourth chapter, we show that functionally original species are more at threat than other species and that they face more threats than functionally common species. The thesis shows the importance of taking into account multiple spatial, temporal and taxonomic scales to characterize functional originality. The diversity of biological models studied in this thesis (herbs, trees, birds, mammals, and coral reef fishes) underline the conceptual interest of functional originality in diverse ecological contexts.

keywords: biogeography, community assembly, scales of diversity, functional traits