



HAL
open science

**Comprendre et protéger la diversité des mammifères :
une approche de biogéographie évolutive et fonctionnelle
à l'échelle du globe.**

Florent Mazel

► **To cite this version:**

Florent Mazel. Comprendre et protéger la diversité des mammifères : une approche de biogéographie évolutive et fonctionnelle à l'échelle du globe.. *Ecologie, Environnement*. Université Grenoble Alpes, 2015. Français. NNT : 2015GREAV063 . tel-01684747

HAL Id: tel-01684747

<https://theses.hal.science/tel-01684747v1>

Submitted on 15 Jan 2018

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Biodiversité Écologie Environnement**

Arrêté ministériel : 7 août 2006

Présentée par

Florent MAZEL

Thèse dirigée par **Wilfried Thuiller**
codirigée par **Sébastien Lavergne**

préparée au sein du **Laboratoire d'Écologie Alpine**
dans l'**École Doctorale de Chimie et des Sciences du Vivant**

Comprendre et protéger la diversité des mammifères : une approche de biogéographie évolutive et fonctionnelle à l'échelle du globe.

Thèse soutenue publiquement le **14 décembre 2015**,
devant le jury composé de :

M. Miguel ARAÚJO

Professeur à l'Université de Madrid, Rapporteur et **président du jury**.

M. Olivier HARDY

Maître de Recherches à l'Université de Bruxelles, Rapporteur

M. Fabien LEPRIEUR

Maître de conférence à l'Université de Montpellier, Examineur

M. Frédéric DELSUC

Chargé de Recherche à l'Université de Montpellier, Examineur

M. Francesco FICETOLA

Chargé de Recherche à l'Université Grenoble Alpes, Examineur

M. Sébastien LAVERGNE

Chargé de Recherche à l'Université Grenoble Alpes, Co-directeur de
thèse, Invité

M. Wilfried THULLER

Directeur de Recherche à l'Université Grenoble Alpes, Directeur de thèse



À mon père

SOMMAIRE GÉNÉRAL

LISTE DES ARTICLES	5
LISTE DES ANNEXES	6
INTRODUCTION GÉNÉRALE	8
PARTIE I. Classer la diversité des diversités phylogénétiques et fonctionnelles	48
Chapitre I. Tentative de synthèse exhaustive	58
Chapitre II. L'importance de l'échelle évolutive	90
PARTIE II. Utiliser les diversités phylogénétiques et fonctionnelles en conservation	118
Chapitre III. Quantification de l'érosion de la diversité phylogénétique	128
Chapitre IV. Les points chauds de diversité fonctionnelle et phylogénétique	140
PARTIE III. Utiliser les diversités phylogénétiques et fonctionnelles pour comprendre certains patrons biogéographiques	170
<i>SOUS PARTIE III.1</i>	
Chapitre V. L'environnement et l'espace modulent la répartition mondiale des oiseaux et mammifères à différentes échelles phylogénétiques	184
Chapitre VI. Le régime alimentaire et la phylogénie de l'hôte modulent la composition des microbiotes à différentes échelles phylogénétiques	198
<i>SOUS PARTIE III.2</i>	
Chapitre VII. La convergence évolutive des assemblages de mammifères	222
CONCLUSION GÉNÉRALE	254
GLOSSAIRE	258
ANNEXES	262

LISTE DES ARTICLES

PARTIE I.

Article 1. C. M. Tucker, M. W. Cadotte, S. B. Carvalho, J. T. Davies, S. Ferrier, S. A. Fritz, R. Grenyer, M. R. Helmus, L. S. Jin, A. O. Mooers, S. Pavoine, O. Purschke, D. W. Redding, D. F. Rosauer, M. Winter, **F. Mazel**. *A unification of phylogenetic metrics for conservation, community ecology and macroecology*. En révision dans *Biological Reviews*

Article 2. **F. Mazel**, Davies, T.J, Gallien L., Groussin, M., Münkemüller T. & Thuiller, W. *Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics*. Accepté dans *Ecography*

Article 3. D.W. Redding, **F. Mazel**, and. A.O. Mooers. *Measuring evolutionary isolation for conservation*. *PloS one*, 9(12), e113490.

PARTIE II.

Article 4. **F. Mazel**, J. Renaud, F. Guilhaumon, D. Mouillot, D. Gravel & W. Thuiller. *Mammalian phylogenetic diversity area relationships at a continental scale (2015)*. *Ecology*, 96 (10), 2814-2822

Article 5. **F. Mazel**, F. Guilhaumon, N. Mouquet, V. Devictor, D. Gravel, J. Renaud, M. V. Cianciaruso, R.D. Loyola, J.A.F. Diniz-Filho, D. Mouillot and W. Thuiller (2014) *Multifaceted diversity-area relationships reveal global hotspots of mammalian species, trait and lineage diversity*. *GEB*, 23, 836–847

Article 6. W. Thuiller, L. Maiorano, L., **F. Mazel**, F. Guilhaumon, G.F. Ficetola, S. Lavergne, S., J. Renaud, J., C. Roquet and D. Mouillot (2015). *Conserving the functional and phylogenetic trees of life of European tetrapods*. *Phil. Trans. of the Royal Society of London B: Biological Sciences*, 370(1662), 2014.

PARTIE III.

Article 7. F. Mazel, R. Wüest, J. Renaud, S. Lavergne and W. Thuiller. Decomposing the global beta diversity of mammals and birds along evolutionary scale reveal the differential effects of space and contemporary climate. En préparation.

Article 8. M. Groussin*, F. Mazel*, J. Sanders, C. Smillie, S. Lavergne, W. Thuiller and E. Alm. Selecting the bacterial lineages living in guts of mammals. En préparation

Article 9. F. Mazel, T. J. Davies, D. Georges, S. Lavergne, W. Thuiller and P. R. Peres-Neto. *Improving phylogenetic regression under complex evolutionary models.* Accepté dans *Ecology*

Article 10. F. Mazel, R. Wüest, J. Renaud, S. Lavergne and W. Thuiller. *The global biogeography of non-neutral trait evolution in mammals.* En préparation

LISTE DES ANNEXES

PARTIE I.

Annexe 1.1. Matériel supplémentaire de l'article 1

Annexe 1.2. Matériel supplémentaire de l'article 2

Annexe 1.3. Article 3

Annexe 1.4. Résumé d'un article additionnel

PARTIE II.

Annexe 2.1. Matériel supplémentaire de l'article 4

Annexe 2.2. Matériel supplémentaire de l'article 5

Annexe 2.3. Article 6

PARTIE III.

Annexe 3.1.1. Matériel supplémentaire de l'article 7

Annexe 3.1.2. Matériel supplémentaire de l'article 8

Annexe 3.2.1. Article 9

Annexe 3.2.2. Matériel supplémentaire de l'article 10

INTRODUCTION GÉNÉRALE

« ... that grand subject,
that almost keystone of the laws of creation
- Geographical Distribution. »

C. Darwin

Dans une lettre à J.D. Hooker, 1845

1. Les origines du questionnement biogéographique

« On ne connaît pas complètement une science tant qu'on n'en sait pas l'histoire »

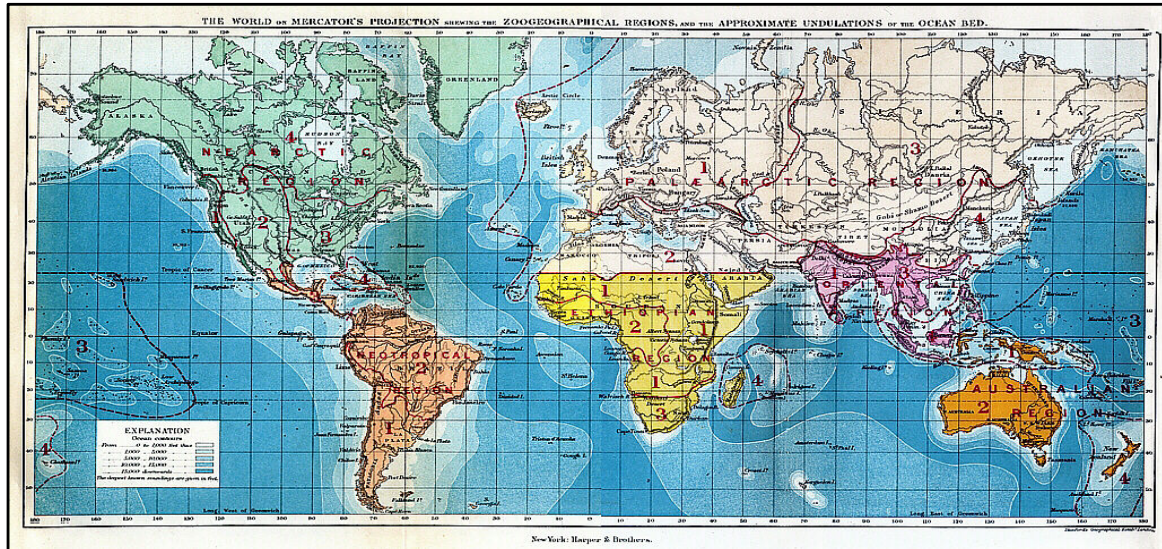
Auguste Comte

Cours de philosophie positive (1830-1842)

Pourquoi des régions géographiques différentes abritent-elles des espèces différentes? Pourquoi certaines en possèdent davantage? Ces questions reflètent les principales problématiques des biogéographes et ont constitué le fil directeur de mon travail de thèse.

La biogéographie décrit et tente de comprendre « la géographie de la nature » (Lomolino et al. 2010), c'est-à-dire la distribution - dans l'espace et dans le temps - des êtres vivant sur terre. Ce vaste objectif n'est pas nouveau. En effet, les populations de chasseurs-cueilleurs du Paléolithique savaient déjà très probablement localiser leur nourriture et leurs prédateurs (Ladle and Whittaker 2011). Au delà de cette origine très lointaine, de nombreux auteurs ont, au cours de l'histoire, décrit la répartition géographique des êtres vivants : c'est par exemple le cas du philosophe grec Théophraste (371 à 228 av. JC) ou encore de celui de l'écrivain et naturaliste romain Pline l'Ancien (23 à 79 ap. JC). De telles descriptions étaient cependant difficilement comparables puisqu'elles ne reposaient pas sur le même système de classification des êtres vivants. Il faudra attendre le XVIII^{ème} siècle et la classification binomiale de Carl von Linné (1707-1778) pour disposer d'une nomenclature solide permettant de mener des études biogéographiques. Un nom d'espèce devra alors se composer de deux mots en latin, le premier donnant le nom de genre auquel l'espèce appartient et le deuxième le nom de l'espèce à l'intérieur du genre (par exemple, les humains appartiennent au genre *Homo* et à l'intérieur de ce genre à l'espèce *Homo sapiens*). C'est finalement sous l'impulsion de grandes campagnes d'exploration biologique à travers le monde (voir les flèches orange de l'encadré 1, page 13) que la biogéographie a connu un essor spectaculaire au XIX^{ème} siècle. Alexander von Humboldt (1769-1859) a par exemple méticuleusement décrit la répartition des plantes sur les flancs du volcan Teide sur l'île de Tenerife ou du Chimborazo en Equateur (Humboldt and Bonpland 1805). Ses études de l'étagement de la végétation ont d'ailleurs fourni des représentations cartographiques emblématiques (voir Figure 1).

1. By Wallace (1876)



2. By Holt *et al.* (2009)

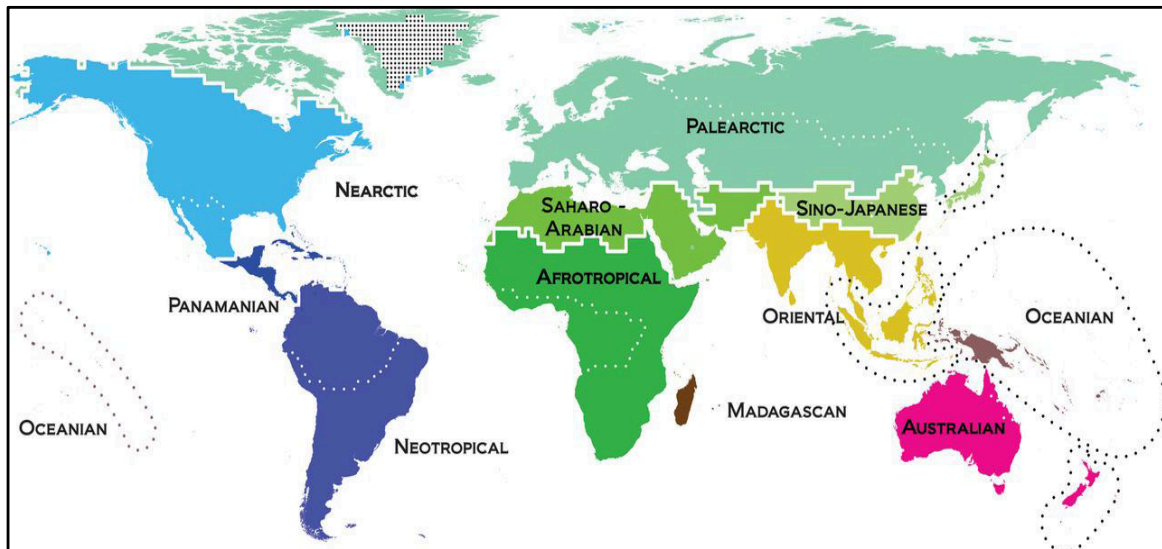


Figure 2. Régions Zoo-géographiques du monde. La figure présente deux versions des grandes provinces biogéographiques du monde basées sur les distributions de certains vertébrés terrestres.

Ses travaux et voyages ont alors inspiré une génération de naturalistes du milieu du XIX^{ème} siècle, notamment Charles Darwin (1809-1882), Alfred Russel Wallace (1823-1913) ou encore Joseph Dalton Hooker (1817-1911). Au cours de nouvelles expéditions scientifiques de plusieurs années, ces derniers ont amassé une quantité considérable de nouvelles données et ont jeté les bases de la biogéographie moderne (Sclater 1858, Darwin 1859, Wallace 1876). Ces découvertes et leurs intégrations à large échelle spatiale ont permis à Alfred Russel Wallace (1876) de décrire deux des plus frappants patrons de la biogéographie mondiale:

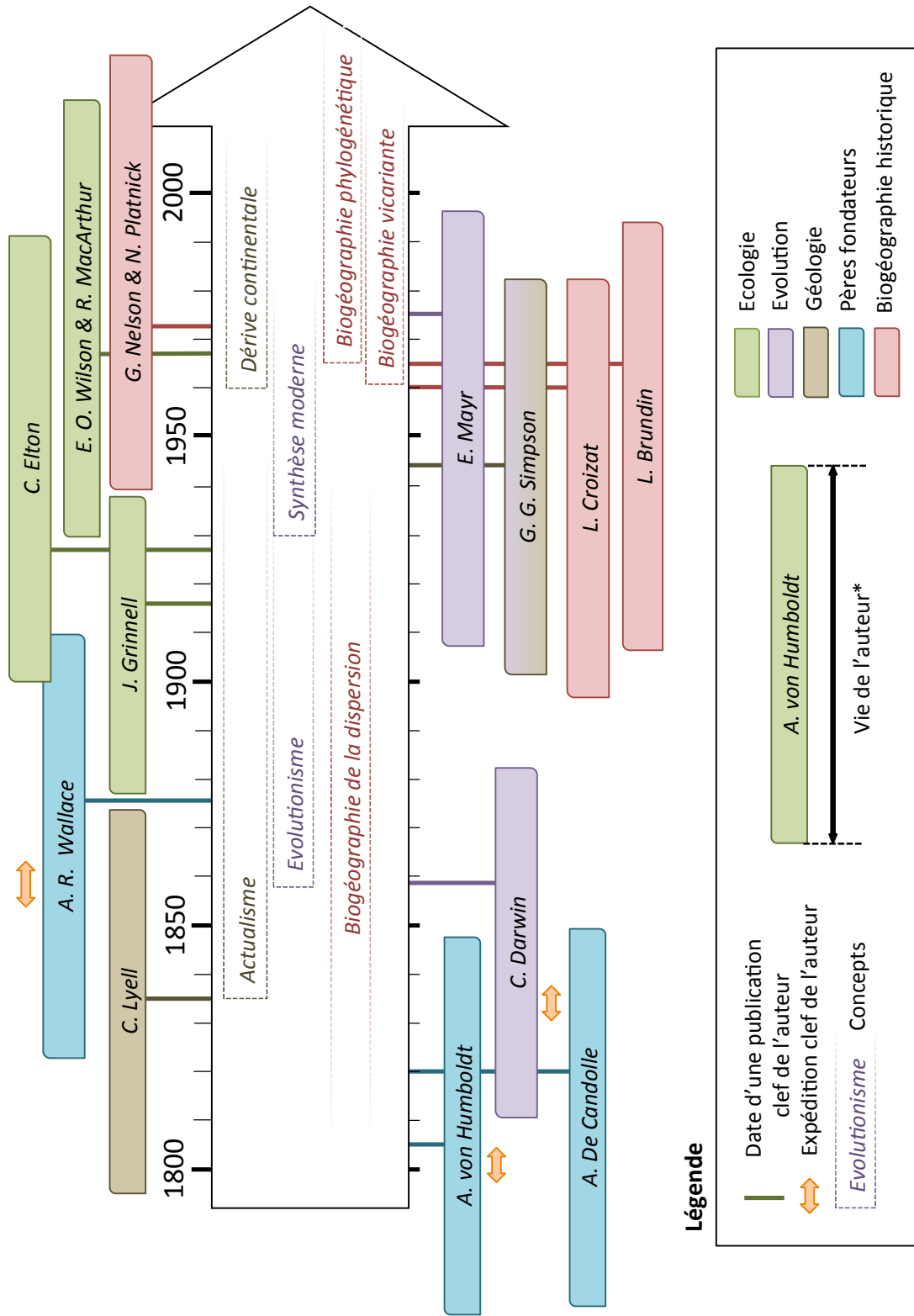
« Animal life is, on the whole, far more abundant and more varied in the tropics than in any other part of the globe »
Wallace, 1876

Le gradient latitudinal de diversité décrit ici par Wallace atteste du fait qu'il y a davantage d'espèces dans les tropiques que dans les régions tempérées. Ce patron, documenté à l'époque pour les animaux, est en fait généralisable à un grand nombre d'êtres vivants (Hillebrand 2004).

« Zoological regions [...] are the most recent natural primary divisions of the earth as regards to its forms of animal life »
Wallace, 1876

Wallace décrit ici un provincialisme à large échelle spatiale, c'est à dire « la présence simultanée d'un grand nombre de formes endémiques bien différenciées dans une région » (Lomolino et al. 2010). Il découpera ainsi le monde en six régions biogéographiques distinctes (Wallace 1876, figure 2.1), en se basant notamment sur les travaux de Sclater (1858). Ce découpage du monde a été utilisé pendant très longtemps et ce n'est que relativement récemment qu'il a été réactualisé (Kreft & Jetz 2007; Holt *et al.* 2013, figure 2.2). On se rend alors compte à quel point Wallace et ses prédécesseurs avaient déjà atteint une précision exceptionnelle pour l'époque (comparez les deux régionalisations de la figure 2).

Encadré 1. Un bref aperçu de l'histoire de la biogéographie



Ce document n'a pas pour objectif de retracer toute l'histoire de la biogéographie (voir Lomolino *et al.* 2010), mais simplement d'en souligner les grandes lignes. * Lorsque deux auteurs sont notés, la durée de vie correspond à celle du premier.

Introduction générale

Cette description de la répartition des espèces est une étape fondamentale de l'enquête biogéographique. Elle ne constitue cependant que le premier pas vers ses deux objectifs majeurs. Le premier – fondamental – réside dans la compréhension des processus sous-tendant ces distributions (Diamond 1975, Lomolino et al. 2010). Le second – appliqué – est d'informer la société sur les conséquences biologiques des changements globaux provoqués par l'homme et, éventuellement, de proposer des solutions pour conserver les espèces (Whittaker et al. 2005).

Dans cette introduction, je propose d'abord de brefs rappels sur les processus fondamentaux contrôlant la distribution des espèces et les bases conceptuelles de la conservation de la biodiversité (points 2 et 3). Je présente ensuite les questions générales de ma thèse et l'approche qui m'a permis d'y répondre (point 4). Puis je décris les données utilisées (point 5) et la logique de l'organisation générale de mon travail (point 6).

2. Comprendre la distribution des espèces : les forçages fondamentaux

La répartition des espèces repose traditionnellement sur deux types de facteurs : les facteurs écologiques (biogéographie écologique, reportée en vert dans l'encadré 1) et les facteurs historiques (biogéographie historique, reportée en rouge dans l'encadré 1) (Morrone and Crisci 1995, Lomolino et al. 2010).

2.1. La niche écologique et la biogéographie écologique

Von Humboldt (1805) avait remarqué que les plantes s'étagaient le long des gradients altitudinaux (figure 1). La formalisation de la relation entre la distribution des espèces et les conditions environnementales n'est pourtant établie qu'au début du XX^{ème} siècle. C'est en effet Grinnell (1917) qui pour la première fois a défini la niche écologique d'une espèce comme l'ensemble des conditions environnementales dans lesquelles elle peut survivre et se reproduire. En effet, des contraintes physiologiques peuvent empêcher le maintien d'une population dans un environnement donné, par exemple le froid pour certains oiseaux nord-américains (Root 1988a, b) ou le taux de dessiccation pour certains crustacés vivant sur les rochers de l'estran (*i.e.* la zone littorale située entre les limites extrêmes des plus hautes et des plus basses marées, Connell 1961).

Une autre vision de la niche a été proposée par Elton (1927) : la niche d'une espèce est considérée comme l'ensemble des fonctions assurées par celle-ci au sein d'un écosystème, c'est en d'autres termes « son statut dans la communauté » (Elton 1927). La niche « Grinnellienne » se concentre donc sur les effets de l'environnement sur l'espèce alors que la niche « Eltonienne » se concentre sur les effets de l'espèce sur l'environnement.

Hutchinson (1957) propose de relier le concept de niche écologique aux ressources nécessaires à la survie et à la reproduction de l'espèce. En incorporant les travaux de Gause (1934) sur la compétition entre espèces pour l'utilisation des ressources du milieu, il définit ainsi deux types de niche. L'ensemble des conditions abiotiques permettant la survie de l'espèce sans interactions biotiques et sans limitation de la dispersion correspond à la **niche fondamentale** de l'espèce. La **niche réalisée** représente, quant à elle, les conditions environnementales dans lesquelles on observe

l'espèce dans la nature. La niche réalisée n'est donc dans certains cas qu'une sous-partie de la niche fondamentale. En effet, les capacités de dispersion des espèces n'étant pas infinies, l'absence d'une espèce dans une région donnée ne signifie pas forcément que les conditions environnementales ne s'y prêtent pas, mais peut simplement être due au fait que l'espèce n'a pas encore atteint la région. Il est aussi admis que les interactions biotiques (prédation, compétition, facilitation) peuvent restreindre ou élargir la niche réalisée. Par exemple, les expériences de transplantation de Connell (1961) montrent qu'une espèce de balane (*Balanus balanoides*) présente naturellement dans la zone haute de l'estran (la niche réalisée de l'espèce) peut survivre dans des zones plus basses en l'absence de compétiteurs (*i.e.* cette zone est incluse dans la niche fondamentale). Cependant elle se trouve exclue par la chthamale (*Chthamalus stellatus*), plus compétitrice, naturellement présente dans cette zone basse. Il est aussi possible que la niche réalisée d'une espèce soit plus étendue que sa niche fondamentale si des interactions positives ou des processus de type source-puits le lui permettent. Par exemple, la présence de certaines plantes en coussin permet l'installation d'autres espèces dans des conditions abiotiques qu'elles n'occupent normalement pas (processus de facilitation).

La nature de la niche écologique des espèces repose souvent sur des adaptations morphologiques, physiologiques ou phénologiques. Par exemple, les espèces de crustacés vivant dans le haut de l'estran possèdent des adaptations physiologiques leur permettant de résister à la dessiccation. Ces adaptations leur permettent donc d'être présentes dans un espace environnemental particulier. Ainsi les traits fonctionnels vont permettre de caractériser la niche écologique des espèces. Un trait fonctionnel est une caractéristique morphologique, physiologique ou phénologique d'un individu qui a un impact indirect sur sa performance (ou « fitness », voir Violle *et al.* 2007). Ces traits permettent donc de faire le lien entre les caractéristiques des individus (ou des espèces) et leur niches écologiques.

En définitive, il est possible de passer de l'espace environnemental occupé (*i.e.* la niche écologique *sensu* Grinnell) à l'espace géographique en connaissant la distribution dans l'espace des facteurs environnementaux. Cette approche ne permet cependant pas de rendre compte de toutes les observations.

2.2. La diversification et la dispersion des organismes : la biogéographie historique

« Dans les eaux, nous ne trouvons ni le castor ni le rat musqué, mais le coypou et le capybara, rongeurs ayant le type sud-américain. [...] Ces faits dénotent l'existence de quelque lien organique intime et profond qui prévaut dans le temps et l'espace, [...] indépendamment des conditions physiques. Il faudrait qu'un naturaliste fût bien indifférent pour n'être pas tenté de rechercher quel peut-être ce lien. »

Darwin (1859)

Dans cette citation tirée de *L'origine de Espèces* (1859) Darwin constate que les facteurs physiques ne permettent pas de rendre compte de certaines distributions disjointes (*i.e.* séparées) de certains taxons. En l'occurrence, la famille des Castoridés se répartit en Eurasie et en Amérique du Nord alors que celle des Cavidés (dont font partie les capybaras) est restreinte à l'Amérique du Sud (les deux familles appartiennent à l'ordre des rongeurs). Il existe en effet de nombreux groupes de mammifères typiques du continent sud-américain qui, selon Darwin, entretiennent un « lien organique intime », c'est-à-dire une évolution et une histoire communes. La biogéographie des mammifères du continent sud-américain en général illustre effectivement très bien comment les explications de nature historique permettent de mieux comprendre la distribution des espèces et notamment d'expliquer le provincialisme de ce continent. Sa faune est aujourd'hui composée par une mosaïque étonnante de lignées propres à l'Amérique du Sud et de lignées beaucoup plus cosmopolites (Encadré 2). Cette structure est le reflet d'une histoire complexe, qui peut être en partie révélée grâce au registre fossile (Encadré 3). On s'aperçoit alors que le continent, splendidement isolé depuis le démembrement de la Pangée permienne (Simpson 1980), a connu des vagues successives de colonisations et de diversifications qui ont contribué à former sa faune actuelle, indépendamment des facteurs physiques (Encadrés 2 & 3). Cependant, dans certain cas, le registre fossile est trop pauvre pour retracer et comprendre l'histoire de ces distributions. La phylogénie représente alors un outil complémentaire permettant

de retracer conjointement l'histoire des lignées et de leur distribution (Brundin 1972, voir Encadré 1). En étudiant les relations de parenté entre êtres vivants, elle permet en effet de reconstruire un arbre phylogénétique qui retrace l'histoire de la formation des lignées (figure 3.1, Lecointre & Le Guyader 2001). Le rapprochement entre la théorie de l'évolution et les méthodes biogéographiques constitue le socle fondamental de toutes les méthodes biogéographiques actuelles (Ronquist and Sanmartín 2011). L'utilisation d'une multitude de lignées biologiques indépendantes – et non pas d'une seule – pour parvenir à une histoire commune des aires de distributions représente une autre avancée méthodologique majeure. Cette approche, introduite par Leon Croizat (1894-1982, voir Encadré 1) dans les années 50-60 (Croizat 1952, 1964) est plus connue sous le nom de panbiogéographie.

Encadré 2. Un bref aperçu de l'originalité des mammifères sud-américains.

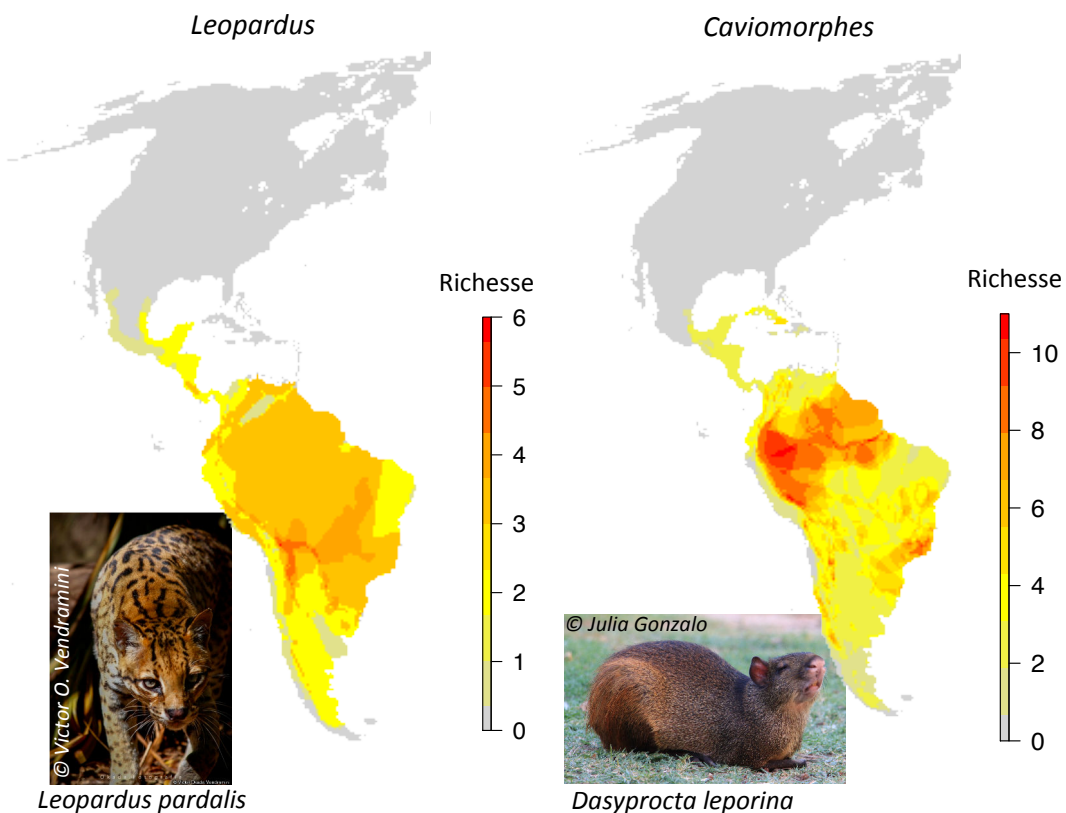
La diversité et la singularité des mammifères sud-américains est fascinante (Darwin 1859; Simpson 1980). En effet les mammifères actuels peuplant le continent représentent une mosaïque étonnante de lignées propres à l'Amérique du Sud et de lignées beaucoup plus cosmopolites. Ces lignées propres (endémiques) ont la particularité de se distribuer sur tout le gradient de la classification taxonomique.

Par exemple, on y retrouve quelques clades proches du niveau de l'ordre :

- les xénarthres dont font partie les tatous, fourmiliers et paresseux.
- les caviomorphes (les Histricognathes d'Amérique du Sud, voir la carte de richesse) dont font partie le capybara ou l'agouti (voir cliché) .
- les singes du nouveau monde (ou Platyrrhiniens) dont font partie le singe hurleur ou les ouistitis.

Et à l'autre bout du spectre de l'échelle taxonomique, on retrouve certains genres très majoritairement confinés à l'Amérique du sud (et éventuellement centrale) n'appartenant pas aux clades ci-dessus mais bien à des clades plus largement répandus :

- le genre *Leopardus* (voir la carte de richesse) qui contient notamment l'ocelot (voir cliché) et qui appartient à la famille (presque) cosmopolite des félidés.
- le genre *Mazama* qui contient des espèces de petits cerfs et qui appartient à la famille largement répandue des cervidés (mais néanmoins absente en Afrique et en Australie).

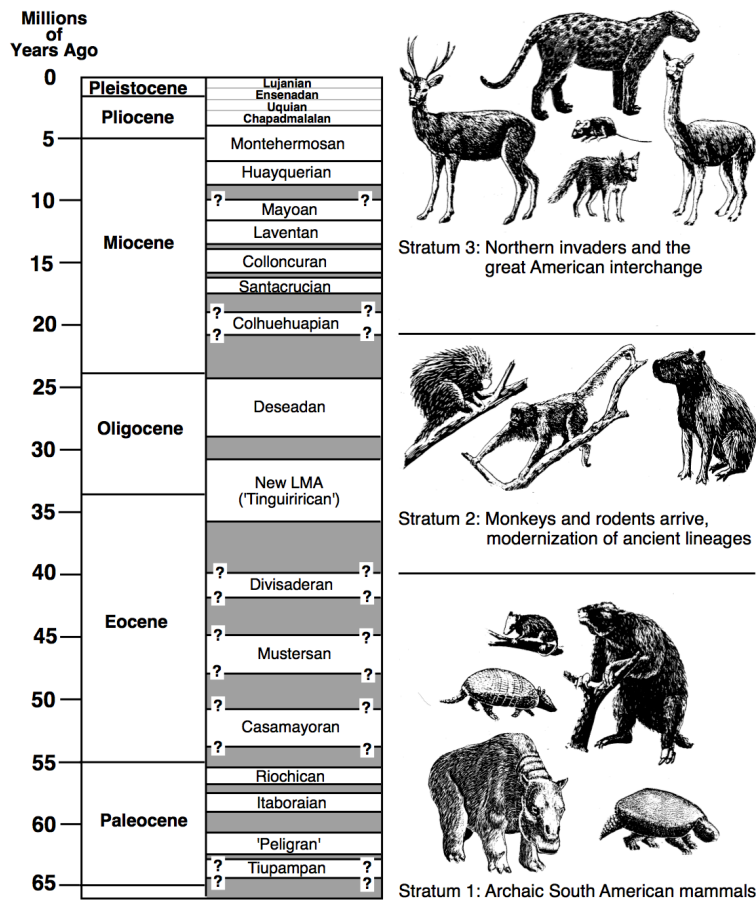


Encadré 3. Un bref résumé de la paléontologie des mammifères sud-américains.

En schématisant, on distingue trois phases (ou « strates ») majeures dans le registre fossile des mammifères sud-américains (Simpson 1980; Flynn 1998). Les groupes appartenant à ces trois phases n'ont pas nécessairement de représentants aujourd'hui et c'est pour cela que le registre fossile est d'une importance capitale.

La première phase contient des faunes dites « archaïques » qui peuplaient le continent au début de son isolation : elle comprend des lignées de xénarthres, de marsupiaux et d'ongulés aujourd'hui disparus. La deuxième phase contient des éléments « modernisés » de la phase 1 ainsi qu'une lignée de rongeurs et une lignée de primates, dont les représentants actuels sont les caviomorphes et les singes du Nouveau Monde (voir Encadré 2). Ces deux événements de colonisation apparaissent être le fruit d'une migration transocéanique, probablement depuis l'Afrique (Poux *et al.* 2006). La troisième et dernière phase contient les faunes dites modernes et correspond à la dernière vague de colonisation. Le GABI (« Great American Biotic Interchange ») représente l'arrivée de nombreux nouveaux ordres (cetartiodactyles, carnivores, lagomorphes et périssodactyles) consécutivement à la formation de l'isthme de Panama entre 13 et 3.5 Ma (Haug & Tiedemann 1998; Coates & Stallard 2013; Bacon *et al.* 2015).

En résumé, l'Amérique du Sud a donc été le théâtre d'un « splendide isolement » (Simpson 1980) et s'est vue colonisée par trois vagues successives. Ces colons se sont alors diversifiés pour former les groupes d'espèces que l'on connaît aujourd'hui, ou se sont éteints.

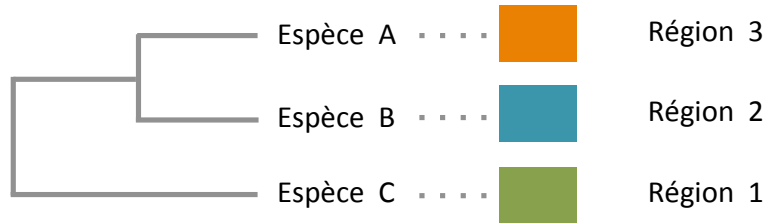


Reproduced from Flynn (1998)

Traditionnellement, deux processus évolutifs sont proposés pour expliquer les distributions disjointes des taxons (figure 3.2) : la dispersion et la vicariance (Lomolino et al. 2010). La vicariance postule que les régions A et B ne formaient qu'une seule et même entité lorsque le clade est apparu et que c'est son morcellement postérieur qui a conduit à la séparation des deux sous-clades (figure 3.3.1). La dispersion postule que le clade est apparu dans une des régions et a ensuite dispersé dans l'autre en donnant naissance à une nouvelle lignée (figure 3.3.2). La séparation des sous-clades est dans les deux cas le fruit d'une spéciation allopatrique.

L'analyse de l'histoire de la biogéographie nous montre combien l'importance paradigmatique accordée aux deux processus a varié au cours du temps, souvent sous forme de révolutions scientifiques (Kuhn 1962) et notamment en fonction de l'état des connaissances dans d'autres disciplines. Linné postulait ainsi que les organismes provenaient tous de l'arche de Noé qui s'était échouée sur le Mont Ararat (dans la Turquie actuelle) et avaient ensuite dispersé sur l'ensemble du globe. Darwin et sa théorie de l'évolution vinrent balayer ces explications et il devint alors plus commun de penser que des lignées pouvaient apparaître indépendamment sur les continents et disperser via des ponts transocéaniques éphémères. Cette vision fut soutenue par de très nombreux naturalistes à différentes époques, notamment par le paléontologue G. G. Simpson (voir Encadré 1) ou le zoogéographe P. Darlington (1957). C'est l'acceptation tardive (dans les années 60) de la théorie de la dérive des continents proposée par un géographe, Alfred Wegener (1912, voir Encadré 1) qui redonnera une plus grande importance au processus de vicariance. La dispersion sera alors considérée comme un bruit de fond parasite et sans véritable intérêt (Nelson 1974, Platnick and Nelson 1978, Lomolino et al. 2010). En fait, le débat ne portait pas tant sur l'importance relative de la dispersion et de la vicariance, mais plutôt sur le fait que la dispersion étant un processus hautement stochastique (contingent), il n'avait pas sa place dans une démarche hypothético-déductive classique *sensu* Popper (1968). Finalement, une synthèse combinant vicariance et dispersion semble émerger depuis quelques années (Sanmartín et al. 2001, Ronquist and Sanmartín 2011), notamment car les événements de dispersion peuvent en fait être prédictibles (par exemple, voir Sanmartín & Ronquist 2004).

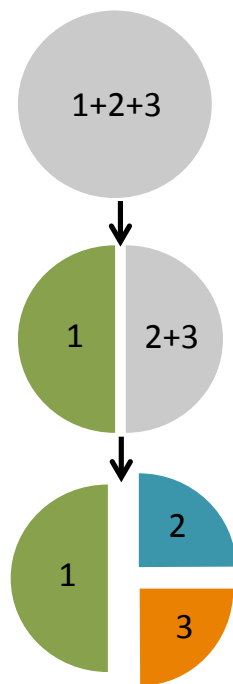
1. Arbre phylogénétique



2. Distributions géographiques

3. Hypothèses biogéographiques

3.1. Vicariance



3.2. Dispersion

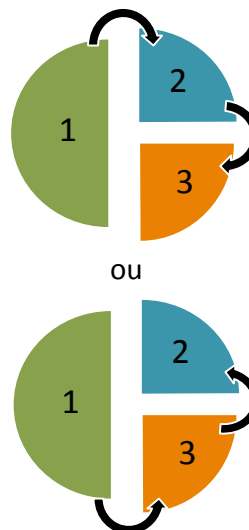


Figure 3. Les processus fondamentaux en biogéographie historique.

La figure présente les données de base utilisées en biogéographie : les relations historiques (1) et la distribution actuelle (2) de trois espèces. Elle décrit ensuite les deux processus fondamentaux (3) permettant d'expliquer ces données : la vicariance (3.1) et la dispersion (3.2).

Synthèse

La diversification et la dispersion des organismes, c'est à dire les dynamiques évolutives de leurs aires de distribution, jouent un rôle majeur dans l'établissement des patrons de répartition actuels et passés (Simpson 1944, 1980, Gould 2002, Sanmartín and Ronquist 2004, Lomolino et al. 2010, Hoorn et al. 2010, Belmaker and Jetz 2015). Cependant, les facteurs de nature historique et les facteurs de nature écologique expliquent conjointement la répartition des êtres vivant et leurs influences semblent varier selon l'échelle d'étude considérée (Levin 1992, Chave 2013).

2.3. L'importance de l'échelle d'étude

« *The very foundation of geography is scaling.* »

J.A. Wiens (1989)

La biogéographie étudie un système biologique donné dans un espace et un temps particuliers. La description d'un patron biogéographique et la compréhension des phénomènes sous-jacents nécessitent donc un positionnement sur trois échelles - spatiale, taxonomique et temporelle - distinctes. Traditionnellement, une échelle se décompose en un grain et une étendue. Chacune de ces deux composantes peut affecter les patrons observés mais aussi l'importance respective des différents processus sous-jacents.

L'échelle spatiale

En écologie, l'échelle spatiale revêt une importance fondamentale (MacArthur and Wilson 1967, Wiens 1989, Levin 1992, Chave 2013). Le grain est l'aire du site dans laquelle les espèces sont répertoriées alors que l'étendue est l'aire recouvrant l'ensemble des sites. Premièrement, le patron biogéographique lui-même dépend très largement du grain de l'étude. Wiens (1989) en donne un exemple d'une saisissante simplicité : deux espèces d'oiseaux sont globalement présentes dans la même partie du monde (*i.e.* en Amérique du Nord-Est, association positive) mais ne coexistent pas à l'échelle locale de quelques hectares (association négative). La description des patrons est donc fondamentalement reliée à l'échelle considérée. Deuxièmement, l'importance

relative de la niche écologique et de l'histoire est aussi contrainte par l'échelle spatiale de l'étude (Rosenzweig 1995, Lomolino et al. 2010, Chase and Myers 2011). Par exemple, les patrons à large échelle semblent influencés par la géographie de la spéciation et les contraintes de dispersion historiques (la présence des océans dans le cas des régions zoogéographiques du monde, figure 2). A l'intérieur de ces régions, c'est-à-dire à une échelle inférieure, l'environnement (et donc la niche écologique des espèces *sensu* Grinnell) a un effet important, par exemple via la répartition des biomes. Au niveau du paysage et de la communauté, les effets relatifs de la stochasticité, des interactions biotiques et des perturbations sont encore beaucoup discutés (Ricklefs 2008). Dans tous les cas, il apparaît que les processus à l'œuvre à différentes échelles sont interdépendants. Ainsi, les patrons à l'échelle locale, par exemple la richesse spécifique d'une communauté, sont en partie déterminés par les processus agissant à cette échelle, comme par exemple l'influence de l'environnement ou de la dispersion (Hubbell 2001, Chase and Myers 2011). Cependant ils sont aussi forcément contraints par la richesse régionale, qui peut être façonnée par des processus macro-évolutifs (e.g. Ricklefs 1987). La dépendance est cependant réciproque : les patrons à larges échelles spatiales se construisent à partir de patrons locaux et sont donc nécessairement sous contrôle des processus agissant à l'échelle locale.

L'échelle biologique et taxonomique

L'ensemble des entités biologiques présentes en un endroit donné à un moment donné et qui composent le système étudié peut lui aussi être défini à plusieurs niveaux. La hiérarchie phylogénétique fournit ici un cadre conceptuel robuste pour ce changement d'échelle. Cette échelle décrit un continuum depuis l'individu jusqu'au clade entier (par exemple l'ensemble des mammifères). Nous pouvons donc progressivement ranger les individus en boîtes concentriques qui sont représentées par les rangs taxonomiques. On peut alors étudier un système biologique aux niveaux des individus, des espèces, des genres, des familles ou mêmes des ordres. Un tel changement d'échelle a aussi démontré son intérêt en biogéographie étant donné que différents processus peuvent jouer à différentes échelles biologiques (e.g. Qian 2009; Thuiller *et al.* 2010).

L'échelle temporelle

L'échelle temporelle de l'observation correspond au temps pendant lequel l'observation est menée. L'abondance d'une espèce va ainsi pouvoir varier de façon drastique sur une courte échelle de temps mais être relativement stable à moyen et long terme. Dans cette thèse, je me suis cependant focalisé sur de longues échelles temporelles uniquement.

2.4. Synthèse

Ces brefs rappels montrent que la biogéographie est un exercice à la confluence d'une multitude de disciplines. L'approche historique se focalise sur l'évolution de la distribution des espèces et par conséquent, s'appuie largement sur l'évolution des espèces (reportée en violet dans l'encadré 1) et la géologie (reportée en marron dans l'encadré 1). La biogéographie écologique est quant à elle influencée par la physiologie, la dynamique des populations ou encore l'écologie des communautés. Pour comprendre l'importance relative des processus historiques et écologiques, une approche biogéographique intégrative se doit d'effectuer des changements d'échelles. Dans cette thèse, j'ai eu à cœur de travailler le long de ces échelles d'étude, par exemple l'échelle taxonomique (I.2 et III.1) mais aussi l'échelle spatiale (II). La compréhension des mécanismes régissant la distribution des espèces nous offre alors des outils pertinents pour prédire le devenir des espèces dans un monde changeant.

3. Protéger les espèces : la biogéographie de la conservation

«The cowman who cleans his range of wolves does not realize that he is taking over the wolf's job of trimming the herd to fit the range. He has not learned to think like a mountain. Hence we have dustbowls, and rivers washing the future into the sea.»

Aldo Leopold (1949)

Comme le fait remarquer Aldo Leopold dans cet extrait de « Thinking like a mountain » (1949), les conséquences des activités humaines sur les écosystèmes peuvent parfois être aussi imprévisibles que désastreuses. En l'occurrence, l'éradication du loup dans certaines zones de l'Amérique du Nord a des conséquences en cascade sur l'ensemble du milieu montagnard. Plus généralement, les changements globaux induits par l'homme semblent avoir profondément modifié la distribution du vivant. Ces forçages sont de nature extrêmement variable : changements climatiques, déforestation, changements d'utilisation des terres ou propagation des espèces invasives (e.g. Barnosky *et al.* 2012). Face à ces menaces, les biogéographes ont un rôle de premier plan : ils se doivent de proposer des prédictions sur les conséquences de ces menaces et éventuellement des solutions viables pour les atténuer. Ainsi la biogéographie de la conservation (Whittaker *et al.* 2005) peut être définie comme la discipline qui s'attache à « appliquer les principes, théories et analyses biogéographiques aux problèmes concernant la biologie de la conservation ». Il s'agit donc de définir les unités de bases sur lesquelles travailler, de prédire leur devenir dans un monde changeant et éventuellement de fournir des conseils pour les protéger.

3.1. Définir les unités de base de la biodiversité : de quoi parle-t-on ?

La variabilité du vivant est exubérante et se trouve souvent qualifiée par le terme un peu fourre-tout de « biodiversité » (contraction de « diversité biologique » en anglais). Le mot lui-même est davantage une construction sociale et politique que scientifique. Il est apparu dans les années 80 dans le but de sensibiliser la société à son érosion. Sa définition est extrêmement large puisqu'elle englobe, selon la Convention sur la diversité biologique (www.cbd.int, introduite au Sommet de la Terre à Rio en 1992) la « variabilité des organismes vivants de toute origine y compris, entre autres, les

écosystèmes terrestres, marins et autres écosystèmes aquatiques et les complexes écologiques dont ils font partie; cela comprend la diversité au sein des espèces et entre espèces ainsi que celle des écosystèmes ». Cette définition étant peu opérationnelle, il faut s'attacher à en préciser la nature.

En conservation, deux approches sont généralement entreprises pour définir ce que l'on veut préserver : l'approche peut-être fonctionnelle ou compositionnelle (Callicott et al. 1999). L'approche fonctionnelle privilégie plutôt les processus des écosystèmes comme les flux de matières (azote, phosphore, etc.) alors que l'approche compositionnelle se concentre d'avantage sur les entités biologiques comme les individus ou les espèces. En ce sens, cette dernière approche se rapproche davantage de la biogéographie écologique et historique. Nous focaliserons donc sur celle-ci dans un premier temps pour l'élargir par la suite dans le développement et la discussion générale.

3.2. Prédire l'érosion de la biodiversité

Dans le cadre d'une approche compositionnelle, l'érosion de la biodiversité se traduit notamment par l'extinction des populations ou des espèces. Pour les biogéographes, il s'agit donc de pouvoir prédire ces extinctions en fonction de différents scénarii futurs. Il existe de nombreuses méthodes de prédiction mais deux sont particulièrement utilisées. Premièrement, l'utilisation des « modèles de distribution » permet de projeter les distributions d'espèces dans le futur en fonction des scénarii de changement climatique fournis par exemple par le groupe d'experts intergouvernemental sur l'évolution du climat (e.g. Thuiller *et al.* 2005). On peut alors évaluer la part d'espèces susceptibles de voir son aire de distribution réduire significativement et donc être potentiellement en danger d'extinction. C'est notamment le cas pour les espèces vivant à haute latitude et haute altitude (Thuiller *et al.* 2005). Dans un deuxième temps, l'utilisation de la relation aire-espèce (c'est-à-dire la relation statistique entre l'aire d'échantillonnage et le nombre d'espèces, voir Chapitre III-IV) permet de prédire les taux d'extinction en fonction de la perte ou de la fragmentation de l'habitat (e.g. Thomas *et al.* 2004; Hanski & Zurita 2013), une des menaces majeures qui pèse sur la biodiversité.

3.3. Protéger la biodiversité ?

La biogéographie de la conservation s'attache aussi à élaborer des solutions pour protéger la biodiversité. Le constat fondamental est le suivant : la protection de l'ensemble des entités biologiques peuplant la terre n'est pas possible, il faut donc choisir et protéger des aires restreintes. La biologie de la conservation vise donc à définir ou à compléter des aires de protection. Ce plan de conservation repose sur quatre critères complémentaires : la représentativité, la persistance, l'efficacité et la flexibilité. Il s'agit d'élaborer un projet flexible socialement et viable économiquement pour protéger un échantillon représentatif de l'ensemble biologique qui sera capable de se maintenir à plus ou moins long terme.

3.4. Synthèse

De la même façon que la biogéographie fondamentale, la biogéographie de la conservation travaille le long de différentes échelles d'étude : échelle de l'entité biologique (populations, espèces, écosystèmes) et échelles spatiales d'intervention (régions, pays, globe). En définitive les deux disciplines manipulent donc des systèmes complexes qu'il s'agit de comprendre et de protéger. Dans les deux cas, il faut dans un premier temps tenter de les décrire de façon synthétique, ce à quoi je vais m'attacher dans le prochain point.

4. Considérer la biodiversité sous toutes ces facettes pour mieux la comprendre et la protéger

En général, on synthétise la structure (ou la diversité) d'un assemblage d'espèce par un ou plusieurs indices quantitatifs. Par exemple, le plus simple est le nombre d'espèces contenues dans l'assemblage (=la richesse). De façon générale, ces mesures représentent différentes façon de synthétiser les données faunistiques de base, par exemple plusieurs milliers de polygones de distribution. Dans le cas de la richesse, la diversité est entrevue par le prisme de l'espèce. Bien qu'étant informative et importante, elle ignore cependant les différences fonctionnelles et/ou phylogénétiques entre espèces et en ce sens ne représente qu'une vue partielle de la biodiversité. La démarche générale de ma thèse propose de (1) décrire les facettes évolutives et les facettes fonctionnelles de la biodiversité pour (2) distinguer entre plusieurs hypothèses concernant l'importance des processus explicatifs sous-jacents (par exemple historiques ou écologiques) et (3) mieux protéger l'histoire et le fonctionnement des écosystèmes. La description de ces facettes via l'utilisation des indices de diversité phylogénétique et fonctionnelle a déjà été entreprise en biogéographie fondamentale (Webb et al. 2002, Petchey and Gaston 2006) et de la conservation (Forest et al. 2007). Je commence donc par décrire les avancées récentes et les intérêts de la démarche (4.1) puis je pose les questions spécifiques de ma thèse qui utiliseront cette démarche (4.2).

4.1. Etat de l'art

4.1.1. Principe et intérêts de la démarche en biogéographie fondamentale

Contexte

Dans cette thèse je me suis intéressé à la description et à la compréhension des patrons biogéographiques à large échelle spatiale et je me suis attaché à étudier les propriétés émergentes de ces systèmes, une approche que l'on nomme parfois la macro-écologie (Brown and Maurer 1989, Brown 1999, Keith et al. 2012). La macro-écologie s'intéresse ainsi à certains grands patrons de la diversité biologique comme le nombre d'espèces (Hillebrand 2004, Belmaker and Jetz 2015), la taille des aires de distribution (Gaston and Blackburn 2008), la taille des organismes (Olson et al. 2009) ou encore les

différences de composition faunistique entre grandes régions du monde (Kreft and Jetz 2010, Holt et al. 2013). Les règles générales en écologie correspondent à des corrélations saisissantes entre ces variables faunistiques et des paramètres géographiques ou climatiques (Lawton 1999). On pourra citer les plus incontournables :

- le nombre d'espèce et l'aire d'échantillonnage, *i.e.* la relation aire-espèces (MacArthur & Wilson 1967; Storch *et al.* 2012, voir Chapitre III-IV).
- le nombre d'espèce et la latitude, *i.e.* le gradient latitudinal de diversité (Wallace 1876, Hillebrand 2004, Belmaker and Jetz 2015).
- la taille des aires de répartition et la latitude, *i.e.* la loi de Rapoport (Lomolino et al. 2010).
- la dissimilarité faunistique et la distance géographique, *i.e.* « the distance-decay relationship » (Rosenzweig 1995; Qian & Ricklefs 2012, voir chapitre VII).
- la masse/taille d'une espèce et la latitude, *i.e.* la règle de Bergman (Lomolino *et al.* 2010, voir Annexe 3).

Ces règles écologiques représentent le terrain d'investigation de la plupart des biogéographes. Après les avoir confirmées, il s'agit de proposer et de tester différents types d'hypothèses explicatives. Ici, le propos n'est pas de détailler ces règles, mais d'introduire l'approche suivie dans cette thèse. Pour comprendre ces grands patrons de la biogéographie, il faut dans un premier temps élaborer des hypothèses explicatives qui produisent des prédictions claires et précises que l'on peut confirmer ou infirmer en les comparant aux observations. Nous avons vu que la biogéographie écologique reposait entre autre sur l'utilisation des traits fonctionnels alors que la biogéographie historique était fondamentalement reliée aux arbres phylogénétiques. Peut-on utiliser conjointement les traits fonctionnels et les phylogénies pour tester des hypothèses précises expliquant les patrons biogéographiques ? En utilisant l'exemple du gradient latitudinal de diversité, je montre dans cette introduction comment une approche utilisant conjointement les indices de diversité fonctionnelle et phylogénétique est prometteuse.

L'apport des données phylogénétiques

Pour expliquer le gradient latitudinal de diversité, l'hypothèse du « time for speciation » (TFS) couplée à celle du « tropical niche conservatism » (TNC, Wiens & Donoghue 2004) postule que (1) la plupart des clades sont d'origine tropicale car ce biome occupait une large partie du globe au cénozoïque, que (2) les adaptations aux conditions tempérées sont rares (hypothèse TNC) et qu'en conséquence les tropiques ont accumulé davantage d'espèces au cours de l'évolution (hypothèse TFS). Ce postulat prédit que les espèces tempérées doivent être relativement « jeunes » et que les tropiques contiennent des espèces « jeunes » et « vieilles », c'est à dire qu'ils constituent un musée et un berceau de diversité. Il est donc nécessaire de décrire plus finement la structure phylogénétique des assemblages d'espèces tropicales et tempérées pour tester si les hypothèses proposées sont valides. La description de la diversité phylogénétique* des assemblages permet alors de tester une telle prédiction, puisqu'elle résume la structure des relations phylogénétiques entre espèces dans un endroit donné. Dans ce cas, elle permet de tester des hypothèses de type historique.

L'apport des données fonctionnelles

Une seconde hypothèse postule que les régions possédant une plus grande disponibilité en niches écologiques permettront la coexistence d'un plus grand nombre d'espèces (Belmaker and Jetz 2015). En effet, la théorie de la coexistence prédit que deux espèces possédant des niches écologiques identiques ne peuvent pas coexister et vont s'exclure mutuellement (Chesson 2000). Il s'agit donc de tester si les zones tropicales possèdent relativement plus de niches disponibles que les zones tempérées. Les traits fonctionnels vont permettre de caractériser la niche écologique des espèces et ainsi de tester cette prédiction. En construisant une mesure de diversité fonctionnelle (Petchey and Gaston 2006) au niveau des assemblages d'espèces, il va donc être possible de tester si les régions riches en espèces possèdent aussi davantage de niches écologiques (Belmaker and Jetz 2015).

4.1.2. Principe et intérêts de la démarche en biogéographie de la conservation

L'utilisation de la diversité phylogénétique et de la diversité fonctionnelle dans le cadre de la biogéographie de la conservation revêt plusieurs intérêts.

Tout d'abord, la dualité traditionnelle entre approche fonctionnelle et compositionnelle peut être unifiée en utilisant simultanément des mesures taxonomiques et des mesures fonctionnelles de la biodiversité. En effet, les mesures de diversités fonctionnelles permettent de décrire de façon plus réaliste certains aspects du fonctionnement des écosystèmes (Díaz & Cabido 2001), par exemple leur productivité (Cadotte et al. 2011). Les services que rendent les écosystèmes aux sociétés humaines pourraient eux aussi être représentés par des indices de diversité fonctionnelle (Díaz *et al.* 2007). En conséquence, la préservation de la diversité fonctionnelle pourrait permettre de mieux protéger le fonctionnement des écosystèmes (Mouillot et al. 2013).

La préservation de la diversité phylogénétique (Faith 1992), c'est-à-dire de l'histoire évolutive des espèces, apparaît être une considération davantage éthique. Est-il également important de protéger une espèce qui est le seul représentant de son ordre (par exemple *Orycteropus afer*, l'oryctérope du Cap) ou une espèce appartenant à un ordre contenant plusieurs milliers d'espèces (par exemple une espèce de rongeur) ? En fait, Vane-Wright *et al.* (1991) montrent que dans un monde où la préservation de toutes les lignées est impossible (la fameuse « agonie du choix »), il est important de considérer l'originalité (= la rareté) évolutive des espèces, c'est-à-dire leur position dans l'arbre de la vie. En effet, la perte de lignées évolutives uniques engendrera la perte définitive d'un grand nombre de caractéristiques uniques alors que la perte de lignées « redondantes » n'aura qu'un effet mineur sur la perte de caractéristiques uniques. Certains auteurs considèrent aussi qu'une forte diversité phylogénétique représente un réservoir de potentiel adaptatif dans un futur instable (Forest et al. 2007) et, potentiellement, un réservoir de diversité fonctionnelle.

En définitive, il s'agit, selon Lecointre (2011), de préserver les espèces pour ce qu'elles font (= approche fonctionnelle) et pour ce qu'elles sont (= approche phylogénétique). Certaines analyses commencent à montrer que les différentes facettes de la biodiversité ne sont pas interchangeable ou « congruentes », c'est à dire qu'en en préservant une, on ne garantit pas forcément la protection des autres (Devictor *et al.* 2010). En conséquence, il apparaît important de considérer les trois facettes de la biodiversité dans les plans de conservation. A large échelle, cette démarche a d'ailleurs été intégrée aux objectifs de la plateforme intergouvernementale sur la biodiversité et les services des écosystèmes (IPBES, Díaz *et al.* 2015).

Introduction générale

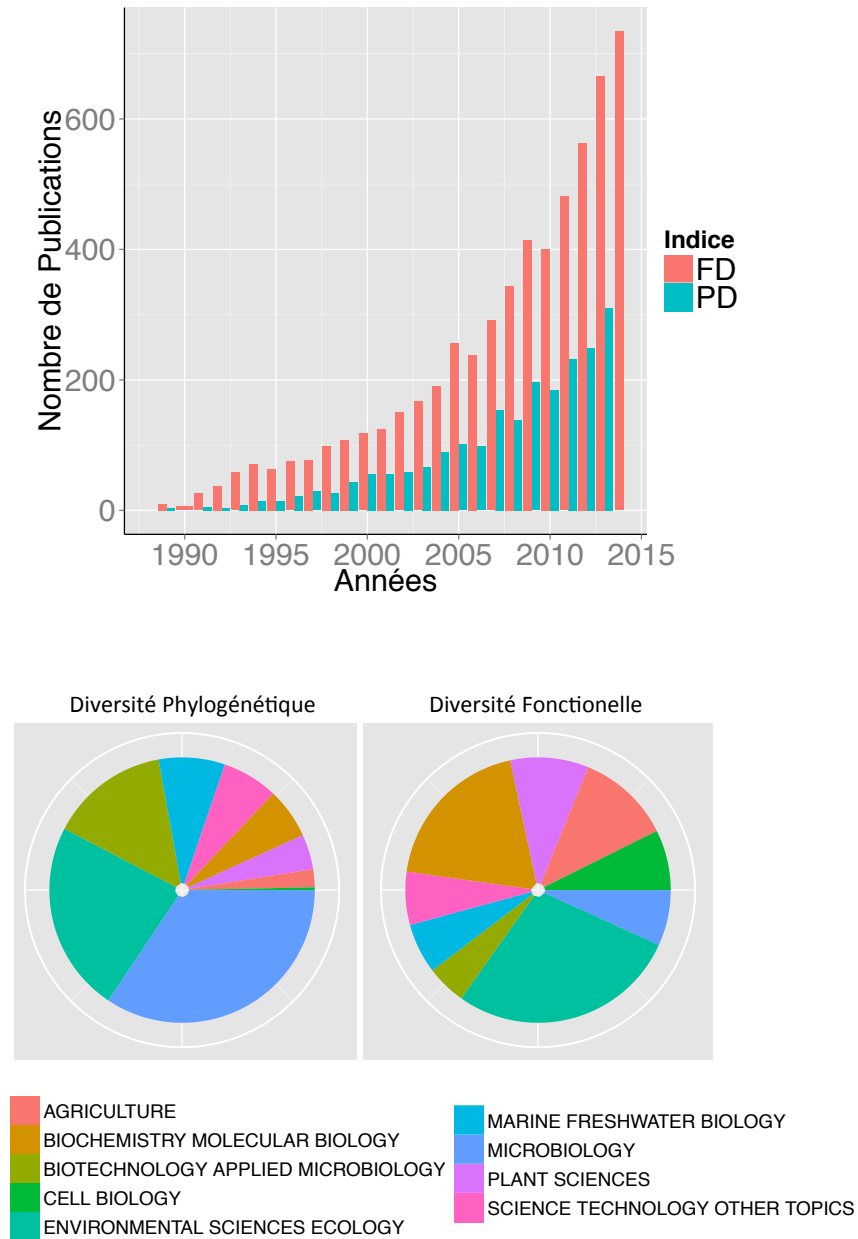


Figure 4. Utilisation des indices de diversité phylogénétique et fonctionnelle dans la littérature (d'après ISI web of knowledge, le 01/09/2015). La figure A présente le nombre d'articles qui utilisent les expressions « diversité phylogénétique » ou « diversité fonctionnelle » dans leurs mots clefs, au cours du temps. La figure B présente la répartition de l'ensemble de ces articles dans différentes disciplines (seules les cinq premières disciplines pour chaque expression sont conservées).

4.2. Questions abordées dans cette thèse

Il apparaît ainsi que l'utilisation des diversités fonctionnelles et phylogénétiques en biologie soit prometteuse. On observe d'ailleurs depuis le début des années 90 une forte augmentation de leur utilisation, notamment en écologie (figure 4) et de nouveaux champs entiers de recherche leur sont maintenant dédiés, par exemple la biogéographie fonctionnelle (Violle *et al.* 2014) ou encore « l'écophylogénétique » (Mouquet *et al.* 2012).

Dans cette thèse, nous proposons d'étudier les grands patrons biogéographiques via la description des différentes facettes de la diversité biologique (taxonomique, fonctionnelle et phylogénétique). En particulier :

(1) Les différentes facettes de diversité des mammifères sont-elles congruentes à l'échelle du globe ? En d'autres termes, protéger le nombre d'espèces suffit-il à protéger la diversité fonctionnelle et évolutive ?

(2) Ces différentes facettes sont-elles également vulnérables aux changements globaux et en particulier à la perte d'habitat ?

(3) Ces différentes facettes sont-elles également couvertes par les zones de protection ?

(4) Que peuvent nous apporter ces différentes facettes dans la compréhension des grands patrons de répartition du vivant ?

En particulier, peut-on décomposer la distribution des lignées sur une échelle phylogénétique pour mieux comprendre les effets relatifs de l'histoire et de la niche écologique ?

5. Le système biologique et l'échelle spatiale de l'étude

Les seuls groupes de taille importante dont on connaît la répartition mondiale, les relations de parenté ainsi qu'un certain nombre de traits fonctionnels sont les mammifères et les oiseaux. Il est d'abord utile de se demander pourquoi ces groupes sont si bien connus, puisqu'ils vont constituer le matériel de base de mon travail.

5.1. Les origines du biais de connaissance

Les mammifères et les oiseaux sont relativement mieux connus que la plupart des autres groupes pour trois raisons majeures. Premièrement, les mammifères et les oiseaux contiennent relativement peu d'espèces (environ 5000 et 10 000 respectivement) contrairement aux plantes (environ 300.000, <http://www.theplantlist.org/>) ou aux arthropodes (5-10.000.000 ; Ødegaard, 2000). Deuxièmement, ils sont relativement gros, ce qui facilite évidemment leur découverte et leur description. Dans le cas des mammifères, une troisième raison majeure apparaît : ils nous ressemblent.

« Ce sont les mammifères qui doivent occuper le premier rang dans le règne animal, sous le rapport du perfectionnement de l'organisation et du plus grand nombre de facultés. »

Jean Baptiste de Lamarck (1809)

Cette citation de Lamarck (1744-1829) peut aujourd'hui paraître dépassée mais elle est pourtant le reflet d'un véritable impact de l'anthropocentrisme en biogéographie et en évolution. Pourtant, Darwin initia bien un radical changement de paradigme chez les naturalistes il y a déjà plus de 150 ans. Il remplaça en effet l'homme au sein des primates et des êtres vivants en général, et non plus aux côtés de Dieu, ce qui lui valut d'ailleurs de nombreuses critiques et moqueries (figure 5). Malgré les preuves avancées par Darwin, l'impact de l'anthropocentrisme en évolution a continué à être fort. Les planches de Ernst Haeckel (1834-1919), notamment celles sur les animaux, en sont une

illustration remarquable (figure 6). En effet, on y voit une représentation en arbre où les mammifères occupent le sommet du houppier, véhiculant ainsi l'idée qu'ils sont de nature « supérieure ». Cet anthropocentrisme chronique explique en partie pourquoi les mammifères sont aujourd'hui mieux connus que les autres groupes et pourquoi ils représentent donc un groupe de choix dans l'étude des patrons biogéographique à large échelle spatiale.

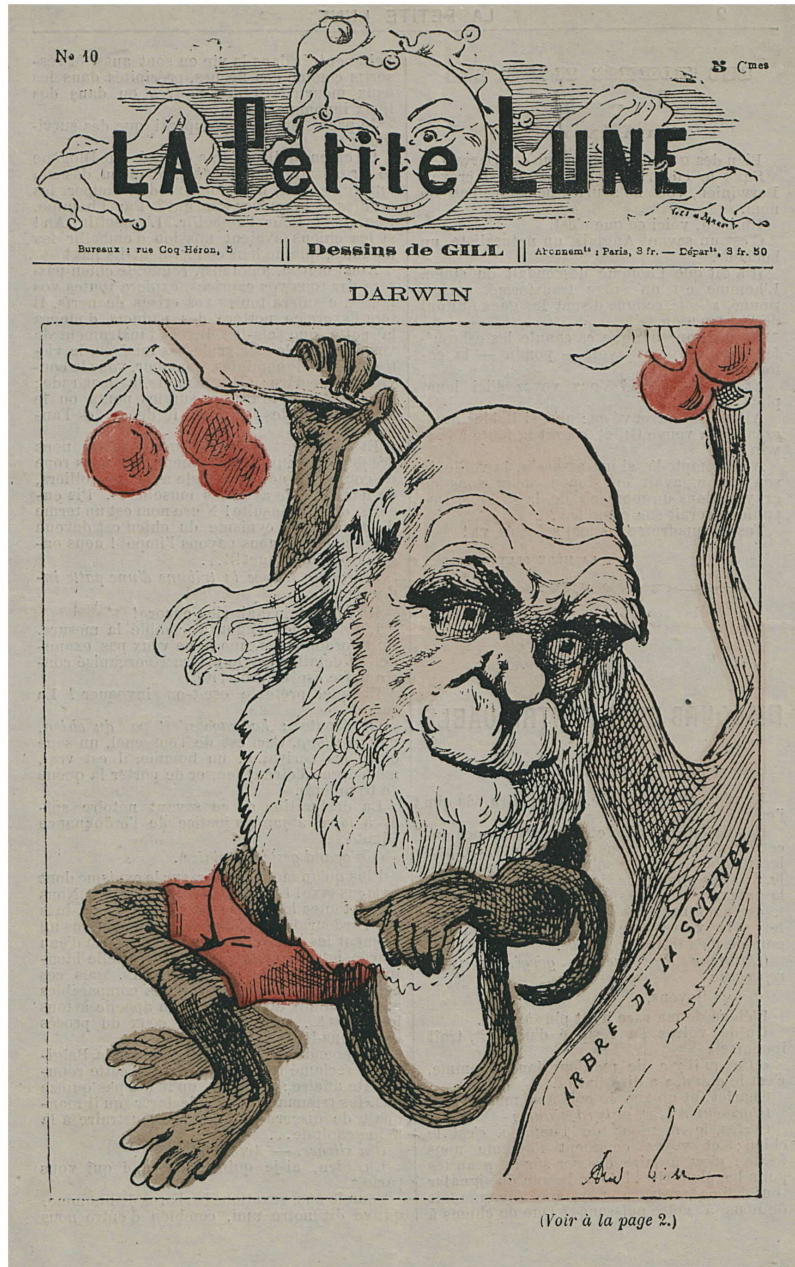


Figure 5. Caricature de Charles Darwin. Cette caricature a été dessinée par André Gill en 1878 pour *La petite lune*. Elle dépeint Darwin en singe (une hérésie !) accroché à l'arbre de la science.

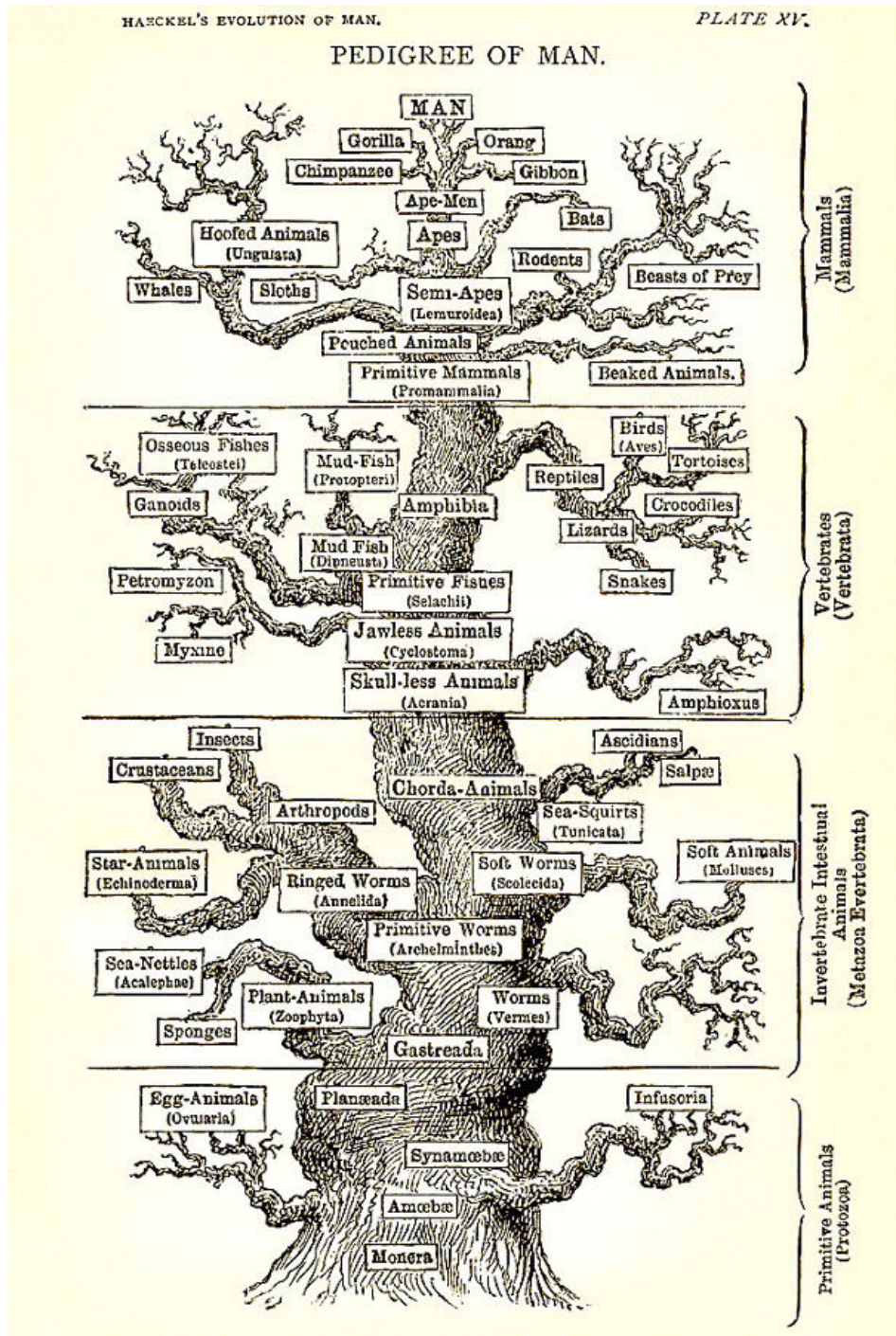


Figure 6. Planche de Ernst Haeckel (1874) représentant le pedigree de l'homme (« pedigree of man »). L'homme y est représenté au sommet de la plus haute branche d'un arbre des animaux dont le tronc est découpé en quatre stades majeurs : des « animaux primitifs » à la base suivis d'invertébrés, des vertébrés et finalement des mammifères. On notera deux points intéressants : (1) les mammifères ne sont pas placés au sein des vertébrés, mais bien au dessus ; (2) certaines branches paraissent redescendre, comme celle de la myxine ou celle des éponges.

5.2. Les données de distribution utilisées

Les données de distribution de ces groupes sont donc relativement anciennes et parmi les meilleures à large échelle spatiale. Elles sont aujourd'hui mises à disposition par l'IUCN pour les mammifères (<http://www.iucnredlist.org/>) et par BirdLife pour les oiseaux (<http://www.birdlife.org/>).

L'étendue de mon échelle d'étude est le globe terrestre et le grain va dépendre de la qualité des données de distribution. Les données brutes disponibles sont des polygones d'étendue d'occurrence (« extent of occurrence », EOO; cf. figure 7 pour le loup, *Canis lupus*). En fait, elles ont été compilées sur la base de dire d'experts nationaux réunis en comités plus larges. Concrètement elles ont été produites en traçant un « contour » autour des localités où l'espèce a été effectivement observée. Ceci implique que l'espèce ne sera pas forcément présente à l'intérieur de l'ensemble du polygone (à la différence des distributions basées sur les aires d'occupations – « area of occupancy » –). A priori il existe aussi de grands écarts de précisions entre espèces et entre régions. La carte de distribution du loup en montre un exemple frappant : la répartition est extrêmement détaillée en Europe de l'ouest, où l'on reconnaît même les massifs montagneux. Au contraire, le loup apparaît réparti sur l'ensemble de la péninsule arabique, ce qui n'est a priori pas le cas (F. Ficetola, communication personnelle).

Pour définir des assemblages d'espèces, il faut définir des aires -souvent des « pixels » - dans lesquelles on note si l'espèce est absente ou présente. Peut-on alors réellement utiliser les données de l'IUCN et de BirdLife sans risquer de surestimer le nombre de pixels dans lesquels l'espèce est présente (étant donné que son EOO ne représente qu'un contour)? Une solution conservatrice est d'utiliser un grain d'étude « assez large », ce qui permettra de minimiser les faux positifs. Ce grain adéquat n'est en fait pas défini et différents auteurs utilisent différents grains : par exemple Jenkins *et al.* (2013) utilise un grain de 10km*10km alors que certaines études montrent qu'on ne peut pas descendre en dessous de 100km*100km sans risquer de faire de graves erreurs (Hurlbert and Jetz 2007). Dans cette thèse, j'adopte donc une démarche conservatrice en utilisant un grain grossier (> 100*100km) dans la plupart des cas.



Figure 7. L'aire de distribution du loup (*Canis lupus*) dans le monde, selon l'IUCN. Les zones en orange montrent la répartition actuelle de l'espèce alors que la zone en rouge montre la zone où l'espèce a disparu.

6. Stratégie de travail et plan de la thèse.

6.1. Principe

L'utilisation des indices de diversité fonctionnelle et phylogénétique en biogéographie se heurte à une première difficulté majeure. La discipline est littéralement parasitée par la multiplication des indices, empêchant ainsi toute tentative de synthèse cohérente. En effet, le grand nombre d'indices de diversité proposés a créé une véritable « jungle » dans laquelle il est compliqué – si ce n'est impossible – de se retrouver (Winter et al. 2012). A titre d'exemple, la discipline comprend plus de 60 indices de diversité phylogénétique (voir le matériel supplémentaire de l'article 1). Puisque différentes études utilisent différents indices et qu'il n'existe pas de synthèse permettant de relier les indices les uns aux autres (mais voir Pavoine & Bonsall 2011), toute tentative de généralisation est compliquée.

La première partie de ma thèse s'attache donc à résoudre ce problème en répondant à une triple nécessité : répertorier, classer et comparer les indices de diversité. Les articles 1 & 3 (Chapitre I et Annexe 1.3, respectivement) listent, classent et comparent les indices de diversité. L'article 2 (Chapitre II) va plus loin en explorant la sensibilité des principaux indices à l'échelle taxonomique/phylogénétique.

La deuxième partie de ma thèse explore la sensibilité des approches de conservation classique, c'est-à-dire basées sur la diversité taxonomique uniquement, à la prise en compte des diversités fonctionnelles et phylogénétiques. En particulier, j'évalue la perte de la diversité phylogénétique consécutive à la destruction d'habitat (Article 4, Chapitre III), la nature des points chauds de diversité (« *biodiversity hotspots* ») à plusieurs facettes (« *multifaceted biodiversity hotspots* », Article 5, Chapitre IV), et la capacité des aires de protection européenne à préserver les diversités fonctionnelles et phylogénétiques (Article 6, Annexe 2.3).

La troisième partie de ma thèse s'attache à mieux comprendre quelques problématiques de la biogéographie fondamentale en utilisant les indices de diversité phylogénétique et fonctionnelle. Notamment, je montre dans un premier temps comment les changements d'échelles le long d'un gradient phylogénétique permettent de séparer les effets de différents processus (sous-partie 1), notamment en biogéographie à grande échelle spatiale (Article 7, Chapitre V) mais aussi en

biogéographie des microbes (Article 8, Chapitre VI). La deuxième sous-partie montre comment une analyse conjointe des diversités fonctionnelles et phylogénétiques permet de mieux comprendre l'évolution morphologique des assemblages de mammifères à large échelle (Articles 9 & 10, Annexe 3.2.1 et Chapitre VII, respectivement).

6.2. Organisation

Chaque partie est constituée d'un sommaire détaillé, d'une introduction, de plusieurs chapitres comportant un ou plusieurs articles, d'une synthèse et d'une bibliographie. Je propose une synthèse générale en guise de conclusion mais le cœur de ma discussion sera développée dans les synthèses de chacune des parties. Les dix articles présentés se retrouvent à l'interface entre la biogéographie fonctionnelle, phylogénétique et de la conservation (figure 8).

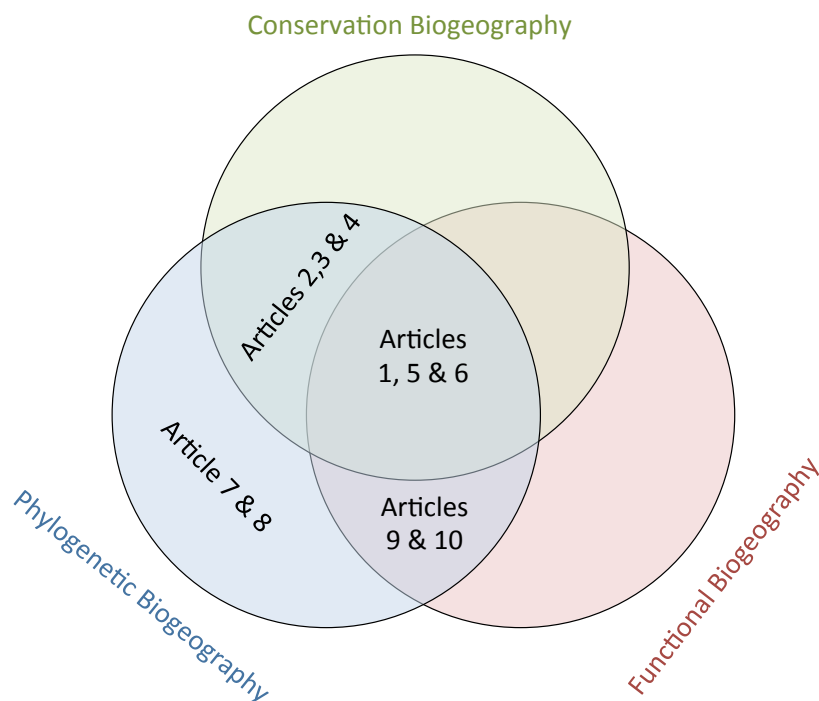


Figure 8. La figure présente la répartition de mes articles de thèse à l'interface de trois disciplines majeures de la biogéographie, illustrant ainsi le caractère pluridisciplinaire de mon travail.

Bibliographie de l'introduction générale

- Barnosky, A. D. et al. 2012. Approaching a state shift in Earth's biosphere. - *Nature* 486: 52–8.
- Belmaker, J. and Jetz, W. 2015. Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients (H Arita, Ed.). - *Ecol. Lett.* 18: 563–571.
- Brown, J. 1999. Macroecology: progress and prospect. - *Oikos* in press.
- Brown, J. and Maurer, B. 1989. Macroecology: the division of food and space among species on continents. - *Science* (80-.). 243: 1145–1150.
- Brundin, L. 1972. Phylogenetics and Biogeography. - *Syst. Biol.* 21: 69–79.
- Cadotte, M. W. et al. 2011. Beyond species: functional diversity and the maintenance of ecological processes and services. - *J. Appl. Ecol.* 48: 1079–1087.
- Callicott, J. B. et al. 1999. Current Normative Concepts in Conservation. - *Conserv. Biol.* 13: 22–35.
- Chase, J. M. and Myers, J. A. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. - *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366: 2351–63.
- Chave, J. 2013. The problem of pattern and scale in ecology: what have we learned in 20 years? - *Ecol. Lett.* 16 Suppl 1: 4–16.
- Chesson, P. 2000. Mechanisms of Maintenance of Species Diversity. - *Annu. Rev. Ecol. Syst.* 31: 343–366.
- Connell, J. H. 1961. The Influence of Interspecific Competition and Other Factors on the Distribution of the Barnacle *Chthamalus Stellatus*. - *Ecology* 42: 710.
- Croizat, L. 1952. Manual of phytogeography, or an account of plant-dispersal throughout the world. - The Hague; W. Junk.
- Croizat, L. 1964. Space, Time, form: The Biological Synthesis. - The author.
- Darlington, P. J. 1957. Zoogeography: The Geographical Distribution of Animals. in press.
- Darwin, C. 1859. The origin of species by means of natural selection, or, The preservation of favored races in the struggle for life. (J Murray, Ed.). - The Werner Company.
- Devictor, V. et al. 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.
- Diamond, J. 1975. Assembly of species communities. - In: Cody ML, D. J. (ed), *Ecology and evolution of communities*. Harvard University Press, pp. pp. 342–444.
- Díaz, S. et al. 2007. Incorporating plant functional diversity effects in ecosystem service assessments. - *Proc. Natl. Acad. Sci. U. S. A.* 104: 20684–9.
- Díaz, S. et al. 2015. The IPBES Conceptual Framework — connecting nature and people. - *Curr. Opin. Environ. Sustain.* 14: 1–16.
- Díaz, S. and Cabido, M. 2001. Vive la différence: plant functional diversity matters to ecosystem processes. - *Trends Ecol. Evol.* 16: 646–655.
- Elton, C. 1927. *Animal Ecology*. - Sedgwick and Jackson.
- Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - *Biol. Conserv.* 61: 1–10.
- Forest, F. et al. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. - *Nature* 445: 757–760.
- Gaston, K. and Blackburn, T. 2008. *Pattern and Process in Macroecology*. - Wiley & Sons.
- Gause, G. 1934. Experimental analysis of Vito Volterra's mathematical theory of the struggle for existence. - *Science* (80-.). 79: 16–17.
- Gould, S. J. 2002. *The Structure of Evolutionary Theory*. - Harvard University Press.
- Grinnell, J. 1917. The Niche-Relationships of the California Thrasher. - *Auk* 34: 427–433.
- Hanski, I. and Zurita, G. 2013. Species–fragmented area relationship. - *Proc. Natl. Acad. Sci. U. S. A.* 110: 12715–12720.
- Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. - *Am. Nat.* 163: 192–211.
- Holt, B. G. et al. 2013. An Update of Wallace's Zoogeographic Regions of the World. - *Science* (80-.). 339: 74–78.
- Hoorn, C. et al. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. - *Science* 330: 927–31.
- Hubbell, S. 2001. *The unified neutral theory of biodiversity and biogeography*. - Princeton University Press.
- Humbolt, A. von and Bonpland, A. 1805. *Essai sur la géographie des plantes : accompagné d'un tableau physique des régions équinoxiales, fondé sur des mesures exécutées, depuis le dixième degré de latitude boréale jusqu'au dixième degré de latitude australe, pendant les années 1799, 1800, 1801.*, - Levrault Schoell, Paris.

Introduction générale

- Hurlbert, A. H. and Jetz, W. 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. - *Proc. Natl. Acad. Sci. U. S. A.* 104: 13384–9.
- Hutchinson, G. E. 1957. Concluding Remarks. - *Cold Spring Harb. Symp. Quant. Biol.* 22: 415–427.
- Jenkins, C. N. et al. 2013. Global patterns of terrestrial vertebrate diversity and conservation. - *Proc. Natl. Acad. Sci. U. S. A.* 110: E2602–2610.
- Keith, S. A. et al. 2012. What is macroecology? - *Biol. Lett.* in press.
- Kreft, H. and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. - *Proc. Natl. Acad. Sci. U. S. A.* 104: 5925–5930.
- Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species distributions. - *J. Biogeogr.* 37: 2029–2053.
- Kuhn, T. S. 1962. *The Structure of Scientific Revolutions.* - University of Chicago Press.
- Ladle, R. J. and Whittaker, R. J. 2011. *Conservation Biogeography.* - John Wiley & Sons.
- Lamarck, J.-B. 1809. *Philosophie zoologique.* - Dentu.
- Lawton, J. H. 1999. Are there general laws in ecology? - *OIKOS* 84: 177–192.
- Lecointre, G. 2011. La biodiversité : au pays des aveugles le borgne est roi. - *Comptes Rendus Palevol* 10: 331–334.
- Lecointre, G. and Le Guyader, H. 2001. *Classification phylogénétique du vivant.* Belin. - Belin.
- Leopold, A. 1949. *A Sand County Almanac: And Sketches Here and There.* - Oxford University press.
- Levin, S. 1992. The problem of pattern and scale in ecology. - *Ecology* 73: 1943–1967.
- Lomolino, M. V. et al. 2010. *Biogeography.* - Sinauer Associates.
- MacArthur, R. and Wilson, E. 1967. *The theory of island biogeography.*
- Morrone, J. J. and Crisci, J. V. 1995. *Historical Biogeography: Introduction to Methods.* - *Annu. Rev. Ecol. Syst.* 26: 373–401.
- Mouillot, D. et al. 2013. Rare species support vulnerable functions in high-diversity ecosystems. - *PLOS Biol.* in press.
- Mouquet, N. et al. 2012. Ecophylogenetics: advances and perspectives. - *Biol. Rev.* 87: 769–785.
- Nelson, G. 1974. *Historical Biogeography: An Alternative Formalization.* - *Syst. Biol.* 23: 555–558.
- Ødegaard, F. 2000. How many species of arthropods? Erwin's estimate revised. - *Biol. J. Linn. Soc.* 71: 583–597.
- Olson, V. a et al. 2009. Global biogeography and ecology of body size in birds. - *Ecol. Lett.* 12: 249–259.
- Pavoine, S. and Bonsall, M. B. 2011. Measuring biodiversity to explain community assembly: a unified approach. - *Biol. Rev.* 86: 792–812.
- Petchey, O. L. and Gaston, K. J. 2006. Functional diversity: back to basics and looking forward. - *Ecol. Lett.* 9: 741–758.
- Platnick, N. I. and Nelson, G. 1978. A Method of Analysis for Historical Biogeography. - *Syst. Zool.* 27: 1.
- Popper, K. R. 1968. *The Logic of Scientific Discovery.* - Routledge.
- Qian, H. 2009. Global comparisons of beta diversity among mammals, birds, reptiles, and amphibians across spatial scales and taxonomic ranks. - *J. Syst. Evol.* 47: 509–514.
- Qian, H. and Ricklefs, R. E. 2012. Disentangling the effects of geographic distance and environmental dissimilarity on global patterns of species turnover. - *Glob. Ecol. Biogeogr.* 21: 341–351.
- Ricklefs, R. E. 1987. Community diversity: relative roles of local and regional processes. - *Science* 235: 167–71.
- Ricklefs, R. E. 2008. Disintegration of the Ecological Community. - *Am. Nat.* 172: 741–750.
- Ronquist, F. and Sanmartín, I. 2011. *Phylogenetic Methods in Biogeography.* - *Annu. Rev. Ecol. Evol. Syst.* 42: 441–464.
- Root, T. 1988a. Environmental factors associated with avian distributional boundaries. - *J. Biogeogr.* 15: 489–505.
- Root, T. 1988b. Energy constraints on avian distributions and abundances. - *Ecology* 69: 330–339.
- Rosenzweig, M. L. 1995. *Species Diversity in Space and Time.* - Cambridge University Press.
- Sanmartín, I. et al. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. - *Biol. J. Linn. Soc.* 73: 345–390.
- Sanmartín, I. and Ronquist, F. 2004. Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. - *Syst. Biol.* 53: 216–43.
- Sclater, P. L. 1858. On the general Geographical Distribution of the Members of the Class Aves. - *J. Proc. Linn. Soc. London. Zool.* 2: 130–136.
- Simpson, G. G. 1944. *Tempo and Mode in Evolution.* - Columbia University Press.
- Simpson, G. G. 1980. *Splendid isolation : the curious history of South American mammals.* - New Haven : Yale University Press.

Introduction générale

- Storch, D. et al. 2012. Universal species-area and endemics-area relationships at continental scales. - *Nature* 488: 78–81.
- Thomas, C. D. et al. 2004. Extinction risk from climate change. - *Nature* 427: 145–148.
- Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. - *Proc. Natl. Acad. Sci. U. S. A.* 102: 8245–8250.
- Thuiller, W. et al. 2010. Resolving Darwin's naturalization conundrum: a quest for evidence. - *Divers. Distrib.* 16: 461–475.
- Vane-Wright, R. I. et al. 1991. What to protect?—Systematics and the agony of choice. - *Biol. Conserv.* 55: 235–254.
- Violle, C. et al. 2007. Let the concept of trait be functional! - *Oikos* 116: 882–892.
- Wallace, A. 1876. *The geographical distribution of animals.* - Cambridge Univ. Press.
- Webb, C. O. et al. 2002. Phylogenies and Community Ecology. - *Annu. Rev. Ecol. Evol. Syst.* 33: 475–505.
- Wegener, A. 1912. Die Entstehung der Kontinente. - *Geol. Rundschau* 3: 276–292.
- Whittaker, R. J. et al. 2005. Conservation Biogeography: assessment and prospect. - *Divers. Distrib.* 11: 3–23.
- Wiens, J. 1989. Spatial scaling in ecology. - *Funct. Ecol.* 3: 385–397.
- Wiens, J. and Donoghue, M. 2004. Historical biogeography, ecology and species richness. - *TRENDS Ecol. Evol.* 19: 639–644.
- Winter, M. et al. 2012. Phylogenetic diversity and nature conservation: where are we? - *Trends Ecol. Evol.* 28: 199–204.

PARTIE I

CLASSER LA DIVERSITÉ DES DIVERSITÉS PHYLOGÉNÉTIQUES ET FONCTIONNELLES

Partie I

*« Je regardai alentour, et, je ne sais pourquoi, je vous assure que jamais, jamais auparavant
cette terre, [...] cette jungle [...] ne m'avaient paru
si privés d'espoir, si sombres, si impénétrables à la pensée humaine. »*

Joseph Conrad

Au coeur des ténèbres, 1899



La jungle vue d'un inselberg, Guyane française © M. Réjou-Méchain

SOMMAIRE DE LA PARTIE I

Articles associés à la partie I	52
Annexes associées à la partie I	52
INTRODUCTION. Les différentes facettes de la biodiversité	53
1. Un constat alarmant	54
2. Une triple nécessité	55
3. Les grandes facettes de la biodiversité	55
CHAPITRE I. Une tentative de synthèse exhaustive des indices phylogénétiques α et β	58
Résumé	61
Article 1	63
CHAPITRE II. L'importance de l'échelle phylogénétique	90
Résumé	91
Article 2	93
SYNTHÈSE & DISCUSSION. Vers une unification ?	108
1. Synthèse du travail	109
2. Utilité de l'approche	111
3. Limites et perspectives du travail	111

ARTICLES ASSOCIÉS À LA PARTIE I

ARTICLE 1

C. M. Tucker, M. W. Cadotte, S. B. Carvalho, J. T. Davies, S. Ferrier, S. A. Fritz, R. Grenyer, M. R. Helmus, L. S. Jin, A. O. Mooers, S. Pavoine, O. Purschke, D. W. Redding, D. F. Rosauer, M. Winter & **F. Mazel**. *A unification of phylogenetic metrics for conservation, community ecology and macroecology*.

En révision dans *Biological Reviews*

ARTICLE 2

F. Mazel, Davies, T.J, Gallien L., Groussin, M., Münkemüller T. & Thuiller, W. *Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics*.

Accepté dans *Ecography*

ARTICLE 3

Redding, D.W., **Mazel, F.**, and A.O. Mooers. *Measuring evolutionary isolation for conservation*.

PloS one, 9(12), e113490.

ANNEXES ASSOCIÉES À LA PARTIE I

ANNEXE 1.1. Matériel supplémentaire de l'article 1

ANNEXE 1.2. Matériel supplémentaire de l'article 2

ANNEXE 1.3. Article 3

ANNEXE 1.4. Résumé d'un article additionnel

INTRODUCTION DE LA PARTIE I

LES DIFFÉRENTES FACETTES DE LA BIODIVERSITÉ

1. Un constat alarmant

L'utilisation des diversités phylogénétiques et fonctionnelles a connu un essor fulgurant ces 25 dernières années (voir la figure 4 de l'introduction générale). En effet, il apparaît que leur utilisation peut permettre de mieux comprendre (1) les processus générant les distributions des espèces à toutes les échelles spatiales (Webb et al. 2002, Graham and Fine 2008, Mouquet et al. 2012) et (2) le fonctionnement et les services des écosystèmes (Díaz et al. 2007, Srivastava and Cadotte 2012). Ce potentiel prometteur est cependant obscurci par les trop nombreux indices (ou métriques) qui constituent actuellement une véritable **jungle** (Mouquet et al. 2012, Winter et al. 2012). Bien que la diversité phylogénétique ou fonctionnelle d'un assemblage d'espèces ne puisse pas se mesurer qu'avec un seul nombre, il paraît inconcevable qu'il y ait des dizaines de façons indépendantes et également intéressantes de mesurer la diversité phylogénétique d'une seule communauté (voir matériel supplémentaire 1 de l'article 1). En fait, une part substantielle de ces indices portent la même information : il existe même des indices qui possèdent des noms différents mais qui sont exactement identiques. Certains auteurs ont déjà pointé ce problème (Swenson 2011) :

« This is therefore another clear example of either authors not being aware of contemporary metrics or not comparing their 'new' metric to known existing metrics first either via simulation or by comparing their equations. »

Il est donc urgent de faire le tri dans cette jungle. Nathan Swenson le fait justement remarquer (2011) :

« As the number of metrics continues to grow, further studies will be needed to show which new metrics actually provide novel information and strengthen the statistical toolkit of the phylogenetic community ecologist. »

La première partie de cette thèse va donc tenter de relever ce défi en se demandant :

***Peut-on synthétiser les indices de diversités
le long de grandes « dimensions » de l'information?***

2. Une triple nécessité

L'objectif est donc de construire une carte simplifiée de la jungle des indices qui pourra éventuellement permettre aux écologistes de ne plus se perdre. La construction d'une telle carte a trois objectifs :

- (1) répertorier chacun des indices de la jungle (c'est-à-dire s'y perdre plus d'une fois)
- (2) tracer des sentiers autour de groupes d'indices similaires
- (3) faire en sorte que ces sentiers délimitent des groupes d'indices certes similaires mais surtout utiles pour tester des hypothèses écologiques précises.

On retrouvera l'inventaire (presque !) exhaustif des indices dans le matériel supplémentaire 1 de l'article 1 et la table 1 de l'annexe 1. Je présente maintenant les grandes lignes des sentiers que j'ai choisi.

3. Les grandes facettes de la diversité

Je pense qu'une classification des indices est possible, tout du moins pour ses composantes majeures. Je pense qu'il existe de grandes dichotomies sur lesquelles on peut tenter une classification :

(1) Dichotomie au niveau des types de distances considérées.

C'est la distinction entre diversité fonctionnelle et diversité phylogénétique. Chacune de ces distances est intéressante en soi et leur comparaison paraît aussi prometteuse (d'où la nécessité de disposer d'un cadre méthodologique et conceptuel identique pour les deux).

(2) Dichotomie au niveau de l'objet biologique considéré.

La diversité fait forcément référence à un objet biologique donné. Cet objet peut se définir à plusieurs échelles biologiques : l'espèce, la communauté ou la paire de communautés.

<i>Objet</i>	<i>Diversité</i>
Espèce	Originalité
Communauté	α
Paire de communautés	β

Cette organisation correspond en partie à la distinction classique proposée par Whittaker (1960) entre diversités α , β et γ . Cette distinction est originellement basée sur la diversité taxonomique, soit le nombre d'espèces d'un assemblage local (diversité α), le nombre d'espèces d'un assemblage régional (diversité γ) et la différenciation moyenne entre les communautés (diversité $\beta = \alpha/\gamma$). Les diversités α et γ sont de nature identique, à l'échelle spatiale près. On considère donc ici deux mesures fondamentales :

- la diversité au niveau d'un seul site : la diversité α
- la différence de composition entre deux sites : la diversité β
- finalement, j'ajouterai ici la quantification d'un autre niveau organisationnel : l'espèce elle-même. En effet, la littérature regorge de mesure « d'originalité » phylogénétique ou fonctionnelle (Isaac *et al.* 2007; Buisson *et al.* 2010, voir Annexe 1).

On peut donc déjà ranger les indices dans six grandes cases distinctes (deux types de distances * trois unités biologiques). Les chapitres I et II explorent les cases phylogénétiques en répertoriant et classant systématiquement chacun des indices de diversité des communautés (c'est-à-dire α et β) alors que l'annexe 1 se concentre sur les originalités phylogénétiques des espèces. J'explore donc ici les mesures phylogénétiques uniquement, mais je précise en discussion les nombreux parallèles avec les diversités et originalités fonctionnelles.

CHAPITRE I

UNE TENTATIVE DE SYNTHÈSE EXHAUSTIVE DES INDICES PHYLOGÉNÉTIQUES α ET β .

« Diversity indices do, however, have a problem. They combine the variable we are trying to measure [...]. I find it better to keep variables separated. That sharpens both questions and analyses. »

Michael L. Rosenzweig (1995)

Ce premier chapitre vise à classifier les indices de diversité phylogénétique α et β en grandes familles ou « variables » comme le précise Rosenzweig.

Résumé de l'article 1 (en révision dans *Biological reviews*)

L'utilisation des phylogénies en conservation, en écologie des communautés et en macro-écologie est devenue commune et a élargi notre compréhension de la diversité biologique. Ces sous-disciplines ont reconnu l'intérêt de l'utilisation des relations évolutives mais ont développé des approches phylogénétiques de façon relativement indépendante. La prolifération des indices de diversité phylogénétique consécutive à ce manque de collaboration empêche aujourd'hui les méta-analyses, les synthèses et les généralisations des résultats existants. Par ailleurs, il est difficile pour les utilisateurs de sélectionner un indice correspondant à une question donnée car les indices ne sont pas bien reliés aux questions de recherche. Pour améliorer ce choix, l'application et l'interprétation des indices de diversité phylogénétique, nous proposons d'organiser les indices existant dans un cadre méthodologique unificateur.

Les questions concernant l'information phylogénétique dans les assemblages d'espèces peuvent se résumer en trois points (les « variables » auxquelles Rosenzweig fait référence dans la citation): quelle quantité ? Quelle différence ? Quelle régularité ? Nous montrons que ces trois questions reflètent trois dimensions d'un arbre phylogénétique : sa richesse, sa divergence et sa régularité. Nous classifions et répertorions 73 indices de diversité phylogénétique. À partir de cette liste, nous identifions trois indices représentatifs de ces dimensions (les « variables » auxquelles Rosenzweig fait référence dans la citation). Dans le cadre de la diversité alpha, ces indices représentatifs sont :

- (1) la PD_{Faith} (Faith 1992) représente la somme des longueurs de branches reliant toutes les espèces de la communauté.
- (2) le MPD (pour « mean pairwise distance » Webb *et al.* 2002) représente la distance phylogénétique moyenne entre toutes les espèces d'un assemblage.
- (3) la VPD (pour « variance of pairwise distance ») représente la variance des distances phylogénétiques d'un assemblage.

Ce cadre méthodologique est dans un premier temps validé avec des simulations et nous permet d'identifier les indices à cheval sur plusieurs dimensions. Nous montrons que le choix de l'indice repose sur le lien entre la question de recherche et la dimension correcte. Le guide proposé dans ce papier aidera les chercheurs à se retrouver dans la jungle actuelle des indices.

Note sur la participation

Cet article est le fruit d'une collaboration initiée par M. W. Cadotte, J. T. Davies, A. Mooers et D. Rosaeur dans le cadre d'un workshop à Leipzig sous la tutelle du centre allemand pour la recherche intégrative sur la biodiversité (iDiv & sDiv) coordonné par M. Winter. C'est mon directeur de thèse, Wilfried Thuiller, qui a soutenu et permis ma participation à ce travail de synthèse, notamment pour partager mes connaissances sur la jungle des indices. Mon apport majeur dans ce travail est la production de la table de classification (Figure 1 de l'article 1) et de celle répertoriant tous les indices utilisés (Matériel supplémentaire 1).

A guide to phylogenetic metrics for conservation, community ecology and macroecology

Caroline M. Tucker¹, Marc W. Cadotte^{2,3}, Silvia B. Carvalho⁴, Jonathan Davies⁵, Simon Ferrier⁶, Susanne A. Fritz⁷, Rich Grenyer⁸, Matthew R. Helmus^{9,10}, Lanna S. Jin¹¹, Arne O. Mooers¹², Sandrine Pavoine^{13,14}, Oliver Purschke^{15,16}, David W. Redding¹⁷, Dan. F. Rosauer¹⁸, Marten Winter¹⁵, **Florent Mazel**¹⁹.

¹ Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado, USA.

² Biological Sciences, University of Toronto-Scarborough, Scarborough, Ontario, Canada.

³ Stake Key Laboratory of Biocontrol, Key Laboratory of Biodiversity Dynamics and Conservation of Guangdong, Higher Education Institutes, College of Ecology and Evolution, Sun Yat-sen University, Guangzhou, PR China.

⁴ CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto, Porto, Portugal.

⁵ Department of Biology, McGill University, Montreal, Quebec, Canada.

⁶ CSIRO Ecosystem Sciences, Climate Adaptation Flagship, Canberra, Australia.

⁷ Biodiversity & Climate Research Centre (BiK-F) and Senckenberg Gesellschaft für Naturforschung, Frankfurt, Germany.

⁸ School of Geography and the Environment, University of Oxford.

⁹ Department of Animal Ecology, Vrije Universiteit, Amsterdam, The Netherlands.

¹⁰ Center for Biodiversity, Department of Biology, Temple University, Philadelphia, PA 19122 USA.

¹¹ Ecology & Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada.

¹² Department of Biology, Simon Fraser University, Vancouver, BC, Canada.

¹³ Centre of Ecology and Conservation Sciences, French Museum of Natural History, Paris, France.

¹⁴ Department of Zoology, University of Oxford, U.K.

¹⁵ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig

¹⁶ Geobotany and Botanical Garden, Institute of Biology, Martin Luther University of Halle-Wittenberg, Halle (Saale), Germany.

¹⁷ Centre for Biodiversity and Environmental Research, Department of Genetics, Evolution and Environment, University College London, London, Britain.

¹⁸ Research School of Biology, Australian National University, Canberra, Australia.

¹⁹ Univ. Grenoble Alpes & CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France.

The use of phylogenies in ecology has become common and has broadened our understanding of biological diversity. Ecological sub-disciplines, particularly conservation, community ecology and macroecology, all recognize the value of evolutionary relationships but the resulting development of phylogenetic approaches has led to a proliferation of phylogenetic diversity metrics. The use of many metrics across the sub-disciplines hampers potential meta-analyses, syntheses, and generalizations of existing results. Further, there is no guide for selecting the appropriate metric for a given question, and different metrics are frequently used to address similar questions. To improve the choice, application, and interpretation of phylo-diversity metrics, we organize existing metrics by expanding on a unifying framework for phylogenetic information.

Generally, questions about phylogenetic relationships within or between assemblages tend to ask three types of questions: How much? How different? or How even? We show that these questions reflect three dimensions of a phylogenetic tree: *richness*, *divergence*, and *regularity*. We classify the 73 existing phylo-diversity metrics based on their mathematical form within these three dimensions and identify 'anchor' representatives: for α -diversity metrics these are PD (Faith's phylogenetic diversity), MPD (mean pairwise distance), and VPD (variation of pairwise distances). By analyzing mathematical formulas and using simulations, we use this framework to identify metrics that mix dimensions, and we provide a guide to choosing and using the most appropriate metrics. We show that metric choice requires connecting the research question with the correct dimension of the framework and that there are logical approaches to selecting and interpreting metrics. The guide outlined in this paper will help researchers to navigate the current jungle of indices.

Key Words. *Biodiversity hotspots, Biogeography, Community assembly, Conservation, Diversity metrics, Evolutionary history, Phylogenetic diversity, Prioritization, Range size*

Contents.

I. Introduction

II. A unifying framework for phylo-diversity metrics

- (1) Describing the framework
- (2) Classifying phylo-diversity metrics using the dimensions framework
 - (2a) *Richness*
 - (2b) *Divergence*
 - (2c) *Regularity*
- (3) Analysing the dimensions framework through simulations

III. Additional complexities in metric choice

- (1) Abundances
- (2) Parametric Indices (Hill numbers and entropies).
- (3) Metrics that depend on species richness

IV. Connecting ecological questions and hypotheses with phylo-diversity metrics

- (1) Applying richness metrics
- (2) Applying divergence metrics
- (3) Applying regularity metrics

V. A guide to phylo-diversity metrics

- (1) Metric selection for ecologists
- (2) An example of metric selection

VI. Moving forward

VII. Conclusion

VII. Acknowledgements

IX. Works Cited

I. Introduction.

Phylogenetic information is a critical component of modern ecology, particularly for the sub-disciplines of macroecology, community ecology, and conservation biology (Beck *et al.*, 2012; Cadotte *et al.*, 2010b; Crozier, 1992; Davies & Cadotte, 2011b; Faith, 1992; Mouquet *et al.*, 2012; Purvis *et al.*, 2000; Vane-Wright, Humphries & Williams, 1991; Webb, 2000; Webb *et al.*, 2002; Winter *et al.*, 2009). The growing use of phylogenies recognizes that the branching pattern on a phylogenetic tree reflects the accumulation of phenotypic, genetic, behavioural, and/or phenological differences between evolutionary lineages (Harvey & Pagel, 1991). These accumulated differences may in turn describe, explain, or predict biological and ecological processes. The potential usefulness of phylogenies to answer ecological questions, coupled with the exponentially growing availability of phylogenies for many taxonomic groups, has given rise to a proliferation of different phylogenetic metrics. Currently, there is an increasing “jungle of [phylogenetic] indices” (Winter, Devictor & Schweiger, 2013) of more than 73 available metrics used to describe phylogenetic diversity (here “phylo-diversity”, Appendix S1). This jungle reflects not only the increasing number of phylo-diversity metrics found in the literature, but also the confusion regarding how the different metrics relate to each other in both mathematical and ecological terms.

Metric choice is often driven by historical precedence, individual experience, and sub-discipline tradition, rather than objective criteria. Phylo-diversity metrics first appeared in conservation biology in response to the perception that, in the face of widespread extinctions, minimizing loss of evolutionary diversity should be a priority (Vane-Wright *et al.*, 1991). Maximizing the evolutionary diversity of a group of species should maximize their feature (i.e. phenotypic, behavioral and/or ecological) diversity, and so phylogenetic measures should be more effective than species-based measures at preserving such diversity (Faith, 1992). Community ecology and macroecology more recently incorporated phylogenies into analyses, using evolutionary relationships to understand observed ecological and macroevolutionary patterns and processes, such as community assembly or biodiversity gradients. Community ecology tends to use phylogenetic relatedness between taxa or communities to infer local ecological processes (Webb *et al.*, 2002) or to

predict ecosystem properties (Mouquet *et al.*, 2012), while macroecology uses phylogenetic information to help disentangle explanations for large-scale patterns of diversity (Fritz & Rahbek, 2012; Jetz *et al.*, 2012; Wiens *et al.*, 2004; Winter *et al.*, 2009). Despite these different foci, there is considerable overlap in sub-discipline approaches and interests. Indeed, some metrics are commonly considered across all ecological sub-disciplines, such as Faith’s PD (Faith, 1992), while others are isolated to particular sub-disciplines, e.g. ED (Isaac, 2007; Redding, 2003) for conservation; MPD and MNTD for community ecology (Clarke & Warwick, 1998; but see Davies & Buckley, 2011a; Webb *et al.*, 2002).

Choosing the most appropriate phylogenetic measure for a particular ecological question is complicated by the vast collection of phylo-diversity metrics available. Recent overviews (Chiu, Jost & Chao, 2014; Mouquet *et al.*, 2012; Vellend *et al.*, 2010; Winter *et al.*, 2013) have called for a unifying framework for phylo-diversity metrics, in order to clarify the conceptual relationships between existing metrics, to highlight their redundancies or complementarity, and to ultimately encourage the correct usage and interpretation of metrics (Chao, Chiu & Jost, 2010; Chao, Chiu & Jost, 2014; Chiu *et al.*, 2014; Faith, 2008; Pavoine & Bonsall, 2011; Pavoine, Love & Bonsall, 2009; Pearse *et al.*, 2014; Rosauer & Mooers, 2013; Schweiger *et al.*, 2008; Swenson, 2011; Tucker & Cadotte, 2013). A recent framework developed by Pavoine & Bonsall (2011) based on preliminary work by Mouillot *et al.* (2005) and Ricotta (2007), provides a broad clarification of the mathematical underpinnings of phylo-diversity metrics, allowing them to be grouped under three mathematical dimensions (*richness, divergence, regularity*). Although the Pavoine-Bonsall framework represents an important step forward in clarifying the conceptual relationships underlying metrics, it lacks (1) comprehensive classification of metrics, since it included only a small subset of published phylo-diversity metrics, and (2) guidance for the correct choice of metrics and connection of research questions with the appropriate dimension. The purpose of this article is to provide a comprehensive and practical guide to understanding and correctly applying phylo-diversity metrics for ecological questions. This should help in selecting from among the 73 metrics currently available, while emphasizing the value in distinguishing between and utilizing the three different dimensions of phylo-diversity metrics.

We establish the connection between the types of ecological questions or hypotheses researchers test and the corresponding dimension identified by the Pavoine-Bonsall framework. Our goals are to incorporate existing phylo-diversity metrics into the framework and verify their fit, analyse redundancy and distinguish among metrics within dimensions, and provide examples to guide their use.

II. A unifying framework for phylo-diversity metrics

(1) Describing the framework.

Despite the vast array of phylo-diversity metrics, a simple set of mathematical underpinnings provides a natural scheme to group the metrics into three conceptual dimensions (Pavoine *et al.*, 2009): *richness*, *divergence*, and *regularity*. These dimensions capture the mathematical operation inherent to a metric, which includes: (a.) the *sum* of accumulated phylogenetic differences among taxa (“richness”); (b.) the *mean* of the phylogenetic relatedness among taxa (“divergence”), representing the average phylogenetic difference between taxa in an assemblage; or (c.) the *variance* in differences among taxa, representing how regular the phylogenetic differences between taxa in an assemblage are (“regularity”, Figure 1). We use *regularity* rather than the similar term “evenness”, because the latter has previously been used to describe how abundances are combined with

measures of evolutionary distances (Pavoine *et al.*, 2011; Pearse *et al.*, 2014).

A second axis of information in this framework reflects whether the metric uses information about a single set of tips of a phylogenetic tree within an assemblage (i.e. questions about a single community or regional species pool, hereafter referred to as α -diversity), or about several sets of tips (i.e. comparisons of assemblages over space or time, hereafter referred to as β -diversity). For the purposes of this paper, *assemblage* simply denotes a group of taxa of interest: such taxa may, but need not, co-occur in space or time. Examples include taxa in a local community, regional species pool, or those selected in a particular conservation strategy. And similarly, though we may refer to species for simplicity, note that the metrics discussed are often applied to diversity below species level, or where species have not been described.

This framework (1) provides an intuitive approach based on the mathematical formulations of the metrics, (2) assesses both within- and between-assemblage diversity components, (3) is analogous to the functional diversity framework, thus aiding comparisons between phylo- and functional diversity (Villéger *et al.*, 2008) and (4) is applicable for both abundance and presence/absence formulations.

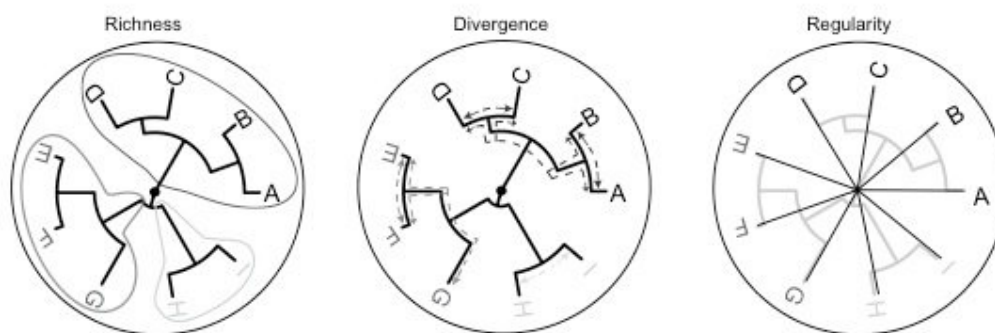


Figure 1. Conceptual diagrams illustrating the calculation of each of the 3 dimensions of phylogenetic information: richness (left), divergence (centre), and regularity (right). The branching diagram in each image is a phylogenetic tree representing the inferred evolutionary relationship between taxa A-I. Taxa A-I are grouped within three assemblages (A-D, dark grey; E-G, mid-grey; and H-I, light grey). Tree branches represent accumulated differences between taxa.

Within each dimension, different phylogenetic metrics can be constructed using various types of phylogenetic components (referred to as *units*): these include branch lengths, pairwise phylogenetic distances between taxa, measures of phylogenetic or evolutionary isolation (e.g. species distinctiveness, fair proportion; Isaac, 2007; Redding, 2003), or other measures of tree topology. Thus within each dimension, we differentiate between metrics based on the phylogenetic unit used for their construction. We refer to groups of metrics in a particular dimension that are constructed using the same units as “families” For example, richness metrics composed using branch lengths (e.g. PD, PE) would be considered a family based on their shared dimension and unit of construction.

(2) Classifying phylo-diversity metrics using the dimensions framework.

Here we use the dimensions framework introduced in (1) to classify phylo-diversity metrics used to answer phylogeny-focused ecological questions. We searched the ecological literature and identified many common metrics not classified within the Pavoine-Bonsall framework, for a total of 73 metrics. These metrics are diverse, but share many common properties (Pavoine *et al.*, 2011; Pearse *et al.*, 2014; Vellend *et al.*, 2010). These metrics are included in Table 1, classified based on the mathematical dimension (richness, divergence, or regularity) and the diversity level (α -diversity, Table 1-A; β -diversity, Table 1-B). Formulas for all metrics and additional details are found in Appendix 1 (metrics from Allen, Kon & Bar-Yam, 2009; Barker, 2002; Bryant *et al.*, 2008; Cadotte *et al.*, 2010b; Chen *et al.*, 2012; Chiu *et al.*, 2014; Clarke *et al.*, 1998; Clarke & Warwick, 2001; Colless, 1982; Dehling *et al.*, 2014; Faith, 1992; Hardy & Jost, 2008; Hardy & Senterre, 2007; Helmus *et al.*, 2007; Isaac, 2007; Ives & Helmus, 2010; Izsák & Papp, 2000; Izsák & Szeidl, 2002; Jost, 2006; Jost, 2007; Kembel *et al.*, 2010; Leinster & Cobbold, 2012; Lozupone *et al.*, 2007; Lozupone & Knight, 2005; Miller, Zanne & Ricklefs, 2013; Mouchet & Mouillot, 2011; Nipperess, Faith & Barton, 2010; Pavoine *et al.*, 2009; Pavoine, Ollier & Pontier, 2005; Pavoine & Ricotta, 2014; Pybus & Harvey, 2000; Rao, 1982; Redding, 2003; Rosauer *et al.*, 2009; Safi *et al.*,

2013; Scheiner, 2012; Swenson, 2011; Vellend *et al.*, 2010; Villéger *et al.*, 2008; Webb, Ackerly & Kembel, 2008; Webb *et al.*, 2002; Weiher & Keddy, 1995). Table 1 also identifies the absence of published phylo-diversity metrics in some categories; for example, β -diversity (row B) in particular lacks metrics in a number of possible categories. However, particularly for α -diversity metrics, most categories in Table 1 include multiple phylo-diversity metrics. Further, while it is likely that additional metrics will be found in the literature and missed here (or will be developed in the future), we believe they can be easily placed within this framework.

(2a) Richness, Table 1-1.

Richness metrics sum up the quantity of phylogenetic differences present in an assemblage, and we can further distinguish metrics according the type of basic units they sum across. Metrics may sum branch lengths (e.g. Faith's phylogenetic diversity or "PD"; Faith, 1992; Table 1-1.1a); pairwise phylogenetic distances (e.g. Phylogenetic Species Richness, "PSR"; Helmus *et al.*, 2007; Table 1-1.2a); or they may sum the evolutionary distinctiveness of the taxa in an assemblage (or total Evolutionary Distinctness, "ED"; Jetz *et al.*, 2014; e.g. phylogenetic endemism, "PE"; Rosauer *et al.*, 2009; Table 1-1.3a). Richness metrics that compare diversity between sets (Table 1-b) may calculate the proportion of shared branch length between two communities (e.g. Unifrac, Lozupone *et al.*, 2005; Table 1-1.1b) or a proportional measure of pairwise phylogenetic similarities among species (Pavoine *et al.*, 2014; Table 1-1.2b). These metrics capture the difference in phylogenetic composition between assemblages.

(2b) Divergence, Table 1-2.

The divergence dimension contains metrics that average the distribution of units extracted from a phylogenetic tree. Divergence metrics that describe a single assemblage may be calculated using branch lengths (Table 1-2.1a), pairwise distances (Table 1-2.2a), or phylogenetic isolation (Table 1-2.3a). Distances may be measured using all pairwise distances for a group of taxa (e.g. Mean Pairwise Distance, "MPD"; Webb *et al.*, 2002; Table 1.2.2.1a) or only a subset

of the possible pairwise distances (Table 1-2.2.2a), in which case generally the shortest distances between taxa are considered (e.g. Mean Nearest Taxon Distance, "MNTD"; Webb *et al.*, 2002). For metrics that compare divergence between assemblages (Table 1-2b), distances may be measured using branch lengths (e.g. H_β , which relies on additive decomposition, Mouchet *et al.*, 2011; Table 1-2.1b), or else all (Table 1-2.2.1b) or the shortest pairwise distances (Table 1-2.2.2b).

(2c) Regularity, Table 1-3.

The regularity dimension contains metrics that characterize how the phylogenetic tree differs from a star phylogeny (i.e. a phylogeny in which all species are equally unrelated). In other words these metrics quantify how regularly species are located along the phylogenetic tree and how evenly distant they are from each other species. They are grouped into three families based on the use of tree topology (Table 1-3.1a), pairwise distances (all or a subset of pairwise distances; Table 1-3.2a), or phylogenetic isolation (Table 1-3.3a). We did not identify any published metrics comparing regularity between assemblages.

(3) Analysing the dimensions framework through simulations.

Table 1 represents consensus built on analysis of mathematical formulations as well author opinions, yet metric behaviour may still vary within any particular subcategory (e.g. a cell in Table 1), since metrics that share general characteristics may still differ in behaviour. Additionally, some metrics may integrate components from more than one dimension of a phylogeny. Given the grouping of metrics into the three conceptual dimensions in Table 1, we predict that the similarity of metrics within a dimension should be higher (e.g. correlations between their values should be higher) than the similarity of metrics in different dimensions. We use simulations (described briefly below, and in detail in Appendix 2) to evaluate the coherence of metrics within their presumed dimension, and to identify any metrics that deviate notably from their dimension.

We simulated 100 phylogenetic trees with 64 taxa, with a wide distribution of branch lengths

(δ statistic) and tree symmetry (I_c statistic) using *sim.bdtree* function in the R package *geiger* (Harmon *et al.*, 2008). For each tree we created 8 types of landscapes (containing 256 communities) that represented simplified outcomes of possible assembly processes, using the *scape* function in the R package *pez* (Pearse *et al.*, 2015). These landscapes varied in whether (1) there was a phylogenetic signal (sensu Blomberg, Garland & Ives, 2003) in species' environmental optima, and whether that signal reflected 'repulsion' (divergence of optima), 'attraction' (convergence of optima), or no phylogenetic correlation, (2) range size (repulsion, attraction, no signal), and (3) spatial autocorrelation in range distribution (Table 2). Thus, assemblages in these landscapes could vary from having spatially autocorrelated ranges and a strong phylogenetic signal for range size and environmental optima, to having random assembly with no phylogenetic structure in range size or environmental optima and no spatial autocorrelation. In total there were 8 landscapes simulated for each tree, for a total of 800 distinct landscapes. For the communities in each of these 800 landscapes, we calculated values for the α -diversity and β -diversity metrics listed in Table 1.

In addition to these analyses using the trees with 64 taxa, we generated 100 trees with 16 taxa and 100 trees with 256 taxa. For each tree size we similarly simulated 800 landscapes, thus constructing a total of 2400 landscapes. We calculated a subset of α -diversity metrics (omitting abundance weighted and parametric metrics, for a total of 27 metrics) from Table 1-A across each of these landscapes to consider briefly whether metric behaviour is sensitive to tree size.

We explored the underlying relationships between α -diversity metrics using principal component analysis (PCA) based on a pairwise Spearman correlation matrix between all α -diversity metrics for each landscape. We use Spearman correlation, as it is robust to nonlinear relationships and outliers. We included all metrics from Table 1-A for the analysis: for visual clarity, Figure 2 presents only presence/absence metrics and excludes parametric metrics and

metrics that include multiple dimensions (these metrics are presented in Figure 5a & 5b). The total explained variance when all metrics were included for analysis was 44.2% for principal component 1, and 20.3% for principal component 2. The remaining axes explain much less variation (PC 3 explained 6.0%), and so we display only the first two axes. Note that we use PCA here as a technique for visualizing the relationships between metrics: principle component axes are orthogonal and independent, and thus not expected to be equivalent to the three dimensions we have identified. Though the dimensions capture different aspects of that phylogeny, ultimately all are dependent on the same underlying processes of evolution. Hence, fewer than three independent PCA axes should be necessary to capture variation related to these three dimensions.

PCA results for the α -diversity metrics (Figure 2) suggest that the richness, divergence, and regularity dimensions are clearly divided within ordination space and to illustrate this we use the PD, MPD, and VPD metrics as anchors or guidelines for the expected position. The majority of the richness, divergence, and regularity metrics cluster with other metrics from the same dimension. In general, richness metrics load on positive values of PC 1 and 2, divergence metrics load on negative values of the second axis, while regularity metrics tend to load on negative values of the first axis. Divergence, as represented by the position of MPD, is also captured by phylogenetic species variability (PSV) and the average taxonomic diversity (AvTD) because mathematically these metrics are identical (Appendix 2). It is notable however, that for the divergence dimension, metrics that rely on nearest neighbour distances (MNTD) or phylogenetic isolation (mean(ED)) do not cluster closely with MPD, which is composed using all pairwise phylogenetic distances. Regularity, as represented by the position of VPD, is closely correlated with a number of similar metrics, including variance in nearest taxonomic distance (VNTD) and the variance in evolutionary distinctiveness (var(ED)). These results confirmed that the similarity of metrics within a dimension is generally greater (e.g. correlations between their values is higher) than the

similarity of metrics that are found in different dimensions, although interesting divergences also occur (see Section III).

β -diversity metrics (Figure 3) capture the dissimilarity between assemblages. The first PC axis explains 22.9% of variance and the second PC axis 15.6%. The first axis captures a gradient from the richness dimension (positive values) to the divergence dimension (negative values). The entropic metrics (Chiu *et al.* (2014), ${}^qD(T)$) vary from being highly correlated with the richness dimension ($q=0$), to being increasingly associated with the divergence metrics as the value of q increases. The second axis captures a separate source of information from the dimensions framework presented here: Swenson (2011) showed that β -diversity metrics can emphasize either differences among communities towards the base (“basal”) or tips (“terminal”) of a phylogenetic tree. In our simulations large values along the second axis appear to capture basal metrics, such as Rao’s D (equivalent to Dpw), while known terminal metrics such as Dnn fall close to zero. This suggests that for β -diversity metrics, metric choice should additionally consider whether it is more of interest to capture internal versus terminal tree structure (Jin, Cadotte & Fortin, 2015; Swenson, 2011).

Tree size of the source pool used to simulate communities influenced the similarity of metrics: the variation among α -diversity metrics increased when trees were small (16 taxa; Figure 4). In general, the multidimensional space occupied by metrics measured on landscapes constructed with 64 and 256 taxa trees overlapped, while the metrics calculated for the trees with 16 taxa tended to occupy separate areas of the ordination. Further, some metrics (e.g. I_c , upper right hand corner, Figure 4) behave similarly regardless of tree size. To completely understand the sensitivity or robustness of the different phylo-diversity metrics to changes in tree size, more in-depth analyses would be required. Nonetheless we feel our conclusions about metric behaviour within the dimensions framework will be generally across a variety of tree sizes with the smallest trees accentuating the small differences in metric calculations because the phylogenetic signal exhibited by

small clades is inherently variable (Blomberg *et al.*, 2003).

III. Additional complexities in metric formulation.

Though the metrics generally group by dimensions, we also consider a number of additional factors that can alter the usage and interpretation of phylo-diversity metrics. Metric behaviour may be complicated by factors such as the inclusion of abundances, underlying correlations with species richness, and the emphasis on rare versus common species.

(1) Abundances, Figure 5a.

Species abundances are often an important source of information: for example, weighting schemes using abundances or range sizes have a long history for conservation prioritization (Vane-Wright *et al.*, 1991). All of the dimensions of phylogenetic information can be weighted using some measure of abundance or other weight that allows information about species' commonness or rarity to be incorporated (Table 1, metrics in red and plotted alongside presence/absence metrics in Figure 5a).

Abundances may be incorporated in several ways. For metrics applied to a local assemblage, species' relative abundances may be incorporated. This is the ratio between species' absolute abundance (e.g. cover, number of individuals, or biomass) and total absolute abundance of the community (e.g. total cover, total number of individuals or total biomass). Relative abundances are then used to weight phylogenetic units such as pairwise distances (e.g. MPD_{Ab} ; Miller *et al.*, 2013; Table 1-2.2.1a) or branch lengths (e.g. PD_{Aw} ; Vellend *et al.*, 2010; Table 1-1.1a). There are two general weighting schemes: 1) those where locally abundant species are de-emphasized relative to locally rare species, which are weighted more highly because they may be important for conservation (Cadotte & Davies, 2010a); versus 2) those that emphasize abundant species, such as when analyzing species contributions to ecosystem function (Cadotte *et al.*, 2009). Such weighting schemes allow the impact of the number and relative abundance of rare and distinctive species in local communities to be considered explicitly. For

questions that consider a larger spatial scale, range sizes or endemism may be alternate sources of abundance information (Isaac, 2007). For example the phylogenetic endemism metric (PE; Rosauer *et al.*, 2009; Table 1-1.1a), weights the length of a particular branch by its whole geographic extent (defined as the union of the distribution of species descending from it). Thus rarity is defined as species with small ranges, rather than low abundances.

Simulation results (Figure 5a) suggest that the incorporation of abundance into metrics in the divergence and regularity dimensions produced metrics that behaved similarly to other metrics in the same dimensions. Therefore, when abundances within an assemblage are of interest for questions about divergence and regularity, appropriate metric choices exist and can be clearly interpreted and reasonably compared to presence-absence metrics in those dimensions. For richness, the abundance weighted metrics AED and PE clustered with other metrics in the richness dimension. However, several richness metrics that include weighting for abundance, such as abundance-weighted phylogenetic diversity (PD_{aw}) or average abundance weighted phylogenetic diversity (ΔnPD), did not cluster with the richness metrics. It may be that sensitivity of these indices to patterns in abundance evenness leads their behaviour to converge with that of indices from the other phylogenetic dimensions, and so the user should use caution when interpreting these particular metrics.

(2) Parametric Indices (Hill numbers and entropies), Figure 5b.

Hill numbers are a group of diversity measures that aim to quantify diversity in units of equivalent numbers of equally abundant species (Hill, 1973). Hill numbers incorporate information about abundances and variance in abundances, and retain constant units ("effective number of species"), and have recently been extended to include phylogenetic information (Chao *et al.*, 2010; Chiu *et al.*, 2014; Leinster *et al.*, 2012). These frameworks rely on a unified formula of phylo-diversity that is adjusted using a single "scaling parameter", q . The value of q determines the influence of rare taxa.

Within the richness/divergence/regularity framework, changes in the value of the q scaling parameter affect the dimension of the parametric index being measured, rather than simply altering the influence of rare species, as is the case for the taxonomic versions of these metrics (Figure 5b). For example Chao's (2010) $q=0$ and $q=2$ metrics correspond to a richness and a divergence metric respectively. In contrast the Leinster framework (Leinster *et al.*, 2012; ${}^qD^Z(p)$) varied very little with q in how it classified communities in our framework. Note that the Scheiner framework (Scheiner, 2012), although reliant on q , differs from the parametric indices in that it sums species level evolutionary distinctiveness (ED) instead of summing across edges in the phylogeny. Parametric indices require further theoretical treatment and applications to fully determine their properties, and so care must be taken in selecting and interpreting them.

(3) Metrics that depend on species richness, Figure 5b.

The original Pavoine-Bonsall framework was defined independently of species richness so that multiplying an index by species richness (e.g. $PSR=PSV \times \text{Species richness}$) would not change its classification (both PSR and PSV were considered in Pavoine-Bonsall framework as divergence indices). Richness is here defined more broadly to include any counting of evolutionary units, be they branch lengths or other phylogenetic distances. As a result, some metrics are classified differently from the original work, and our richness dimension is intrinsically more influenced by species richness than the other two dimensions.

Several indices that we *a priori* classified in other dimensions were found to be correlated strongly with the richness dimension, as a result of their underlying relationships with species richness. We suggest this explains the behaviour of the parametric indices based on Hill numbers, discussed above, as these are by definition dependent on species richness. H_{ED} , H_{AED} , and metrics in the Scheiner framework combine phylogenetic regularity with species richness, which leads their behaviour to be strongly

correlated with the richness dimension (PD), as the result of their strong underlying correlation with species richness (since PD and species richness are also often highly correlated). Rao's Quadratic Entropy (Rao's QE; Table 1-2.2.1a) is primarily an index of divergence (Clarke *et al.*, 1998; Rao, 1982) as most of its variation occurs along the divergence dimension—but it is also slightly correlated with the richness dimension since it includes the diagonal of the distance matrix. E_{ED} behaves differently from both its component dimensions because it quantifies the deviation from even ED and so is uncorrelated with species richness.

For some metrics, simple transformations can be applied to remove this dependence. For example, indices based on Hill numbers can be divided by species richness to remove its effect. The effects of species richness and abundance evenness can be removed by using appropriate null models (Pavoine *et al.*, 2013), which can also ease comparisons with functional diversity, since species richness and abundance evenness may artificially exaggerate correlations between phylogenetic diversity and functional diversity.

IV. Connecting ecological questions and hypotheses with phylo-diversity metrics.

The dimensions classification framework unites metrics developed in across ecological sub-disciplines and used for different purposes. However, the framework does not easily resolve the problem of choosing among metrics for a particular analysis. Ecological questions, whether from conservation, community ecology, or macroecology, all consider how accumulated differences between species (reflected by divergences along a phylogeny) may relate to biological processes or patterns. Evolutionary history is considered an outcome or predictor of processes of interest. We suggest that questions about these processes or patterns can be simplified and unified to recognize the three general themes of *How much total diversity is present in an assemblage (or among assemblages)*, *How different, on average, are taxa in an assemblage (or among assemblages)*, and/or *How regular or variable are the differences between taxa in an assemblage (or among assemblages)* (Figure 6). We review below the types of

questions asked by ecologists using phylogenies, identify commonalities, and connect these questions with appropriate phylo-diversity metrics.

(1) Applying richness metrics (Figure 6, column 1).

Richness metrics can be used to measure or describe observed patterns of diversity; these values may also be compared to equivalent taxonomic and functional measures. As richness metrics sum the quantity of phylogenetic differences in an assemblage, they are often assumed to capture ‘feature diversity’ under some models of trait evolution (Kelly, Grenyer & Scotland, 2014). In this context, measures of phylogenetic richness may be used to answer questions about the quantity and distribution of extant biodiversity, arguably better than species-based metrics (Rosauer *et al.*, 2013). Phylogenetics offers metrics which are relatively insensitive to taxonomic inflation (Isaac, 2007), and which can easily incorporate taxa (or other evolutionary units) for which there is little information, other than their placement on the tree of life. Feature diversity may be considered a valuable indicator of either future utility or future evolutionary potential (Forest *et al.*, 2007; Mace, Gittleman & Purvis, 2003) and so conservation biologists have been interested in the protection of total feature diversity for questions of prioritization of taxa and/or areas (e.g. Bennett *et al.*, 2014; Forest *et al.*, 2007; Isaac, 2007; Jetz *et al.*, 2014; Purvis, 2008; Rodrigues *et al.*, 2011). For example, Tucker *et al.* (2012) asked how Proteaceae phylogenetic diversity was distributed spatially in the Cape Floristic Province. To capture the total evolutionary richness in a spatial unit, they considered two richness metrics to consider – PD, and the sum of abundance-weighted ED (BED) – and compared the distributions of these metrics with Proteaceae species richness in the region.

Phylogenetic richness (either α - or β -diversity) has also been used as a predictor or response variable in numerous studies, across multiple spatial or temporal scales and for diverse natural systems. Variation in phylogenetic richness through space and time is often hypothesized to

be an outcome of different ecological and evolutionary processes (Cavender-Bares *et al.*, 2009; Mouquet *et al.*, 2012). For example, as invasive species represent a non-random combination of traits, and phylogenetic metrics can be used to capture such ‘feature diversity’, it may be hypothesized that invasion should lead to differential changes in phylogenetic richness (α - or β -diversity) compared to species richness. Winter *et al.* (2009) tested this by comparing taxonomic and phylogenetic richness metrics in invaded assemblages (ultimately showing that alien species led to a decrease in phylogenetic distinctness (i.e. divergence) rather than richness). In a separate application Thuiller *et al.* (2011) found that species’ vulnerability to climate change clustered weakly along the phylogeny, and used this relationship to predict how the amount and distribution of phylogenetic richness will change in the future.

(2) Applying divergence metrics (Figure 6, column 2).

Questions about ecological communities have frequently considered phylogenetic distance to be a proxy for differences in functional traits (Ackerly, 2009; reviewed in Freckleton, Harvey & Pagel, 2002; Mouquet *et al.*, 2012; Srivastava *et al.*, 2012), with the assumption that closely related species are more functionally similar, and thus overlap more in their ecological niche, than those that are more distantly related (Connolly *et al.*, 2011; Gerhold *et al.*, 2015; but see Narwani *et al.*, 2013; Purschke *et al.*, 2013; Violle *et al.*, 2011). Underlying this are additional assumptions that closely related species occur in sympatry and that trait evolution is divergent, so the most similar taxa are the most closely related (Gerhold *et al.*, 2015). When these assumptions hold, it is often hypothesized that if environmental filtering drives community assembly, taxa within an assemblage will be more related on average than expected in a random or null assemblage (Cavender-Bares & Wilczek, 2003; but see Mayfield & Levine, 2010; Webb *et al.*, 2002). Alternatively, if competitive interactions are important, it may be hypothesized that co-occurring taxa will be less related (i.e. more divergent) than expected on average. Divergence indices, particularly MPD and MNTD indices, have been used to test these

types of hypotheses about the mean relatedness of taxa within an assemblage. For example, Helmus *et al.* (2010) considered whether disturbed communities tended to contain more closely related species, reflecting the role of environmental filtering in selecting disturbance-tolerance taxa. They hypothesized that more closely related species might have similar traits, and so be similarly adapted to disturbance conditions. To test this, the authors used the PSV metric, which is closely related to MPD, and compared the average relatedness of species in disturbed communities versus non-disturbed communities.

Note that although these are frequently expressed hypotheses in community ecology, there are many possible relationships between phylogenetic relatedness and co-occurrence that can be tested using divergence metrics. Gerhold *et al.* (2015) provide alternative scenarios that may preclude the interpretations described above – for example, trait similarity may actually facilitate coexistence (see also Mayfield *et al.*, 2010), competitive exclusion may be incomplete in assemblages, and regional species pools and processes, rather than local processes, may determine local assembly. Thus, testing questions about evolutionary history requires both identifying the correct type of metric for a given question as well as considering the assumptions that might relate patterns to processes.

The phylogenetic topology of species' assemblages can further provide information about processes structuring regional species pools (Heard & Cox, 2007; Purvis *et al.*, 2011), and the likelihood that these will be invaded or altered (Gerhold *et al.*, 2011). Macroecological studies have incorporated information about divergences in phylogenies to compare phylogenetic distances separating sister lineages and capture variation in diversification rates (e.g. Ackerly, 2009; Weir & Schluter, 2007), to identify geographical centres of diversification (e.g. Jetz *et al.*, 2012), or the drivers of niche evolution or conservation (e.g. Dormann *et al.*, 2009; Wiens *et al.*, 2004). Such macroecological approaches allow tests of whether diversification rates differ between biogeographical regions, across latitudes or at different times through history. In

addition, patterns can be compared to null expectations generated from models that integrate the processes of speciation, extinction and colonization (Pigot & Etienne, 2015) providing more powerful tests of the mechanisms structuring regional species assemblages.

(3) Applying regularity metrics (Figure 6, column 3).

The regularity metrics appear less frequently in the literature, and we identified no published examples for β -diversity. They are typically used for questions about how evenly evolutionary history is distributed between taxa in an assemblage, and as with divergence metrics, are often applied with the assumption that phylogenetic distance is a proxy for differences in functional traits. Under such a framework, one might hypothesize that greater evenness in the distribution of similarity among species should result in lower competition (Kraft, Valencia & Ackerly, 2008). Cadotte (2013) manipulated phylogenetic relatedness and species richness in plant communities, and tested whether the selection effect (the dominance of highly competitive or productive species, one putative mechanism underlying the diversity-ecosystem function correlation), might be related to the topology of the phylogenetic tree. In fact, the selection effect was correlated with a regularity metric, IAC. As regularity metrics reflect evenness in the distribution of dissimilarity among species, this finding suggests that the selection effect is strongest when closely related species are present. In the field of macroecology, Davies and Buckley (2011a) considered variation in pairwise distances (VPD) to explore unevenness in the distribution of PD globally for terrestrial mammals, which provided insight into the historical processes behind global patterns of species richness.

V. A guide to phylo-diversity metrics.

Here we provide a robust and intuitive framework to guide researchers and practitioners on the selection and matching of phylo-diversity metrics to their research questions. We have shown that the different metrics align with three dimensions of phylo-diversity: richness, divergence, and regularity;

dimensions that themselves align naturally with common research questions (Figure 6). In highlighting this natural linkage between research questions and associated hypotheses, phylogenetic dimensions, and appropriate metrics, we hope to facilitate the growing usage of phylogenies in ecology. Further, we hope that this work will encourage researchers to choose amongst existing metrics rather than formulate new metrics that may have similar properties to those already in existence.

(1) Metric selection for ecologists.

Importantly, our classification framework predicts metric behaviour based on their mathematical properties. Although particular metrics have become entrenched in particular fields, we show there is mathematical redundancy among them and alternative metrics may be equally able to address similar questions. This suggests that the choice of metric could be simplified. Researchers must first specify whether they are most interested in describing properties within a set or between sets, and then determine whether their research question(s) necessitate the use of “how much”, “how different” or “how even” dimensions. By placing their questions within these dimensions, the researcher can then identify the set of most appropriate metrics to choose from (Figures 7). The choice of metric need not depend on discipline, or whether the taxa of interest represent those found in an experimental sample, ecological community, biogeographical region, or clade of conservation interest.

It is sometimes suggested that the choice of statistical analyses should be made *a priori*, during experimental design and before performing actual experiments (Underwood, 1997). Although this is frequently unfeasible, especially for studies using observational data, this is meant to prevent issues related to multiple comparisons or bias in variable selection. Such issues are similarly problematic as many phylo-diversity metrics are used interchangeably in analyses. Although simply comparing models for all possible metrics and selecting the one with the best explanatory power has been employed in the past (see, e.g. Cadotte, 2013), it results in poorly justified analyses and potentially

confusing inference and we do not recommend it. The general “phylogeny should be important” hypothesis that accompanies multi-metric analyses obscures interpretation. As we have presented throughout this paper, metrics from different dimensions should not be treated interchangeably as they represent different types of information and effectively test different hypotheses.

A recommended strategy is to find the most appropriate metric through *a priori* identification of the key components of the research hypothesis. If our question is about the total evolutionary diversity contained within a reserve (richness), or say, the average species distinctness (divergence) within assemblages across an environmental gradient, then the correct dimension should be straightforward to identify (Figure 6). Once the dimension associated with the question is identified, researchers can restrict their choices to those the associated column in Table 1. They can also reduce the possible metric choices to those for either α -diversity or β -diversity, depending on which they are asking questions about.

In some cases there can be validity in comparing multiple metrics within a dimension, particularly if the metrics have different properties (e.g. units, formulation) of interest to the researcher, and they are all appropriate to the particular question or conclusion. In addition, high redundancy amongst metrics within a dimension can make selection among them somewhat arbitrary. As a result it may usually be easier to select the ‘anchor’ metric that represents a dimension (for α -diversity, PD for richness, MPD or perhaps MNTD for divergence, and VPD for regularity), given their ease of interpretation and precedence in the literature. One might use MPD when questions relate to branching occurring deep within a tree versus MNTD for questions related to terminal branching. Users may consider alternatives to the anchor metrics if these provide better fit to a specific question and analysis. For example, parametric indices (i.e. Hill numbers across a range of values of q) allow users to fully consider the impact of rare and common species on evolutionary diversity. In addition, they may be used for comparative

analysis of parametric indices across phylogenetic, taxonomic, and functional diversity. Alternatively, users may want to account for abundances in some form (range sizes, rarity, etc.), or compare results for presence-absence and abundance-weighted versions of metrics (e.g. MPD vs. MPD_{ab}). In other cases, it may be reasonable to select a metric if direct comparison with previously published values is desired. Even in these cases, we stress that our simulation results can be used to guide interpretation of alternative metrics, including through comparison of their behaviour with that of anchor metrics.

The choice of phylogenetic units merits additional discussion. Metrics may be constructed using different countable units, including using branch lengths, pairwise phylogenetic distances between taxa, or evolutionary isolation. Each of these components can be measured similarly (i.e. time of evolution in millions of years), and should rather be understood as differing in their “targets” or objectives. If one wants to conserve as many evolutionary units as possible, branch lengths should be the target of interest. If users wish to consider competition among species, pairwise distances may be an appropriate choice of unit. Phylogenetic isolation should be the target if users want to conserve unique species or relate conservation prioritization or ecological processes to differences in the rate of diversification across the tree. However, it is important to note that metric behaviour often did not differ significantly between metrics within the same dimension but constructed using different units (Figure 2).

There could be some cases in which a researcher wishes to consider the effect of multiple dimensions on a process or variable (for example, how does primary productivity in a site relate to richness and divergence in its plant community?). It is possible to use an approach that matches questions and inference to the metric, while also assessing how the different dimensions influence the variable of interest. For this approach the analysis would consist of creating a statistical model that includes metrics from multiple dimensions. The logical way to do

this would be to select a metric within each dimension following the recommendations in the previous section and then investigate nested models with subsets of the metrics. In the example above, if model selection suggests that a model containing both richness and divergence metrics is best, a researcher might conclude that both the total amount of history and the relative spacing of species are important predictors of productivity.

(2) An example of metric selection.

To illustrate the process of metric selection with a more realistic scenario, we provide the following example. Consider an island mainland system with a number of plant species (Figure 7) for which a researcher is interested in how evolutionary diversity varies between the mainland and island (here, sites and islands will be referred to interchangeably). The researcher first asks whether there are different amounts of evolutionary history represented by the plant communities at each site. As this is a question about the amount or sum of units of evolutionary history, metric choice should focus on the richness dimension (see questions in Figure 6). They may hypothesize that the most distant site (Island C) will share the least amount of evolutionary history with the remaining sites, since its distance should decrease the probability that species arrive from the mainland or potentially encourage diversification *in situ* (MacArthur & Wilson, 1967; Wiens *et al.*, 2004). To compare the evolutionary history between the three sites, the researcher chooses a β -diversity richness metric (Table 1-1B, red metrics in Figure 3), perhaps a branch length based metric such as Faith’s PD $_{\beta}$ that captures the amount of shared evolutionary history. Figure 3 shows that a number of metrics cluster closely and appear to capture the richness dimension, and a choice between these (Unifrac, S_{Jaccard}, S_{Sorensen}) should yield very similar results.

Using the same system and phylogeny, the researcher asks additional questions regarding whether there are differences in how evenly evolutionary history is distributed within the plant assemblages on each of sites A-C. Given that they found that Island C shares the least amount of total evolutionary history with the other sites

(Figure 8-lower panel 1), and Island C contains a number of endemic species, the researchers may now be interested in whether diversification rates differ between sites and if the endemic species on Island C reflects a recent radiation. The researchers recognize that this relates to the evenness or regularity of the distribution of evolutionary history, and so they choose a metric from the regularity dimension (Table 1-3A). They hypothesize that assemblages should be least even on Island C because the biota might derive from independent colonization events by evolutionary distinct lineages (i.e. lineages separated by large phylogenetic distances) which subsequently radiated *in situ*, giving rise to clusters of species separated by short evolutionary distances. The researchers calculate VPD for each site, and can then compare these with the VPD expected were island biota randomly assembled. This hypothetical example illustrates how a researcher's various questions can be simply connected with phylo-diversity metrics and how careful choice leads to clear interpretation of results.

VI. Moving forward

One result of our categorization of metrics is that researchers may be encouraged to look beyond a single oft-used metric and dimension and compare different dimensions, as this is already done with the analogous framework for functional diversity (Villéger *et al.*, 2008). For example, PD and, less often, abundance-weighted versions such as PE, have been the dominant measure of phylogenetic information in conservation biology research. Although we question whether the multiplicity of phylogenetic metrics in the literature has in general advanced research, conservation biology, more than for the other fields, has limited its perspective to the richness dimension alone. Metrics from the divergence or regularity dimensions might provide complementary information about the distribution of biodiversity across taxa in a site, for example.

We hope that this paper stimulates broader thinking and discussion about the use of phylogenies in ecology, conservation and macroecology. While this paper represents a starting point for deciding which metrics to apply

to data, and more analyses and work is required to advance our understanding of these metrics.

In our analyses we simulated a large number of landscape types and randomly generated trees in order to capture typical or average metric behaviour. However, parameter space is vast and future simulations should consider additional phylogeny and landscape attributes that may also influence metric behaviour. For example, it is understood that PD is strongly correlated with SR, though less so when trees are unbalanced or distinct species are also spatially restricted (Rodrigues, Brooks & Gaston, 2005; Tucker *et al.*, 2013). Thus, more complete consideration of parameter space is required to fully assess metric behaviour across a range of phylogenetic topologies and branch length distributions, as well as the phylogenetic signal strength in niche position and range size. Repeating tests with real, rather than simulated trees would also provide important information on expected values for metrics in natural systems. In addition, we still must determine whether metric correlations are robust to all types of phylogenies and landscapes, or whether certain types of perturbations inflate the importance of subtle metric differences.

We believe that our framework supplies guidance to researchers and practitioners on how to use and interpret results from phylogenetic analyses. As noted previously, the three dimensions we employ are simplifications, although their use in both functional and phylogenetic approaches suggests they have utility. In addition, simulation results largely support the classification of metrics and so we suggest that this framework should serve as a starting point for choosing metrics, applying questions and interpreting results.

VII. Conclusion.

(1) The use of phylogenies in community ecology, macroecology, and conservation biology reflects the shared recognition that accumulated evolutionary differences may explain or predict biological and ecological processes. Phylogenetic approaches have revolutionized these disciplines.

(2) The rapid growth of new phylogenetic metrics has limited the development of

phylogenetic methods in ecology and conservation, and prevents meta-analysis and clear interpretation of metrics.

(3) We suggest the intuitive, unifying framework of the phylogenetic dimensions – richness, divergence, and regularity – is very useful, since it applies to biological questions at multiple ecological scales, for single or multiple groups of species, and across fields.

(4) We encourage appropriate metric selection by highlighting links between research questions and metrics falling in the appropriate phylogenetic dimensions; interpretation is made simple by understanding the relationship between a metric's dimension and the mathematical basis of that dimension.

(5) Informed metric selection and interpretation will allow the use of published results across subfields and applications and encourage future work.

VIII. Acknowledgements.

This paper is a joint effort of the sPhy working group and an outcome of a workshop kindly supported by sDiv, the Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (DFG FZT 118). CMT, TJD, MWC, and AOM acknowledge NSERC Canada; MWC thanks the endowed TD Professor of Urban Forest Conservation and Biology chair. SBC was funded by a postdoctoral grant from Fundação para a Ciência e Tecnologia (FCT) (SFRH/BPD/74423/2010), and through the project PTDC/BIA-BIC/118624/2010-FCOMP-01-0124-FEDER-019676, supported by FEDER funds through the Operational Programme for Competitiveness Factors – COMPETE and by National Funds through FCT. MRH is supported by the Netherlands Organisation for Scientific Research (858.14.040). FM received funding from the European Research Council under the European Community's Seven Framework Programme FP7/2007-2013 Grant Agreement no. 281422 (TEEMBIO). SAF acknowledges funding by the LOEWE funding program, of Hesse's Ministry of Higher Education, Research, and the Arts

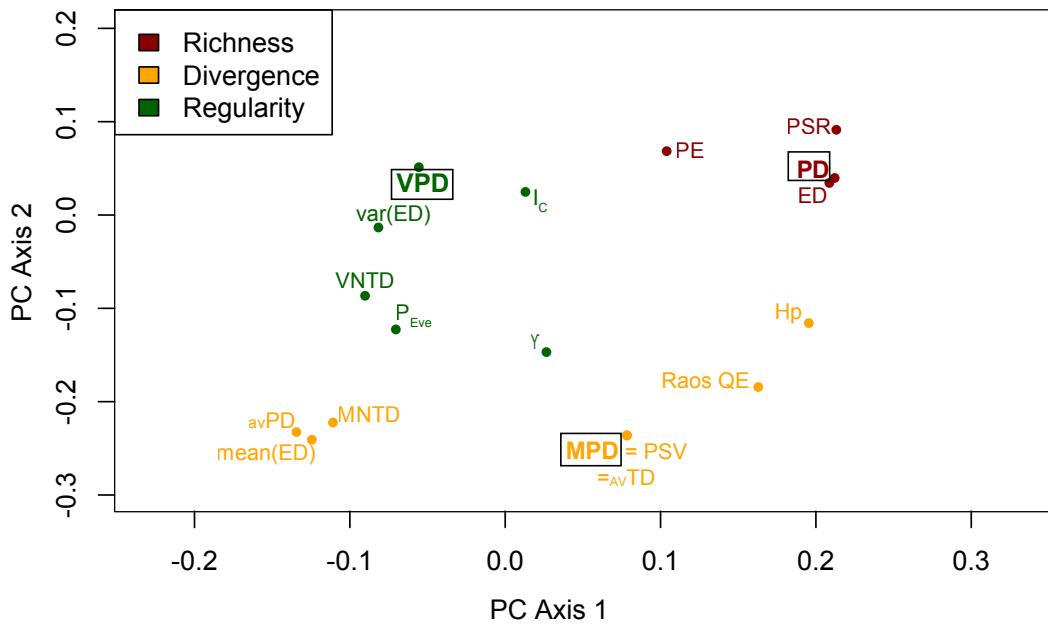


Figure 2. Principle Components Analysis for the Spearman correlation between the alpha-diversity metrics shown in Table 1A. Results represent measures taken from 800 simulated landscapes, based on 100 simulated phylogenetic trees and 8 landscape types defined in Table 2 (detailed methods in Appendix 2). Excluded from this plot are abundance-weighted metrics and those classified as parametric indices (see Figure 4 for their position). X-and Y-axes are scaled to reflect explained variance. Bolded, boxed metrics reflect ‘anchor’ metrics (PD, MPD and VPD) that align most closely with the richness, divergence and regularity dimensions, respectively.

DIMENSION => Unit => <=Level	1. Richness: How much? Sum of...		2. Divergence: How different? Mean of...		3. Regularity: How regular? Variance of...				
	1.1 Branch lengths	1.2 Pairwise distances	1.3 Phylogenetic isolation	2.1 Branch lengths	2.2 Pairwise distances/similarities	2.3 Phylogenetic isolation	3.1 Tree topology	3.2 Pairwise distances	3.3 Phylogenetic isolation
A. Diversity within sets	1.1a Amount of evolutionary history: Across species PD Across individuals AnPD Effective ${}^qD(T) * \&$ ${}^0PD *$ Per species PD _{sw} Endemic PE	1.2a Sum of pairwise distances: PSR F	1.3a Sum of evolutionary distinctiveness: sum(EDFP) AED	2.1a Sum of branch lengths divided by species richness: avPD Effective number of species given phylogenetic balance and abundance evenness: $q > 0$ ${}^qD(T) * \&$ & ${}^qPD *$ and associated entropies: H _p *, I _c *	2.2a Effective number of species given phylogenetic balance and abundance evenness: ${}^qD(p) *$ Mean of all distances including zero intra-species distances: Rao QE, PSE, J Mean inter-species distances: MPD, MPD _{sw} , AvTD, PSV	2.3a Mean shortest distance between a species and all others: MNTD MNTD _{ab}	3.1a Branching symmetry and distribution: I _c ; Y IAC	3.2a Variance of nearest neighbour distances: VNTD VNTD _{ab} PhyloEVE	3.3a Variance of species isolation metrics: Var(EDFP) E _{ed} HED HAED ${}^qD(P) *$ ${}^qD(AT) *$
	1.1b Amount of (un)shared evolutionary history: Phylosor, Unifrac Faith's PD _p Phylosor _{abr} Unifrac _{ab}	1.2b Sum of inter-set distances relative to intra-set distances: S _{Jaccard} , S _{Sokal-Sneath} , S _{Sørensen} , S _{Ochial} S _{secondary} S _{Sokal-Sneath} S _{Sørensen} S _{Ochial} (with absolute abundances)	NA	2.1b Effective number of sets: ${}^qD_p(T) *$ Amount of entropy gained on average by merging the sets: I _{sp} *, H _p *	2.2b Average inter-set relative to intra-set distances: S _{Jaccard} , S _{Sokal-Sneath} , S _{Sørensen} , S _{Ochial} (with rel. abundances) Average inter-set distances conditionally to intra-set (dis)similarities: PCD Amount of average distances gained on average by merging the sets: Rao DISC Proportion of average distances gained on average by merging the sets: Pst Proportion of average inter-species distances gained on average by merging the sets: Bst, Nst Average inter-set distances: COMDIST, Rao Q _{ab} , D _{sw}	2.2b Average minimum distance between a species of a set and all species in another set: COMDIS TNT	NA	NA	NA
B. Diversity between sets									

Table 1 (previous page). The dimensions classification system for phylo-diversity indices.

The row entries distinguish between α -(within a set) and β -(between sets) diversities while the column entries represent the three proposed dimensions of phylo-diversity. Within each of these three dimensions (namely **richness** (red), **divergence** (yellow), and **regularity** (green)), we distinguish between different phylogenetic units with which diversity can be computed (i.e. branch lengths, tree topology, pairwise phylogenetic distances or phylogenetic isolation). Species isolation indices measure taxa originality within a set (e.g. species distinctiveness, etc.) while pairwise distances represent a given set of patristic distances (we distinguish between pairwise distances that include all distances and those that include only the nearest distances). For each category we provide verbal descriptions and corresponding published indices. For example, the indices belonging to ‘ α -diversity/divergence indices/based on nearest pairwise distances’ can be defined as the “Mean shortest distance between a species and all others” with the corresponding published indices - MNTD and MNTD_{ab}. Empty categories indicate that there are no corresponding published indices (of which we are aware). Indices coloured in red can incorporate abundances. Starred indices (*entropies and Hill numbers) are classified as ‘parametric indices’ in Appendix S1, for their shared reliance on a single “scaling parameter”, q , that determines the influence of rare taxa (see also Figure 4b). Finally, some β -diversity indices cited are dedicated to pairwise comparisons between assemblages while others, like the parametric indices introduced below, can handle multiple comparisons among more than two assemblages.

Community Type	Phylogenetic signal, environmental optima	Environmental optima, signal type	Phylogenetic signal, range size	Range size, signal type	Spatial autocorrelation range
1	TRUE	Repulsion (0.2)	FALSE	N/A	TRUE
2	TRUE	Attraction (20)	FALSE	N/A	TRUE
3	TRUE	Repulsion (0.2)	TRUE	Repulsion (0.2)	TRUE
4	TRUE	Attraction (20)	TRUE	Attraction (20)	TRUE
5	FALSE	N/A	FALSE	N/A	TRUE
6	FALSE	N/A	FALSE	N/A	FALSE
7	FALSE	N/A	TRUE	N/A	TRUE
8	FALSE	N/A	TRUE	N/A	FALSE

Table 2. Landscape types used for simulations of metrics: numbers indicate parameter values in the scape function of pez. For a trait, values of 1 reflect Brownian motion, values less than 1 reflect rates of evolution that accelerate, values greater than 1 reflect rates of evolution that decelerate through time. See Appendix 2 for additional details.

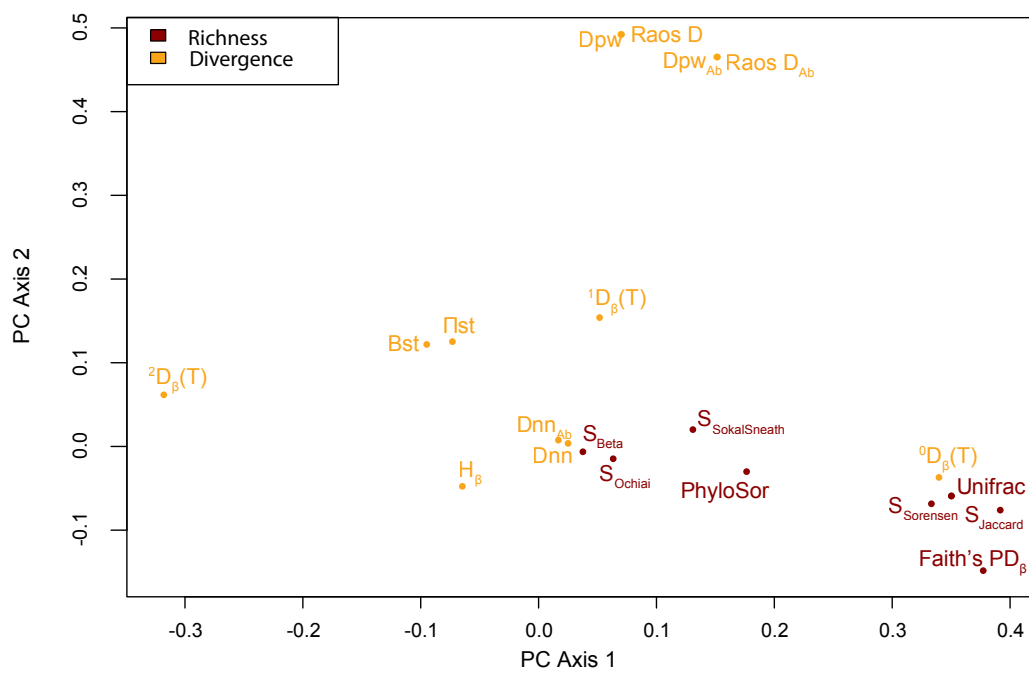


Figure 3. Principle Components Analysis of the Spearman correlations between the beta-diversity metrics shown in Table 1B. Results represent measures taken from 800 simulated landscapes and 8 landscape types defined in Table 2 (see Appendix 2 for details). X-and Y-axes are scaled to reflect explained variance.

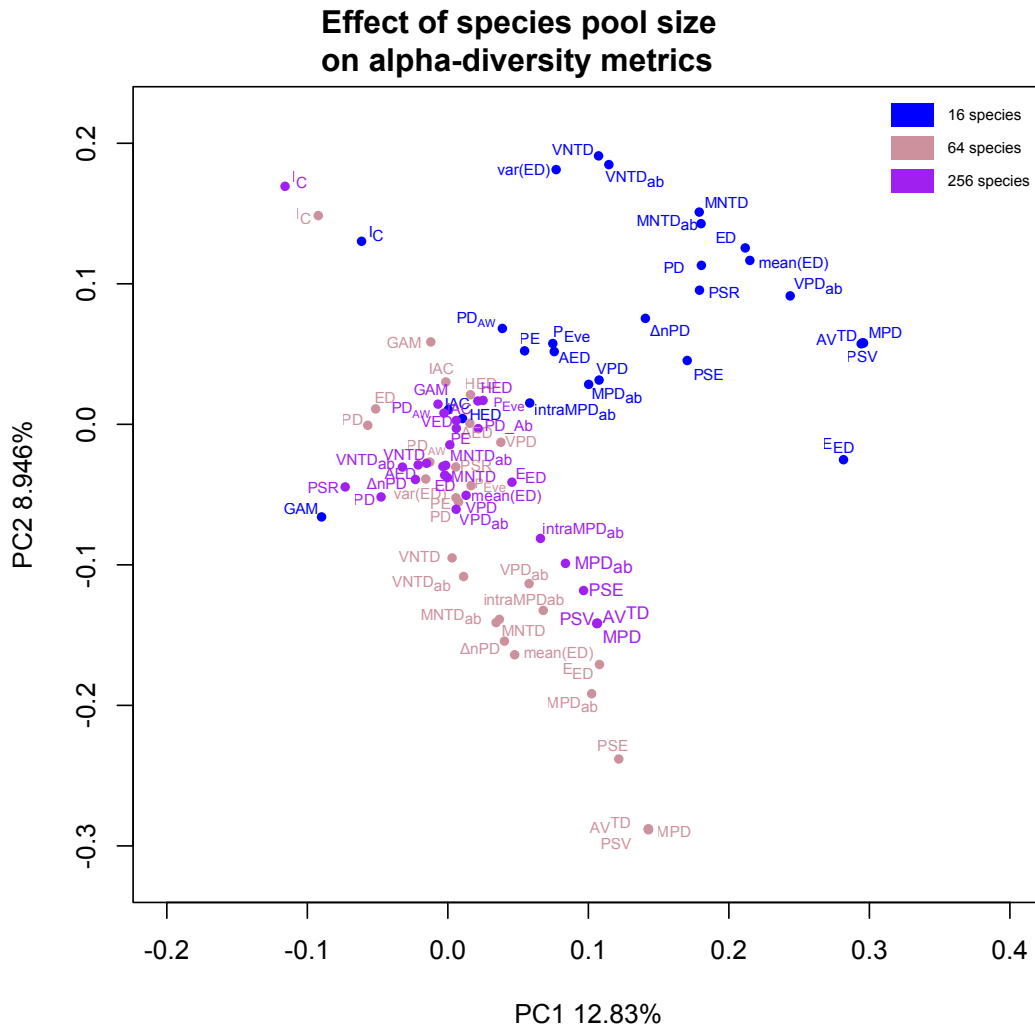


Figure 4. Effect of tree size on relationships between a subset of alpha-diversity metrics. Principle Components Analysis for the Spearman correlation between 27 commonly used α -diversity metrics shown in Table 1. Results represent measures taken from 2400 simulated landscapes, based on 300 simulated phylogenetic trees with three different sizes (16, 64 and 256 taxa, 100 trees of each size) and 8 landscapes types defined in table 2 (detailed methods in Appendix 2). Excluded from this plot are abundance-weighted metrics and those classified as parametric indices (see Figure 4 for their position). X- and Y-axes are scaled to reflect explained variance. Bolded, boxed metrics reflect ‘anchor’ metrics (PD, MPD and VPD) that align most closely with the richness, divergence and regularity dimensions, respectively.

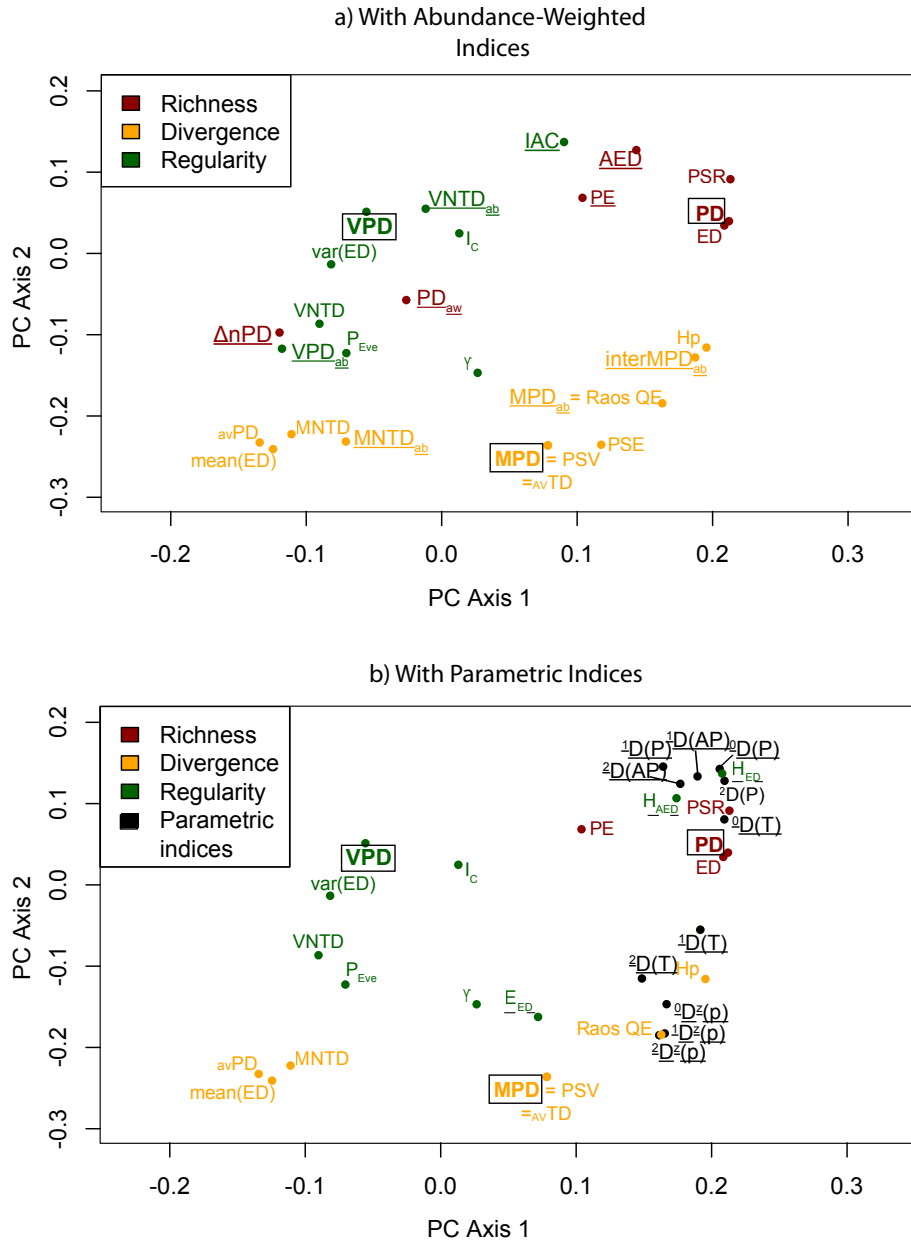


Figure 5. Principal Components Analysis showing the Spearman correlation for the alpha-diversity metrics shown in Table 1, results represent measures taken from 800 simulated landscapes (Appendix 2), including **A)** abundance-weighted metrics (underlined), or **B)** parametric indices (underlined and shown in black). X-and Y-axes are scaled to reflect explained variance (PC1 = 44.2%, PC2 = 20.3% for the PCA performed with all metrics). Bolded, boxed metrics reflect ‘anchor’ metrics (PD, MPD and VPD) that align with the richness, divergence and regularity dimensions, respectively.

Dimension	Richness: <i>How much?</i>	Divergence: <i>How different?</i>	Regularity: <i>How even?</i>
Definition	How much evolutionary history is associated with a set of tips?	How closely related (on average) are tips within a set?	How evenly is evolutionary history distributed between tips within a set?
Interpretation	<i>What is the total evolutionary history within (or between) assemblages?</i>	<i>How similar are species within (or between) assemblages?</i>	<i>How is similarity distributed within (or between) assemblages?</i>
Example Questions	(i) Evolutionary diversity as dependent variable		
	Is evolutionary history correlated with resource availability? How does compositional similarity decay with geographic distance? Do current protected areas adequately protect evolutionary history?	Is environmental filtering more important in high elevation communities compared to low elevation communities? Do geologically unstable regions have more recently diverged species? Where are concentrations of endemic evolutionary diversity?	Does biotic competition drive community structure? What is the relationship between evolutionary history and variation in function?
	(ii) Evolutionary diversity as independent variable		
	Are more diverse communities more productive? Does evolutionary diversity predict the spatial distribution of diversity? Do more diverse communities supply greater ecosystem services?	Are distinct species more invasive relative to community composition? Do island radiations influence bird co-occurrence patterns?	Do patterns of niche space occupancy affect community stability? Does a community with evenly distributed evolutionary history have greater potential to adapt to future habitats?

Figure 6. Connecting the dimensions framework to ecological questions. Practical definition of each dimension and example questions for each from community ecology, macroecology, and conservation biology. Included are questions for which evolutionary history is considered as a (a.) response or (b.) predictor of the processes of interest. Column colours correspond to the three dimensions from Table 1: richness (sum, red); divergence (mean distance, yellow); and regularity (variance, green)

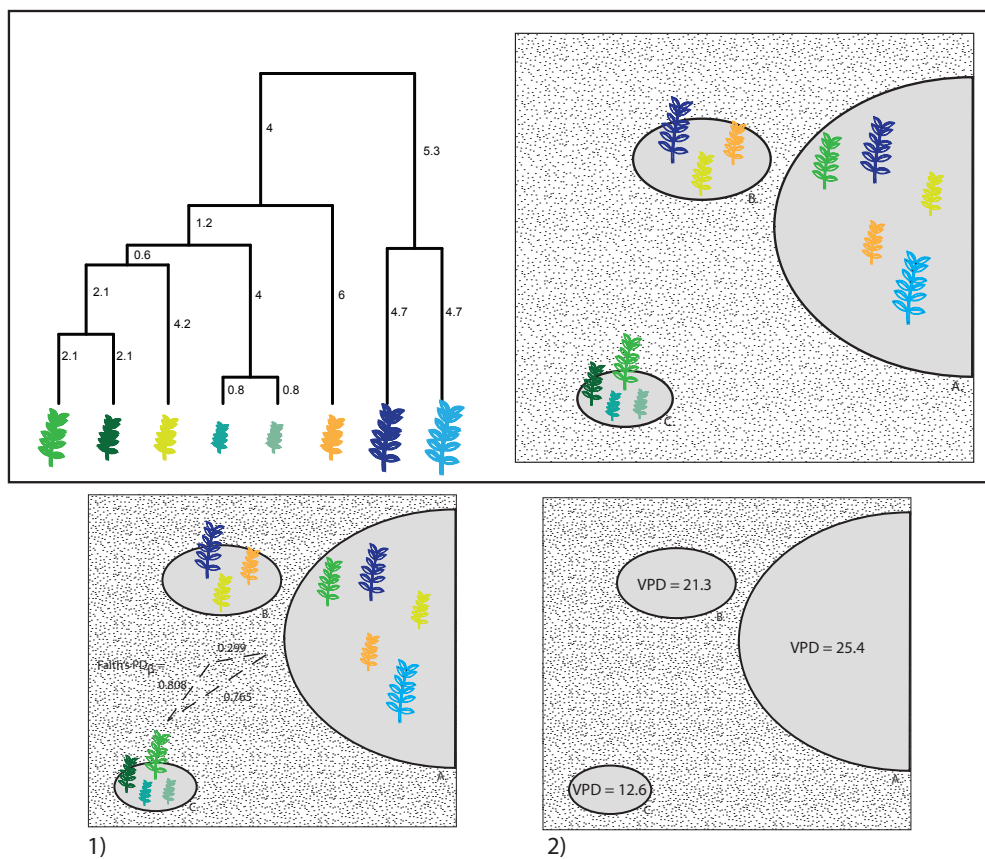


Figure 7. Example involving the flora in an hypothetical island mainland system (top panel), for which a researcher wishes to choose appropriate phylo-diversity metrics to test how evolutionary diversity varies between the mainland and island sites. Details in Section V-2. The top panel shows the distribution of species among the sites, and their phylogenetic relatedness. Values on the phylogeny represent hypothetical distances between species (e.g. branch lengths, etc). Lower panel 1 shows how evolutionary history is shared between sites (richness metric, β -diversity, Faith's PD_{β}). Lower panel 2 shows how evenly evolutionary history is distributed at each site (regularity metric, VPD).

IX. Works Cited.

- 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.
- ALLEN B. L., KON, M. & BAR-YAM, Y. (2009). A new phylogenetic diversity measure generalizing the shanno index and its application to phyllostomid bats. *American Naturalist* **174**, 236.
- BARKER G. M. (2002). Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation. *Biological Journal of the Linnean Society* **76**, 165-194.
- BECK J., BALLESTEROS-MEJIA, L., BUCHMANN, C. M., DENGLER, J., FRITZ, S. A., GRUBER, B., HOF, C., JANSEN, F., KNAPP, S. & KREFT, H. (2012). What's on the horizon for macroecology? *Ecography* **35**, 673-683.
- BENNETT J. R., ELLIOTT, G., MELLISH, B., JOSEPH, L. N., TULLOCH, A. I. T., PROBERT, W. J. M., DI FONZO, M. M. I., MONKS, J. M., POSSINGHAM, H. P. & MALONEY, R. (2014). Balancing phylogenetic diversity and species numbers in conservation prioritization, using a case study of threatened species in New Zealand. *Biological conservation* **174**, 47-54.
- BLOMBERG S. P., GARLAND, T. & IVES, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717-745.
- BRYANT J. A., LAMANNA, C., MORLON, H., KERKHOFF, A. J., ENQUIST, B. J. & GREEN, J. L. (2008). Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences* **105**, 11505-11511.
- CADOTTE M. W. (2013). Experimental evidence that evolutionarily diverse assemblages result in higher productivity. *Proceedings of the National Academy of Sciences* **110**, 8996-9000.
- CADOTTE M. W., CAVENDER-BARES, J., TILMAN, D. & OAKLEY, T. H. (2009). Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *Plos One* **4**, e5695.
- CADOTTE M. W. & DAVIES, T. J. (2010a). Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. *Diversity and Distributions* **16**.
- CADOTTE M. W., DAVIES, T. J., REGETZ, J., KEMBEL, S. W., CLELAND, E. & OAKLEY, T. H. (2010b). Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecology Letters* **13**, 96-105.
- CAVENDER-BARES J., KOZAK, K. H., FINE, P. V. A. & KEMBEL, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters* **12**.
- CAVENDER-BARES J. & WILCZEK, A. (2003). Integrating micro- and macroevolutionary processes in community ecology. *Ecology* **84**, 592-597.
- CHAO A., CHIU, C.-H. & JOST, L. (2010). Phylogenetic diversity measures based on Hill numbers. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 3599-3609.
- CHAO A., CHIU, C.-H. & JOST, L. (2014). Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers. *Annual Review of Ecology, Evolution, and Systematics* **45**, 297-324.
- CHEN J., BITTINGER, K., CHARLSON, E. S., HOFFMANN, C., LEWIS, J., WU, G. D., COLLMAN, R. G., BUSHMAN, F. D. & LI, H. (2012). Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **28**, 2106-2113.
- CHIU C.-H., JOST, L. & CHAO, A. (2014). Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecological Monographs* **84**, 21-44.
- CLARKE K. & WARWICK, R. (1998). Quantifying structural redundancy in ecological communities. *Oecologia* **113**.
- CLARKE K. & WARWICK, R. (2001). A further biodiversity index applicable to species lists: variation in taxonomic distinctness. *Marine Ecology Progress Series* **216**, 265-278.
- COLLESS D. (1982). Review of phylogenetics: the theory and practice of phylogenetic systematics. *Systematic Zoology* **31**, 100-104.
- CONNOLLY J., CADOTTE, M. W., BROPHY, C., DOOLEY, Á., FINN, J., KIRWAN, L., ROSCHER, C. & WEIGELT, A. (2011). Phylogenetically diverse grasslands are associated with pairwise interspecific processes that increase biomass. *Ecology* **92**, 1385-1392.
- CROZIER R. H. (1992). Genetic diversity and the agony of choice. *Biological conservation* **61**, 11-15.
- DAVIES T. J. & BUCKLEY, L. B. (2011a). Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2414-2425.
- DAVIES T. J. & CADOTTE, M. W. (2011b). Quantifying biodiversity - does it matter what we measure? In *Biodiversity Hotspots* (ed. F. E. Zachos and J. C. Habel). Springer.
- DEHLING D., FRITZ, S. A., TOPFER, T. & PACKERT, M. (2014). Functional and phylogenetic diversity and assemblage structure of frugivorous birds along an elevational gradient in the tropical Andes. *Ecography* **In Press**.
- DEVICOR 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. *Ecology Letters* **13**, 1030-1040.
- DORMANN C. F., GRUBER, B., WINTER, M. & HERRMANN, D. (2009). Evolution of climate niches in European mammals? *Biology Letters* **6**, 229-232.
- FAITH D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological conservation* **61**, 1-10.
- FAITH D. P. (2008). Phylogenetic diversity and conservation, pp. 99-115. Oxford University Press: New York, NY, USA.
- FOREST F., GRENYER, R., ROUGET, M., DAVIES, T. J., COWLING, R. M., FAITH, D. P., BALMFORD, A., MANNING, J., PROCHES, S., VAN DER BANK, M., REEVES, G., HEDDERSON, T. A. J. & SAVOLAINEN, V. (2007). Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**, 757-760.
- FRECKLETON R. P., HARVEY, P. H. & PAGEL, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist* **160**, 712-726.

Partie I - Chapitre I

- FRITZ S. A. & RAHBEK, C. (2012). Global patterns of amphibian phylogenetic diversity. *Journal of Biogeography* **39**, 1373-1382.
- GERHOLD P., CAHILL JR, J. F., WINTER, M., BARTISH, I. & PRINZING, A. (2015). Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Functional Ecology* **29**, 600-614.
- GERHOLD P., PÄRTEL, M., TACKENBERG, O., HENNEKENS, S. M., BARTISH, I., SCHAMINÉE, J. H., FERGUS, A. J., OZINGA, W. A. & PRINZING, A. (2011). Phylogenetically poor plant communities receive more alien species, which more easily coexist with natives. *The American Naturalist* **177**, 668-680.
- HARDY O. J. & JOST, L. (2008). Interpreting and estimating measures of community phylogenetic structuring. *Journal of Ecology* **96**, 849-852.
- HARDY O. J. & SENTERRE, B. (2007). Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *Journal of Ecology* **95**, 493-506.
- HARMON L. J., WEIR, J. T., BROCK, C., GLOR, R. E. & CHALLENGER, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129-131.
- HARVEY P. H. & PAGEL, M. (1991). *The comparative method in evolutionary biology*. Oxford University Press, Oxford, UK.
- HEARD S. B. & COX, G. H. (2007). The shapes of phylogenetic trees of clades, faunas, and local assemblages: exploring spatial pattern in differential diversification. *The American Naturalist* **169**, E107-E118.
- HELMUS M. R., BLAND, T. J., WILLIAMS, C. K. & IVES, A. R. (2007). Phylogenetic measures of biodiversity. *The American Naturalist* **169**, E68-E83.
- HELMUS M. R., KELLER, W., PATERSON, M. J., YAN, N. D., CANNON, C. H. & RUSAK, J. A. (2010). Communities contain closely related species during ecosystem disturbance. *Ecology Letters* **13**, 162-174.
- HILL M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology* **54**, 427-431.
- HUBBELL S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.
- ISAAC N. J. B. (2007). Mammals on the edge: conservation priorities based on threat and phylogeny. *Plos One* **2**.
- IVES A. R. & HELMUS, M. R. (2010). Phylogenetic metrics of community similarity. *The American Naturalist* **176**, E128-E142.
- IZSÁK J. & PAPP, L. (2000). A link between ecological diversity indices and measures of biodiversity. *Ecological Modelling* **130**, 151-156.
- IZSÁK J. & SZEIDL, L. (2002). Quadratic diversity: Its maximization can reduce the richness of species. *Environmental and Ecological Statistics* **9**, 423-430.
- JETZ W., THOMAS, G., JOY, J., HARTMANN, K. & MOOERS, A. (2012). The global diversity of birds in space and time. *Nature* **491**, 444-448.
- JETZ W., THOMAS, G., JOY, J., REDDING, D. W., HARTMANN, K. & MOOERS, A. O. (2014). Global Distribution and Conservation of Evolutionary Distinctness in Birds. *Current Biology* **24**, 919-930.
- JIN L. S., CADOTTE, M. W. & FORTIN, M.-J. (2015). Phylogenetic turnover implicates niche conservatism in montane plant species. *Journal of Ecology* **103**, 742-749.
- JOST L. (2006). Entropy and diversity. *Oikos* **113**, 363-375.
- JOST L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427-2439.
- KELLY S., GRENYER, R. & SCOTLAND, R. W. (2014). Phylogenetic trees do not reliably predict feature diversity. *Diversity and Distributions* **20**, 600-612.
- 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.
- KRAFT N. J., VALENCIA, R. & ACKERLY, D. D. (2008). Functional traits and niche-based tree community assembly in an Amazonian forest. *Science* **322**, 580-582.
- LEINSTER T. & COBBOLD, C. A. (2012). Measuring diversity: the importance of species similarity. *Ecology* **93**, 477-489.
- LOZUPONE C. A., HAMADY, M., KELLEY, S. T. & KNIGHT, R. W. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and environmental microbiology* **73**, 1576-1578.
- LOZUPONE C. A. & KNIGHT, R. W. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology* **71**, 8228-8235.
- LOZUPONE C. A. & KNIGHT, R. W. (2008). Species divergence and the measurement of microbial diversity. *FEMS Microbiology Review* **32**, 557-578.
- MACARTHUR R. H. & WILSON, E. O. (1967). *The theory of island biogeography*. Princeton University Press, Princeton.
- MACE G. M., GITTLEMAN, J. L. & PURVIS, A. (2003). Preserving the tree of life. *Science* **300**, 1707-1709.
- MAYFIELD M. M. & LEVINE, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters* **13**, 1085-1093.
- MILLER E. T., ZANNE, A. E. & RICKLEFS, R. E. (2013). Niche conservatism constrains Australian honeyeater assemblages in stressful environments. *Ecology Letters* **16**, 1186-1194.
- MOUCHET M. A. & MOUILLOT, D. (2011). Decomposing phylogenetic entropy into α , β and γ components. *Biological Letters* **7**, 205-209.
- MOUILLOT D., MASON, N. W., DUMANY, O. & WILSON, J. B. (2005). Functional regularity: a neglected aspect of functional diversity. *Oecologia* **142**, 353-359.
- MOUQUET N., DEVICTOR, V., MEYNARD, C. N., MUNOZ, F., BERSIER, L. F., CHAVE, J., COUTERON, P., DALECKY, A., FONTAINE, C. & GRAVEL, D. (2012). Ecophylogenetics: advances and perspectives. *Biological Reviews* **87**, 769-785.
- NARWANI A., ALEXANDROU, M. A., OAKLEY, T. H., CARROLL, I. T. & CARDINALE, B. J. (2013). Experimental evidence that evolutionary relatedness does not affect the ecological mechanisms of coexistence in freshwater green algae. *Ecology Letters* **16**, 1373-1381.
- NIPPERESS D. A., FAITH, D. P. & BARTON, K. (2010). Resemblance in phylogenetic diversity among ecological assemblages. *Journal of Vegetation Science* **21**, 809-820.
- PAVOINE S. & BONSALE, M. (2011). Measuring biodiversity to explain community assembly: a unified approach. *Biological Reviews* **86**, 792-812.
- PAVOINE S., GASC, A., BONSALE, M. & MASON, N. W. (2013). Correlations between phylogenetic and functional diversity: mathematical artefacts or true ecological and evolutionary processes. *Journal of Vegetation Science* **24**, 781-793.
- PAVOINE S., LOVE, M. S. & BONSALE, M. B. (2009). Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically-structured species assemblages: application to rockfish (genus: *Sebastes*) in

Partie I - Chapitre I

- the Southern California Bight. *Ecology Letters* **12**, 898-908.
- PAVOINE S., OLLIER, S. & PONTIER, D. (2005). Measuring diversity from dissimilarities with Rao's quadratic entropy: Are any dissimilarities suitable? *Theoretical Population Biology* **67**, 231-239.
- PAVOINE S. & RICOTTA, C. (2014). Functional and phylogenetic similarity among communities. *Methods in Ecology and Evolution* **In press**.
- PEARSE W. D., CADOTTE, M. W., CAVENDER-BARES, J., IVES, A. R., TUCKER, C. M., WALKER, S. & HELMUS, M. R. (2015). pez: Phylogenetics for the Environmental Sciences. *Bioinformatics* **In press**.
- PEARSE W. D., CAVENDER-BARES, J., PURVIS, A. & HELMUS, M. R. (2014). Metrics and models of community phylogenetics. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*. Springer-Verlag, Berlin.
- PIGOT A. L. & ETIENNE, R. S. (2015). A dynamic null model for phylogenetic community structure. *Ecology Letters* **18**, 153-163.
- PURSCHE O., SCHMID, B., SYKES, M. T., POSCHLOD, P., MICHALSKI, S. G., DURKA, W., KUHN, I., WINTER, M. & PRENTICE, H. C. (2013). Contrasting changes in taxonomic, phylogenetic and functional diversity during a long-term succession: insights into assembly processes. *Journal of Ecology* **101**, 857-866.
- PURVIS A. (2008). Phylogenetic approaches to the study of extinction. *Annual Review of Ecology, Evolution, and Systematics* **39**, 301-319.
- PURVIS A., AGAPOW, P. M., GITTLEMAN, J. L. & MACE, G. M. (2000). Nonrandom extinction and the loss of evolutionary history. *Science* **288**, 328-330.
- PURVIS A., FRITZ, S. A., RODRIGUEZ, J., HARVEY, P. H. & GRENYER, R. (2011). The shape of mammalian phylogeny: patterns, processes and scales. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2462-2477.
- PYBUS O. & HARVEY, P. H. (2000). Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society B-Biological Sciences* **267**, 2267-2272.
- RAO R. C. (1982). Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology* **21**, 24-43.
- REDDING D. W. (2003). Incorporating genetic distinctness and reserve occupancy into a conservation prioritisation approach, University of East Anglia.
- RICOTTA C. (2007). A semantic taxonomy for diversity measures. *Acta Biotheoretica* **55**, 23-33.
- RODRIGUES A. S. L., BROOKS, T. M. & GASTON, K. J. (2005). Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In *Phylogeny and conservation* (ed. A. Purvis, J. L. Gittleman and T. M. Brooks), pp. 101-199. Cambridge University Press, Cambridge, UK.
- RODRIGUES A. S. L., GRENYER, R., BAILLIE, J. E. M., BININDA-EMONDS, O. R. P., GITTLEMAN, J. L., HOFFMANN, M., SAFI, K. & AL., E. (2011). Complete, accurate, mammalian phylogenies aid conservation planning, but not much. *Philosophical Transactions of the Royal Society, London, B* **1579**, 2652-2660.
- 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.
- ROSAUER D. F. & MOOERS, A. O. (2013). Nurturing the use of evolutionary diversity in nature conservation. *Trends in Ecology & Evolution* **28**, 322-323.
- SAFI K., ARMOUR-MARSHALL, K., BAILLIE, J. E. M. & ISAAC, N. J. B. (2013). Global patterns of evolutionary distinct and globally endangered amphibians and mammals. *Plos One* **8**.
- SCHEINER S. M. (2012). A metric of biodiversity that integrates abundance, phylogeny, and function. *Oikos* **121**, 1191-1202.
- SCHWEIGER O., KLOTZ, S., DURKA, W. & KUHN, I. (2008). A comparative test of phylogenetic diversity indices. *Oecologia* **157**, 485-495.
- SRIVASTAVA D. S., CADOTTE, M. W., MACDONALD, A. A. M., MARUSHIA, R. G. & MIROTCHEV, N. (2012). Phylogenetic diversity and the functioning of ecosystems. *Ecology Letters* **15**, 637-648.
- SWENSON N. G. (2011). Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. *Plos One* **6**, e21264.
- THUILLER W., LAVERGNE, S., ROQUET, C., BOULANGEAT, I., LAFOURCADE, B. & ARAUJO, M. B. (2011). Consequences of climate change on the tree of life in Europe. *Nature* **470**, 531-534.
- TUCKER C. M. & CADOTTE, M. W. (2013). Unifying measures of biodiversity: understanding when richness and phylogenetic diversity should be congruent. *Diversity and Distributions* **19**, 845-854.
- TUCKER C. M., CADOTTE, M. W., DAVIES, T. J. & REBELO, A. G. (2012). The distribution of biodiversity: linking richness to geographical and evolutionary rarity in a biodiversity hotspot *Conservation Biology* **25**.
- UNDERWOOD A. J. (1997). *Experiments in Ecology*. Cambridge University Press, Cambridge, UK.
- 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.
- VELLEND M., CORNWELL, W. K., MAGNUSON-FORD, K. & MOOERS, A. Ø. (2010). Measuring phylogenetic biodiversity. *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, UK, 194-207.
- VILLÉGER S., MASON, N. W. & MOUILLOT, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* **89**, 2290-2301.
- VIOLLE C., NEMERGUT, D. R., PU, Z. & JIANG, L. (2011). Phylogenetic limiting similarity and competitive exclusion. *Ecology Letters* **14**, 782-787.
- WEBB C. O. (2000). Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *The American Naturalist* **156**, 145-155.
- WEBB C. O., ACKERLY, D. D. & KEMBEL, S. W. (2008). Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* **24**, 2098-2100.
- WEBB C. O., ACKERLY, D. D., MCPEEK, M. A. & DONOGHUE, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology, Evolution, and Systematics* **33**, 475-505.
- WEIHER E. & KEDDY, P. A. (1995). Assembly rules, null models, and trait dispersion: new questions from old patterns. *Oikos* **74**, 159-164.
- WEIR J. T. & SCHLUTER, D. (2007). The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* **315**, 1574-1576.

Partie I - Chapitre I

- WIENS J. J. & DONOGHUE, M. J. (2004). Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* **19**, 639-644.
- WINTER M., DEVICTOR, V. & SCHWEIGER, O. (2013). Phylogenetic diversity and nature conservation: where are we? *Trends in Ecology & Evolution* **28**, 199-204.
- WINTER M., SCHWEIGER, O., KLOTZ, S., NENTWIG, W., ANDRIOPOULOS, P., ARIANOUTSOU, M., BASNOU, C., DELIPETROU, P., DIDŽIULIS, V., HEJDA, M., HULME, P. E., LAMBON, P. W., PERGL, J., PYSEK, P., ROY, D. P. & KUHN, I. (2009). Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. *Proceedings of the National Academy of Sciences* **106**, 21721-21725.

CHAPITRE II

L'IMPORTANCE DE L'ÉCHELLE PHYLOGÉNÉTIQUE

Ce deuxième chapitre se focalise sur les trois indices de diversité phylogénétique α les plus couramment utilisés dans la littérature. Je tente ici de comparer les propriétés de ces trois indices en utilisant un modèle nul courant.

Résumé de l'article 2 (accepté dans Ecography)

Durant les dernières décennies, la description, l'analyse et la compréhension de la structure phylogénétique des assemblages d'espèces ont représenté un thème central en écologie et en macro-écologie. Parmi la grande variété des indices de diversité phylogénétique, trois ont été majoritairement utilisés :

- (1) la PD_{Faith} (Faith 1992) représente la somme des longueurs de branches reliant toutes les espèces d'un assemblage.
- (2) le MPD (pour « mean pairwise distance » Webb *et al.* 2002) représente la distance phylogénétique moyenne entre toutes les espèces d'un assemblage.
- (3) la MNTD (pour « mean nearest pairwise distance » Webb *et al.* 2002) représente la distance phylogénétique moyenne entre les espèces sœurs d'un assemblage.

Les comparaisons entre les études sont difficiles car il n'existe pas à l'heure actuelle d'analyse contrastant la structure phylogénétique que chaque indice capture. En particulier, il est encore mal connu comment PD_{Faith} est relié aux deux autres indices. En conséquence, il est possible que des patrons apparemment contradictoires entre études soient le fait d'une simple différence entre les propriétés des indices. Dans cette étude, nous proposons de résoudre ce problème en comparant de façon systématique ces indices en utilisant des simulations mais aussi des données réelles. Premièrement, nous utilisons des simulations pour tester l'influence de la structure et la taille de la communauté sur les différences entre indices, tout en faisant varier la forme et la taille de l'arbre phylogénétique. Dans un deuxième temps, nous étudions ces différences pour deux jeux de données empiriques (microbiotes* d'intestins et assemblages de carnivores à l'échelle planétaire). Nous montrons que le MNTD et la PD_{Faith} quantifient en fait une dimension récente (c'est-à-dire près des feuilles de l'arbre) de la structure phylogénétique. Au contraire, le MPD montre un comportement totalement différent et

semble très sensible à la structure phylogénétique profonde (c'est-à-dire près de la racine de l'arbre). Nous suggérons qu'il est important de combiner les indices qui se focalisent à différentes échelles phylogénétiques pour obtenir une meilleure compréhension de la structure évolutive des assemblages.

Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics

Mazel, F.^{1,2}, Davies, T.J.^{3,4}, Gallien L.⁵, Renaud J.^{1,2}, Groussin, M.⁶, Münkemüller T.^{1,2} & Thuiller, W.^{1,2}

¹ Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France.

² CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France.

³ Department of Biology, McGill University, 1205, Avenue Docteur Penfield, Montreal, Quebec, Canada.

⁴ African Centre for DNA Barcoding, University of Johannesburg, PO Box 524, Johannesburg, South Africa.

⁵ Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland.

⁶ Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract.

During the last decades, describing, analysing and understanding the phylogenetic structure of species assemblages has been a central theme in both community ecology and macro-ecology. Among the wide variety of phylogenetic structure metrics, three have been predominant in the literature: Faith's phylogenetic diversity (PD_{Faith}), which represents the sum of the branch lengths of the phylogenetic tree linking all species of a particular assemblage, the mean pairwise distance between all species in an assemblage (MPD) and the pairwise distance between the closest relatives in an assemblage (MNTD). Comparisons between studies using one or several of these metrics are difficult because there has been no comprehensive evaluation of the phylogenetic properties each metric captures. In particular it is unknown how PD_{Faith} relates to MPD and MNTD. Consequently, it is possible that apparently opposing patterns in different studies might simply reflect differences in metric properties. Here, we aim to fill this gap by comparing these metrics using simulations and empirical data. We first used simulation experiments to test the influence of community structure and size on the mismatch between metrics whilst varying the shape and size of the phylogenetic tree of the species pool. Second we investigated the mismatch between metrics for two empirical datasets (gut microbes and global carnivoran assemblages). We show that MNTD and PD_{Faith} provide similar information on phylogenetic structure, and respond similarly to variation in species richness and assemblage structure. However, MPD demonstrate a very different behaviour, and is highly sensitive to deep branching structure. We suggest that by combining complementary metrics that are sensitive to processes operating at different phylogenetic depths (i.e. MPD and MNTD or PD_{Faith}) we can obtain a better understanding of assemblage structure.

Introduction

During the last decades, the phylogenetic structure of species assemblages has received much attention in community ecology and macro-ecology since it holds promise to help unravel the drivers of species coexistence at various spatial scales (Webb et al. 2002, Lozupone and Knight 2005, Mouquet et al. 2012, Warren et al. 2014).

Community ecologists have considered phylogenetic distances between species as a substitute for niche differences and have used phylogenetic structure to disentangle the relative effects of biotic and abiotic environments in shaping present day species distributions (Webb et al. 2002, Mouquet et al. 2012). Coexistence theory predicts that species sharing the same niches compete more

strongly than dissimilar species (HilleRisLambers et al. 2012), it is therefore commonly expected that competition-driven coexistence will generate patterns of phylogenetic 'over-dispersion' (i.e. distantly related species with less niche overlap co-occur). Conversely, phylogenetic 'clustering' is thought to indicate the coexistence of closely related species because of shared environmental niches (Webb et al. 2002, Mouquet et al. 2012, O'Dwyer et al. 2012). Although the link between pattern and process has been a matter of some debate (for example, see Mayfield & Levine 2010), non-random phylogenetic community structure appears common.

Recently, macro-ecologists have also used the phylogenetic structure of assemblages to understand the effects of historical processes on

large scale biodiversity distribution (Davies et al. 2011, Kissling et al. 2012). For example, explosive radiation of species within a given area may result in the co-occurrence of closely related species resulting in phylogenetic clustering, while multiple allopatric speciation events may lead to phylogenetic over-dispersion (Warren et al. 2014). Among the plethora of phylogenetic structure metrics that have been developed and used in both fields (Pavoine and Bonsall 2011), the three most commonly used are Faith's phylogenetic diversity (named PD_{Faith} hereafter), which represents the sum of the branch lengths of the phylogenetic tree linking all species of a particular assemblage (Faith 1992), the mean pairwise distance between all species in an assemblage (MPD) and the pairwise distance between the closest relatives in an assemblage (MNTD) (Webb et al. 2002). As PD_{Faith} correlates closely and positively with species richness (SR, e.g. Tucker and Cadotte 2013), the use of a null model that keeps SR constant while randomizing phylogenetic relationships allows comparisons of assemblages with different SR. Using this null model, standard effect size (SES, Eq. 1) and relative position of observed values with respect to the null distribution can be calculated. SES can be defined as:

$$ses.Metric = \frac{Metric_{obs} - mean(Metric_{null})}{sd(Metric_{null})} \quad (\text{Eq. 1})$$

where $Metric_{obs}$ is the observed metric in a given assemblage, and $Metric_{null}$ is the same metric but calculated n times with n randomised assemblages. The relative position of the observed value with respect to the null distribution is computed as the proportion of null values that are lower than the observed values. It represents the probability to draw the observed value from the null distribution and thus corresponds to a p -value (H_{null} being the null model). For normally distributed data, significance at p -value < 0.05 is equivalent to a standard effect size > 1.96 (or < -1.96).

The standard effect size (and associated p -value) of MPD (ses.MPD hereafter) and MNTD (ses.MNTD hereafter) are commonly used in the community phylogenetic literature (also referred to as NRI and NTL, respectively, Webb et al. 2002), and can be directly compared to the standard effect size of PD_{Faith} (ses. PD_{Faith} hereafter). All three standardized metrics quantify the relative excess (overdispersion) or deficit (clustering) in

phylogenetic diversity for a given species set relatively to the species pool (whose phylogenetic relationships between species are depicted by the "regional tree"). As a consequence, a negative standardized metric reflects a relative clustering of species while a positive standardized metric reflects a relative overdispersion of species.

MPD and MNTD highlight phylogenetic structure of assemblages at different evolutionary depths since MPD is more strongly influenced by the 'basal' structure of the phylogenetic tree while MNTD describes the more 'terminal' structure of the phylogenetic tree (Webb et al. 2002). This is an important aspect since different processes may act at different evolutionary time scales. Some processes may produce basal clustering, while others may create terminal over-dispersion, generating "clusters of overdispersion" (see e.g. blue assemblages in Fig. 1). For example, a cluster of overdispersion could be due to environmental filtering at large evolutionary time scales and competition between close relatives at fine evolutionary time scale (Hardy and Senterre 2007). Conversely, "overdispersion of clusters" would correspond to basal overdispersion and terminal clustering (see e.g. pink assemblages in Fig. 1). Consequently, owing to their property to detect phylogenetic structure on restricted phylogenetic scales, existing metrics may be limited in their ability to capture complex structural patterns and may suffer from decreased power (i.e. inflated false negative) in case clustering and overdispersion occur in concert at different phylogenetic scales. In other words, if clustering and overdispersion occur at different phylogenetic scales, using a metric that averages the information over the entire tree may mask these two opposing patterns.

While differences in the performance of ses.MPD and ses.MNTD are widely recognised, ses. PD_{Faith} has mostly been considered independently or as interchangeable with ses.MPD and ses.MNTD (see Table 1 and Supplementary Material 1 for examples and their detailed justifications, respectively). To our knowledge, this assumed equivalence has never been validated empirically. Comparisons between studies using either jointly MPD and MNTD or PD_{Faith} (either SES of metrics and/or corresponding p -values) are thus difficult to interpret, and it is possible that apparently opposing patterns of phylogenetic structure simply reflect differences in metric properties. While a number of macro/micro ecological studies have explored patterns of

phylogenetic metrics, making unique and important contributions to the literature (e.g. Morlon et al. 2011, Fritz and Rahbek 2012, O'Dwyer et al. 2012, see Table 1), for the most part they have focussed only on PD_{Faith} , limiting our understanding of the phylogenetic scale at which processes structuring species assemblages operate. In contrast, community phylogenetic studies frequently measure both MPD and MNTD, but rarely consider PD_{Faith} . Choice of metric, however, is often poorly justified, and we currently lack a comparative analysis comparing the performance of these three widely used indices.

Here, we perform the first comprehensive comparative analysis of MPD, MNTD and PD_{Faith} using both simulations and empirical data. We first used simulation experiments to understand the relative effect of (1) the size and the shape of the regional tree, and, (2) the richness and structure of the observed community, on the mismatches between the three metrics. Second, we evaluated metric behaviour on real world datasets for gut microbes and terrestrial mammals. Overall, we show that MNTD and PD_{Faith} , are both 'terminal' metrics, and thus could be used interchangeably, while MPD, a 'basal' metric sensitive to deeper branching structure, captures a distinct, but complementary dimension of phylogenetic structure. Their combined use allow for a better understanding of assemblage structure, especially when different processes operates at different phylogenetic depths simultaneously, while the use of only one metric could lead to incomplete interpretations and biased conclusions.

Material and Methods

Simulation experiments

In a first set of simulations, we illustrated the sensitivity of PD_{Faith} , MPD and MNTD to tree shape and phylogenetic structure using a straightforward toy example. We generated four balanced phylogenetic trees with different basal vs. terminal branch lengths (see Fig. 1 column 1). For each tree, we compared the response of the three metrics to four extreme assemblage structures: phylogenetic clustering and overdispersion, clustering of overdispersion and overdispersion of clusters (Fig. 1)

In a second set of simulations, we tested the influence of regional pool sizes, regional tree shapes and species richness of the assemblage on the three metrics. We first created four regional species pools

of different sizes ($n = 20, 40, 100$ and 200 species). For each regional species pool, we simulated 100 regional phylogenetic trees of size n using a pure birth model (function *sim.bdtree* of the *geiger* R package; Harmon et al. 2008) for which we reported two tree shape statistics: imbalance (Colless 1982) and Gamma values, quantifying the "tippiness" vs. "stemminess" of the tree, respectively (Pybus and Harvey 2000). For each of these 400 regional pools (100 regional trees * 4 tree sizes), we constructed local species assemblages by randomly drawing without replacement m species from the regional pool (m thus equals the species richness of the corresponding community). We repeated this procedure varying m from 2 to $n-1$ and randomly drawing five assemblages for a given species richness m . To test the influence of the size of the species assemblage on our three metrics, we then grouped assemblages according to their SR and reported the mismatches between metrics for each group independently. For example, with a regional species pool of 200 species, we created 10 sets of species assemblages with different ranges of SR (from 10 to 30 species for the first set, 30 to 50 for the second, and so on until 170 to 190 species for the last set). For each tree and each set of assemblages we calculated the strength of the relationship between $ses.PD_{Faith}$ and $ses.MPD$ or $ses.MNTD$ as the R^2 of the linear regression between them (we also checked for more complex models, see Results).

Real world empirical data

To compare the three metrics in realistic examples, we compiled two empirical datasets that differed in spatial scale and assemblage structure: mammal gut microbial assemblages and global terrestrial carnivore assemblages.

Mammal gut microbial assemblages

Species assemblages were here defined as the set of microbial Operational Taxonomic Units (OTUs, a taxonomic concept based on DNA sequence similarities commonly used for microbes) living in the gut of different mammalian species. We used the 16S rRNA genes dataset from Muegge et al. (2011) to derive occurrences of 2,820 OTUs (defined at 97% of similarity to create species-like entities) in the gut of 33 mammalian species (i.e. 33 bacterial species assemblages from a regional pool of $n = 2,820$ OTUs) and to reconstruct a regional phylogenetic tree (see Supp. Mat.2. for details).

Global carnivore assemblages

We used the distribution maps provided by the Mammal Red List Assessment (<http://www.iucnredlist.org/>) to derive the occurrence of 241 terrestrial carnivores (regional species pool). We then defined a species assemblage as all carnivores co-occurring in a given 50x50km grid cell. Analyses were thus carried out over 52,346 assemblages around the world. For phylogenetic relationships, we used the recent update of the carnivoran phylogeny proposed by Nyakatura and Bininda-Emonds (2012) as the regional tree.

Metric calculation

For both simulated and empirical datasets, we computed the observed PD_{Faith} , MPD and MNTD for each assemblage using the function *pd*, *mpd* and *mntd* in the R package *picante* (Kembel et al. 2010). We then randomly shuffled the tips of the regional phylogeny and calculated PD_{Faith} , MPD and MNTD for the random assemblages, and repeated this procedure 100 times to obtain a null distribution of values for each assemblage and for each metric. We calculated standard effect sizes using eq.1, as well as the relative position of the observed index in the null distribution to derive *p-values* (e.g. Fig 1). All analysis were performed using R (R Development Core Team 2014).

Results & Discussion

In our first set of simulations (see Material & Methods and Figure 1), we used extreme assemblage structures to illustrate the different phylogenetic signal captured by PD_{Faith} , MPD and MNTD, respectively. All three metrics produce very similar predictions in the case of simple patterns where a single process structures species' assemblages (Fig. 1, clustered patterns in orange and over-dispersed patterns in green). We also explored more complex phylogenetic structures such as clusters of overdispersion or overdispersion of clusters. These patterns might arise when different processes acting at different depth of the tree jointly influence assemblage phylogenetic structure. In these more complex cases, the different metrics suggested different phylogenetic structures, with these differences being further affected by phylogenetic tree shape (Fig. 1).

In the case of clusters of overdispersion (blue line in Fig. 1) on a very 'tippy' tree (i.e. one with relatively

short internal to tip branch lengths; tree A in Fig. 1) the relative rank of observed MPD versus null MPD suggests clustering while PD_{Faith} (and even more MNTD) tends towards over-dispersion (see Fig. 1). This difference can be explained by the sensitivity of MPD to deep branching structure in the tree, which are counted multiple times when computing pairwise distance between all species in the assemblage (Webb et al. 2002). Thus, MPD tends to emphasise the basal clustering rather than the overdispersion within the cluster. In contrast, PD_{Faith} and MNTD are less influenced by internal branches (they are completely ignored by MNTD and counted only once for PD_{Faith}), and therefore tend to be more sensitive to patterns occurring at the tips of the tree. When keeping the same community structure (a cluster of overdispersion, blue curve) but switching to a "stemmy" tree (i.e. one with relatively long internal to tip branch lengths; tree D in Fig. 1), MPD still suggests clustering and MNTD over-dispersion but PD_{Faith} shifts towards clustering. In this case only, PD_{Faith} is closer to MPD than to MNTD because, even though PD_{Faith} counts internal branches only once, in our simple simulation they have a profound effect because they are disproportionately (and perhaps unrealistically) long compared to tip lengths. When considering overdispersion of clusters (red curve) on the same stemmy tree, PD_{Faith} and MNTD were more sensitive to the signal of clustering, whereas MPD was more sensitive to the signal for overdispersion. This simple example illustrates that different metrics can identify apparently contrasting patterns in the same dataset, and that changes in tree structures can alter metric behaviour, and thus the inferences we might draw from them. Overall, PD_{Faith} and MNTD tend to identify similar structure (although not always), but are decoupled from MPD.

Our second set of simulations aimed at investigating whether the observed mismatch between PD_{Faith} and MPD was still apparent with a less-extreme range of community structures, but more realistic regional tree shapes and richer species assemblages. We found that *ses.PD_{Faith}* and *ses.MPD* were still largely decoupled while *ses.PD_{Faith}* and *ses.MNTD* showed congruent results (Fig 2 and Supp. Mat. 3-4). While regional phylogenetic tree shape and species pool size did not influence the discrepancies between *ses.PD_{Faith}* and *ses.MPD* (Supp. Mat. 5-6) correlation strength between metrics decreased with increasing species richness

(mean SR) of assemblages (Fig 2 and 3-4). The richer the community, the more complex they are, and the more divergent are the metrics. Conversely, species poor communities likely have less complex structures that are easily detected by all three metrics.

The analyses of empirical data match to the results from the simulations. Phylogenetic pattern of both carnivore diversity and gut microbiomes revealed a clear mismatch between $\text{ses.PD}_{\text{Faith}}$ and ses.MPD ($R^2 = 0.36$ and 0.19 , for carnivores and microbes respectively, Figs 3-4), and greater congruence between $\text{ses.PD}_{\text{Faith}}$ and ses.MNTD ($R^2 = 0.59$ and 0.96 , for carnivores and microbes respectively, Figs 3-4). These results were almost identical when testing for linear or non-linear relationships between the metrics (see Supp. Mat. 7-8). The clear mismatch between indices further reveals complex structures that cannot be summarized by a single number (i.e. a single metric).

Comparing phylogenetic structure using metrics sensitive to process operating at different evolutionary scales sheds new light on the distribution of diversity. Indeed, for carnivores, the differences between metrics were strongly spatially structured (Fig. 4). In some parts of South America, both $\text{ses.PD}_{\text{Faith}}$ and ses.MNTD suggested phylogenetic clustering which were not evident from patterns of ses.MPD . Recent radiations of particular clades within the neotropics following the Great American Biotic Interchange (Webb 2006, Woodburne 2010) - for example, the ocelot genus *leopardus* (Johnson et al. 2006) - likely results in the co-occurrence of closely related species (Pedersen et al. 2014) at the continental scale. Niche partitioning via fine-scale habitat preferences may then have allowed these close relative species to co-occur at the scale of our analysis (Araújo and Rozenfeld 2014), leading to 'terminal' phylogenetic clustering (as detected by $\text{ses.PD}_{\text{Faith}}$ or ses.MNTD). At the same time, because South American carnivore assemblages also contain species from the two major clades of carnivores (i.e. dog- and cat-like clades: caniformia and feliformia, respectively), which are evolutionarily distinct from each other, but contain approximately equal numbers of species (see Pedersen et al. 2014, and Supp. Mat. 9), regional assemblages tend to show random phylogenetic patterns (as indicated by ses.MPD scores).

In contrast to patterns in South America, the Congo basin and some parts of Eurasia and North America

show significant 'basal' clustering (as indicated by a significant negative ses.MPD values), but little structure towards the tips (as suggested by non-significant patterns of $\text{ses.PD}_{\text{Faith}}$ and ses.MNTD). This pattern may reflect the very unbalanced distribution of caniformia and feliformia in these regions (see Sup. Mat. 9) suggested to be the outcome of the interrelation between large-scale competition and historical biogeography (Pedersen et al. 2014). In Eurasia and North America, caniformia are over-represented, while feliformia dominate in the Congo basin (Pedersen et al. 2014; see also Supp. Mat. 9). Thus, these regions show taxonomic dominance by a single subclade, which is apparent in the "basal" clustering pattern, while there is little regional phylogenetic structure within clades. Taken together these findings demonstrate that one single phylogenetic diversity metric is not able to fully describe the complex structure of assemblage at macro-ecological scales where different processes may operate in different geographical regions and are evident at different phylogenetic depths. It is thus important to use multiple metrics that focus on different evolutionary scales, to be able to draw more precise and robust interpretations of phylogenetic diversity patterns (see e.g. Davies and Buckley 2011). Our results further stress the importance of using multiple metrics in the analysis of large datasets, as complex macro-ecological phylogenetic structures can emerge when studying relatively large groups (see O'Meara 2012, for a macro-evolutionary perspective on this subject).

Our findings at large spatial scale also extend to micro-biology, as we observe very similar trends for gut microbial assemblages (Fig. 4 and Sup. Mat. 9). For example, rodent microbial gut assemblages appear clustered with $\text{ses.PD}_{\text{Faith}}$ or ses.MNTD , but over-dispersed with ses.MPD (Fig. 4). Therefore, when using only metrics of $\text{ses.PD}_{\text{Faith}}$ or ses.MNTD , one would have concluded that rodent microbial gut assemblages mainly consist of some closely related lineages. Whereas, when using only the ses.MPD metric, one would have concluded the opposite: microbial guts assemblages consist of distantly related lineages.

Again, these mismatches among metrics seem to be caused by an over-dispersion of clusters (see Supp. Mat. 10); the capybara (*Hydrochoerus hydrochaeris*) is an interesting example. Because it is a hindgut-fermenter herbivore, plant residuals reaching the fermentative part of the gut are complex to digest

(e.g. cellulose and lignin, Ley et al. 2008) and a variety of plant secondary metabolites need to be broken-down (Dearing et al. 2000). Consequently, very different bacterial clades with different enzymatic equipment inhabit the gut of hindgut herbivores, possibly generating observed patterns of 'basal' overdispersion, i.e. co-occurrence of distantly related bacterial lineages. However, within these clades, host-specific processes structuring bacterial assemblages, such as mucus barriers, oxygen concentration and the innate and adaptive immune systems, may select for specific bacterial lineages (Salzman et al. 2010, Bevins & Salzman 2011, Hooper et al. 2012) favouring clustering of bacterial lineages. Thus, similarly to the case for global carnivoran communities discussed above, we show that a set of complementary metrics is needed to accurately describe the complex phylogenetic structure of micro-biomes.

We have classified phylogenetic metrics as either 'terminal' or 'basal', and we show that $ses.PD_{Faith}$ is a 'terminal' metric that reflects the phylogenetic structure that is dominant near the tip of the tree. Recent reviews on metrics of phylogenetic beta diversity (Swenson 2011) and evolutionary isolation (i.e. phylogenetic "structure" at the species level, Redding *et al.* 2014) have followed a similar classification, placing emphasis on the phylogenetic depth at which patterns emerge. For example, Swenson (2011) classified Unifrac (Lozupone and Knight 2005) and PhyloSor (Bryant et al. 2008), the beta diversity equivalents of PD_{Faith} , as 'terminal' while Dpw (Webb et al. 2008), the beta diversity equivalent of MPD, as 'basal'. Similarly, Redding et al. (2014) suggested that the fair proportion evolutionary distinctiveness metric (Isaac et al. 2007), the equivalent of PD_{Faith} at the species level, best captures 'terminal' isolation, while the 'average phylogenetic distance' (Ricotta 2007), the equivalent of MPD at the species level, is more closely linked to the 'basal' isolation of species.

Our analysis focused on the three most widely used metrics in (macro) ecology. There are, however, a large number of available metrics in the literature that have been proposed over the last decades (see Pavoine and Bonsall, 2011, for a synthetic review). It is beyond the scope of the paper to review and compare all available metrics, and additional studies are needed to extend our results to other metrics. For example, the Rao Quadratic Entropy (Rao QE, e.g. Pavoine and Bonsall 2011) is also widely used to describe phylogenetic structure (e.g.

Devictor et al. 2010). Interestingly, when using presence/absence data, Rao QE has a non-linear relationship with MPD ($Rao\ QE = (m-1)/m * MPD$, with m being the species richness of the assemblage) so that the two metrics essentially carry the same information. As a consequence, our results suggest that Rao QE would also represent a 'basal' metric of phylogenetic diversity, although this remains to be verified.

To better link macro-ecology, community ecology and microbial ecology, our analysis only focused on presence-absence data since abundance data is often not available at macro- and micro-ecological scales. However, we suggest that issues raised here for presence-absence data are likely to also occur when metrics are computed with abundance data. It is nonetheless not trivial to extend our simulation framework to incorporate abundances because additional assumptions have to be made on the relative abundances of species within the simulations, for which many possible scenarios exist. Additionally, since other metrics and recent developments have been proposed to specifically include abundances (e.g. Chao et al. 2010, Faith 2013), more comprehensive tests need to be conducted to evaluate the influence of species' abundance on detected patterns.

Conclusion

The use of PD_{Faith} , MPD and MNTD to characterize the phylogenetic structure of assemblages and regional species assemblages is common in the ecological literature. However, the choice of one metric over another seems more the product of historical contingencies rather than methodological properties. For example, studies linking phylogenetic diversity to area (Morlon et al. 2011), productivity (Cadotte et al. 2008) or functional diversity (Safi et al. 2011) have adopted PD_{Faith} as the natural extension of species richness (SR). This is an interesting avenue but explores only one single –recent- dimension of the phylogenetic structure of communities. Adding metrics that detect deeper phylogenetic structures (such as MPD or Rao QE) may in fact reveal different processes and thus complete our understanding of diversity distribution.

Here, we have demonstrated that the choice of metric can significantly impact inferences on dominant patterns and thus interpretation with regard to potential underlying processes. We show that MNTD and PD_{Faith} behave similarly, but that

MPD is more sensitive to deeper branching structures. Our results extend the findings of Swenson (2011) for beta diversity metrics to alpha diversity metrics in distinguishing between relatively 'basal' or 'terminal' metrics. We call for the joint use of complementary metrics (i.e. MPD and PD_{Faith} or MNTD) to better understand patterns of species assemblage.

Acknowledgments.

The research leading to these results had received funding from the European Research Council under the European Community's Seven Framework Programme FP7/2007-2013 Grant Agreement no. 281422 (TEEMBIO). FM, TM and WT belong to the Laboratoire d'Écologie Alpine, which is part of Labex OSUG@2020 (ANR10 LABX56). All computations were performed using the CiGRI platform of the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA) and France-Grille (<http://www.france-grilles.fr>). TJD was supported by a team grant from the Fonds de recherche du Québec – Nature et technologies

TABLE 1 Hypotheses tested, key papers and corresponding metrics used across fields to depict the phylogenetic structure of species assemblages. The table depicts for each of the two considered field of research (column 1) the hypotheses tested behind classic phylogenetic patterns (column 2) and key papers using either MPD/MNTD or PD_{Faith} (column 3).

	Hypotheses related to the two patterns:		Metrics used to quantify the pattern; examples of key publications		
Fields	Clustering	Over-dispersion	Biotas	MPD/MNTD	PD _{Faith}
Community ecology	Environmental filtering*	Competition	Micro-biotas	Goberna <i>et al.</i> 2014 Bryant <i>et al.</i> 2008	O'Dwyer <i>et al.</i> 2012
			Macro-Biotas	Webb <i>et al.</i> 2002; Graham <i>et al.</i> 2009;	Cadotte <i>et al.</i> 2009; Morlon <i>et al.</i> 2011
Macro-ecology	In-situ speciation	Biogeographic contact zones; Vicariance; Competition at large scale	Cardillo 2011b; Davies & Buckley 2011		Fritz & Rahbek 2012

*but see Mayfield & Levine 2010

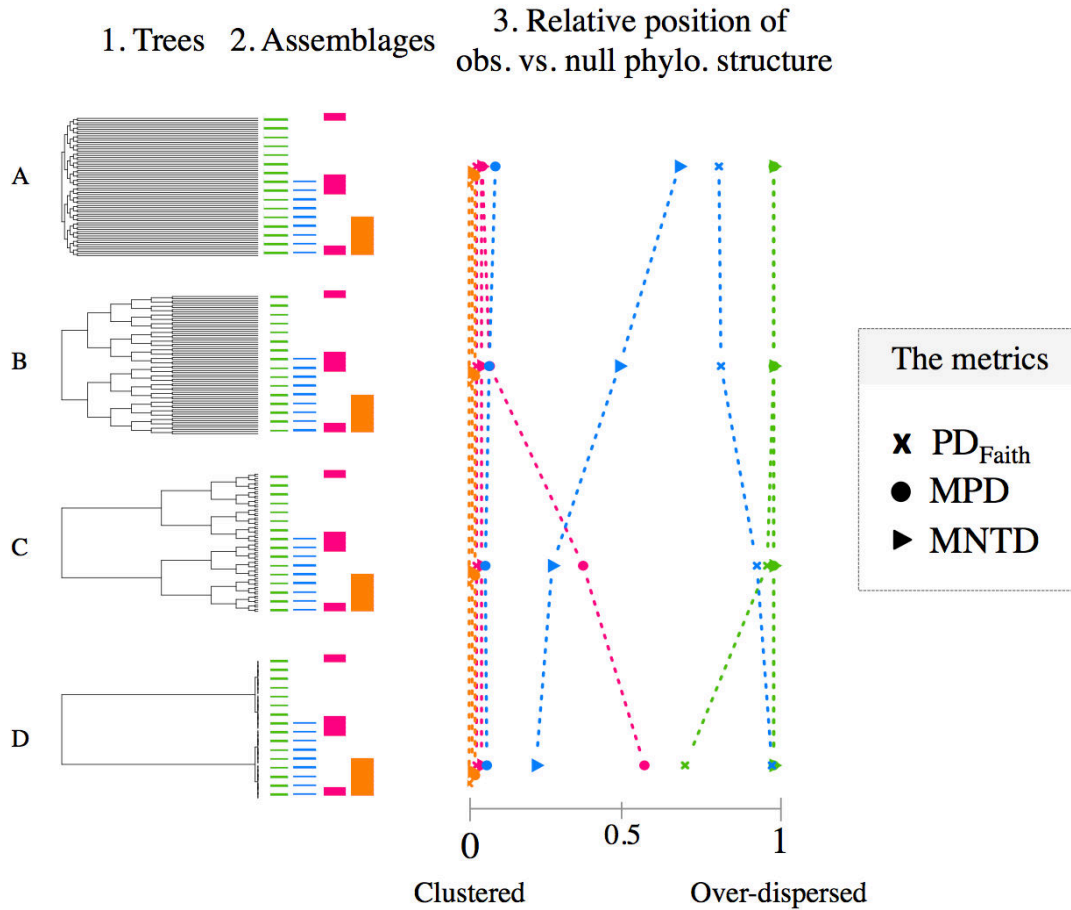


Figure 1. Sensitivity of PD_{Faith} , MNTD and MPD to different assemblage structures and tree shapes. The figure depicts four balanced trees (column ‘Tree’) along with four potential assemblage (column ‘Assemblage’): coloured segments indicate that the corresponding tips on the phylogeny are present in the assemblages (i.e. species presence); from left to right: simple simple overdispersion in green; clustering of overdispersion in blue; overdispersion of clusters in red; and simple clustering in orange. The corresponding diversities are presented in the third column. Values on the X-axis correspond to the relative position of observed value compared to a null distribution (interpreted as a *p-value*). For example, a value of 0 indicates that the observation is lower than all null expectations (‘clustering’) while a value of 0.5 indicates that the observation equals to the median of the null distribution. The different metrics are represented by different symbols (see legend).

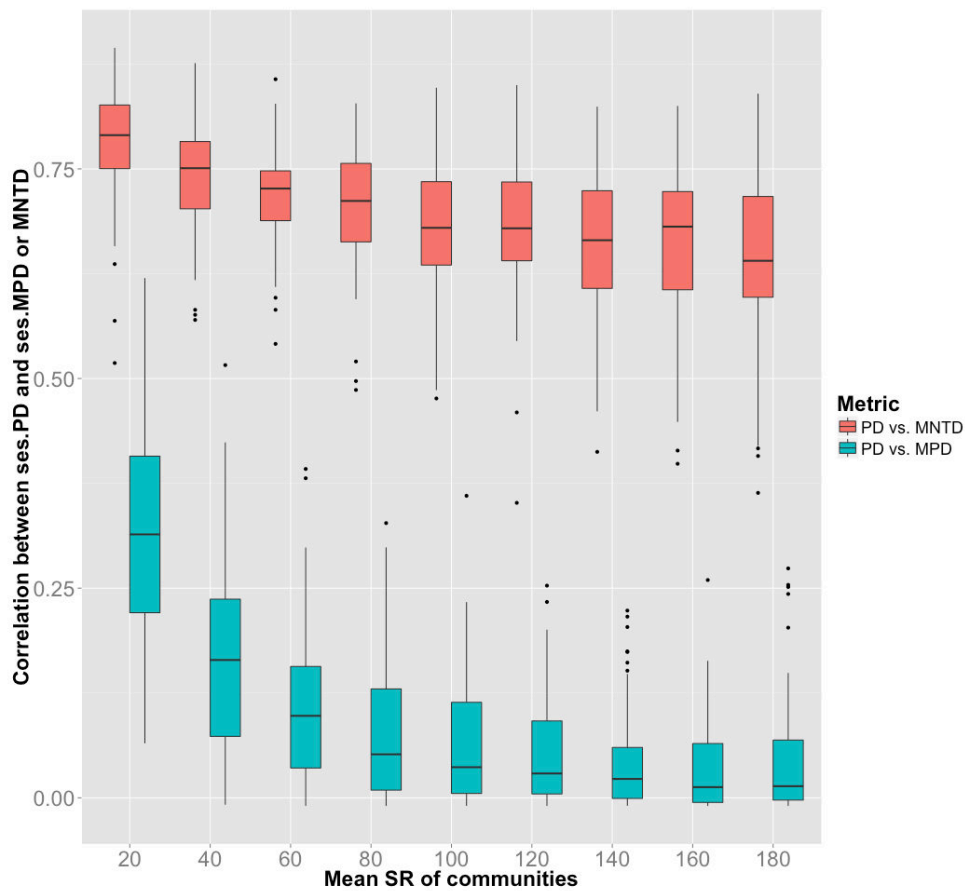


Figure 2. Influence of assemblage species richness on diversity metrics. Boxplots represent R^2 of the linear regressions between ses. PD_{Faith} and ses.MPD, and between ses. PD_{Faith} and ses.MNTD for 10 sets of simulated assemblages that differ in species richness (SR): from 10 to 30 species for the first set, 30 to 50 for the second, and so on, until 170 to 190 species for the last set. For each of these sets, we computed the standard effect size of each metric and the R^2 of the linear relationship between the metrics. We repeated the whole procedure 100 times to obtain a distribution of R^2 for each level of SR.

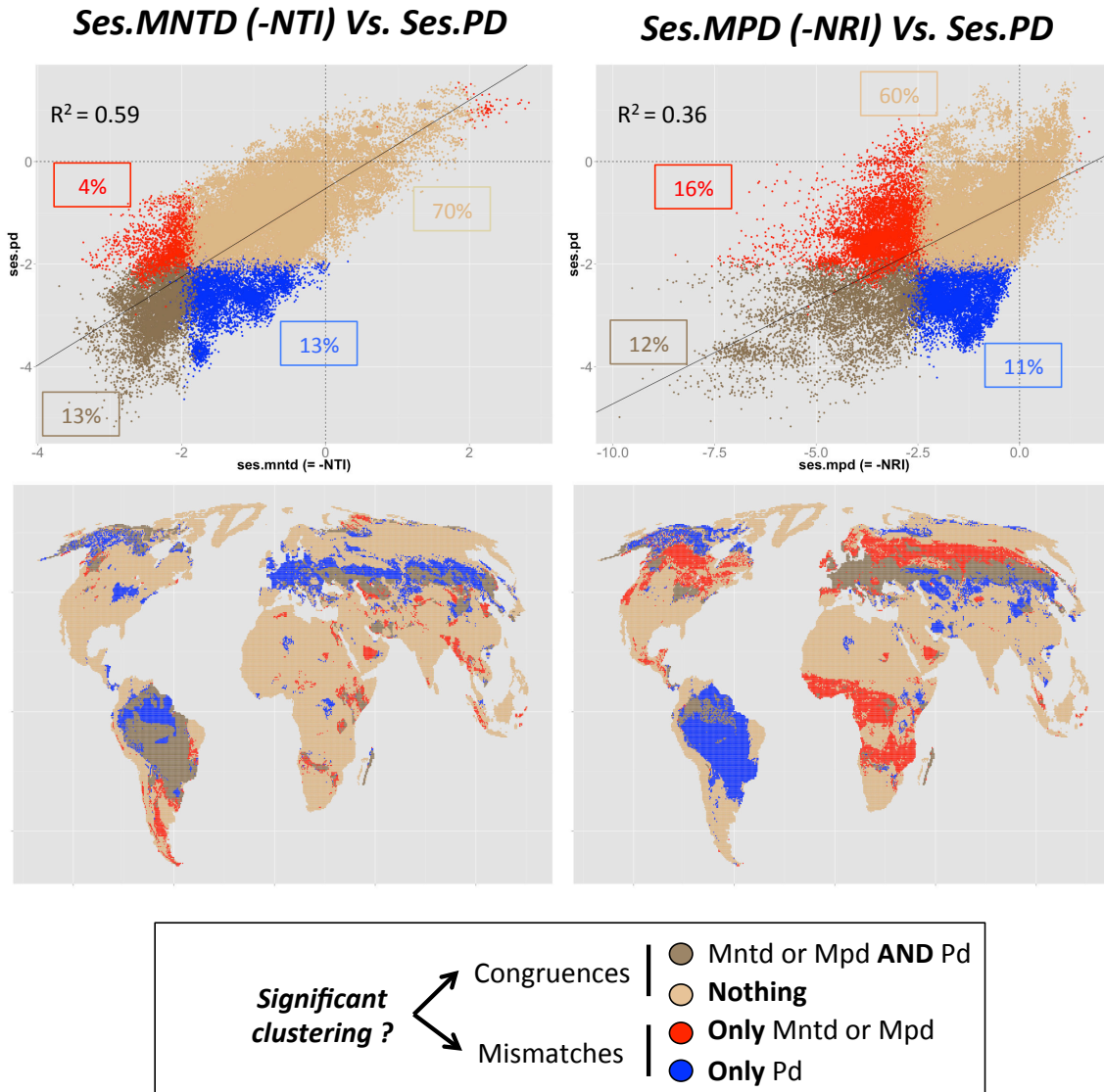


Figure 3. Phylogenetic structure of carnivoran assemblages. The top graphs represent the relationships between $\text{ses.PD}_{\text{Faith}}$ and ses.MPD or ses.MNTD together with the R^2 of their linear relationship. The proportion of congruence and divergence are also represented in four groups: whether PD_{Faith} and MPD (or MNTD) show congruent (significant/non-significant clustering: dark and light brown dots, respectively) or diverging results (only significant PD_{Faith} / only significant MPD (or MNTD); blue and red dots, respectively). Bottom graphs show the spatial structure of these four types of assemblages.

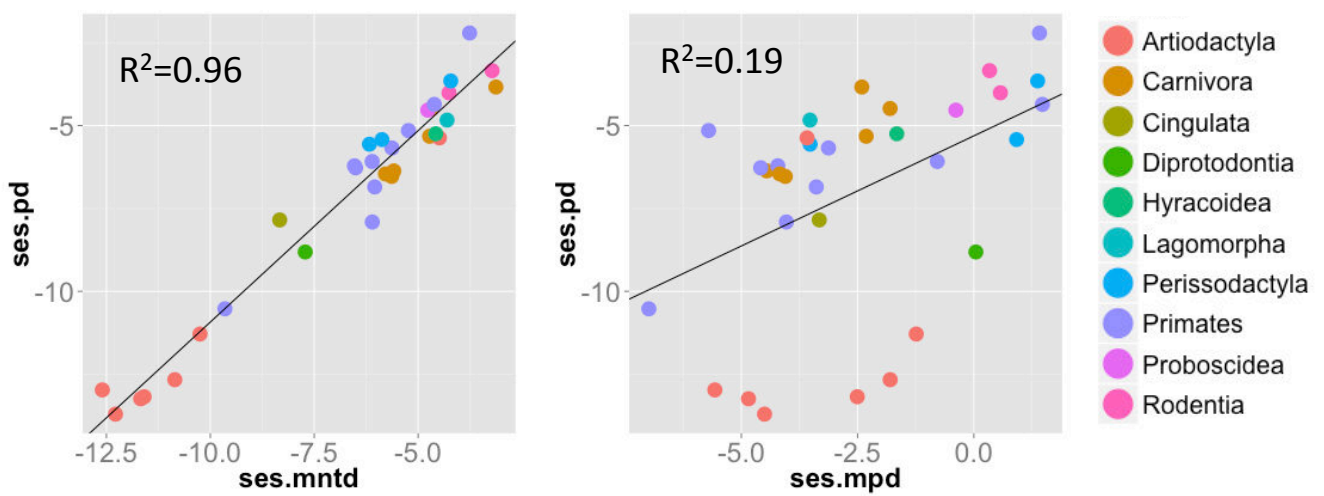


Figure 4. Mammalian guts phylogenetic structure. Raw relationship between $\text{ses.PD}_{\text{Faith}}$ (Y-axis) and $\text{ses.MPD}/\text{ses.MNTD}$ (X-axis) across 33 mammalian gut assemblages. Host orders are represented by different colours (see legend). The R^2 of linear models between metrics are reported on each graph.

References

- Araújo, M. and Rozenfeld, A. 2014. The geographic scaling of biotic interactions. - *Ecography (Cop.)*. 37: 406–415.
- Bryant, J. A. et al. 2008. Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. - *Proc. Natl. Acad. Sci. U. S. A.* 105: 11505–11.
- Cadotte, M. W. et al. 2008. Evolutionary history and the effect of biodiversity on plant productivity. - *Proc. Natl. Acad. Sci. U. S. A.* 105: 17012–7.
- Chao, A. et al. 2010. Phylogenetic diversity measures based on Hill numbers. - *Philos. Trans. R. Soc. London, Ser. B* 365: 3599–3609.
- Colless, D. 1982. Review of phylogenetics: the theory and practice of phylogenetic systematics. - *Syst. Zool.* 31: 100–104.
- Davies, T. J. and Buckley, L. B. 2011. Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. - *Philos. Trans. R. Soc. London, Ser. B* 366: 2414–2425.
- Davies, T. J. et al. 2011. The influence of past and present climate on the biogeography of modern mammal diversity. - *Philos. Trans. R. Soc. London, Ser. B* 366: 2526–2535.
- Dearing, M. D. et al. 2000. Diet breadth of mammalian herbivores: nutrient versus detoxification constraints. - *Oecologia* 123: 397–405.
- Devictor, V. et al. 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.
- Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - *Biol. Conserv.* 61: 1–10.
- Faith, D. P. 2013. Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework. - *Ann. N. Y. Acad. Sci.* 1289: 69–89.
- Fritz, S. and Rahbek, C. 2012. Global patterns of amphibian phylogenetic diversity. - *J. Biogeogr.* 38: 1373–1382.
- Hardy, O. J. and Senterre, B. 2007. Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. - *J. Ecol.* 95: 493–506.
- Harmon, L. J. et al. 2008. GEIGER: investigating evolutionary radiations. - *Bioinformatics* 24: 129–31.
- HilleRisLambers, J. et al. 2012. Rethinking Community Assembly through the Lens of Coexistence Theory. - *Annu. Rev. Ecol. Evol. Syst.* 43: 227–248.
- Hooper, L. V. et al. 2012. Interactions between the microbiota and the immune system. - *Science* 336: 1268–73.
- Isaac, N. J. B. et al. 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. - *PLoS One* 2: e296.
- Johnson, W. et al. 2006. The late Miocene radiation of modern Felidae: a genetic assessment. - *Science (80-.)*. 311: 73–77.
- Kembel, S. W. et al. 2010. Picante: R tools for integrating phylogenies and ecology. - *Bioinformatics* 26: 1463–4.
- Kissling, W. et al. 2012. Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide. - *Proc. Natl. Acad. Sci. U. S. A.* 109: 7379–7384.
- Ley, R. E. et al. 2008. Evolution of mammals and their gut microbes. - *Science* 320: 1647–51.
- Lozupone, C. and Knight, R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. - *Appl. Environ. Microbiol.* 71: 8228–8235.
- Morlon, H. et al. 2011. Spatial patterns of phylogenetic diversity. - *Ecol. Lett.* 14: 141–149.
- Mouquet, N. et al. 2012. Ecophylogenetics: advances and perspectives. - *Biol. Rev.* 87: 769–785.
- Muegge, B. D. et al. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. - *Science* 332: 970–4.
- Nyakatura, K. and Bininda-Emonds, O. R. P. 2012. Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. - *BMC Biol.* 10: 12.
- O'Dwyer, J. et al. 2012. Phylogenetic diversity theory sheds light on the structure of microbial communities. - *PLoS Comput. Biol.* 8: e1002832.
- O'Meara, B. 2012. Evolutionary inferences from phylogenies: a review of methods. - *Annu. Rev. Ecol. Evol. Syst.* 43: 267–285.
- Pavoine, S. and Bonsall, M. B. 2011. Measuring biodiversity to explain community assembly: a unified approach. - *Biol. Rev.* 86: 792–812.
- Pedersen, R. Ø. et al. 2014. Macroecological evidence for competitive regional-scale interactions between the two major clades of mammal carnivores (Feliformia and Caniformia). (D Russo, Ed.). - *PLoS One* 9: e100553.
- Pybus, O. and Harvey, P. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. - *Proc. R. Soc. London, Ser. B* 267: 2267–2272.
- R Development Core Team 2014. R: A Language and Environment for Statistical Computing.
- Redding, D. W. et al. 2014. Measuring Evolutionary Isolation for Conservation. (WJ Murphy, Ed.). - *PLoS One* 9: e113490.
- Ricotta, C. 2007. A semantic taxonomy for diversity measures. - *Acta Biotheor.* 55: 23–33.
- Safi, K. et al. 2011. Understanding global patterns of mammalian functional and phylogenetic diversity. - *Philos. Trans. R. Soc. London, Ser. B* 366: 2536–2544.
- Swenson, N. G. 2011. Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. - *PLoS One* 6: e21264.
- Tucker, C. and Cadotte, M. 2013. Unifying measures of biodiversity: understanding when richness and phylogenetic diversity should be congruent. - *Divers. Distrib.* 19: 845–854.
- Warren, D. L. et al. 2014. Mistaking geography for biology: inferring processes from species distributions. - *Trends Ecol. Evol.* 29: 572–580.
- Webb, S. 2006. The great American biotic interchange: patterns and processes. - *Ann. Missouri Bot. Gard.*: 245–257.
- Webb, C. O. et al. 2002. Phylogenies and Community Ecology. - *Annu. Rev. Ecol. Evol. Syst.* 33: 475–505.
- Webb, C. O. et al. 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. - *Bioinformatics* 24: 2098–2100.
- Woodburne, M. O. 2010. The Great American Biotic Interchange: Dispersals, Tectonics, Climate, Sea Level and Holding Pens. - *J. Mamm. Evol.* 17: 245–264.

Partie I

Partie I

SYNTHÈSE ET DISCUSSION DE LA PARTIE I

Cette première partie de ma thèse visait à répondre à la question suivante :

***Peut-on synthétiser les indices de diversités
le long de grandes dimensions structurelles?***

1. Synthèse du travail

Je proposai en introduction de construire une classification des indices de diversité phylogénétique en commençant par différencier l'unité de base que les indices utilisent :

- (1) la paire de communautés (diversité phylogénétique β)
- (2) la communauté (diversité phylogénétique α)
- (3) l'espèce (originalité phylogénétique).

Les chapitres I et II ainsi que l'annexe 1 montrent qu'une carte simplifiée de la jungle des indices de diversité phylogénétique existe et nous identifions des sentiers pour faciliter l'exploration de la jungle. Ces sentiers n'ont pas été tracés à l'aveugle mais regroupent des indices aux propriétés similaires, c'est-à-dire mesurant une même dimension de la diversité phylogénétique. J'espère que ce type de cartographie évitera à l'avenir un grand nombre des marches aléatoires contre-productives dans la jungle des indices de diversité.

1.1. Les indices de diversité au niveau des communautés

Le premier chapitre se base sur les trois dimensions structurelles fondamentales identifiées par Pavoine & Bonsall (2011) : la richesse, la divergence et la régularité. Cette grille de lecture nous permet de (1) jalonner le sentier de panneaux d'orientation indiquant l'appartenance des indices à l'une ou l'autre des dimensions et (2) de fournir un manuel d'orientation qui relie les questions de recherche des écologistes avec les

dimensions indiquées sur le sentier. Le deuxième chapitre est complémentaire du premier car il identifie une dimension transversale, applicable à chacune des dimensions proposées : c'est l'échelle phylogénétique avec laquelle les structures sont mesurées. Par exemple, nous montrons que la PD_{Faith} quantifie en fait une structure phylogénétique « terminale » (c'est à dire près des feuilles de l'arbre) et que, lorsqu'il est utilisé tout seul il ne renseigne pas sur la structure phylogénétique profonde des assemblages. Finalement, nous pouvons résumer ces deux chapitres en trois groupes d'axes orthogonaux :

- (1) l'échelle phylogénétique ou fonctionnelle considérée
- (2) la dimension structurale considérée (richesse, divergence, régularité)
- (3) l'intégration des abondances (locales, régionales ou mondiales).

En définitive, nous avons montré dans ces deux premiers chapitres comment les indices de diversité phylogénétique pouvaient être classés de façon rigoureuse. Il nous manque cependant la « diversité » d'une identité biologique : l'originalité des espèces. Ceci fait l'objet de l'annexe 1.3, résumé ci-dessous.

1.2. Les indices d'originalité phylogénétique des espèces

L'annexe 1.3 est le fruit d'une collaboration avec David Redding et Arne Mooers, que j'ai rencontrés au workshop sur la diversité phylogénétique à Leipzig. L'étude était déjà largement avancée (notamment les simulations) lorsque j'ai été associé au projet et ma contribution a donc été limitée (corrections des formules des indices, liens avec les diversités des communautés, correction du manuscrit).

Dans cette analyse, nous montrons que les indices d'originalité phylogénétique s'évaluent en fait sur une échelle phylogénétique. Du côté « récent » du spectre, l'indice phare représente l'âge de l'espèce (c'est-à-dire l'âge de l'ancêtre commun à l'espèce et à son espèce sœur, le « pendant edge », PE) alors que du côté « interne » du spectre, on retrouve la distance moyenne de l'espèce à toutes les autres espèces de l'arbre. Ainsi, on s'aperçoit que l'échelle phylogénétique est une caractéristique fondamentale de tous les indices proposés dans la littérature et permet de tester des hypothèses écologiques différentes.

2. Utilité de l'approche

Cette carte des sentiers, accompagnée de son manuel d'orientation, devrait permettre aux chercheurs de ne plus se perdre dans la jungle et donc de choisir un indice pour quantifier une facette de la structure phylogénétique d'un assemblage ou de l'originalité d'une espèce. Une fois choisi, l'indice peut être utilisé directement, par exemple dans le cas des communautés:

- (1) en le reliant à des caractéristiques environnementales ou taxonomiques dans le cadre d'une recherche fondamentale (Cadotte et al. 2008, Safi et al. 2011, Belmaker and Jetz 2015)
- (2) en maximisant sa valeur dans le cadre de la conservation (Rodrigues and Gaston 2002, Rodrigues et al. 2011, Pollock et al. 2015).

Dans le cadre des originalités phylogénétiques, on peut notamment y voir un outil intéressant en biologie des invasions (Thuiller et al. 2010) puisqu'il permet de quantifier l'originalité d'une espèce invasive à la communauté résidente.

3. Limites et perspectives du travail

3.1. Lien avec la diversité fonctionnelle

Le travail de synthèse effectué se concentre sur les mesures de diversité phylogénétiques. Il est cependant important de faire le lien avec les indices de diversités fonctionnelles car les deux approches sont souvent complémentaires. En ce sens, ils permettent un rapprochement entre les facettes fonctionnelles et phylogénétiques et donc un rapprochement de la biogéographie évolutive et de la biogéographie fonctionnelle.

Nous avons constamment proposé de faire le lien, particulièrement dans l'article 1 et l'annexe 1.3, entre indices phylogénétiques et indices fonctionnels. Ainsi, la carte de classification que nous proposons pour les indices de l'article 1 est aisément transposable à l'univers fonctionnel. Dans le cas le plus simple, les dissimilarités

fonctionnelles sont représentées sous forme de dendrogramme (Petchey and Gaston 2007) et dans ce cas, tous les indices de diversités phylogénétiques ont un équivalent fonctionnel, qui d'ailleurs ont, dans certains cas, un nom (malheureusement) différent. Par exemple, l'équivalent de PD_{Faith} en fonctionnel est un indice très utilisé, il s'agit de la FD sensu Petchey & Gaston (2007). Même dans les cas où les distances fonctionnelles entre espèces ne sont pas représentées sous forme de dendrogramme, les liens entre indices sont criants : le MPD ou l'entropie quadratique de Rao sont utilisables autant dans un espace fonctionnel que phylogénétique.

3.2. Proposer une revue synthétique des indices d'originalité phylogénétiques et fonctionnels

Dans le cadre des originalités d'espèces, un travail de synthèse entre phylogénie et fonctions semble urgent. En effet, on dénombre de très nombreux indices dans les deux domaines, mais aucune synthèse ne permet d'en faire le lien, ni même de les ranger dans des cases compréhensibles et facilement utilisables par les écologistes. Pourtant les parallèles semblent clairs et je pense que l'on peut décomposer, encore une fois, les indices le long (1) d'une échelle phylogénétique ou fonctionnelle et (2) d'un ensemble d'espèces de référence et (3) d'une considération de l'abondance de l'espèce :

(1) Les indices d'originalité fonctionnelle s'étalent eux aussi le long d'un continuum d'échelle. Par exemple, Buisson *et al.* (2008) définissent « l'originalité » et « l'unicité » fonctionnelles des espèces comme la distance moyenne d'une espèce à l'ensemble de référence et la distance moyenne à l'espèce la plus proche dans l'espace fonctionnel. Ces indices correspondent tout simplement aux deux extrémités du continuum d'originalité phylogénétique mais possèdent – évidemment - des noms différents, ce qui ne facilite pas une vision intégrative.

(2) L'ensemble d'espèces de référence auquel une espèce donnée est comparée permet là aussi d'unifier des littératures qui ne communiquent pas forcément.

a. Échelle locale de la communauté

Les indices d'originalités à l'échelle de la communauté sont utilisés en biologie de l'invasion pour comprendre si les caractéristiques de la

communauté résidente influencent le succès invasif de l'espèce exotique (voir Thuiller *et al.* 2010 pour une synthèse du domaine et Gallien *et al.* 2015 pour une application, dont je reporte le résumé en annexe 1.4).

b. Échelle plus large de l'aire de distribution de l'espèce

C'est ce qui a été appelé les « champs phylogénétiques » des espèces (« phylogenetic fields », Villalobos *et al.*, 2013) et qui a notamment été utilisé pour comprendre la structure de cooccurrence des oiseaux (Barnagaud *et al.* 2014).

c. Échelle regroupant toutes les espèces d'un taxon donné

Ceci correspond aux mesures « classiques » d'originalité.

(3) La pondération de l'originalité des espèces par leur rareté (par exemple la taille de leurs aires de distribution) est de plus en plus employée, notamment dans une optique de conservation (Jetz *et al.* 2014).

Je pense que ce travail de synthèse pourra clarifier les relations entre indices et donc permettre aux écologues de plus facilement choisir les indices en fonction de leur question.

3.3. Incorporer l'effet des modèles nuls

Dans la synthèse des indices de diversité de communauté proposée au chapitre I, nous avons utilisé la valeur brute des indices de diversité pour les classer. Cependant, la valeur des indices de diversités est souvent comparée à une distribution de valeurs « nulles » qui dérivent d'un modèle nul particulier (Gotelli and Colwell 2001, Gotelli and McGill 2006, Hardy 2008, Chalmandrier *et al.* 2013). Ce modèle nul permet de tester si une structure de communauté est aléatoire relativement à une série d'hypothèses précises. Les indices sont souvent standardisés par la moyenne et l'écart-type de la distribution nulle. Nous n'avons pas intégré cette dimension dans le chapitre I par manque de place et car les indices peuvent aussi s'utiliser sans modèles nuls. Cependant, cette dimension est explorée dans le chapitre II. Il apparaît que lorsqu'un modèle nul est utilisé, les dimensions mises à jour dans le premier chapitre sont moins apparentes. En effet, la PD_{Faith} considérée comme un indice de richesse phylogénétique se trouve très

proche du MNTD, un indice de divergence phylogénétique (voir le chapitre II). Je pense que cette différence apparente ne diminue pas l'intérêt de mon premier chapitre car l'utilisation des modèles nuls n'est pas systématique, notamment en conservation. Par ailleurs, il faudrait probablement explorer le comportement des indices bruts et standardisés par le modèle nul de façon systématique, ce qui pourra faire l'objet d'études futures.

4. Transition

La seconde et la troisième partie de ma thèse s'attachent à utiliser quelques unes de ces avancées conceptuelles en biogéographie de la conservation et en biogéographie fondamentale. En particulier, la partie II s'intéresse à la richesse phylogénétique et à sa congruence avec la richesse taxonomique. Je propose aussi de comparer différentes dimensions (richesse et divergence) pour construire des points chauds (« hotspots ») de biodiversité. Finalement, la troisième partie est bâtie sur l'utilisation de l'échelle phylogénétique. En effet, nous montrons dans ces trois premières analyses l'importance de l'échelle phylogénétique pour décrire et comprendre la structure des entités biologiques. J'utilise cette démarche pour distinguer l'effet de différents processus dans l'assemblage de communautés.

BIBLIOGRAPHIE GÉNÉRALE
DE LA PARTIE I
(HORS ARTICLES)

Partie I - Bibliographie

- Barnagaud, J.-Y. et al. 2014. Ecological traits influence the phylogenetic structure of bird species co-occurrences worldwide. (D Mouillot, Ed.). - *Ecol. Lett.* 17: 811–20.
- Belmaker, J. and Jetz, W. 2015. Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients (H Arita, Ed.). - *Ecol. Lett.* 18: 563–571.
- Buisson, L. E. T. et al. 2008. Climate change hastens the turnover of stream fish assemblages. - *Glob. Chang. Biol.* 14: 2232–2248.
- Buisson, L. et al. 2010. Uncertainty in ensemble forecasting of species distribution. - *Glob. Chang. Biol.* 16: 1145–1157.
- Cadotte, M. W. et al. 2008. Evolutionary history and the effect of biodiversity on plant productivity. - *Proc. Natl. Acad. Sci. U. S. A.* 105: 17012–7.
- Chalmandrier, L. et al. 2013. A family of null models to distinguish between environmental filtering and biotic interactions in functional diversity patterns. - *J. Veg. Sci. Off. organ Int. Assoc. Veg. Sci.* 24: 853–864.
- Díaz, S. et al. 2007. Incorporating plant functional diversity effects in ecosystem service assessments. - *Proc. Natl. Acad. Sci. U. S. A.* 104: 20684–9.
- Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - *Biol. Conserv.* 61: 1–10.
- Gallien, L. et al. 2015. Contrasting the effects of environment, dispersal and biotic interactions to explain the distribution of invasive plants in alpine communities. - *Biol. Invasions* 17: 1407–1423.
- Gotelli, N. and Colwell, R. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. - *Ecol. Lett.* 4(4), 379–391.
- Gotelli, N. and McGill, B. 2006. Null versus neutral models: what's the difference? - *Ecography (Cop.)*. 29: 793–800.
- Graham, C. H. and Fine, P. V. A. 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. - *Ecol. Lett.* 11: 1265–77.
- Hardy, O. 2008. Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. - *J. Ecol.* 96: 914–926.
- Isaac, N. J. B. et al. 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. - *PLoS One* 2: e296.
- Jetz, W. et al. 2014. Global distribution and conservation of evolutionary distinctness in birds. - *Curr. Biol.* 24: 919–30.
- Mouquet, N. et al. 2012. Ecophylogenetics: advances and perspectives. - *Biol. Rev.* 87: 769–785.
- Pavoine, S. and Bonsall, M. B. 2011. Measuring biodiversity to explain community assembly: a unified approach. - *Biol. Rev.* 86: 792–812.
- Petchey, O. L. and Gaston, K. J. 2007. Dendrograms and measuring functional diversity. - *Oikos* 116: 1422–1426.
- Pollock, L. J. et al. 2015. Phylogenetic diversity meets conservation policy: small areas are key to preserving eucalypt lineages. - *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370: 20140007.
- Rodrigues, A. S. L. and Gaston, K. J. 2002. Maximising phylogenetic diversity in the selection of networks of conservation areas. - *Biol. Conserv.* 105: 103–111.
- Rodrigues, A. S. L. et al. 2011. Complete, accurate, mammalian phylogenies aid conservation planning, but not much. - *Philos. Trans. R. Soc. London, Ser. B* 366: 2652–2660.
- Safi, K. et al. 2011. Understanding global patterns of mammalian functional and phylogenetic diversity. - *Philos. Trans. R. Soc. London, Ser. B* 366: 2536–2544.
- Srivastava, D. and Cadotte, M. 2012. Phylogenetic diversity and the functioning of ecosystems. - *Ecol. Lett.* 15(7), 637–648.
- Swenson, N. G. 2011. Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. - *PLoS One* 6: e21264.
- Thuiller, W. et al. 2010. Resolving Darwin's naturalization conundrum: a quest for evidence. - *Divers. Distrib.* 16: 461–475.
- Villalobos, F. et al. 2013. Phylogenetic fields of species : cross-species patterns of phylogenetic structure and geographical coexistence. *Proc. R Soc Lond B*: 280(1756), 20122570.
- Webb, C. O. et al. 2002. Phylogenies and Community Ecology. - *Annu. Rev. Ecol. Evol. Syst.* 33: 475–505.
- Whittaker, R. 1960. Vegetation of the Siskiyou mountains, Oregon and California. - *Ecol. Monogr.* 30(3), 279–338.
- Winter, M. et al. 2012. Phylogenetic diversity and nature conservation: where are we? - *Trends Ecol. Evol.* 28: 199–204.

PARTIE II

UTILISER LES DIVERSITÉS PHYLOGÉNÉTIQUES ET FONCTIONNELLES EN CONSERVATION

SOMMAIRE DE LA PARTIE II

Articles associés à la partie II	120
Annexes associées à la partie II	120
INTRODUCTION. État de l'art	121
1. Contexte	122
2. La biologie de la conservation	122
3. L'apport potentiel des facettes fonctionnelles et phylogénétiques	124
4. Synthèse et stratégie de travail	125
CHAPITRE III. Quantification de l'érosion de la diversité phylogénétique	128
Résumé	129
Article	130
CHAPITRE IV. Les points chauds de diversité fonctionnelle et phylogénétique	140
Résumé	141
Article	142
SYNTHÈSE & DISCUSSION	155
1. Évaluer les menaces	156
2. La protections des différentes facettes	158
3. Perspectives	162
BIBLIOGRAPHIE	164

ARTICLES ASSOCIÉS À LA PARTIE II

ARTICLE 4

Mazel, F., Renaud, J., Guilhaumon, F., Mouillot, D., Gravel D., & Thuiller, W. Mammalian phylogenetic diversity area relationships at a continental scale, 2015.

Ecology, 96 (10), 2814-2823.

ARTICLE 5

Mazel, F., Guilhaumon, F., Mouquet, N., Devictor, V., Gravel, D., Renaud, J., Cianciaruso, M.V., Loyola, R.D., Diniz-Filho, J.A.F., Mouillot, D. & Thuiller, W. (2014) Multifaceted diversity-area relationships reveal global hotspots of mammalian species, trait and lineage diversity, 2014.

Global Ecology and Biogeography, 23, 836–847.

ARTICLE 6

Thuiller, W., Maiorano, L., **Mazel, F.**, Guilhaumon, F., Ficetola, G.F., Lavergne, S., Renaud, J., Roquet, C. and Mouillot, D. 2015. Conserving the functional and phylogenetic trees of life of European tetrapods, 2014.

Philosophical Transactions of the Royal Society of London B: Biological Sciences, 370(1662).

ANNEXES ASSOCIÉES À LA PARTIE II

ANNEXE 2.1. Matériel supplémentaire de l'article 4

ANNEXE 2.2. Matériel supplémentaire de l'article 5

ANNEXE 2.3. Article 6

INTRODUCTION DE LA PARTIE II

1. Contexte

Il est aujourd'hui largement admis que les activités humaines menacent l'avenir de la biodiversité (Pimm et al. 1995, Pimm and Raven 2000, Ceballos and Ehrlich 2002, Brook et al. 2003, Barnosky et al. 2011, 2012, De Vos et al. 2015). Ces menaces sont de différentes natures (appropriation des ressources, fragmentation des habitats, introduction d'espèces exotiques, déplacement de pathogènes, destruction directe des espèces, changements climatiques) et ont pour conséquences le déclin des populations, la modification du fonctionnement des écosystèmes et l'extinction accélérée des espèces par rapport aux taux « normaux » (c'est-à-dire mesurés sur des temps géologiques).

La prise de conscience de telles menaces n'est en fait pas nouvelle et date de la fin du XIX^e siècle (Jepson and Whittaker 2002). Elle s'est accompagnée de la naissance d'un mouvement de conservation de la nature, surtout dans les élites occidentales dans un premier temps, puis de toute la société à partir des années 50 et les décennies suivantes (à titre d'exemple Greenpeace et WWF sont fondés en 1961 et 1971, respectivement). Ce mouvement a été ensuite rejoint par les scientifiques dont l'expertise était vitale pour (1) prédire l'impact des menaces et (2) protéger les espèces le plus efficacement possible.

2. La biologie de la conservation

La biologie de la conservation en tant que discipline scientifique prend ses racines dans les années 50 et propose de décrire ces changements, en comprendre les mécanismes pour éventuellement les limiter et ainsi préserver cette biodiversité (Soulé 1985, Whittaker et al. 2005).

2.1. Les motivations

Les motivations de cette préservation peuvent être d'ordre éthique (c'est-à-dire que l'espèce humaine a un devoir moral de protéger les autres êtres vivants) ou d'ordre utilitariste (Ladle and Whittaker 2011). En effet les écosystèmes sont de plus en plus considérés comme des fournisseurs de « services » pour l'espèce humaine. Ces services des écosystèmes peuvent être des services de régulation (régulation de l'eau, du climat,

des nutriments, pollinisation, formation et rétention des sols, etc.) de production (nourriture, bois, ressources médicales, ornementales et génétiques, etc.) ou de « bien-être » (valeur esthétique, culturelle, historique) (de Groot et al. 2002). Dans l'économie actuelle, on tente alors d'assigner une valeur monétaire aux services des écosystèmes afin de pouvoir les intégrer au marché (par exemple, voir Costanza *et al.* 1997).

2.2. Les approches

On distingue en général deux approches en biologie de la conservation : l'approche fonctionnelle et l'approche compositionnelle (Callicott et al. 1999). L'approche fonctionnelle privilégie plutôt les propriétés générales des écosystèmes comme la taille des compartiments et les flux de matières alors que l'approche compositionnelle se concentre davantage sur les entités biologiques comme les individus ou les espèces. Les deux approches ont leurs avantages et leurs inconvénients et il est donc généralement admis qu'elles sont toutes les deux importantes (Ladle and Whittaker 2011).

Dans le cadre de l'approche compositionnelle, l'entité biologique de choix à préserver est traditionnellement l'espèce (Matthews et al. 2001) et/ou la richesse taxonomique (c'est-à-dire le nombre d'espèces, Rodrigues *et al.* 2000), et les choix sont très souvent dictés par des critères patrimoniaux. En particulier, les espèces rares localement ou endémiques (c'est-à-dire dont l'aire de distribution est restreinte à la zone d'étude) ont souvent un poids plus important dans les plans de conservation. Un des outils importants de la biologie de la conservation est d'aménager des aires de protection des espèces où elles seront moins menacées. Le choix de l'emplacement de telles zones se base sur quatre critères. Il s'agit d'élaborer un projet flexible socialement et viable économiquement pour protéger un échantillon représentatif de l'ensemble biologique qui sera capable de se maintenir à plus ou moins long terme. La planification systématique de la conservation (« systematic conservation planning », SCP) est une discipline récente (Margules and Pressey 2000) qui propose des outils quantitatifs et transparents pour réaliser ces objectifs. Par exemple, l'approche dites en « cœur de zone » (« Core Area » in Zonation, Moilanen *et al.* 2009) permet de sélectionner un ensemble optimal de zones à protéger, en donnant un poids plus important aux espèces qui ont une aire de distribution réduite.

L'approche fonctionnelle privilégie plutôt les propriétés fonctionnelles des écosystèmes comme la taille des compartiments de matières et les flux entre ces compartiments (par exemple stock et flux de carbone, respectivement). Ce fonctionnement est souvent vital pour préserver les services que fournissent les écosystèmes aux sociétés humaines (Loreau 2010, Cardinale et al. 2012). Il s'agit donc de prédire les conséquences des changements globaux sur le fonctionnement et les services des écosystèmes et éventuellement d'en restaurer certaines composantes (Ladle and Whittaker 2011).

3. L'apport potentiel des facettes fonctionnelles et phylogénétiques

3.1. La diversité fonctionnelle

Il est maintenant bien établi que ce sont les traits fonctionnels des individus qui assurent le rôle de l'espèce dans l'écosystème et donc à terme son impact sur le fonctionnement et ses services (Díaz et al. 2007, Lavorel et al. 2011). En effet, la diversité fonctionnelle apparaît souvent comme un meilleur prédicteur du fonctionnement des écosystèmes que la richesse spécifique (Díaz and Cabido 2001, Petchey and Gaston 2006, Díaz et al. 2007, Mouillot and Villéger 2011, Paquette and Messier 2011, Lavorel et al. 2011). Il semble donc que la perte d'une espèce unique et importante fonctionnellement puisse avoir des conséquences plus importantes sur le fonctionnement de l'écosystème que la perte d'une espèce redondante (Mouillot et al. 2013, 2014). Ainsi, la préservation d'une telle facette pourrait représenter une manière efficace de conserver le fonctionnement d'un écosystème.

3.2. La diversité phylogénétique

L'importance de l'utilisation de la diversité phylogénétique en conservation est reconnue depuis longtemps (Vane-Wright et al. 1991, Faith 1992) et a été réaffirmée à de nombreuses reprises (Rodrigues and Gaston 2002, Forest et al. 2007, Thuiller et al. 2011, Pollock et al. 2015). Cependant, son utilisation dans la pratique reste

extrêmement limitée (Winter et al. 2012) mis à part la mise en place du projet 'Edge of Existence' (EDGE; Isaac *et al.* 2007) qui se focalise sur la conservation des espèces distinctes évolutivement et globalement en danger d'extinction (<http://www.edgeofexistence.org/>). Pourtant la diversité phylogénétique semble avoir plusieurs avantages en conservation.

Tout d'abord, si les traits sont conservés le long de la phylogénie, la diversité phylogénétique pourrait être un proxy intéressant du fonctionnement des écosystèmes. Ceci a été montré dans certains cas, notamment pour la productivité (Cadotte et al. 2008, 2009) et il apparaît que l'utilisation conjointe de la diversité phylogénétique et fonctionnelle soit prometteuse en ce sens (Flynn et al. 2011, Srivastava and Cadotte 2012, Cardinale et al. 2012). Cependant, ces expériences sont relativement restreintes et des évidences empiriques plus robustes sont nécessaires pour prouver le lien entre diversité phylogénétique et fonctionnement. Ceci a fait dire à certains auteurs que, dans l'état actuel des connaissances, l'intérêt de l'utilisation de la diversité phylogénétique en conservation serait limité (Winter et al. 2012). C'est oublier un aspect fondamental de l'information portée par cette facette : l'histoire (Rosauer and Mooers 2013). En effet, au-delà de l'intérêt utilitariste des services des écosystèmes, préserver une histoire unique revêt une dimension éthique fondamentale (Vane-Wright et al. 1991, Rosauer and Mooers 2013). En ce sens, la diversité phylogénétique est une composante fondamentale de la biodiversité et doit donc participer à l'objectif de représentativité.

4. Synthèse et stratégie de travail

En définitive, il s'agit, selon Lecointre (2011), de préserver les espèces pour ce qu'elles font (=approche fonctionnelle) et pour ce qu'elles sont (=approche phylogénétique). Dans un premier temps, je vais tenter de comprendre les conséquences des changements globaux sur certaines facettes encore mal explorées de la biodiversité. Le chapitre III tente ainsi de quantifier l'érosion de la diversité phylogénétique avec la perte d'habitat à une échelle mondiale. La prédiction de la perte de diversité n'est cependant que le premier pas vers la protection de cette diversité à plusieurs facettes. Il est alors alarmant de voir que certaines analyses commencent à montrer que les différentes facettes de la biodiversité ne sont pas interchangeables ou « congruentes », c'est-à-dire qu'en en préservant une, on ne garantit pas forcément la protection des

autres (p. ex. Devictor *et al.* 2010). Le deuxième chapitre (chapitre IV) de cette deuxième partie vise donc à vérifier ce résultat à une échelle mondiale (Devictor *et al.* se focalisent à l'échelle d'un seul pays) en utilisant les mammifères comme exemple. Finalement, l'annexe 2.3 teste si les aires de protection actuelles protègent efficacement ou non la diversité phylogénétique et fonctionnelle des tétrapodes en Europe.

CHAPITRE III

QUANTIFICATION DE L'ÉROSION DE LA DIVERSITÉ PHYLOGÉNÉTIQUE DES MAMMIFÈRES À L'ÉCHELLE DU GLOBE SUITE À LA PERTE D'HABITAT

Ce chapitre pose la question suivante :

La diversité phylogénétique est-elle plus ou moins robuste que la richesse spécifique à la perte d'habitat ?

Résumé de l'article

De la même façon que la relation aire-espèce (SAR pour « species area relationship »), la relation aire-diversité phylogénétique (PDAR pour « phylogenetic diversity area relationship ») décrit la tendance de la diversité phylogénétique à augmenter avec l'aire d'échantillonnage. La recherche sur les PDARs pourrait permettre de mieux comprendre les processus influençant l'assemblage d'espèces à différentes échelles spatiales et de prédire la perte de diversité phylogénétique (PD pour 'diversité phylogénétique') des suites de la perte d'habitat. Cependant ces PDARS sont encore très peu étudiés. Étudier la PD a de nombreux avantages par rapport à la richesse spécifique (SR pour 'species richness') puisque la PD nous renseigne sur des processus comme l'extinction et la spéciation, les règles d'assemblage et le fonctionnement des écosystèmes. Dans cette analyse, nous étudions l'universalité de la forme des PDARs à une échelle continentale en utilisant les mammifères terrestres. Nous définissons une mesure relative (comparée à la SR) de la robustesse de la PD à la perte d'habitat et nous montrons qu'elle est toujours supérieure à la robustesse de la SR mais qu'elle varie selon les continents. Nous utilisons ensuite un modèle nul pour évaluer la part de robustesse expliquée par la nature hiérarchique de l'arbre et la distribution spatiale des lignées. Nous montrons que cette robustesse est relativement bien expliquée par la forme de l'arbre mais que dans certains cas la distribution spatiale des lignées est aussi importante.

Mammalian phylogenetic diversity–area relationships at a continental scale

FLORENT MAZEL,^{1,2,7} JULIEN RENAUD,^{1,2} FRANÇOIS GUILHAUMON,³ DAVID MOUILLOT,^{4,5} DOMINIQUE GRAVEL,⁶
AND WILFRIED THULLER^{1,2}

¹Université Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France

²CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France

³IRD, MARBEC Université de Montpellier, Montpellier, France

⁴MARBEC Université de Montpellier, Montpellier, France

⁵ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland 4811 Australia

⁶Université du Québec à Rimouski, Département de Biologie, Chimie et Géographie, Québec, Québec G5L 3A1 Canada

Abstract. In analogy to the species–area relationship (SAR), one of the few laws in ecology, the phylogenetic diversity–area relationship (PDAR) describes the tendency of phylogenetic diversity (PD) to increase with area. Although investigating PDAR has the potential to unravel the underlying processes shaping assemblages across spatial scales and to predict PD loss through habitat reduction, it has been little investigated so far. Focusing on PD has noticeable advantages compared to species richness (SR), since PD also gives insights on processes such as speciation/extinction, assembly rules and ecosystem functioning. Here we investigate the universality and pervasiveness of the PDAR at continental scale using terrestrial mammals as study case. We define the relative robustness of PD (compared to SR) to habitat loss as the area between the standardized PDAR and standardized SAR (i.e., standardized by the diversity of the largest spatial window) divided by the area under the standardized SAR only. This metric quantifies the relative increase of PD robustness compared to SR robustness. We show that PD robustness is higher than SR robustness but that it varies among continents. We further use a null model approach to disentangle the relative effect of phylogenetic tree shape and nonrandom spatial distribution of evolutionary history on the PDAR. We find that, for most spatial scales and for all continents except Eurasia, PDARs are not different from expected by a model using only the observed SAR and the shape of the phylogenetic tree at continental scale. Interestingly, we detect a strong phylogenetic structure of the Eurasian PDAR that can be predicted by a model that specifically account for a finer biogeographical delineation of this continent. In conclusion, the relative robustness of PD to habitat loss compared to species richness is determined by the phylogenetic tree shape but also depends on the spatial structure of PD.

Key words: conservation biogeography; habitat loss; null models; phylogenetic diversity; species–area relationship; strict nested design.

INTRODUCTION

The species–area relationship (SAR) describes the tendency of species richness (SR) to increase with area (Rosenzweig 1995). This relationship is documented for a wide range of taxonomic groups and ecosystems (Guilhaumon et al. 2008, Triantis et al. 2012) and its understanding is central to ecology and conservation biogeography (Rosenzweig 1995, Whittaker et al. 2005). For instance, the SAR is a key tool to estimate species extinctions from habitat destruction and climate change (Pimm and Raven 2000, Thomas et al. 2004, Pereira et al. 2010, Matias et al. 2014). Nevertheless, a SAR approach reduces biological diversity to species richness only and fails to include the amount of evolutionary

history in species assemblages (Mouquet et al. 2012). To fill this gap, the phylogenetic diversity–area relationship (PDAR; Morlon et al. 2011) can help unravel the processes assembling communities across spatiotemporal scales and provides complementary tools for conserving the Tree of Life (Mazel et al. 2014). For instance, translating SAR into PDAR allows to predict the loss of PD through habitat destruction. This prediction is essential since the loss of a given amount of PD or the loss of an entire lineage could have strong negative ecological consequences since distinct lineages are likely to perform different functions (Cadotte et al. 2008, Mouquet et al. 2012).

Several mechanisms have been proposed to explain the SAR, such as sampling effects (Rosenzweig 1995), the effect of habitat size on extinction rates (MacArthur and Wilson 1967), the scaling of environmental heterogeneity with area (Kadmon and Allouche 2007) or dispersal limitation (Hubbell 2001). In complement to

Manuscript received 2 October 2014; revised 17 March 2015; accepted 9 April 2015. Corresponding Editor: N. J. B. Kraft.

⁷ E-mail: flo.mazel@gmail.com

the SAR, the PDAR brings unique information about the different processes structuring biodiversity at different spatial scales, helping, for example, to quantify the effects of biotic interactions at small scales vs. biogeographical processes at large scale. A particular feature of the PDAR is that the shape of the phylogenetic tree ultimately drives its relative position to the SAR. A star phylogeny would produce a PDAR proportional to the SAR, while a complete and recent polytomy at the tips of the tree would produce an extreme PDAR that would reach its maximum from the smallest area (see Fig. 1A). In addition to those mechanisms, biogeographic history together with ecological processes should also influence the PDAR (see Fig. 1B). Allopatric speciation and/or competition between close relative species would result in a relatively higher PD than expected for a given SR (overdispersion; Webb et al. 2002, O'Dwyer et al. 2012). Coexistence theory indeed predicts that similar species will compete more strongly than dissimilar species, leading to the exclusion of one of the similar species (HilleRisLambers et al. 2012). If we assume that niche differences are properly portrayed by phylogenetic differences, we predict a phylogenetic overdispersion (i.e., distantly related species co-occur) under competitive interactions (Webb et al. 2002, but see Mayfield and Levine 2010). Reciprocally, low PD may be expected if close relative species tend to co-occur because of shared environmental niches and/or geographic isolation of land mass (phylogenetic clustering; Webb et al. 2002, O'Dwyer et al. 2012; see Fig. 1B). Overall the difference between SAR and PDAR curves is thus very informative on the way the phylogenetic structure of assemblages varies across spatial scales.

In summary, the PDAR is ultimately influenced by (1) the shape of the SAR that depends on species range placement over space (e.g., either clumped or random), (2) the structure of the phylogenetic tree, and (3) the species range placement in regard to the phylogeny (that ultimately depends on eco-evolutionary processes). Since the pioneering work by Morlon et al. (2011), that first introduced PDAR, no study has tried to explain large scale PDARs and to disentangle the relative influence of these three factors.

In this paper, we report the first large-scale analysis of PDAR over the globe for mammal assemblages. We used the calibrated and dated ultrametric phylogenetic tree updated by Fritz et al. (2009) from Bininda-Emonds et al. (2007). We extracted the distribution maps provided by the Mammal Red List Assessment (*available online*)⁸ for 4616 terrestrial species to obtain occurrence data on worldwide grid cells of approximately 110 × 110 km and used a strictly nested design recently published (SNQ; Storch et al. 2012) to produce median SAR and PDAR at a continental scale (i.e., we computed median SR and PD over each spatial scale to

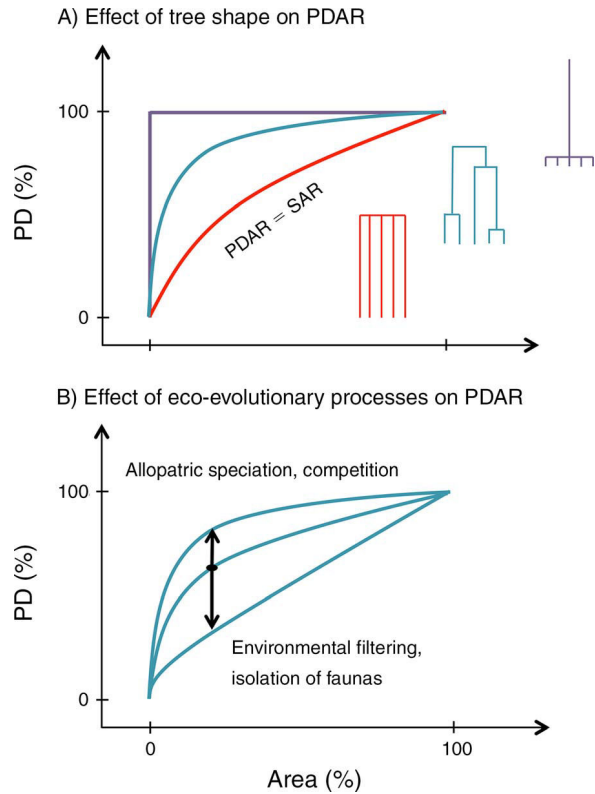


FIG. 1. Expected variation of the standardized phylogenetic diversity–area relationship (PDAR) given (A) different tree shapes and (B) different eco-evolutionary processes. (A) The three standardized PDARs correspond to the three trees depicted above the graph. Note that the red PDAR also corresponds to the observed species–area relationship (SAR) as the red tree is a star phylogeny. (B) Different eco-evolutionary processes may change the PDAR if they act differently among spatial scales. We expect that competition and/or allopatric speciation may relatively increase the phylogenetic diversity (PD) at small scale while environmental filtering and/or geographic isolation of biotas may relatively decrease the PD at small scale.

produce median SAR and PDAR, respectively). First we ask whether PDAR differs from the SAR at a continental scale and how this difference may affect the robustness of PD to habitat loss. To do so we define the relative robustness of PD (compared to SR) to habitat loss as the area between the standardized PDAR and standardized SAR (i.e., standardized by the diversity of the largest spatial window) divided by the area under the standardized SAR only (called the relative area under the curve, AUC_r). This metric quantifies the relative increase of PD robustness compared to SR robustness (Fig. 2). Second, we ask whether PDARs are a simple consequence of the observed SAR and a random sampling of species on the phylogenetic tree or if they also depend on eco-evolutionary processes. Assuming that the continental SAR for mammals can be adequately modelled by a random placement model of species ranges (Storch et al.

⁸ <http://www.iucnredlist.org/>

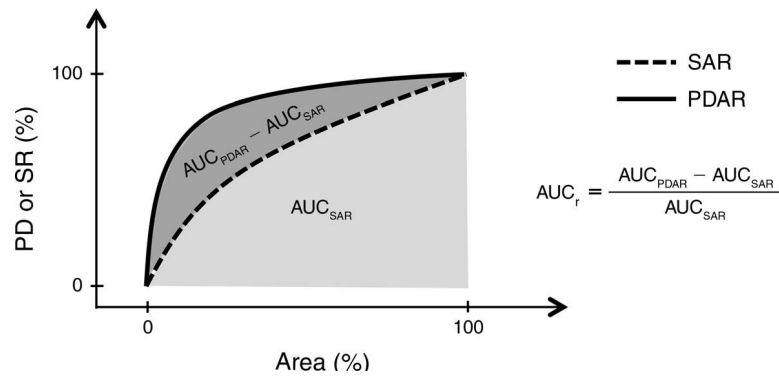


FIG. 2. Hypothetical example to quantify the relative robustness of PD (compared to species richness [SR]) to habitat loss (relative area under the curve [AUC_r]) using PDAR and SAR. The example shows how to quantify the relative PDAR shape by measuring the area between the two curves (SR, PD, and Area are expressed as percentages) and computing AUC_r .

2012), we derive PDAR expectations that only rely on the phylogenetic tree shape. To do so, we use a tip-shuffling null model that keeps the observed species range distribution, SAR and phylogenetic tree shape while shuffling the phylogenetic relationships among species. Third, we ask whether AUC_r depends on the phylogenetic tree shape only (see Fig. 1A) or if it is also an outcome of eco-evolutionary processes (see Fig. 1B). To do so we take advantage of our null model approach to produce null AUC_r expectations. More specifically we estimate the effect of tree structure (see Fig. 1A) on the relative robustness of PD to habitat loss across continents. Our analyses confirm that PD might be more robust than SR to habitat loss but that this higher robustness differs across continents for different reasons.

METHODS

Data sets and diversity metrics

We used the distribution maps provided by the Mammal Red List Assessment (see footnote 8) for 4616 terrestrial species to obtain occurrence data on worldwide grid cells of approximately 110×110 km. The best resolution to use the IUCN maps is still under discussion in the literature (Storch et al. 2012, Jenkins et al. 2013). We here used the resolution commonly used at global scale (Belmaker and Jetz 2011, Storch et al. 2012). This was our basic unit to construct SAR and PDAR. Domestic, aquatic, and semiaquatic mammals were excluded from the analysis.

We used the calibrated and dated ultrametric phylogenetic tree updated by Fritz et al. (2009) from Bininda-Emonds et al. (2007).

To characterize the PD of an assemblage we used the Faith's measure (Faith 1992). This metric represents a richness or volume of diversity (Pavoine and Bonsall 2011) and simply sums up branch lengths of the given species assemblage phylogeny (Rodrigues and Gaston 2002). Faith's measure is an intuitive and relatively simple measure of PD. It is also, by construction, generally highly correlated with SR (Huang et al. 2012,

Tucker and Cadotte 2013). The use of any PD metric theoretically linked with SR in the description of PDAR has been criticized because PDAR would be biased by "spurious artefacts of a statistical relationship between species richness and area" (Helmus and Ives 2012). Here the comparison of the SAR and the PDAR we propose, i.e., with the standardization and the null model that removes the effect of SR on PD (see *Analyzing SAR and PDR: Understanding the absolute value of PDAR*) avoid this artefact while it allows a simple interpretation of the results.

Constructing SAR and PDAR

Median and median absolute deviance (MAD) of SR and PD were reported for each spatial scale (from 110×110 km up to 2200×2200 km) by using the framework proposed by Storch et al. (2012). We do not use mean SR and PD as the data was highly non-normal (see Appendix A for examples of distributions of diversity). It uses a strictly nested quadrat design where a moving window (Leitner and Rosenzweig 1997, Lennon et al. 2001) reports the SR and PD of all possible windows of a given size within a continent. The median and MAD of SR and PD are then computed for each spatial scale. This procedure implies that some cells are counted several time for a given spatial windows and thus some pseudo-replication is inevitably introduced. Nevertheless all designs have their own drawbacks and SNQ have several important advantages (Storch et al. 2012). We implemented the algorithm within a reduced subset of the five continents (see Appendix B for further details) to avoid some border effect, i.e., for each scale (whatever its size), all pixels of the selected area of the continent will be sampled at least one time. The spectrum of spatial scale analyzed was set between 1×1 to 14×14 cells for Australia (i.e., from approximately 110×110 km to 1540×1540 km) and from 1×1 to 20×20 cells for North and South America, Africa, and Eurasia (i.e., from approximately 110×110 km to 2200×2200 km) following Storch et al. (2012). Note that the resulting

curve corresponds to a type I curve in the terminology proposed by Scheiner (2003).

Analyzing SAR and PDAR

Comparing the relative shape of SAR and PDAR.—We compared the shape of PDAR and SAR using two complementary approaches. First we fitted a power model (Rosenzweig 1995) to each SAR and PDAR. We then reported the slope (z) value of the linear model in a log-log space. These values were then used to depict in a simple way the relative shape of PDAR and SAR. Because PDAR and SAR are not necessarily best modelled by a power function (Guilhaumon et al. 2008, Mazel et al. 2014) we also (1) fitted alternative statistical models (see Appendix C) and (2) directly compared PDAR and SAR without an a priori function. As PD is expressed in units of times while SR in number of species, they are not directly comparable. Therefore we used a basic standardization procedure by rescaling each sampling windows PD and SR value by the value reached at the maximum sample size (Mazel et al. 2014). This gives a relative diversity value for each sampling windows, compared to the largest one (that thus represents 100%). We define the relative robustness of PD (compared to SR) to habitat loss as the area between the standardized PDAR and standardized SAR divided by the area under the standardized SAR only (Eq. 1 and Fig. 2).

$$AUC_r = \frac{AUC_{PDAR} - AUC_{SAR}}{AUC_{SAR}}. \quad (1)$$

If we define the absolute robustness of PD and SR as the AUC under the PDAR and SAR, our metric quantifies the relative increase of absolute PD robustness compared to absolute SR robustness (Fig. 2). To study the relative increase of PD and SR with area we simply computed the local slope (or derivative) of PDAR and SAR on the standardized coordinates assuming a first point of null diversity and area. Indeed when area tends to zero, diversity also necessarily tends to zero (as sampled area becomes smaller than a single individual).

Understanding the absolute value of PDAR.—Second we used a null model approach to describe and investigate the absolute value of PDAR. This approach allows to compare null expectations with the observed PDAR and avoids the bias caused by the correlation between PD and SR.

We chose to use the observed SAR as a starting point because it has already been shown to be modelled by a simple null model where species ranges are randomly distributed within the continent (Storch et al. 2012). Assuming the SAR, we computed a null PDAR expectation by randomly shuffling the tips of the phylogeny within a given pool of species. This procedure breaks the link between species range size/position and phylogenetic relationships but keep unchanged the

distribution of range size and the local species richness (Hardy 2008). By repeating this procedure n times, we were able to assess the significance of the observed PDAR relative to our null expectation (using a two-sided test).

In other words, for each randomization, we (1) shuffled the tips of the phylogeny within a given species pool. The resulting randomized phylogeny was used to (2) compute null PD values for each basic grid cells (approximately 110×110 km) and we (3) applied the methodology described above to compute the resulting null median and MAD PDARs. Such null model may help unravel the determinants of the PDAR. For example we may expect competition (Pigot and Tobias 2013) or environmental filtering to occur at smaller scale, potentially leading to phylogenetic overdispersion or clustering respectively (Webb et al. 2002; but note that competition may also lead to clustering, see Mayfield and Levine [2010]). Also, phylogenetic clustering could be detected at larger scales because of biogeographical effect (Rosenzweig 1995). We used two null models that use either a continental or a biogeographic pool of species. We restricted the analysis of the biogeographic pool of species for Eurasia only because it is the only continent in our design that is a mix of distantly related zoogeographic regions (Wallace 1876, Holt et al. 2013).

Continental pool of species

This null model simply shuffles the tips within the entire continental phylogeny. The significance of the observed values of PDAR was assessed by comparing observed values with 1000 randomized PDARs. We further confirmed this simulation approach by using analytical expectations of PD based on the framework of Nipperess and Matsen (2013; Appendix D)

Biogeographical pool of species

First we defined zoogeographic regions following the methodology of Holt et al. (2013). To do so we computed phylogenetic beta diversity values between each pair of grid cells from the Eurasian continent by using an index independent of species richness (Lennon et al. 2001, Holt et al. 2013):

$$\beta = 1 - \frac{a}{\min(b, c) + a} \quad (2)$$

where a = the branch lengths shared by the two grid cells and b and c represent the branch lengths unique to each grid cell, with $\min(b, c)$ representing the minimum value between b and c .

Then we identified groups of grid cells (zoogeographic regions) using the unweighted pair group method with arithmetic mean (UPGMA, function `hclust` in R; R Development Core Team 2014). We varied the number of delimited zoogeographic regions from 1 to 30 (see Appendix E for examples). We then used these regions to construct a biogeographical null model of the PDAR.

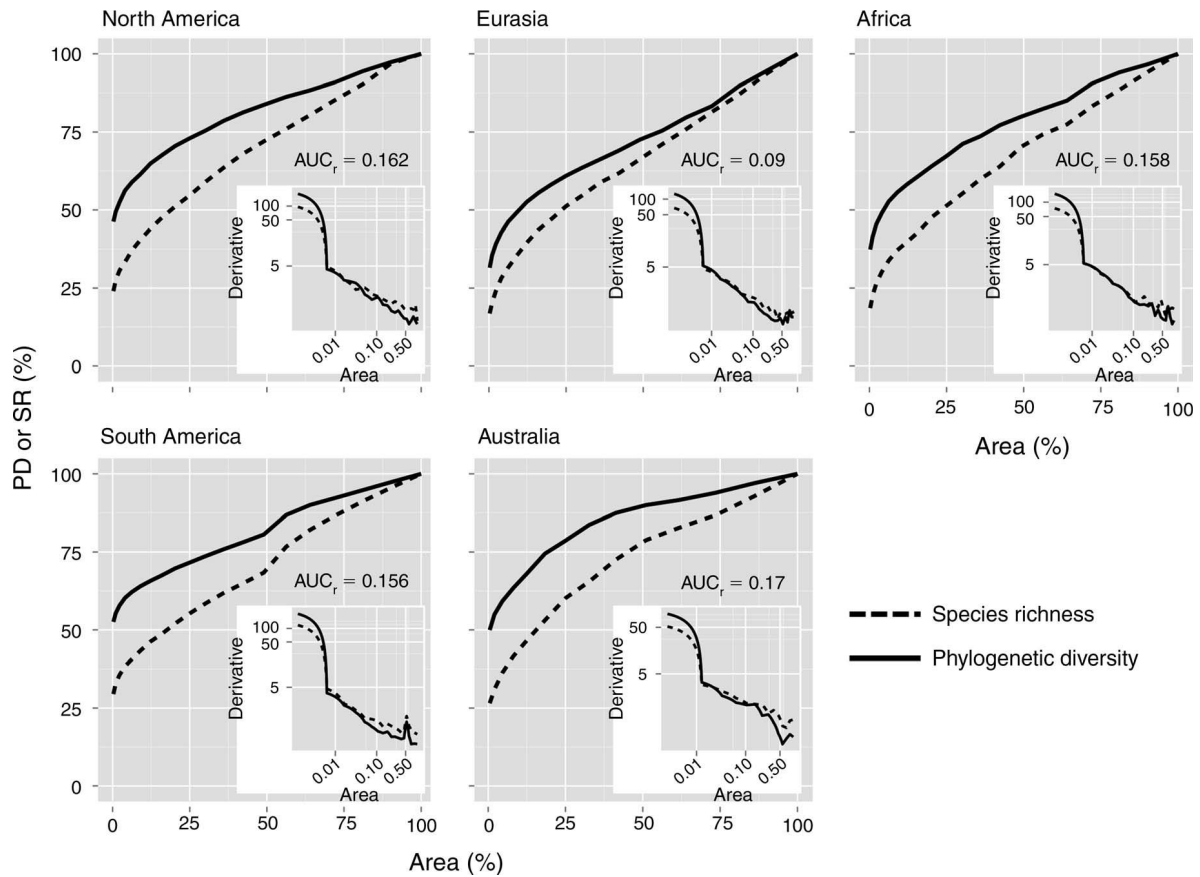


FIG. 3. Observed rescaled median SARs and median PDARs. For each continent, we report the SAR and the PDAR rescaled by the value of the maximum SR and PD, respectively. The two curves are both expressed in percentage of maximum diversity and thus are directly comparable. We also report the corresponding AUC_r values (see Fig. 2). In the lower-right corner subplots, we show the corresponding local derivatives.

While we were shuffling species within the entire continental pool of species in the previous null model, we shuffled here species within the pool of species belonging to a specific zoogeographic region. Because all species are not restricted to one unique zoogeographic region, we adopted a probabilistic approach where, for each randomization independently, a zoogeographic region k is assigned to a species i with a probability of $P_{i,k}$ depending on its coverage $C_{i,k}$ in this region with respect to its total coverage across all regions

$$P_{i,k} = \frac{C_{i,k}}{\sum_K C_{i,k}} \tag{3}$$

where K represents the entire set of regions (from 2 to 30). For each randomization, we computed a null PDAR and tested significance by comparing the observed PDAR and 100 null PDAR for each number of zoogeographic regions defined.

RESULTS AND DISCUSSION

To visually compare the PDAR and the SAR of mammals we standardized the two curves by the

maximal diversity reached in the data set. The two resulting curves are thus expressed as the percentage of maximal diversity and are directly comparable (Fig. 3). We show that PDARs approach their maximum faster than SARs for all continents (Fig. 3). To describe the rate of PD and SR accumulation as a function of area, we estimated local derivatives and show that PDARs accelerate much faster than SARs for small areas and that this tendency reverses for large areas (see subplots of Fig. 3). We show that the power model is among the best models to fit the data set (Appendix F) but fails to model the upward acceleration of PD on a log-log scale (e.g., Fig. 4). The slope of the power model is lower for the PDARs than for the SAR ($z_{PDAR} < z_{SAR}$, see Appendix G). The last point has been previously reported at this scale (Mazel et al. 2014) but with a different PDAR/SAR reconstruction based on non-overlapping ecoregions (Olson et al. 2001). This result is thus independent of the sampling procedure and approaches, and is, overall, not surprising. Indeed the SAR fully represents the PDAR in the case of a star phylogeny (in this case, the PD is proportional to SR). As the phylogenetic tree departs from a star phylogeny

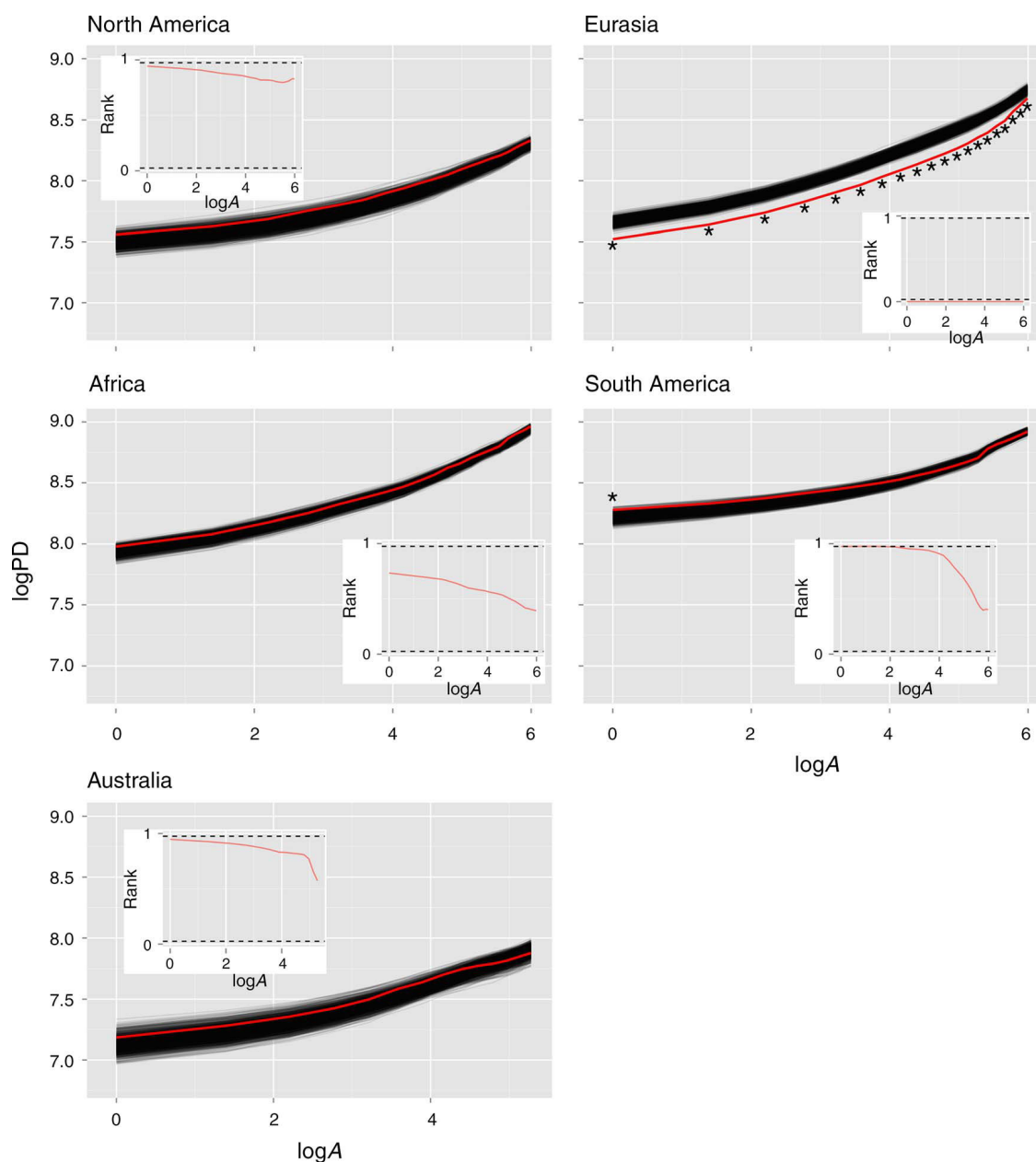


FIG. 4. Median PDARs obtained from the continental null model. For each continent, the envelope corresponding to 1000 null continental PDARs is shown in black while the observed PDAR is in red. In the corner of each panel, we plot the relative rank of observed PD value within the null PD distribution as a function of $\log(\text{area})$ ($\log A$). For each spatial scale, it is computed as the proportion of null PD values that are lower than the observed value (a value of 0.5 indicates that observed PD equals the median of the null distribution). The dashed lines correspond to a relative rank of 2.5% and 97.5%. When the computed relative ranks fall out of this 95% envelope, a star is reported in the main panel (see Appendix H for the relative ranks associated with power model parameters).

(i.e., as some branches start to be shared between species), the relationship between PD and SR becomes concave and the PDAR deviates from the SAR (because redundancies between species are introduced, see Fig. 1A).

As a consequence, AUC_r values are positive but we find that they differ across continents (see Figs. 2 and 3). The use of the SAR to predict species extinction from

habitat loss (Pimm and Raven 2000, Thomas et al. 2004, Halley et al. 2014) has been questioned (He and Hubbell 2011) but remains useful (e.g., Axelsen et al. 2013, Hanski and Zurita 2013, Matias et al. 2014), especially when species ranges are randomly distributed (He and Hubbell 2011). If we assume that the SAR and the PDAR can be used to predict the loss of species and PD, respectively, through habitat loss (Mazel et al. 2014), the

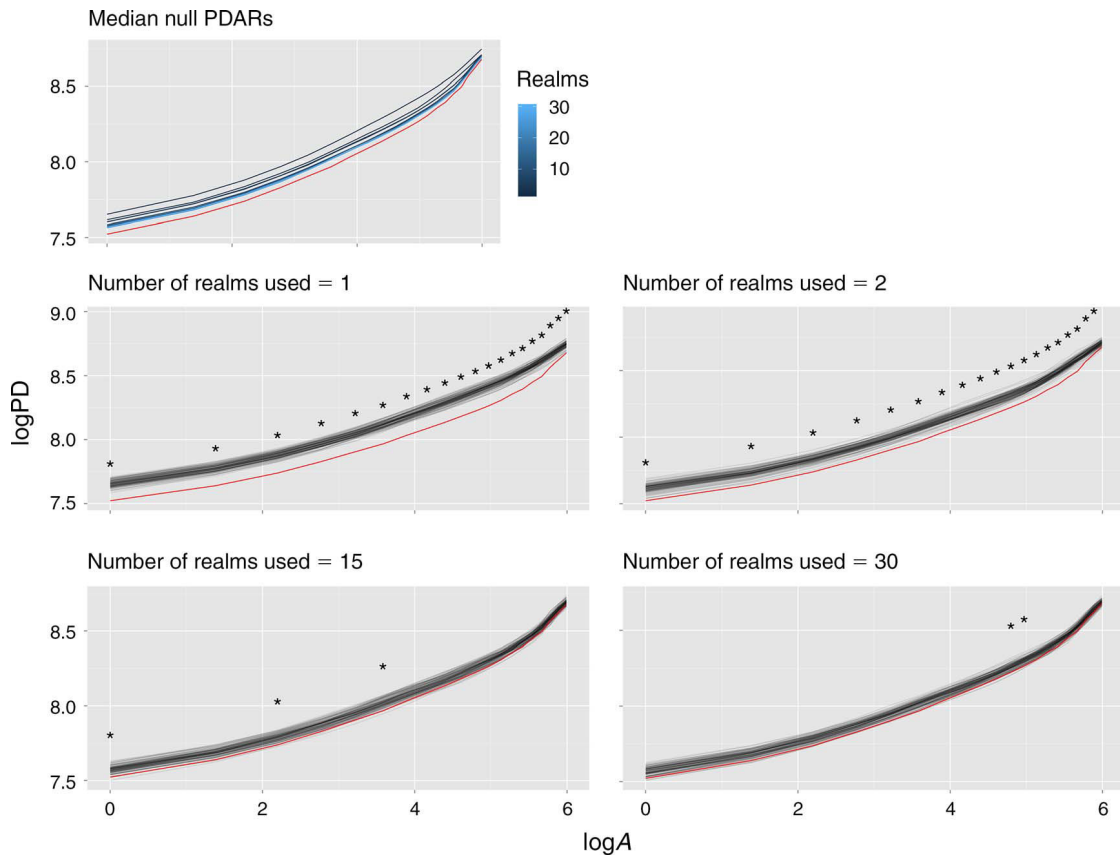


FIG. 5. Median PDARs obtained from the Eurasian biogeographical null models. The biogeographical null models shuffle the tips of the phylogeny according to biogeographical origin (see *Methods*). We present the results from null models containing different numbers of biogeographic regions. The top panel presents the median PDAR obtained for different numbers of biogeographic regions. The four other panels represent the details of four biogeographic null models that used 1 (continental null model), 2, 15, or 30 biogeographic regions, respectively. The star indicates that the relative rank of the observed PD value within the null PD distribution is lower (or higher) than 0.025 (or 0.975) for a given area.

AUC_r (Fig. 2) then represents the relative robustness of PD (compared to SR) to habitat destruction. In this case, we show, for example, that the Australian PD will be relatively more robust, at the continental scale, to habitat loss than the Eurasian PD (note that, at the global scale, the evolutionary history of Australia is unique; Holt et al. 2013). This difference could be explained by the different structure of the two trees (see Fig. 1A) and/or by different eco-evolutionary processes leading to a different spatial pattern of PD (e.g., either clustered, random, or overdispersed; see Fig. 1B). Teasing apart these two mechanisms called for using appropriate null models.

For all continents except Eurasia, and for most of the spatial scales, we find that PD values are not significantly different from those obtained with the null model randomizing the phylogenetic relationships among species (Fig. 4 and Appendix H). This means that the only significant phylogenetic effect that influences the PDAR is the shape of the observed continental tree. Several non-mutually exclusive hypotheses can be proposed to explain this result. First, we use here an

averaged phylogenetic structure across a whole continent and different assembly processes may have been mixed. In North America, for example, we could expect a phylogenetic clustering at high elevation in the Rocky Mountains (e.g., due to environmental filtering) while overdispersion could be found in the lowland forest (e.g., due to competition; Graham et al. 2009), resulting in higher MAD of PD than expected by chance (see Appendix I). Such distribution may likely compensate each other during the sampling process, resulting in a null random distribution when averaged across assemblages. Second, the spatial and phylogenetic scale of our analysis is perhaps too large to detect any effect of repulsion/attraction of species. Indeed, at the smallest resolution we have used (110×110 km), co-occurring species do not necessarily interact with each other (probably because this scale is still very large) and may, for example, use different habitats (Araújo and Rozenfeld 2014, but see Cardillo 2011). Also, specific group of mammals may show repulsion or attraction while others not, blurring the overall pattern (see, e.g., Pedersen et al. 2014).

The Eurasian PDAR is, however, much lower than expected by chance at all spatial scales, indicating phylogenetic clustering (Fig. 4 and Appendix H). This continent is a mosaic of biogeographic realms with diverging biogeographic history (Wallace 1876, Holt et al. 2013) and thus mixes very different faunas: there are many strict Palearctic species (e.g., the wolverine, *Gulo gulo*) and strict oriental species (e.g., the Asiatic elephant, *Elephas maximus*). The continental null model mixes all these faunas and thus tends to overestimate the expected median PD of assemblages. For example, *Elephas maximus* represents the only Afrotherian species present in our data set so it has a very high distinctiveness at the continental scale and will considerably increase relative local PD. We consequently develop a biogeographical null model that takes into account the historical origin of taxa. We show that this null model progressively decreases the random PD expectations (Fig. 5) and that 15 realms were sufficient to correctly predict most of the spatial scale median PD (Fig. 5 and Appendices J–K). Our approach may sound circular at first glance because we use spatial and phylogenetic data (to define zoogeographic regions) to explain spatial and phylogenetic data (the PDAR). Nevertheless, the aim of any null models is rather to ask how much synthetic information we need from the initial data to parsimoniously explain this data. The null model is necessarily constrained by the initial data but if this constrain is too high (i.e., a lot of the initial data is used) the null model will necessarily be plausible (the “narcissus effect”; Gotelli 2001). Here our aim is to quantify how much synthetic information is needed to parsimoniously explain the Eurasian PDAR. Finally, it is not directly possible to conclude from the biogeographic null model that the effect we detected is purely historic, i.e., that it derives from the mix of fauna due to continental drift. Indeed Eurasia has also a steep north-south gradient in term of climate (and SR) and we cannot reject the hypothesis of a phylogenetic clustering due to environmental filtering (Webb et al. 2002). One potential future avenue could be to contrast the relative predictive power of biogeographic null model vs. a climate based null model.

We also find that PD at small scale tends to be higher than expected by chance for South America (phylogenetic overdispersion, Fig. 4), leading to a significantly low z value for the PDAR ($P < 0.005$, Appendix H). This observation may result from allopatric speciation events and/or competition at the smallest scale (Pigot and Tobias 2013) but more work is needed to test these hypotheses. For Eurasia, on the contrary, we observe that small-scale PD tends to be relatively lower than expected by chance (phylogenetic clustering, Fig. 4) leading to a high z value ($P < 0.1$, Appendix H). This could be a possible product of environmental filtering occurring only at small scale.

These last two results shed light on the discrepancy between observed and null AUC_r (Appendix L). Indeed

the robustness of PD to habitat loss depends on the structure of the phylogenetic tree (represented here by the null model mean expectation) and also on the spatial pattern of phylogenetic structure (represented by the departure of the observed PDAR from the mean null model, see subplots in Fig. 4). We indeed find that Australian observed AUC_r is higher than predicted by the continental tree structure because small scale Australian PD tends to be relatively higher than expected by chance (see Appendix L), providing a buffer against the loss of PD. In contrast, observed Eurasia AUC_r is lower than expected by the tree shape (see Appendix L). This is because Eurasian PD is disproportionately low at small scale, reducing the area between the PDAR and the SAR compared to random expectations and thus being more vulnerable to habitat reduction. Overall, we demonstrate that the additional robustness of PD to habitat loss compared to species richness is determined by the phylogenetic tree shape but also depends on the spatial structure of PD.

ACKNOWLEDGMENTS

F. Mazel would like to thank Loïc Chalmandrier for helpful discussions on null models, Petr Keil for discussions on the PDARs, and two anonymous referees for helpful comments on a previous version of the manuscript. The research leading to these results had received funding from the European Research Council under the European Community's Seven Framework Programme FP7/2007-2013 Grant Agreement no. 281422 (TEEMBIO). F. Mazel, W. Thuiller, and J. Renaud belong to the Laboratoire d'Écologie Alpine, which is part of Labex OSUG@2020 (ANR10 LABX56).

LITERATURE CITED

- Araújo, M., and A. Rozenfeld. 2014. The geographic scaling of biotic interactions. *Ecography* 37:406–415.
- Axelsen, J., U. Roll, L. Stone, and A. Solow. 2013. Species–area relationships always overestimate extinction rates from habitat loss: comment. *Ecology* 94:761–763.
- Belmaker, J., and W. Jetz. 2011. Cross-scale variation in species richness–environment associations. *Global Ecology and Biogeography* 20:464–474.
- Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. *Nature* 446:507–512.
- Cadotte, M. W., B. J. Cardinale, and T. H. Oakley. 2008. Evolutionary history and the effect of biodiversity on plant productivity. *Proceedings of the National Academy of Sciences USA* 105:17012–17017.
- Cardillo, M. 2011. Phylogenetic structure of mammal assemblages at large geographical scales: linking phylogenetic community ecology with macroecology. *Philosophical Transactions of the Royal Society B* 366:2545–2553.
- Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61:1–10.
- Fritz, S. A., O. R. P. Bininda-Emonds, and A. Purvis. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology Letters* 12:538–549.
- Gotelli, N. 2001. Research frontiers in null model analysis. *Global Ecology and Biogeography* 10:337–343.
- Graham, C. H., J. L. Parra, C. Rahbek, and J. A. McGuire. 2009. Phylogenetic structure in tropical hummingbird communities. *Proceedings of the National Academy of Sciences USA* 106 (Supplement):19673–19678.

- Guilhaumon, F., O. Gimenez, K. J. Gaston, and D. Mouillot. 2008. Taxonomic and regional uncertainty in species-area relationships and the identification of richness hotspots. *Proceedings of the National Academy of Sciences USA* 105: 15458–15463.
- Halley, J., V. Sgardeli, and K. Triantis. 2014. Extinction debt and the species–area relationship: a neutral perspective. *Global Ecology and Biogeography* 23:113–123.
- Hanski, I., and G. Zurita. 2013. Species–fragmented area relationship. *Proceedings of the National Academy of Sciences USA* 110:12715–12720.
- Hardy, O. 2008. Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology* 96:914–926.
- He, F., and S. Hubbell. 2011. Species–area relationships always overestimate extinction rates from habitat loss. *Nature* 473: 368–371.
- Helmus, M. R., and A. R. Ives. 2012. Phylogenetic diversity–area curves. *Ecology* 91:31–43.
- HilleRisLambers, J., P. B. Adler, W. S. Harpole, J. M. Levine, and M. M. Mayfield. 2012. Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology, Evolution, and Systematics* 43:227–248.
- Holt, B. G., et al. 2013. An update of Wallace’s zoogeographic regions of the world. *Science* 339:74–78.
- Huang, S., P. R. Stephens, and J. L. Gittleman. 2012. Traits, trees and taxa: global dimensions of biodiversity in mammals. *Proceedings of the Royal Society B* 279:4997–5003.
- Hubbell, S. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, New Jersey, USA.
- Jenkins, C. N., S. L. Pimm, and L. N. Joppa. 2013. Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences USA* 110: E2602–2610.
- Kadmon, R., and O. Allouche. 2007. Integrating the effects of area, isolation, and habitat heterogeneity on species diversity: a unification of island biogeography and niche theory. *American Naturalist* 170:443–454.
- Leitner, W., and M. Rosenzweig. 1997. Nested species–area curves and stochastic sampling: a new theory. *Oikos* 503–512.
- Lennon, J. J., P. Koleff, J. J. D. Greenwood, and K. J. Gaston. 2001. The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology* 70:966–979.
- MacArthur, R., and E. Wilson. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey, USA.
- Matias, M., D. Gravel, F. Guilhaumon, P. Desjardins-Proulx, M. Loreau, T. Münkemüller, and N. Mouquet. 2014. Estimates of species extinctions from species–area relationships strongly depend on ecological context. *Ecography* 37: 001–012.
- Mayfield, M., and J. Levine. 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters* 13:1085–1093.
- Mazel, F., F. Guilhaumon, N. Mouquet, V. Devictor, D. Gravel, J. Renaud, R. L. Marcus Vinicius Cianciaruso, J. A. F. Diniz-Filho, D. Mouillot, and W. Thuiller. 2014. Multifaceted diversity–area relationships reveal global hotspots of mammalian species, trait and lineage diversity. *Global Ecology and Biogeography* 23:836–847.
- Morlon, H., D. W. Schilck, J. A. Bryant, P. A. Marquet, A. G. Rebelo, C. Tauss, B. J. M. Bohannan, and J. L. Green. 2011. Spatial patterns of phylogenetic diversity. *Ecology Letters* 14: 141–149.
- Mouquet, N., et al. 2012. Ecophylogenetics: advances and perspectives. *Biological Reviews* 87:769–785.
- Nipperess, D., and F. Matsen. 2013. The mean and variance of phylogenetic diversity under rarefaction. *Methods in Ecology and Evolution* 4:566–572.
- O’Dwyer, J., S. Kembel, and J. Green. 2012. Phylogenetic diversity theory sheds light on the structure of microbial communities. *PLoS Computational Biology* 8:e1002832.
- Olson, D. M., et al. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience* 51:933–938.
- Pavoine, S., and M. B. Bonsall. 2011. Measuring biodiversity to explain community assembly: a unified approach. *Biological Reviews* 86:792–812.
- Pedersen, R. Ø., B. Sandel, and J.-C. Svenning. 2014. Macroecological evidence for competitive regional-scale interactions between the two major clades of mammal carnivores (Feliformia and Caniformia). *PLoS ONE* 9:e100553.
- Pereira, H. M., et al. 2010. Scenarios for global biodiversity in the 21st century. *Science* 330:1496–1501.
- Pigot, A. L., and J. A. Tobias. 2013. Species interactions constrain geographic range expansion over evolutionary time. *Ecology Letters* 16:330–338.
- Pimm, S., and P. Raven. 2000. Biodiversity: extinction by numbers. *Nature* 403:843–845.
- R Development Core Team. 2014. *R: a language and environment for statistical computing*. R Project for Statistical Computing, Vienna, Austria. www.r-project.org
- Rodrigues, A. S. L., and K. J. Gaston. 2002. Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biological Conservation* 105:103–111.
- Rosenzweig, M. L. 1995. *Species diversity in space and time*. Cambridge University Press, Cambridge, UK.
- Scheiner, S. M. 2003. Six types of species–area curves. *Global Ecology and Biogeography* 12:441–447.
- Storch, D., P. Keil, and W. Jetz. 2012. Universal species–area and endemics–area relationships at continental scales. *Nature* 488:78–81.
- Thomas, C. D., et al. 2004. Extinction risk from climate change. *Nature* 427:145–148.
- Triantis, K. A., F. Guilhaumon, and R. J. Whittaker. 2012. The island species–area relationship: biology and statistics. *Journal of Biogeography* 39:215–231.
- Tucker, C., and M. Cadotte. 2013. Unifying measures of biodiversity: understanding when richness and phylogenetic diversity should be congruent. *Diversity and Distributions* 19:845–854.
- Wallace, A. 1876. *The geographical distribution of animals*. Cambridge University Press, Cambridge, UK.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology, Evolution, and Systematics* 33: 475–505.
- Whittaker, R. J., M. B. Araújo, J. Paul, R. J. Ladle, J. E. M. Watson, K. J. Willis, and P. Jepson. 2005. Conservation biogeography: assessment and prospect. *Diversity and Distributions* 11:3–23.

SUPPLEMENTAL MATERIALS

Ecological Archives

Appendices A–L are available online: <http://dx.doi.org/10.1890/14-1858.1.sm>

CHAPITRE IV

LES POINTS CHAUDS DE DIVERSITÉ PHYLOGÉNÉTIQUE ET FONCTIONNELLE CHEZ LES MAMMIFÈRES

Ce chapitre pose la question suivante :

Les points chauds de biodiversité fonctionnelle et phylogénétique sont-ils identiques aux points chauds de diversité spécifique ?

Résumé de l'article

Le but de l'étude est de définir des points chauds de biodiversité fonctionnelle et phylogénétique à l'échelle des biomes terrestres et de les comparer aux points chauds « classiques » basés sur la richesse spécifique uniquement. Nous utilisons les mammifères sur l'ensemble globe terrestre comme cas d'étude. Nous avons calculé la richesse spécifique, la diversité fonctionnelle et phylogénétique pour 782 écorégions terrestres en utilisant les aires de distributions de 4616 mammifères. Nous avons utilisé un ensemble d'indices de diversité qui incorpore les couverts relatifs des espèces dans chaque écorégion. Nous avons ensuite construit des relations aire-diversité à plusieurs facettes afin d'ordonner les écorégions selon leur niveau de diversité. Nous avons finalement défini les points chauds comme les écorégions avec le plus de diversité. Nous montrons que les points chauds de diversité phylogénétique et fonctionnelle ne sont en général ni congruents entre eux ni avec les points chauds classiques. Ceci démontre que les plans de conservation qui se focalisent sur la richesse spécifique uniquement peuvent passer à côté des autres facettes de la diversité. En conséquence, il est d'une importance vitale d'adopter une perspective à plusieurs facettes de diversité pour solidifier l'approche de conservation intégrative.

RESEARCH
PAPER

Multifaceted diversity–area relationships reveal global hotspots of mammalian species, trait and lineage diversity

Florent Mazel^{1*}, François Guilhaumon², Nicolas Mouquet³, Vincent Devictor³, Dominique Gravel⁴, Julien Renaud¹, Marcus Vinicius Cianciaruso⁵, Rafael Loyola⁵, José Alexandre Felizola Diniz-Filho⁵, David Mouillot^{2,6} and Wilfried Thuiller¹

¹Laboratoire d'Ecologie Alpine, Grenoble, France, ²Laboratoire ECOSYM Université Montpellier 2, Montpellier, France, ³Institut des Sciences de l'Evolution, UMR 5554, CNRS, Université Montpellier 2, Montpellier, France, ⁴Département de Biologie, Chimie et Géographie, Université du Québec à Rimouski, Québec, Canada, ⁵Departamento de Ecologia, ICB, Universidade federal de Goiás, Goiânia, Brasil, ⁶ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Qld 4811, Australia

ABSTRACT

Aim To define biome-scale hotspots of phylogenetic and functional mammalian biodiversity (PD and FD, respectively) and compare them with 'classical' hotspots based on species richness (SR) alone.

Location Global.

Methods SR, PD and FD were computed for 782 terrestrial ecoregions using the distribution ranges of 4616 mammalian species. We used a set of comprehensive diversity indices unified by a recent framework incorporating the relative species coverage in each ecoregion. We built large-scale multifaceted diversity–area relationships to rank ecoregions according to their levels of biodiversity while accounting for the effect of area on each facet of diversity. Finally we defined hotspots as the top-ranked ecoregions.

Results While ignoring relative species coverage led to a fairly good congruence between biome-scale top ranked SR, PD and FD hotspots, ecoregions harbouring a rich and abundantly represented evolutionary history and FD did not match with the top-ranked ecoregions defined by SR. More importantly PD and FD hotspots showed important spatial mismatches. We also found that FD and PD generally reached their maximum values faster than SR as a function of area.

Main conclusions The fact that PD/FD reach their maximum value faster than SR could suggest that the two former facets might be less vulnerable to habitat loss than the latter. While this point is expected, it is the first time that it has been quantified at a global scale and should have important consequences for conservation. Incorporating relative species coverage into the delineation of multifaceted hotspots of diversity led to weak congruence between SR, PD and FD hotspots. This means that maximizing species number may fail to preserve those nodes (in the phylogenetic or functional tree) that are relatively abundant in the ecoregion. As a consequence it may be of prime importance to adopt a multifaceted biodiversity perspective to inform conservation strategies at a global scale.

Keywords

Conservation biogeography, diversity indices, functional diversity–area relationship, Hill's numbers, mammals, phylogenetic diversity–area relationship, species–area relationship.

*Correspondence: Florent Mazel, CNRS, LECA, 2233 Rue de la Piscine, Grenoble, Isère, 38041, France.
E-mail: flo.mazel@gmail.com

INTRODUCTION

Understanding the ecological and evolutionary processes driving the distribution of life on Earth is essential from both applied and theoretical perspectives. The quantification of biodiversity, central to conservation science, has recently moved from a focus on pure species counting (e.g. species richness, SR) to a more integrative approach. Assessments of biodiversity now consider the overall evolutionary history embedded within a set of taxa (i.e. phylogenetic diversity, PD) along with the diversity of ecological traits (i.e. functional diversity, FD). In conservation science, this novel approach has redefined the identification of species of conservation interest by taking their high evolutionary or functional distinctiveness into consideration (Isaac *et al.*, 2007; Mouillot *et al.*, 2013) and has also made it possible to detect unique macroecological assemblages (Forest *et al.*, 2007), for example 'cradles' and 'museums' of life (Chown & Gaston, 2000). Furthermore, the loss of FD or PD per unit of habitat loss is likely to be a better predictor of ecosystem vulnerability than the loss of single species. Indeed, the loss of a given amount of FD or PD, often assumed to be related to particular combinations of functional traits or of a certain lineage, respectively, may threaten the functioning of the ecosystem, whereas the loss of a given single species might not be noticeable if redundant species persist (Loreau *et al.*, 2002; Srivastava *et al.*, 2012).

This new perspective also provides fundamental insights into community assembly at multiple spatial scales (Mouquet *et al.*, 2012). A multifaceted approach may help unravel the different drivers of community structure (e.g. competition or environmental filtering; Webb *et al.*, 2002) or ecosystem functioning (Cadotte *et al.*, 2009; Gravel *et al.*, 2012). In macroecology, contrasting SR, PD and FD offers a potential means for disentangling the processes shaping large-scale diversity distribution (Davies & Buckley, 2011; Safi *et al.*, 2011; Huang *et al.*, 2012). For example, the global latitudinal diversity gradient has recently been re-interpreted from a novel evolutionary perspective, merging Earth's climatic history, phylogenetic diversity and species richness in a unified and testable framework (Hawkins *et al.*, 2012). A multifaceted perspective thus represents a promising avenue for exploring the distribution of diversity because it is at the crossroads between ecology, evolution and conservation biology but also palaeontology and palaeoclimatology (Hawkins *et al.*, 2006).

One of the most striking features of biodiversity is the spatial heterogeneity of its distribution, with some regions harbouring extraordinary levels of biodiversity: the so-called biodiversity hotspots (Reid, 1998; Ceballos & Ehrlich, 2006; Guilhaumon *et al.*, 2008). These have not only fascinated macroecologists, who try to understand their origins (e.g. the historical perspective; Wiens *et al.*, 2011), but also conservationists seeking the best opportunities to allocate the limited resources available for global-scale conservation. For example the biodiversity hotspots concept has been proposed to prevent the extinction of large numbers of endangered species, by protecting places 'where exceptional concentrations of endemic species are undergoing exceptional loss of habitat' (Myers *et al.*, 2000).

The most recent comparisons of the world-wide distribution of hotspots have been limited to different taxonomic groups and components of SR for a given taxon (e.g. endemic, total, endangered; Orme *et al.*, 2005; Ceballos & Ehrlich, 2006; Lamoreux *et al.*, 2006) or when carried out in a multifaceted context, have included only limited functional information [e.g. Huang *et al.* (2012) used only geographic range size and body mass as descriptors of mammal FD to define hotspots]. This lack of relevant trait information makes it difficult to adequately represent the spatial distribution of FD because geographic range size may not properly portray species niches, rather it is mostly influenced by historical biogeography and macroevolution (Gaston, 2003).

Here we identified global hotspots of mammalian taxonomic diversity (TD), PD and FD. We based our analyses on the updated version (Fritz *et al.*, 2009) of the dated phylogeny of Bininda-Emonds *et al.* (2007) and a set of functional traits encompassing important aspects of mammal resource use, selected to represent independent and informative niche dimensions (Safi *et al.*, 2011). We used the world's ecoregions (Olson *et al.*, 2001) to define geographical units harbouring unique species assemblages and ecosystems. Ecoregions have proven valuable for addressing a range of questions in macroecology and more applied conservation issues (Lamoreux *et al.*, 2006; Guilhaumon *et al.*, 2008).

To account for expected area effects on TD (Triantis *et al.*, 2012), PD (Morlon *et al.*, 2011) and FD (Cumming & Child, 2009) we constructed diversity–area relationships (DARs hereafter) for 13 terrestrial biomes (global-scale regions gathering ecoregions experiencing similar environmental conditions such as tundra or mediterranean forests). We used a model-averaging approach that fits 19 mathematical functions to the data (Guilhaumon *et al.*, 2008; Triantis *et al.*, 2012) and then computed an Akaike information criterion (AIC)-weighted average of the 19 predicted curves. To quantify the different types of diversity, we used a set of unified TD, PD and FD indices that weigh species coverage differently (Chao *et al.*, 2010) and correspond to modified versions of Faith PD (Faith, 1992), Phylogenetic entropy (Allen *et al.*, 2009) and Rao quadratic entropy (Rao, 1992) (see Methods). For each diversity index we identified as hotspots those ecoregions with the largest positive deviations from, respectively, SARs (species–area relationships), PDARs (phylogenetic diversity–area relationships) and FDARs (functional diversity–area relationships) and investigated their spatial congruences.

Our global exploration of mammals SARs, PDARs and FDARs reveals important mismatches between the spatial scaling and the geographical extremes of SR, PD and FD, calling for integrative approaches.

METHODS

Dataset

Mammal assemblages

We used the distribution maps provided by the Mammal Red List Assessment (<http://www.iucnredlist.org/>) for 4616

terrestrial species (for which we have functional traits; see below) to obtain occurrence data for each of the 827 ecoregions defined by Olson *et al.* (2001). We retained 782 ecoregions (mean number of ecoregions per biome = 60.1, SD = 53.3, min. = 17, max. = 223). Ecoregions are a valuable tool for studying multifaceted hotspots because they also serve as the basis of World Wildlife Fund conservation planning (Olson & Dinerstein, 1998), the international efforts of Nature Conservancy (Groves, 2003) and the delineation of Conservation International's Biodiversity Hotspots (Mittermeier *et al.*, 2004) and High Biodiversity Wilderness Areas (Mittermeier *et al.*, 2003). Furthermore, ecoregions have commonly been used to define taxonomic hotspots (Lamoreux *et al.*, 2006; Guilhaumon *et al.*, 2008) because they encompass relatively homogeneous biological systems. We retained ecoregions harbouring more than one mammal species and excluded mangrove ecoregions and large uninhabited parts of Greenland and Antarctica because of low data reliability or availability for these areas (Lamoreux *et al.*, 2006). Domestic mammals were also excluded from the analysis.

For each ecoregion and species, species coverage (C_i) was calculated as the intersected surface (in km²) between the range of the species and the ecoregion. We then computed, for each species i , the following relative coverage (RC_i , equation 1)

$$RC_i = \frac{C_i}{\sum_i C_i}. \quad (1)$$

Basically a species will have low relative coverage in a given ecoregion if its distribution range is small. The relative coverage was used to calculate diversity indices incorporating relative abundance (see below). By doing this we were able to differentiate a species that is poorly represented in an ecoregion, but with a unique evolutionary history (e.g. a monotreme species) or with unique functional traits (e.g. a top predator), from species with a similar evolutionary history (or functional traits) but with greater occupancy in the ecoregion. This weighting scheme emphasizes species that are well distributed in the ecoregion. Establishing how our measure of relative coverage is important for conservation and ecosystem functioning is not straightforward. Nevertheless we believe that the evolutionary history/functional characteristics of a species that shows a very small distribution range in a given ecoregion should not have the same theoretical influence on PD/FD as a widespread species in this ecoregion. Although it is unlikely that our measure of relative coverage represents a direct measure of local species abundance, it has been shown that a positive relationship between range size and local abundance is common (Gaston *et al.*, 2000). Nevertheless departure from this relationship probably exists. First, we did not use the complete range size of the species but only its extent in the ecoregion. Second, we acknowledge that the potential important residual variation that exists around the relationship may depend on species life-history traits. For example species with high dispersal abilities (or a species at a high level in the trophic hierarchy) may have a large range size but be relatively rare at the local scale. It is also possible that a species with a narrow range may exhibit a high

local abundance, for example because it uses an abundant resource that is restricted to a small area of the ecoregion. Nevertheless, we believe that our measure of species coverage was a needed first step to incorporate abundances into the definition of PD/FD hotspots.

Phylogeny and functional traits

We used the calibrated and dated ultrametric phylogenetic tree updated by Fritz *et al.* (2009) from Bininda-Emonds *et al.* (2007). We computed functional diversity indices using body mass (log-transformed), diet (vertebrates, invertebrates, foliage, stems and bark, grass, fruits, seeds, flowers, nectar and pollen, roots and tubers), habits (aquatic, fossorial, ground-dwelling, above-ground-dwelling), activity period (diurnal, nocturnal, cathemeral, crepuscular) and litter size (data from Safi *et al.*, 2011). These traits encompass important aspects of mammal resource use, including the temporal and spatial windows used to get their food. They represent independent and informative niche dimensions for evaluating variability in mammal traits related to important ecosystem processes, such as decomposition and seed dispersal, as well as trophic control (Sekercioglu, 2010).

Diversity indices

A myriad of methods have been proposed in the last years to include species traits in diversity indices (Pavoine & Bonsall, 2011). Here we follow the comprehensive framework from Chao *et al.* (2010), which unifies a set of TD, PD and FD indices based on Hill numbers. There were three reasons for this. First it unifies most of the TD, PD and FD indices used in the literature (see below and Table 1). Second it represents equivalent numbers of species to satisfy the replication principle that ensures intuitive results for ecologists and conservation biologists (Jost, 2006; Chao *et al.*, 2010). For example, if the PD of an assemblage equals d (d being a real positive number), it has the same diversity as a community consisting of d equally abundant and maximally distinct species (i.e. with the maximum distance observed in the phylogeny). Third, we present here one of the only comprehensive and intuitive frameworks that incorporates relative species coverage (or abundance) into biodiversity indices.

Phylogenetic diversity

Consider a phylogenetic tree composed of a set B_i of i branches. PD can be defined as the 'mean diversity of order q over T years' (Chao *et al.*, 2010):

$${}^q\bar{D}(T) = \left\{ \sum_{i \in B_T} \frac{L_i}{T} a_i^q \right\}^{1/(1-q)} \quad (2)$$

where L_i is the length of branch i in the set B_i , a_i is the total abundance descended from branch i (i.e. the summed

Table 1 The set of diversity indices used in the analysis.

		Type of indices		
Original version				
Without species differences		SR	Shannon	Simpson
With species differences	Phylo.	PD (Faith, 1992)	Hp	QE*
	Funcio.	FD (Petchey & Gaston, 2006)	Not named	QE*
Hill numbers version				
Without species differences		SR	exp (Shannon)	1 / Simpson
With species differences	Phylo.	Faith _{cor} PD	Allen _{cor} PD	Rao _{cor} PD
	Funcio.	Faith _{cor} FD	Allen _{cor} FD	Rao _{cor} FD
Link between original and Hill numbers version		Faith _{cor} PD = PD / T	Allen _{cor} PD & FD = exp (Hp/T)	Rao _{cor} PD & FD = 1/ (1- QE)
<i>q</i> value		0	1	2
Weighting by species' coverage		No	Yes	Yes

For this study we used the Hill numbers version with species differences. These transformed versions obey the replication principle and can be grouped in a unified formula using the *q* parameter (see equation 2 and Chao *et al.*, 2010). The table gives the abbreviations used in the text. It also provides the link between original and transformed indices and indicates if coverage is used in the calculation of the indices.

*QE can be calculated with any distances (phylogenetic or functional) between species.

T represents the height of the phylogenetic tree or the functional dendrogram. PD, Phylogenetic diversity; FD, Functional diversity; Hp, Phylogenetic entropy (Allen *et al.* 2009); QE, Rao quadratic entropy (Rao, 1982); SR, Species richness; Phylo, Phylogenetic; Funcio, Functional.

abundance or relative coverage of species descending from this branch) and *T* is the height of the tree. The parameter *q* affects the influence of node (or branch segment) abundance on the diversity index: a high *q*-value gives more weight to nodes with high relative abundances. This general formula encompasses a set of well-known diversity indices. With *q* = 0, Faith_{cor} PD = PD_{Faith}/T, PD_{Faith} being the phylogenetic diversity defined by Faith (1992) and Faith_{cor} PD being the corrected version of the Faith PD. With *q* = 1, Allen_{cor} PD = exp(*H_p*/T), *H_p* being the phylogenetic entropy as defined by Allen *et al.* (2009) and Allen_{cor} PD being the corrected version of *H_p*. With *q* = 2, Rao_{cor} PD = 1/(1 - QE), QE being the quadratic entropy defined by Rao (1982) and Rao_{cor} PD being the corrected version of QE (see Table 1 for details). To summarize, *q* influences the relative weight of widespread versus rare species in the computation of the diversity index. It gives progressively more weight to widespread species and progressively ignores rare species. This point could be problematic if a species is rare and endemic in this ecoregion because we will progressively ignore this unique species. Nevertheless our study aims to characterize the evolutionary history and the functional characteristics that are widespread in a given ecoregion (and somehow representative), which justifies the use of Chao *et al.*'s (2010) framework.

Functional diversity

We adapted previous indices for FD. First, we calculated functional distance among pairs of species using the Gower distance, which can mix categorical and continuous traits with equal weight and can cope with missing values (some traits were missing for 80 species, representing less than 2% of our dataset). We then applied a hierarchical cluster algorithm to convert the functional distance matrix into a functional dendrogram ensur-

ing the ultrametric property (note that using non-ultrametric functional distances did not change our conclusions) using the unweighted pair group method with arithmetic mean (UPGMA) (function *hclust* in R; R Development Core Team, 2010). The corresponding FD indices were named Faith_{cor} FD, Allen_{cor} FD and Rao_{cor} FD (see Table 1). Note that Faith_{cor} FD is equivalent to the Petchey & Gaston (2006) definition of FD (i.e. 'the total branch length of a functional dendrogram'). Like Faith_{cor} PD, Faith_{cor} FD is intrinsically correlated to SR (Huang *et al.*, 2012). This is the case for all the dendrogram-based approaches for estimating functional volume. Nevertheless it is interesting to use it here because it is directly comparable with Faith_{cor} PD and represents a diversity volume (or 'richness' *sensu* Pavoine & Bonsall, 2011). In addition we computed FD using body mass only to test to what extent the use of multiple traits influences our results. Note also that the expected correlation between SR and FD/PD based on dendrogram becomes weaker when moving from *q* = 0 to *q* = 2.

Species diversity

For maximally distinct species (i.e. a star phylogeny or star functional dendrogram), these indices actually constitute species diversity indices, namely SR when *q* = 0, the exponential of Shannon entropy when *q* = 1 and the inverse of Simpson diversity when *q* = 2 (see Table 1; Chao *et al.*, 2010). We used these TD indices to compare appropriately PDAR/FDAR with the corresponding SAR (i.e. comparing DARs that are built with diversity indices based on the same *q*). However, we always compared PD/FD hotspots with those based on SR (and not Simpson or Shannon indices) since a list of hotspots based only on TD indices (i.e. only quantifying evenness in abundances) might not be appropriate in a conservation context.

Constructing DARs

To account for expected area effects on SR, PD and FD, we provide a construction of DARs for 13 terrestrial biomes (Olson *et al.*, 2001). Such DARs correspond to a non-overlapping design, i.e. they are built from single data points, which corresponds to a type IV curve in Scheiner's (2003) terminology.

Set of models

A wide range of statistical models have been proposed to describe SARs (Tjørve, 2009). Here, 19 models were selected to fit SAR, PDAR and FDAR following Triantis *et al.* (2012) (see Appendix S1 in Supporting information). Recent attempts to model PDAR only used the power model (Morlon *et al.*, 2011) but, given the uncertainty regarding the shape of PDAR and FDAR, we tested a large spectrum of models. These models were chosen because they vary in form (e.g. sigmoid or convex, including asymptotic relationships) and complexity (two to four parameters).

Model fitting

We constructed 117 datasets (9 indices \times 13 biomes) and fitted 19 models to each dataset, for a total of 2223 DARs. We carried out our analyses using another dataset that also adds an 'artificial' point of null diversity and null area (0.001 and 0.001 to avoid computing problems).

Models were fitted using nonlinear regression with minimization of the residual sum of squares. Models were further evaluated by examining the normality and homoscedasticity of residuals. To do so, we applied the Lilliefors's test for normality and a Pearson correlation between squared residuals and area for homoscedasticity. Previous studies (e.g. Guilhaumon *et al.*, 2008) considered a model valid when the *P*-value associated with the normality and homoscedasticity tests exceeded the arbitrary threshold of 5%. All DAR analyses were carried out using an updated version of the 'mmsAR' package (Triantis *et al.*, 2012) for the R statistical and programming environment (R Development Core Team, 2010).

Model averaging

For each dataset, we discriminated between different models using an information-theory framework designed to evaluate multiple working hypotheses (Burnham & Anderson, 2002). The AIC can be used to evaluate the goodness of fit of different non-nested models on a given dataset. The weights of evidence were then derived from the AIC values to evaluate the relative likelihood of each model given the data and the set of models (Burnham & Anderson, 2002). Using these weights we derived averaged DARs for each biome and each diversity index.

Model standardization and comparison

For each DAR, we divided each predicted diversity value by that of the largest ecoregion in the considered biome in order to

report the percentage of maximal diversity reached in this largest ecoregion. The resulting standardized DAR therefore ranges between 0 and 1, which makes it possible to compare the scaling of different diversity facets with area (see below).

Hotspot lists and spatial congruence

Hotspot selection

Averaged residuals were calculated from the standardized averaged model (as defined above). A positive residual for a given ecoregion means that observed diversity is higher than expected given its area. Hotspots were defined as those ecoregions with the highest residuals. We ranked ecoregions according to their averaged residuals: the higher the residuals, the higher the concentration of biodiversity in the ecoregion. Note that ranking in terms of original or standardized curve/observed diversity gives exactly the same results because standardization is linear (see Appendix S2). We also derived an averaged rank across SR, PD and FD hotspots to provide an integrative definition of a hotspot by summing up the ranking for each ecoregion across the biodiversity facets (i.e. SR, PD and FD).

Impact of DAR shape on hotspot lists

We investigated whether PDARs and FDARs were different enough from corresponding SAR to deeply modify the hotspot rankings. In other words we wanted to test whether PDARs/FDARs are needed to define hotspots or if SAR is a good proxy for FDAR/PDAR when defining hotspots. SAR, PDAR and FDAR were directly comparable thanks to the standardization procedure explained above (they are all expressed as a proportion of the maximal diversity predicted for the largest ecoregion and thus vary between 0 and 100%). We computed the difference between the standardized PD/FD in each ecoregion and the proportion of diversity predicted by the area using the SAR (and not PDAR/FDAR as previously done) and ranked these differences to compute lists of hotspots. Then, we compared the congruence between PD/FD hotspot lists derived from SAR and the 'natural' PD/FD hotspot lists derived from the PDAR/FDAR (as explained in the previous section). If SARs correctly model the scaling of PD/FD with area, the lists of hotspots should be very similar. In this case, SARs would be well suited to direct modelling of the spatial scaling of PD/FD to define hotspots and it would not be necessary to construct explicit PDAR/FDAR.

RESULTS

We start first by reporting the general results of the statistical procedures related to the DAR estimations and then by describing the outcomes of this procedure for the hotspot lists.

DAR modelling

Convergence, homoscedasticity and normality

One of the 19 models showed unrealistic fits (Epm2, see Appendix S1) and was not considered in the analysis. Of the remaining

2106 model fits (13 biomes × 18 models × 9 indices), 1895 (90%) converged. Amongst the different indices, models fitting Rao-based DAR showed the highest non-normality of residuals (homoscedasticity was not the limiting factor; results not shown). Indeed at the 1% level test of homoscedasticity and normality of the residuals, 53, 68 and 75% of the models were valid for the Rao-, Allen- and Faith-based indices, respectively.

Relative model fit

The variation in diversity indices explained by area was generally high (the median R^2 of the best function in each dataset was 0.5; see Appendix S3) but was quite variable. The R^2 of the best model for each dataset ranged from $R^2 = 0.0001$ (asymptotic model fitting SR in Montane grasslands and shrublands) to $R^2 = 0.95$ (the P2 function fitting Rao_{cor} PD in temperate coniferous forest; see Appendix S3). No single best model outperformed across all data sets, with model selection varying

markedly across biomes and diversity indices and revealing substantial levels of uncertainty with different models showing equivalent levels of support (see Appendix S4).

Model shape

To illustrate the difference between the rate of increase in SAR and FDAR/PDAR, we plotted the difference between predicted PDAR/FDAR and the corresponding predicted SAR for four biomes that cover the latitudinal gradient (Fig. 1, Appendix S5). The starting value of the curve was zero in most cases, while differences between PDAR/FDAR tended to zero as area increased. This means that PDAR/FDAR and SAR have the same proportion of diversity when area tends to zero (generally it was 0% of maximum diversity) and also end at the same point because of the standardization (their respective maximum 100%). In the intermediate area between the two extremes, PDAR and FDAR were in general higher than SAR (i.e. a positive difference), indicating that PDAR and FDAR reached their

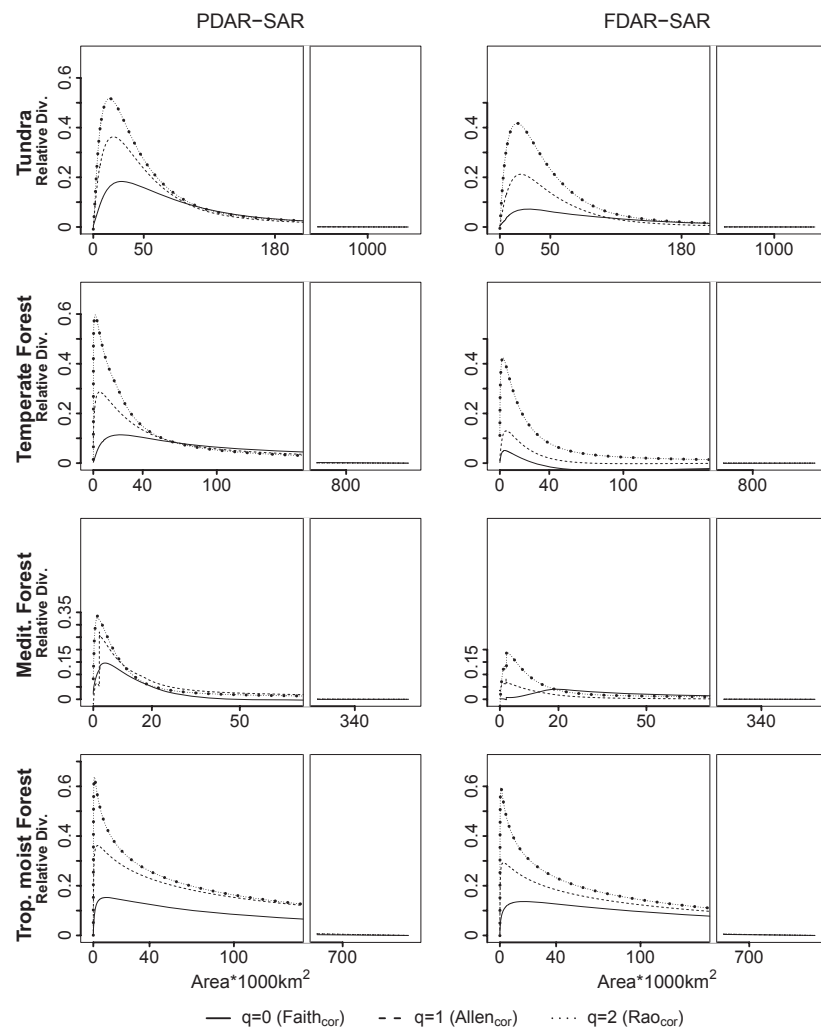


Figure 1 Differences between predicted phylogenetic diversity–area relationship (PDAR)/functional diversity–area relationship (FDAR) values and corresponding predicted species–area relationship (SAR) values. Rows correspond to different biomes, while columns represent differences between diversity–area relationship (DAR): PDAR–SAR and FDAR–SAR. For each plot, the differences between PDAR/FDAR and SAR are represented for three values of q : 0 (Faith_{cor} index), 1 (Allen_{cor} index) and 2 (Rao_{cor} index). Positive differences mean that PDAR or FDAR are higher than SAR. Area is given in km². Trop. moist Forest, tropical moist forest; Medit. F., mediterranean forest; Relative Div., relative diversity.

maximum diversity faster than SAR. This difference increased with the q parameter defining the weight given to species coverage in the diversity indices (Faith_{cor} to Rao_{cor}). PDAR and FDAR had a similar shape in most cases. Results were qualitatively equivalent when fitting DARs without artificial zeros except when area tended to zero: PDAR/FDAR started at a relatively higher percentage of diversity than SARs and thus the difference between PDAR/FDAR and SAR curves tended to start with positive values for some biomes (see Appendix S5).

Functional and phylogenetic hotspots

We extracted residuals (i.e. observed minus predicted diversity) from each averaged DAR and ranked them to define hotspots of diversity. As an example, we mapped diversity ranks for tropical moist forests (Fig. 2; but also see Appendix S6 for all biomes and indices) considering Allen_{cor} PD and FD hotspots as well as the traditional SR hotspots. SR rankings were relatively well distributed in the three continents whereas PD Allen_{cor} hotspots were much more concentrated in the Afrotropics (and Central America) or in the Afrotropical and Indomalaysian realms for FD Allen_{cor} hotspots. Interestingly, when focusing on the five hottest hotspots for this biome (Table 2), two important results emerged: (1) the list of SR hotspots shared few ecoregions with the lists of Allen_{cor} PD, FD and integrative hottest hotspots (i.e. two, one and three ecoregions, respectively); and (2) the hottest FD and PD hotspots shared only two ecoregions.

The same hotspot mismatches were found across all biomes (Fig. 3). For example, with the cut-off point for defining a hotspot set at the 5% richest ecoregions we found that congruence ranged from 5% (FD Rao_{cor} hotspots versus SR hotspots) to 74% (PD Faith_{cor} hotspots versus SR hotspots). Interestingly, when compared with hotspots defined with SR, hotspots defined with the Faith_{cor} index strongly matched, while those defined with Rao_{cor} strongly mismatched; those defined with Allen_{cor} fell in between. In other words, the hotspot rankings were significantly correlated – but were not equal – across different indices (see Appendix S7). These differences were robust against the threshold used to define valid DAR models (i.e. the P -value threshold used to reject a model based on the non-normality and/or homoscedasticity of its residuals; see Appendix S8) due to the weak influence of this threshold on the definition of hotspots (see Appendix S9). We also explored to what extent the use of multiple traits influences the definition of hotspots. We show that FD hotspots lists based on body mass only differ from FD hotspots defined with our complete set of traits (Appendix S9).

On average, using SAR instead of PDAR/FDAR to define PD/FD hotspots marginally modified the hotspot list (Appendix S10). However, it turns out that there is still high variability between biomes. For some of these, using SAR instead of PDAR/FDAR dramatically changes the hotspot lists. Note also that the nonlinear fit of the power model alone gave fairly similar results to those obtained when using model averaging to define hotspots (Appendix S11).

DISCUSSION

We found considerable geographical mismatches between global mammal hotspots of SR, PD and FD and, quite importantly, found that the magnitude of the mismatches depends on the index considered, which highlights the importance of considering a variety of indices (Huang *et al.*, 2012). Mismatches were higher with Rao-based indices (Rao_{cor}), lower when using the Faith_{cor} indices and in between for the Allen_{cor} indices, whatever the facet considered. This is not entirely surprising given the correlation between SR and PD/FD (high with Faith_{cor}, medium with Allen_{cor} and weak with Rao_{cor}).

Rodrigues *et al.* (2011) have already pointed out a high congruence between Faith_{cor} PD hotspots and SR hotspots. As a result, they concluded that incorporating phylogenetic information is not a major concern in conservation. Nevertheless, incorporating relative species coverage into the definition of multifaceted hotspots alters this conclusion. Faith_{cor} indices do not incorporate species abundance or coverage and give equal weight to rare and dominant evolutionary history in a given location (Chao *et al.*, 2010). However, it seems appropriate to give less weight to particular evolutionary histories (i.e. particular branch paths) that are rare in a given ecoregion because they are less representative than a widespread species in this ecoregion. Allen_{cor} and Rao_{cor} indices give more weight to a given branch if it is long and well represented in an ecoregion. For instance, hotspots defined using PD Rao_{cor} are mostly concentrated in the Australasian realm because of the presence of marsupials. This group has a unique evolutionary history since they diverged 147 million years ago from placentals (extant eutherians, containing the majority of mammals) and are widely distributed (i.e. have large coverage) through the Australasian ecoregions (and not in South American ecoregions). These results are congruent with those of a recent study revealing important mismatches between global hotspots of mammal trait variance and SR (Huang *et al.*, 2012). Although these authors used a different approach (they used grid cells as geographical units and did not use information about species coverage), these close results are probably explained by the fact that Rao_{cor} indices are linked to a measure of variance (Pavoine & Bonsall, 2011). We also showed that the use of body mass alone to define FD hotspots is not sufficient to match the FD hotspots defined with our complete set of traits, but it still represents an acceptable approximation. We also showed that PD and FD hotspots are not always congruent, suggesting that PD is not necessarily a good surrogate for FD (at least for the functional traits selected here).

As well as defining hotspots, DARs have been shown to be useful in both applied and fundamental ecology. We found that Faith_{cor} PDAR and FDAR generally reach their maximum faster than SAR (Cumming & Child, 2009). This result was expected, since Faith_{cor} PD and FD explicitly account for redundancy between species while SR does not. More specifically, it is possible that small sample areas already contain a broad set of phylogenetic history and FD (e.g. a mouse and an elephant), whereas large sample areas contain relatively more redundant

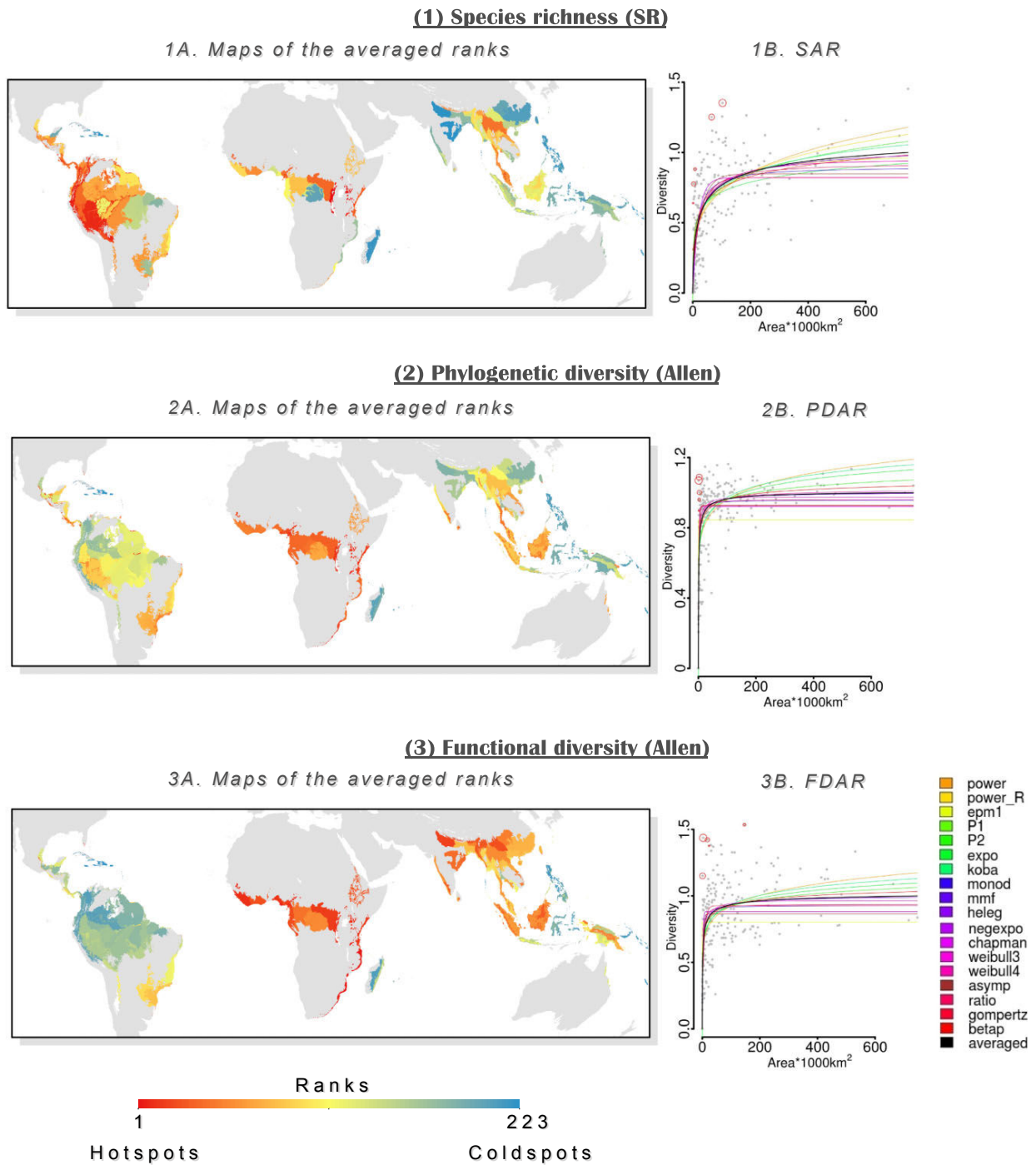


Figure 2 Taxonomic, phylogenetic and functional mammal hotspot selection for tropical moist forests. For each biodiversity facet (1, species richness; 2, phylogenetic diversity (Allen_{cor} PD); and 3, functional diversity (Allen_{cor} FD)) a map (a) and a diversity area relationship (b) are presented. Graphs (b) represent the species–area relationship (SAR), phylogenetic diversity–area relationship (PDAR) and functional diversity–area relationship (FDAR). Model fits are shown with a coloured curve (see legend) and the averaged fit is presented in black. Red circles indicate hotspots, the larger the diameter, the higher the ranking. Maps (a) represent the derived ranks from the residuals of the averaged model presented in (b).

Rank	Ecoregions	Area (km ²)	REALM
Traditional hotspots (SR)			
1	Albertine Rift montane forests	103,403	AT
2	East African montane forests	65,199	AT
3	Eastern Panamanian montane forests	3031	NT
4	Atlantic Coast restingas	7850	NT
5	Mount Cameroon and Bioko montane forests	1141	AT
Phylogenetic hotspots (Allen _{cor} PD)			
1	Mount Cameroon and Bioko montane forests	1141	AT
2	Knysna–Amatole montane forests	3061	AT
3	Peninsular Malaysian peat swamp forests	3610	IM
4	Eastern Panamanian montane forests	3031	NT
5	Chimalapas montane forests	2077	NT
Functional hotspots (Allen _{cor} FD)			
1	Knysna–Amatole montane forests	3061	AT
2	Mount Cameroon and Bioko montane forests	1141	AT
3	KwaZulu–Cape coastal forest mosaic	17,779	AT
4	Southern Zanzibar–Inhambane coastal forest mosaic	146,463	AT
5	Eastern Arc forests	23,556	AT
Integrative hotspots (Allen _{cor} PD and FD and SR)			
1	Mount Cameroon and Bioko montane forests	1141	AT
2	Eastern Arc forests	23,556	AT
3	East African montane forests	65,199	AT
4	Albertine Rift montane forests	103,404	AT
5	Peninsular Malaysian peat swamp forests	3610	IM

AT, Afrotropics; IM, Indomalaysian; NT, Neotropics; PD, phylogenetic diversity; FD, functional diversity; SR, species richness. Allen_{cor} PD and FD correspond to modified version of Allen entropy.

species (e.g. several species of mice) and thus PDAR/FDAR reach their maximum faster than SAR.

Morlon *et al.* (2011) obtained a similar result for PDAR on nested Mediterranean plant communities ranging from 6.25 to 400 m² of spatial extent. They used a power law (see Appendix S1) to model PDAR and SAR and found that the rate of increase in Faith PD with area (z_{PDAR}) was slower than in SR (z_{SAR}). When standardizing DARS, PDAR is above SAR if $z_{PDAR} < z_{SAR}$. They showed that protected areas in Australian mediterranean-like regions (representing 13% of the regions) capture 72% of PD, but only 56% of SR, indicating that PDAR accumulates total diversity faster than SAR.

Our results show that if only a fraction of the total biome area is protected, the percentage of remaining PD (compared with the initial PD) will be higher than the percentage of remaining species. If we consider that PD or FD are better predictors of ecosystem functioning, resistance or resilience (Cadotte *et al.*, 2009; Gravel *et al.*, 2012) than SR, it means that ecosystem features might be more robust to species loss than previously predicted (but see Moullot *et al.*, 2013).

We also found that a key feature of a comprehensive measure of diversity is that when rarely represented evolutionary history is progressively removed (i.e. using different values of q), the differences between PDAR/FDAR and SAR increase. In other words, PD/FD of abundant lineages reaches its maximum faster than when considering all lineages having the same coverage.

Table 2 The five hottest hotspots of tropical and subtropical moist broadleaf forest.

This result suggests that the evolutionary history or functional traits of well-represented taxa are relatively more rapidly sampled when area increases. For example, branches of major mammal lineages (e.g. bats, rodents or carnivores) are probably already well sampled in small ecoregions and thus PD or FD reach their maximum faster than TD in larger sample areas. It follows that well-represented functional/phylogenetic biodiversity might be robust to habitat loss, a point that is not detected when considering all lineages having the same coverage.

Although SARs have been thoroughly investigated (Scheiner, 2003), we have shown that there is not a single best model that fits all the data. Thus the automatic use of a single model (traditionally the linear version of the power model) is not justified. Conversely, to date PDAR and FDAR have been subject to very little investigation (but see Cumming & Child, 2009; Morlon *et al.*, 2011; Wang *et al.*, 2011; Helmus & Ives, 2012). Here, given the important variability across biomes and indices, we also show that a single best model does not exist for PDAR and FDAR. Nevertheless the model-averaging framework allows these uncertainties to be taken into account and we used an averaged prediction to remove the area effect on PD/FD. We also asked whether the averaged SAR could be a good proxy of the averaged PDAR/FDAR to remove this area effect to define PD/FD hotspots. We demonstrated that there is a notable difference between PD/FD hotspot lists defined using PDAR/FDAR and those defined using SAR, suggesting that the construction

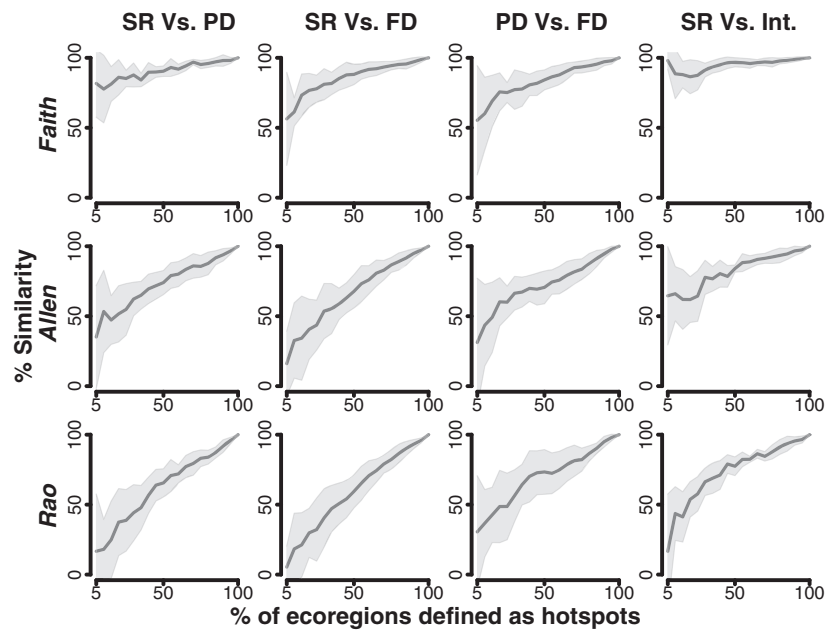


Figure 3 Relationships between the threshold used to define hotspots (expressed in percentage of ecoregion defined as a hotspot) and the similarity between corresponding hotspot lists. From the left to the right columns we compared species richness and phylogenetic diversity hotspots (SR Vs. PD, species richness and functional diversity hotspots (SR Vs. FD), phylogenetic and functional diversity hotspots (PD Vs. FD), species richness and integrative hotspots (agreement between the three facets, SR Vs. Int.). From the top to the bottom row we used $Faith_{cor}$, $Allen_{cor}$ and Rao_{cor} as PD and FD. The dark continuous line represents the mean percentage of congruence of hotspot lists averaged across biomes; the shaded polygon is the associated standard error of the mean. The relative congruence among hotspot lists of two biodiversity facets was determined as the number of ecoregions identified as hotspots by both, divided by the total number of ecoregions in a group.

of PDAR/FDAR is required to define functionally or phylogenetically based hotspots and that SAR alone cannot be used for this purpose.

We constructed DARs using a particular experimental design (Scheiner, 2003) but we are aware that all methods for constructing DARs have their own drawbacks and we suggest that the next challenge in the study of large-scale multifaceted DARs is to test different methodological designs. For example, the strictly nested design (SNQ) of Storch *et al.* (2012) seems particularly interesting to analyse. Nevertheless, since our work was about delineating hotspots of diversity, we had to construct DARs using a non-overlapping design.

CONCLUSION

Here we used a unified framework for building large-scale DARs for each facet of mammal diversity. The spatial scaling of each facet revealed that PD/FD reach their maximal diversity faster than SAR, suggesting that PD/FD might be less vulnerable than SR to habitat loss. In addition, we extracted the area effect on the diversity of individual terrestrial ecoregions to identify multifaceted hotspots of diversity. We showed that multifaceted hotspots are not necessarily congruent and, thus, that SR, PD and FD are not necessarily good surrogates for each other, especially when considering relative species coverage. Although the identification of global hotspots is important as an initial

coarse-scale assessment of the conservation value of different regions (Lamoreux *et al.*, 2006), several challenges would need to be addressed before our results could be directly transferred into conservation planning actions.

ACKNOWLEDGEMENTS

We thank J. Lawler, G. Cumming and an anonymous referee, who provided useful comments on an earlier version of this manuscript. The research leading to these results received funding from the European Research Council under the European Community's Seven Framework Programme FP7/2007–2013 grant agreement no. 281422 (TEEMBIO). M.V.C., R.D.L. and J.A.F.D.F. received productivity research grants from CNPq, Brazil. D.M. has received funding from the Marie Curie International Outgoing Fellowship (FISHECO) with agreement number IOF-GA-2009-236316. FM, WT and JR belong to the Laboratoire d'Écologie Alpine, which is part of Labex OSUG@2020 (ANR10 LABX56).

REFERENCES

- Allen, B., Kon, M. & Bar-Yam, Y. (2009) A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats. *The American Naturalist*, **174**, 236–243.

- Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L. & Purvis, A. (2007) The delayed rise of present-day mammals. *Nature*, **446**, 507–512.
- Burnham, K.P. & Anderson, D.R. (2002) *Model selection and inference: a practical information-theoretic approach*. Springer-Verlag, New York.
- Cadotte, M., Cavender-Bares, J., Tilman, D. & Oakley, T. (2009) Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *PLoS ONE*, **4**, e5695.
- Ceballos, G. & Ehrlich, P.R. (2006) Global mammal distributions, biodiversity hotspots, and conservation. *Proceedings of the National Academy of Sciences USA*, **103**, 19374–19379.
- Chao, A., Chiu, C.-H. & Jost, L. (2010) Phylogenetic diversity measures based on Hill numbers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3599–3609.
- Chown, S. & Gaston, K. (2000) Areas, cradles and museums: the latitudinal gradient in species richness. *Trends in Ecology and Evolution*, **15**, 311–315.
- Cumming, G.S. & Child, M.F. (2009) Contrasting spatial patterns of taxonomic and functional richness offer insights into potential loss of ecosystem services. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 1683–1692.
- Davies, T.J. & Buckley, L.B. (2011) Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2414–2425.
- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, **61**, 1–10.
- Forest, F., Grenyer, R., Rouget, M., Davies, T.J., Cowling, R.M., Faith, D.P., Balmford, A., Manning, J.C., Procheş, S., van der Bank, M., Reeves, G., Hedderson, T.A.J. & Savolainen, V. (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, **445**, 757–760.
- Fritz, S.A., Bininda-Emonds, O.R.P. & Purvis, A. (2009) Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology Letters*, **12**, 538–549.
- Gaston, K.J. (2003) *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford.
- Gaston, K.J., Blackburn, T.M., Greenwood, J.J.D., Gregory, R.D., Quinn, R.M. & Lawton, J.H. (2000) Abundance–occupancy relationships. *Journal of Applied Ecology*, **37**, 39–59.
- Gravel, D., Bell, T., Barbera, C., Combe, M., Pommier, T. & Mouquet, N. (2012) Phylogenetic constraints on ecosystem functioning. *Nature Communications*, **3**, art. 1117. doi:10.1038/ncomms2123
- Groves, C. (2003) *Drafting a conservation blueprint: a practitioner's guide to planning for biodiversity*. Island Press, Washington, DC.
- Guilhaumon, F., Gimenez, O., Gaston, K.J. & Mouillot, D. (2008) Taxonomic and regional uncertainty in species–area relationships and the identification of richness hotspots. *Proceedings of the National Academy of Sciences USA*, **105**, 15458–15463.
- Hawkins, B.A., Diniz-Filho, J.A.F., Jaramillo, C.A. & Soeller, S.A. (2006) Post-Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *Journal of Biogeography*, **33**, 770–780.
- Hawkins, B.A., McCain, C.M., Davies, T.J., Buckley, L.B., Anacker, B.L., Cornell, H.V., Damschen, E.I., Grytnes, J.-A., Harrison, S., Holt, R.D., Kraft, N.J.B. & Stephens, P.R. (2012) Different evolutionary histories underlie congruent species richness gradients of birds and mammals. *Journal of Biogeography*, **39**, 825–841.
- Helmus, M.R. & Ives, A.R. (2012) Phylogenetic diversity–area curves. *Ecology*, **91**, 31–43.
- Huang, S., Stephens, P.R. & Gittleman, J.L. (2012) Traits, trees and taxa: global dimensions of biodiversity in mammals. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 4997–5003.
- 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.
- Jost, L. (2006) Entropy and diversity. *Oikos*, **113**, 363–375.
- Lamoreux, J.F., Morrison, J.C., Ricketts, T.H., Olson, D.M., Dinerstein, E., McKnight, M.W. & Shugart, H.H. (2006) Global tests of biodiversity concordance and the importance of endemism. *Nature*, **440**, 212–214.
- Loreau, M., Naeem, S. & Inchausti, P. (2002) *Biodiversity and ecosystem functioning: synthesis and perspectives*. Oxford University Press, New York.
- Mittermeier, R.A., Mittermeier, C.G., Brooks, T.M., Pilgrim, J.D., Konstant, W.R., da Fonseca, G.A.B. & Kormos, C. (2003) Wilderness and biodiversity conservation. *Proceedings of the National Academy of Sciences USA*, **100**, 10309–10313.
- Mittermeier, R.A., Gil, P.R., Hoffman, M., Pilgrim, J., Brooks, T., Mittermeier, C.G., Lamoreux, J. & Da Fonseca, G.A.B. (2004) *Hotspots revisited*. CEMEX, Mexico City.
- Morlon, H., Schwilk, D.W., Bryant, J.A., Marquet, P.A., Rebelo, A.G., Tauss, C., Bohannan, B.J.M. & Green, J.L. (2011) Spatial patterns of phylogenetic diversity. *Ecology Letters*, **14**, 141–149.
- Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-vivien, M., Kulbicki, M., Lavergne, S., Lavorel, S., Mouquet, N., Paine, T.C.E., Renaud, J. & Thuiller, W. (2013) Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biology*, **11**(5), e1001569. doi: 10.1371/journal.pbio.1001569
- Mouquet, N., Devictor, V., Meynard, C.N. *et al.* (2012) Ecophylogenetics: advances and perspectives. *Biological Reviews*, **87**, 769–785.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Olson, D.M. & Dinerstein, E. (1998) The Global 200: a representation approach to conserving the Earth's most biologically valuable ecoregions. *Conservation Biology*, **12**, 502–515.

- Olson, D.M., Dinerstein, E., Wikramanayake, E.D. *et al.* (2001) Terrestrial ecoregions of the world: a new map of life on earth. *BioScience*, **51**, 933–938.
- Orme, C.D.L., Davies, R.G., Burgess, M., Eigenbrod, F., Pickup, N., Olson, V.A., Webster, A.J., Ding, T.-S., Rasmussen, P.C., Ridgely, R.S., Stattersfield, A.J., Bennett, P.M., Blackburn, T.M., Gaston, K.J. & Owens, I.P.F. (2005) Global hotspots of species richness are not congruent with endemism or threat. *Nature*, **436**, 1016–1019.
- Pavoine, S. & Bonsall, M.B. (2011) Measuring biodiversity to explain community assembly: a unified approach. *Biological Reviews*, **86**, 792–812.
- Petchey, O.L. & Gaston, K.J. (2006) Functional diversity: back to basics and looking forward. *Ecology Letters*, **9**, 741–758.
- R Development Core Team (2010) *R: a language and environment for statistical computing*. R Development Core Team, Vienna, Austria.
- Rao, R.C. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*, **21**, 24–43.
- Reid, W.V. (1998) Biodiversity hotspots. *Trends in Ecology and Evolution*, **13**, 275–280.
- Rodrigues, A.S.L., Grenyer, R., Baillie, J.E.M., Bininda-Emonds, O.R.P., Gittlemann, J.L., Hoffmann, M., Safi, K., Schipper, J., Stuart, S.N. & Brooks, T. (2011) Complete, accurate, mammalian phylogenies aid conservation planning, but not much. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2652–2660.
- 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 B: Biological Sciences*, **366**, 2536–2544.
- Scheiner, S.M. (2003) Six types of species–area curves. *Global Ecology and Biogeography*, **12**, 441–447.
- Sekercioglu, C.H. (2010) Ecosystem functions and services. *Conservation biology for all* (ed. by P.R. Sodhi and N.S. Ehrlich), pp. 45–72. Oxford University Press, Oxford.
- Srivastava, D.S., Cadotte, M.W., MacDonald, A.A.M., Marushia, R.G. & Mirotnick, N. (2012) Phylogenetic diversity and the functioning of ecosystems. *Ecology Letters*, **15**, 637–648.
- Storch, D., Keil, P. & Jetz, W. (2012) Universal species–area and endemics–area relationships at continental scales. *Nature*, **488**, 78–81.
- Tjørve, E. (2009) Shapes and functions of species–area curves (II): a review of new models and parameterizations. *Journal of Biogeography*, **36**, 1435–1445.
- Triantis, K.A., Guilhaumon, F. & Whittaker, R.J. (2012) The island species–area relationship: biology and statistics. *Journal of Biogeography*, **39**, 215–231.
- Wang, X., Wiegand, T., Wolf, A., Howe, R., Davies, S.J. & Hao, Z. (2011) Spatial patterns of tree species richness in two temperate forests. *Journal of Ecology*, **99**, 1382–1393.
- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology, Evolution, and Systematics*, **33**, 475–505.
- Wiens, J.J., Pyron, R.A. & Moen, D.S. (2011) Phylogenetic origins of local-scale diversity patterns and the causes of Amazonian megadiversity. *Ecology Letters*, **14**, 643–652.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1 Diversity–area relationship model list.

Appendix S2 Proof of rank invariance after standardization.

Appendix S3 Variance explained by the diversity–area relationships.

Appendix S4 Species–area relationship, phylogenetic diversity–area relationship and functional diversity–area relationship model selection patterns.

Appendix S5 Analysing model shape by plotting differences in diversity–area relationships.

Appendix S6 Maps of taxonomic, phylogenetic and functional mammal hotspots.

Appendix S7 Correlation of ranks between facets.

Appendix S8 Relationships between the threshold used to define hotspots and the congruence between hotspots.

Appendix S9 Robustness of the hotspots lists to choice of diversity–area relationship model.

Appendix S10 Congruence between functional diversity hotspots defined with different sets of functional traits.

Appendix S11 Importance of diversity–area relationship construction for defining multifaceted hotspots.

Appendix S12 Comparing model averaging and the power model alone to define hotspots.

BIOSKETCH

F. Mazel is a PhD student mostly interested in macroecology and macroevolution. Specifically, he aims to describe and understand the distribution of biodiversity in the light of ecology, evolution, palaeontology and palaeoclimatology.

Author contributions: F.M., F.G., N.M., V.D., D.G., D.M. and W.T. designed the study. J.R. collected and formatted distribution data. M.V.C., R.L. and J.A.F.D.F. provided the functional traits database. F.M. ran the analysis and wrote the first draft of the manuscript; all authors contributed substantially to revisions.

Editor: Joshua Lawler

SYNTHÈSE ET DISCUSSION DE LA PARTIE II

Cette deuxième partie part du principe que les diversités phylogénétiques et fonctionnelles représentent des facettes importantes de la biodiversité biologique et qu'elles doivent donc être prises en compte dans les plans de conservation. Elles apportent en effet des informations uniques que la diversité spécifique ne peut pas représenter puisque cette dernière ignore les différences entre espèces (Vane-Wright et al. 1991, Lecomte 2011). Les intérêts de leurs utilisations sont d'ordre éthique (Rosauer and Mooers 2013) et utilitariste (Díaz et al. 2007, Srivastava et al. 2012). En admettant une telle approche, il faut donc s'attacher à comprendre si ces diversités ont les mêmes répartitions et les mêmes comportements face aux changements globaux. Ceci pourrait alors permettre de confirmer les plans de conservation actuels si les facettes sont congruentes, c'est-à-dire que la protection des espèces et de la richesse spécifique (le paradigme actuel) permet dans le même temps de protéger les autres facettes. Au contraire, si les facettes ne sont pas congruentes, il faudra essayer d'aménager les plans de conservation (par exemple la création ou l'extension d'aires protégées) pour prendre en compte ces nouvelles facettes.

1. Évaluer les menaces qui pèsent sur l'arbre de la vie (de certains vertébrés)

Les changements globaux qui menacent la biodiversité sont de natures multiples. Les changements climatiques par exemple semblent menacer les espèces (Thomas *et al.* 2004; Thuiller *et al.* 2005; Bellard *et al.* 2012). Il a ainsi été récemment montré que de tels changements pouvaient aussi avoir des effets négatifs sur la PD (Thuiller et al. 2011) et sur la FD (Thuiller et al. 2014). La destruction et la fragmentation de l'habitat représentent aussi une menace majeure pour la biodiversité. Morlon *et al.* (2011) ont ainsi proposé d'utiliser la PDAR pour estimer la perte de PD consécutive à la perte d'habitat.

1.1. Apport de l'analyse

En utilisant cet outil, nous avons montré qu'à l'échelle continentale la PD est plus robuste à la perte d'habitat que la SR. Cette découverte n'est en fait pas nouvelle mais nous le quantifions pour la première fois à cette échelle. En particulier, nous

développons un modèle nul qui permet de différencier l'effet de la forme de l'arbre (et donc de « l'assurance » que représentent les branches internes, voir l'article fondateur de Nee & May, 1997) de l'effet de la distribution spatiale de la PD. Nous montrons pour la première fois que les deux processus sont importants.

1.2. Limites de l'analyse

Notre analyse est cependant assez simplifiée et plusieurs limites sont à noter.

(1) Tout d'abord notre échelle d'étude est déconnectée des échelles de destruction et de fragmentation de l'habitat, qui sont souvent inférieures à notre grain minimal de 110*110 km. Cependant, le développement de l'indice de robustesse relative de la PD ainsi que notre approche du modèle nul peuvent être théoriquement appliqués à toutes les échelles spatiales, démontrant l'intérêt de notre analyse.

(2) Nous faisons l'hypothèse que toutes les espèces sont également vulnérables à la perte d'habitat, ou que les extinctions sont aléatoires dans la phylogénie. Cette hypothèse est en fait loin d'être vérifiée et il semble plutôt que les risques d'extinction montrent un signal phylogénétique (Purvis 2000, 2008). Ceci peut être un effet indirect du signal phylogénétique que montrent certains traits liés au risque d'extinction, par exemple la masse moyenne de l'espèce (Fritz et al. 2009). On notera cependant que les risques d'extinction liés aux changements climatiques ne semblent pas impacter de façon disproportionnée la diversité phylogénétique (Pio et al. 2011, Thuiller et al. 2011).

(3) Enfin, l'utilisation de la relation aire-espèce pour prédire les taux d'extinction reste controversée (May et al. 1995, Lewis 2006, He and Hubbell 2011, Matias et al. 2014). En particulier, il n'y a pas de raison a priori que l'on puisse inverser le SAR pour produire des estimations des taux d'extinctions (notamment car il s'est construit sur des temps évolutifs longs or on l'applique sur des temps courts). Par ailleurs, la perte d'habitat (1) n'est jamais totale mais peut simplement représenter une modification, et (2) n'est jamais uniforme mais peut prendre des formes complexes. En définitive, l'application du SAR est donc limitée mais reste pourtant un des seuls outils disponibles.

2. Les différentes facettes de la diversité sont-elles congruentes et également protégées ?

2.1 Les points chauds : des aires à conserver ?

Quelques études commencent à montrer que les différentes facettes de la biodiversité ne sont pas congruentes. C'est le cas pour les oiseaux à l'échelle de la France (Devictor et al. 2010), les vertébrés à l'échelle de l'Europe (Zupan et al. 2014) ou les poissons à l'échelle du monde (Stuart-Smith et al. 2013). Cependant, une telle étude n'a pas été menée à l'échelle du globe pour les mammifères terrestres. Nous développons une telle analyse et montrons que les différentes facettes ne sont pas congruentes entre elles, mais ceci dépend de l'indice de diversité phylogénétique ou fonctionnelle considéré. Nous comparons en fait trois indices de diversité fonctionnelle et phylogénétique à la richesse spécifique. Ces trois indices proviennent du cadre méthodologique proposé par Chao *et al.* (2010) et sont censés représenter des mesures de diversité (phylogénétique ou fonctionnelle) étalées sur un continuum d'importance donnée à l'abondance des espèces. En effet, les indices sont unifiés par les nombres de Hill (Hill 1973) et se définissent tous avec une même formule qui dépend d'un paramètre q . Ce paramètre est censé, selon les auteurs, faire varier le poids donné aux espèces rares. Ainsi le premier indice ($q=0$, PD_{Faith}) n'est pas sensible à l'abondance alors que le troisième ($q=2$, Rao QE) donne un poids important aux espèces abondantes. Nous proposons d'utiliser le couvert relatif de chaque espèce dans chaque écorégion comme mesure « d'abondance ». Ainsi l'indice censé donner plus de poids aux espèces abondantes va quantifier la diversité phylogénétique (ou fonctionnelle) qui est représentée de façon importante dans une écorégion alors que l'indice insensible aux abondances ($q=0$) donnera un poids identique à toutes les abondances. Nous voulons ainsi tester si le couvert relatif de chaque espèce dans les écorégions impacte la nature des points chauds. Dans l'article, nous concluons par l'affirmative puisque l'indice à $q=0$ est très proche de la SR alors que l'indice $q=2$ est très fortement découplé de SR.

Note importante sur les indices utilisés dans l'article

En fait, je me suis rendu compte au cours de ma thèse que le cadre méthodologique proposé par Chao *et al.* (2010), bien qu'apparemment très élégant

mathématiquement, était, d'une certaine façon, un peu trompeur. L'interprétation des indices proposée par ces auteurs me paraît incomplète. Ils expliquent en effet que le paramètre q de leur cadre méthodologique fait uniquement varier l'importance donnée aux abondances relatives des espèces. Ceci paraît discutable. En effet, si ce paramètre ne faisait varier que l'importance de l'abondance, l'utilisation des données de présence/absence (c'est-à-dire sans information d'abondance) devrait conduire à retrouver les mêmes indices, peu importe la valeur de q (voir Chen *et al.* 2012 pour un raisonnement similaire et le matériel supplémentaire 1 de l'article 2 pour une justification plus précise). Ceci ne semble pas être vérifié. En effet nous montrons dans le chapitre I que PD_{Faith} ($q=0$) est un indice de richesse alors que Rao QE ($q=2$) est un indice de divergence. Ces indices ne sont donc a priori pas comparables en soi et le cadre méthodologique de Chao est donc à réinterpréter : le paramètre q impacte effectivement le poids donné aux abondances mais aussi la dimension de la structure phylogénétique décrite. En ce sens, l'interprétation consistant à dire que les différents résultats trouvés pour différentes valeurs de q proviennent de différences d'abondances est incomplète. Je note que ceci est particulièrement alarmant pour les études fondamentales essayant d'expliquer les processus responsables de l'assemblage des communautés.

Réinterprétation et discussion des résultats.

Mais en quoi cette découverte impacte-t-elle notre analyse des points chauds de biodiversité?

En fait, nous montrons donc que les points chauds de richesse phylogénétique sont très proches (mais pas identiques) de ceux de richesse spécifique. Ceci est cohérent avec les résultats de Rodrigues *et al.* (2011) et montrent que les deux facettes sont très congruentes. Néanmoins notre étude ne prend pas en compte la complémentarité de richesse phylogénétique entre les sites (voir Margules & Pressey 2000; Rodrigues & Gaston 2002) et ne peut donc pas directement être utilisée en conservation. Ainsi, les points chauds ne représentent pas nécessairement des zones à conserver absolument car elles sont potentiellement redondantes. Cependant, notre analyse constitue un pas intéressant dans la comparaison des facettes de la biodiversité. Un aspect novateur de notre étude est qu'elle compare différentes dimensions de la structure phylogénétique des assemblages (voir Chapitre 1). Ainsi, Rao QE représente une divergence

phylogénétique ou fonctionnelle et nous montrons que cette dimension est découplée de la richesse spécifique. En quoi une telle mesure de diversité est-elle importante en conservation ? L'indice de choix en conservation est généralement la richesse (qu'elle soit phylogénétique, fonctionnelle ou spécifique) car elle représente une quantité qu'il est aisé de maximiser dans le cadre d'une recherche d'aire à protéger. En revanche, les indices de divergence ont des comportements très inattendus en conservation si l'on désire les maximiser. En effet, l'ajout d'une espèce dans un assemblage peut faire diminuer la valeur de l'indice, ce qui est totalement contre-intuitif en conservation ! Cependant, je pense que de tels indices peuvent être extrêmement intéressants, et cela pour différentes raisons.

- (1) Dans le cadre d'un objectif de représentativité (Margules and Pressey 2000, Ladle and Whittaker 2011), les indices de divergences peuvent nous renseigner sur les assemblages évolutivement uniques, par exemple les musées (rassemblant des espèces « anciennes ») ou les berceaux de la vie (centres de diversification), ce que ne peuvent pas faire les indices de richesse (Devictor et al. 2010, Stuart-Smith et al. 2013, Zupan et al. 2014). Je note ici que les points 'froids' de divergence phylogénétique représentent donc potentiellement des berceaux de diversité mais que je ne les ai pas pris en compte dans l'analyse.

- (2) Dans le cadre d'un objectif utilitariste nous avons vu en introduction que la diversité fonctionnelle (ou phylogénétique) pouvait potentiellement être un meilleur prédicteur du fonctionnement des écosystèmes et donc de leurs services. En fait la divergence fonctionnelle est effectivement un indice de choix (Lavelle et al. 2007). Il est donc intéressant de montrer que de tels indices ne sont pas congruents avec la richesse spécifique.

Nous avons montré qu'il existe un découplage entre les différentes facettes de la diversité de certains vertébrés alors que la plupart du fonctionnement et surtout des services des écosystèmes sont assurés par les producteurs primaires comme les plantes et les décomposeurs comme les champignons. Ainsi, il semble qu'il existe un certain découplage entre notre analyse et sa justification dans le cadre du maintien du fonctionnement des écosystèmes et de leurs services. Néanmoins, plusieurs fonctions

essentielles semblent être assurées par les tétrapodes en général (mammifères mais aussi oiseaux, voir Sekercioglu et al. 2004, Wenny et al. 2011, Luck et al. 2012). Par exemple, certaines espèces permettent la dispersion des graines et la pollinisation de certaines espèces végétales (Sekercioglu 2011) alors que d'autres peuvent façonner les paysages (les espèces « ingénieurs »). Aussi, il est généralement admis que la perturbation d'un seul niveau trophique peut avoir des effets importants pour l'écosystème dans son ensemble (voir Daskin *et al.* (2015) pour un exemple récent). En particulier, les modifications de la structure des niveaux trophiques supérieurs (herbivores et prédateurs) peuvent avoir des conséquences en cascade sur tous les niveaux de la chaîne alimentaire (Schmitz et al. 2000, Shurin et al. 2002, Myers et al. 2007). Ainsi, notre analyse représente un premier pas vers l'inclusion des facettes fonctionnelles de la biodiversité en conservation mais ne représente pas un outil directement utilisable si l'on veut inclure les relations trophiques.

2.2 Les aires de protection préservent-elles les arbres de la vie?

Les points chauds de biodiversité sont des éléments importants en biologie de la conservation (même si leur définition précise est sujette à débat, Rodrigues 2013). Cependant, ils ne sont pas directement utilisables en conservation, contrairement aux approches de conservation systématique (Margules and Pressey 2000). Ces approches permettent de créer des espaces protégés pour répondre à des critères précis pré-établis (par exemple représentativité, complémentarité). Dans le cadre de la protection de la diversité à plusieurs facettes, il s'agit donc dans un premier temps de tester si les aires de protection existantes protègent ou non les diversités fonctionnelles et phylogénétiques. C'est ce que nous avons fait à l'échelle de l'Europe pour les tétrapodes (mammifères, oiseaux, squamates et amphibiens, voir Annexe 1.3). Une analyse de trouée ('gap analysis') nous permet d'estimer dans quelle mesure l'organisation géographique des aires de protection actuelle (parcs et réserves nationales et réseau Natura 2000) coïncident avec la distribution des espèces, et en particulier celle des espèces rares. En effet, nous définissons dans un premier temps le pourcentage de distribution ('objectif de protection') permettant d'assurer le maintien d'une espèce. Ce pourcentage est défini en fonction de la taille de l'aire de distribution : ainsi une espèce

rare aura un objectif de protection de 100% alors qu'une espèce très répandue aura un objectif inférieur. Ensuite, nous calculons la différence (le 'gap') entre l'objectif théorique et la réalité de terrain (réalisation de l'objectif de protection, qui oscille entre 30 et 50% selon les groupes de notre étude). Nous évaluons ensuite dans quelle mesure ce degré de protection des espèces se répercute sur les arbres phylogénétiques et fonctionnels des tétrapodes. Nous montrons qu'en général l'arbre fonctionnel est bien mieux conservé que l'arbre phylogénétique, mais ceci dépend du groupe étudié. Ainsi, il semble important que l'extension future des aires protégées européennes tente de maximiser la représentation de la diversité phylogénétique et/ou fonctionnelle. Une première approche pourrait par exemple utiliser la méthodologie proposée par Pollock *et al.* (2015) et de l'étendre au contexte global en prenant plusieurs facettes de diversité et plusieurs groupes (travail en cours dans l'équipe EMABIO).

3. Perspectives futures

Nous montrons dans cette partie que les facettes phylogénétiques et fonctionnelles de la biodiversité ne sont pas nécessairement géographiquement congruentes avec la facette taxonomique. En d'autres termes, une focalisation excessive sur la protection d'une des facettes ne pourra pas protéger les autres efficacement. Si l'on considère que ces trois facettes doivent être également protégées, il est alors urgent de les incorporer toutes les trois dans les plans de conservation. Cependant, on peut se demander à quel point la protection des trois facettes est indispensable. Je rappelle ici que les deux arguments majeurs pour la protection de toutes les facettes peuvent être d'ordre éthique ou d'ordre utilitariste. Je suis convaincu que, d'un point de vue purement éthique, protéger la diversité fonctionnelle et surtout la diversité phylogénétique est essentiel. D'un point de vue utilitariste, la discussion me paraît plus ouverte. Un des arguments utilitaristes central est de dire que la perte de diversité (provoquée par l'action de l'homme) va engendrer des dysfonctionnement des écosystèmes (d'où la prolifération des études entre fonctionnement et diversité) et, à terme, une diminution potentielle des services. Cependant, l'excellente revue de Srivastava & Vellend (2005) semble mettre en doute certains points de ce raisonnement. En effet, les auteurs montrent qu'il existe un fossé important entre les évidences empiriques de la relation diversité-fonctionnement et les attendus, les mesures et les

moyens dont dispose la biologie de la conservation. Une première série de limites provient du fait que la diversité spécifique ne permet pas de correctement décrire le fonctionnement des écosystèmes. Cette limite apparaît justement contournable si la diversité fonctionnelle est utilisée (ou la diversité phylogénétique si les traits sont conservés), et c'est tout le sens de cette seconde partie. Cependant, plusieurs autres limites persistent, notamment :

- (1) la perte de diversité semble être une cause (indirecte) mineure du dysfonctionnement des écosystèmes, notamment par rapport aux effets directs (par exemple, la destruction d'une forêt).
- (2) les indices de diversité fonctionnelle actuels ne prennent en compte qu'indirectement la structure des réseaux trophiques (mais voir Poisot *et al.*, 2012).

Ces limites pourront être dépassées en combinant des approches empiriques et des avancées théoriques, je pense notamment à la structure et la dynamique des réseaux trophiques. En particulier, je considère qu'il serait important de mieux décrire la structure de tels réseaux (voir par exemple Morales-Castilla *et al.* 2015), de mieux comprendre les conséquences de la perte de certains maillons de la chaîne sur le fonctionnement des écosystèmes et de mieux relier les indices de diversité fonctionnelle et la structure du réseau.

BIBLIOGRAPHIE GÉNÉRALE
DE LA PARTIE II
(HORS ARTICLES)

Partie II – Bibliographie

- Barnosky, A. D. et al. 2011. Has the Earth's sixth mass extinction already arrived? - *Nature* 471: 51–7.
- Barnosky, A. D. et al. 2012. Approaching a state shift in Earth's biosphere. - *Nature* 486: 52–8.
- Bellard, C. et al. 2012. Impacts of climate change on the future of biodiversity. - *Ecol. Lett.* 15: 365–377.
- Brook, B. W. et al. 2003. Catastrophic extinctions follow deforestation in Singapore. - *Nature* 424: 420–6.
- Cadotte, M. W. et al. 2008. Evolutionary history and the effect of biodiversity on plant productivity. - *Proc. Natl. Acad. Sci. U. S. A.* 105: 17012–7.
- Cadotte, M. W. et al. 2009. Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. - *PLoS One* 4: e5695.
- Callicott, J. B. et al. 1999. Current Normative Concepts in Conservation. - *Conserv. Biol.* 13: 22–35.
- Cardinale, B. J. et al. 2006. Effects of biodiversity on the functioning of trophic groups and ecosystems. - *Nature* 443: 989–92.
- Cardinale, B. J. et al. 2007. Impacts of plant diversity on biomass production increase through time because of species complementarity. - *Proc. Natl. Acad. Sci. U. S. A.* 104: 18123–8.
- Cardinale, B. J. et al. 2012. Biodiversity loss and its impact on humanity. - *Nature* 486: 59–67.
- Ceballos, G. and Ehrlich, P. R. 2002. Mammal population losses and the extinction crisis. - *Science* 296: 904–7.
- Chao, A. et al. 2010. Phylogenetic diversity measures based on Hill numbers. - *Philos. Trans. R. Soc. London, Ser. B* 365: 3599–3609.
- Chen, J. et al. 2012. Associating microbiome composition with environmental covariates using generalized UniFrac distances. - *Bioinformatics* 28: 2106–13.
- Costanza, R. et al. 1997. The value of the world's ecosystem services and natural capital. - *Nature* 387: 253–260.
- Daskin, J. H. et al. 2015. Ecological legacies of civil war: 35-year increase in savanna tree cover following wholesale large-mammal declines (L Gomez-Aparicio, Ed.). - *J. Ecol.*: n/a–n/a.
- De Groot, R. S. et al. 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. - *Ecol. Econ.* 41: 393–408.
- De Vos, J. M. et al. 2015. Estimating the normal background rate of species extinction. - *Conserv. Biol.* 29: 452–62.
- Devictor, V. et al. 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.
- Díaz, S. et al. 2007. Incorporating plant functional diversity effects in ecosystem service assessments. - *Proc. Natl. Acad. Sci. U. S. A.* 104: 20684–9.
- Díaz, S. and Cabido, M. 2001. Vive la différence: plant functional diversity matters to ecosystem processes. - *Trends Ecol. Evol.* 16: 646–655.
- Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - *Biol. Conserv.* 61: 1–10.
- Flynn, D. F. B. et al. 2011. Functional and phylogenetic diversity as predictors of biodiversity–ecosystem–function relationships. - *Ecology* 92: 1573–81.
- Forest, F. et al. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. - *Nature* 445: 757–760.
- Fritz, S. A. et al. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. - *Ecol. Lett.* 12: 538–49.
- Gause, G. F. 1937. Experimental Populations of Microscopic Organisms. - *Ecology* 18: 173.
- He, F. and Hubbell, S. 2011. Species-area relationships always overestimate extinction rates from habitat loss. - *Nature* 473: 368–371.
- Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. - *Ecology* 54: 427–432.
- Hutchinson, G. E. 1957. Concluding Remarks. - *Cold Spring Harb. Symp. Quant. Biol.* 22: 415–427.
- Isaac, N. J. B. et al. 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. - *PLoS One* 2: e296.
- Jepson, P. and Whittaker, R. J. 2002. Histories of Protected Areas: Internationalisation of Conservationist Values and their Adoption in the Netherlands Indies (Indonesia). - *Environ. Hist. Camb.* 8: 129–172.
- Ladle, R. J. and Whittaker, R. J. 2011. Conservation Biogeography. - John Wiley & Sons.
- Lavergne, S. et al. 2010. Biodiversity and Climate Change: Integrating Evolutionary and Ecological Responses of Species and Communities. - *Annu. Rev. Ecol. Evol. Syst.* 41: 321–350.
- Lavorel, S. et al. 2007. Assessing functional diversity in the field – methodology matters! - *Funct. Ecol.* 22: 134–147.
- Lavorel, S. et al. 2011. Using plant functional traits to understand the landscape distribution of multiple ecosystem services. - *J. Ecol.* 99: 135–147.

Partie II – Bibliographie

- Lecointre, G. 2011. La biodiversité : au pays des aveugles le borgne est roi. - *Comptes Rendus Palevol* 10: 331–334.
- Lewis, O. T. 2006. Climate change, species-area curves and the extinction crisis. - *Philos. Trans. R. Soc. B Biol. Sci.* 361: 163–171.
- Loreau, M. 2010. Linking biodiversity and ecosystems: towards a unifying ecological theory. - *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365: 49–60.
- Loreau, M. and Hector, A. 2001. Partitioning selection and complementarity in biodiversity experiments. - *Nature* 412: 72–76.
- Luck, G. W. et al. 2012. Improving the application of vertebrate trait-based frameworks to the study of ecosystem services. - *J. Anim. Ecol.* 81: 1065–76.
- Maestre, F. T. et al. 2012. Plant species richness and ecosystem multifunctionality in global drylands. - *Science* 335: 214–8.
- Margules, C. R. and Pressey, R. L. 2000. Systematic conservation planning. - *Nature* 405: 243–53.
- Matias, M. et al. 2014. Estimates of species extinctions from species–area relationships strongly depend on ecological context. - *Ecography (Cop.)*. 37: 001–012.
- Matthews, J. . et al. 2001. The encyclopaedic dictionary of environmental change. - Arnold.
- May, R. M. et al. 1995. - In: Lawton, J. H. and May, R. M. (eds), *Extinction Rates*. Oxford Univ. Press, pp. 1–24.
- Moilanen, A. et al. 2009. The Zonation framework and software for conservation prioritization. - In: *Spatial conservation prioritisation: Quantitative methods and computational tools*. in press.
- Morales-Castilla, I. et al. 2015. Inferring biotic interactions from proxies. - *Trends Ecol. Evol.* 30: 347–56.
- Morlon, H. et al. 2011. Spatial patterns of phylogenetic diversity. - *Ecol. Lett.* 14: 141–149.
- Mouillot, D. and Villéger, S. 2011. Functional structure of biological communities predicts ecosystem multifunctionality. - *PLoS One* 6(3), e17476
- Mouillot, D. et al. 2013. Rare species support vulnerable functions in high-diversity ecosystems. - *PLOS Biol.* 11, e1001569.
- Mouillot, D. et al. 2014. Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. - *Proc. Natl. Acad. Sci.* 111: 13757–13762.
- Myers, R. A. et al. 2007. Cascading effects of the loss of apex predatory sharks from a coastal ocean. - *Science* 315: 1846–50.
- Nee, S. and May, R. M. 1997. Extinction and the loss of evolutionary history. - *Science* (80-.). 288: 328–330.
- Paquette, A. and Messier, C. 2011. The effect of biodiversity on tree productivity: from temperate to boreal forests. - *Glob. Ecol. Biogeogr.* 20: 170–180.
- Petchey, O. L. and Gaston, K. J. 2006. Functional diversity: back to basics and looking forward. - *Ecol. Lett.* 9: 741–758.
- Pimm, S. and Raven, P. 2000. Biodiversity: extinction by numbers. - *Nature* 403: 843–845.
- Pimm, S. L. et al. 1995. The futur of biodiversity. - *Science* (80-.). 269: 347–350.
- Pio, D. V et al. 2011. Spatial predictions of phylogenetic diversity in conservation decision making. - *Conserv. Biol.* 25: 1229–1239.
- Poisot, T. et al. 2012. The dissimilarity of species interaction networks. - *Ecol. Lett.* 15: 1353–61.
- Pollock, L. J. et al. 2015. Phylogenetic diversity meets conservation policy: small areas are key to preserving eucalypt lineages. - *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370: 20140007.
- Purvis, A. 2000. Nonrandom Extinction and the Loss of Evolutionary History. - *Science*. 288: 328–330.
- Purvis, A. 2008. Phylogenetic Approaches to the Study of Extinction. - *Annu. Rev. Ecol. Evol. Syst.* 39: 301–319.
- Rodrigues, A. S. . 2013. Hotspots. - In: Levin, S. A. (ed), *Encyclopedia of Biodiversity* - 2nd Edition. Academic Press, New York, NY, pp. 127–136.
- Rodrigues, A. S. L. and Gaston, K. J. 2002. Maximising phylogenetic diversity in the selection of networks of conservation areas. - *Biol. Conserv.* 105: 103–111.
- Rodrigues, A. S. et al. 2000. Robustness of reserve selection procedures under temporal species turnover. - *Proc. Biol. Sci.* 267: 49–55.
- Rodrigues, A. S. L. et al. 2011. Complete, accurate, mammalian phylogenies aid conservation planning, but not much. - *Philos. Trans. R. Soc. London, Ser. B* 366: 2652–2660.
- Rosauer, D. F. and Mooers, A. O. 2013. Nurturing the use of evolutionary diversity in nature conservation. - *Trends Ecol. Evol.* 28: 322–3.
- Schmitz, O. et al. 2000. Trophic Cascades in Terrestrial Systems: A Review of the Effects of Carnivore Removals on Plants. - *Am. Nat.* 155: 141–153.
- Sekercioglu, C. H. 2011. Functional extinctions of bird pollinators cause plant declines. - *Science*. 334: 1639.

Partie II – Bibliographie

- Sekercioglu, C. H. et al. 2004. Ecosystem consequences of bird declines. - *Proc. Natl. Acad. Sci.* 101: 18042–18047.
- Shurin, J. B. et al. 2002. A cross-ecosystem comparison of the strength of trophic cascades. - *Ecol. Lett.* 5: 785–791.
- Soulé, M. E. 1985. What Is Conservation Biology? - *Bioscience* 35: 727–734.
- Srivastava, D. S. and Vellend, M. 2005. Biodiversity-Ecosystem Function Research: Is It Relevant to Conservation? - *Annu. Rev. Ecol. Evol. Syst.* 36: 267–294.
- Srivastava, D. and Cadotte, M. 2012. Phylogenetic diversity and the functioning of ecosystems. - *Ecol. Lett.* 15: 637–648.
- Srivastava, D. S. et al. 2012. Phylogenetic diversity and the functioning of ecosystems. - *Ecol. Lett.* 15: 637–48.
- Stuart-Smith, R. D. et al. 2013. Integrating abundance and functional traits reveals new global hotspots of fish diversity. - *Nature* 501: 539–42.
- Thomas, C. D. et al. 2004. Extinction risk from climate change. - *Nature* 427: 145–148.
- Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. - *Proc. Natl. Acad. Sci. U. S. A.* 102: 8245–8250.
- Thuiller, W. et al. 2011. Consequences of climate change on the tree of life in Europe. - *Nature* 470: 531–534.
- Thuiller, W. et al. 2014. The European functional tree of bird life in the face of global change. - *Nat. Commun.* 5: 3118.
- Tilman, D. et al. 1996. Productivity and Sustainability Influenced by Biodiversity in Grassland Ecosystems. - *Nature* 379: 718–720.
- Vane-Wright, R. I. et al. 1991. What to protect?—Systematics and the agony of choice. - *Biol. Conserv.* 55: 235–254.
- Wenny, D. G. et al. 2011. The Need to Quantify Ecosystem Services Provided by Birds. - *Auk* 128: 1–14.
- Whittaker, R. J. et al. 2005. Conservation Biogeography: assessment and prospect. - *Divers. Distrib.* 11: 3–23.
- Winter, M. et al. 2012. Phylogenetic diversity and nature conservation: where are we? - *Trends Ecol. Evol.* 28: 199–204.
- Zupan, L. et al. 2014. Spatial mismatch of phylogenetic diversity across three vertebrate groups and protected areas in Europe. - *Divers. Distrib.* 20: 674–685.

PARTIE III

UTILISER LES DIVERSITÉS PHYLOGÉNÉTIQUES ET FONCTIONNELLES POUR MIEUX COMPRENDRE CERTAINS PATRONS BIOGÉOGRAPHIQUES

SOMMAIRE DE LA PARTIE III

Articles associés à la partie III	172
Annexes associées à la partie III	172
INTRODUCTION GÉNÉRALE	173
<u>III.1. IMPORTANCE DE L'ÉCHELLE PHYLOGÉNÉTIQUE</u>	176
INTRODUCTION SOUS-PARTIE III.1	177
CHAPITRE V. L'environnement et l'espace modulent la répartition mondiale des oiseaux et des mammifères à différentes échelles évolutives	184
CHAPITRE VI. Le régime alimentaire et la phylogénie de l'hôte modulent la composition des micro-biomes à différentes échelles évolutives	198
SYNTHÈSE SOUS-PARTIE III.1	210
<u>III.2. LA GÉOGRAPHIE DE LA CONVERGENCE</u>	216
INTRODUCTION SOUS-PARTIE III.2	217
CHAPITRE VII. La convergence évolutive des assemblages de mammifères	222
SYNTHÈSE SOUS-PARTIE III.2	238
SYNTHÈSE & DISCUSSION GÉNÉRALE	240
BIBLIOGRAPHIE	248

ARTICLES ASSOCIÉS À LA PARTIE III

ARTICLE 7

F. Mazel, R. Wüest, J. Renaud, S. Lavergne and W. Thuiller. *Decomposing the global beta diversity of mammals and birds along a phylogenetic scale reveals the differential effects of space and contemporary climate*. En préparation

ARTICLE 8

M. Groussin*, **F. Mazel***, J. Sanders, C. Smillie, S. Lavergne, W. Thuiller and E. Alm. *Phylogenetic decomposition of mammalian gut microbiota reveals the evolution of community symbiosis*. En préparation

* Contribution égale.

ARTICLE 9

F. Mazel, T. J. Davies, D. Georges, S. Lavergne, W. Thuiller and P. R. Peres-Neto. *Improving phylogenetic regression under complex evolutionary models*. Accepté dans *Ecology*

ARTICLE 10

F. Mazel, R. Wüest, J. Renaud, S. Lavergne and W. Thuiller. *The global biogeography of non-neutral trait evolution in mammals*. En préparation

ANNEXES ASSOCIÉES À LA PARTIE III

ANNEXE 3.1.1. Matériel supplémentaire de l'article 7

ANNEXE 3.1.2. Matériel supplémentaire de l'article 8

ANNEXE 3.2.1. Article 9

ANNEXE 3.2.2. Matériel supplémentaire de l'article 10

INTRODUCTION GÉNÉRALE

DE LA PARTIE III.

Si l'on veut résumer, les grandes questions biogéographiques se concentrent sur :

- **la diversité α** , notamment le gradient latitudinal de diversité
- **la diversité β** , notamment les variations de composition spécifique en fonction des régions du monde.

Je fais l'hypothèse que l'utilisation des facettes phylogénétiques et fonctionnelles de la biodiversité va permettre, en décrivant plus finement la structure des assemblages, de tester des hypothèses précises concernant l'origine de ces deux grands patrons. J'ai notamment donné un exemple d'une telle application pour la compréhension du gradient latitudinal de diversité avec l'étude de Belmaker & Jetz (2015). Dans cette troisième partie, je vais en fait uniquement me concentrer sur les patrons de diversité β en abordant la facette phylogénétique dans la première sous-partie (III.1) et en combinant les facettes phylogénétiques et fonctionnelles dans la deuxième partie (III.2).

Nous avons montré dans le chapitre II l'intérêt potentiel de l'utilisation de l'échelle évolutive pour mieux décrire et comprendre les patrons biogéographiques. En effet, il apparaît clair que différents processus peuvent influencer la répartition de la biodiversité à différentes échelles évolutives, et même que des patrons opposés peuvent exister à différentes échelles. Le chapitre II se concentrait sur la diversité, α mais il est probable que de tels résultats soient observables avec la diversité β . Je m'attache donc à tester cette hypothèse dans deux systèmes d'études emboîtés : les

mammifères à l'échelle du globe (Chapitre V) ainsi que les communautés bactériennes ectosymbiotiques vivant dans l'intestin de certains de ces mammifères (microbiotes* intestinaux, Chapitre VI). En particulier, je teste si les processus liés à l'histoire et ceux liés à la niche agissent à la même échelle évolutive.

L'approche phylogénétique précédente n'explicite pas les différences fonctionnelles entre les espèces. Or, ces différences peuvent permettre de tester de façon plus robuste l'importance des processus de niche dans la distribution de la biodiversité. En effet, la niche écologique des espèces est le reflet de phénotypes* particuliers qui sont adaptés à certaines conditions abiotiques ou biotiques. Ainsi, l'utilisation des traits fonctionnels pourrait permettre de mieux comprendre l'importance relative des processus de niche. Dans cette deuxième sous-partie (Chapitre VII), j'utilise les traits fonctionnels des mammifères pour tester leur relation - à l'échelle de la communauté - avec les caractéristiques environnementales dans une perspective évolutive.

SOUS-PARTIE III.1.

L'IMPORTANCE DE L'ÉCHELLE PHYLOGÉNÉTIQUE

INTRODUCTION DE LA SOUS-PARTIE III.1

« The normal human organism can be said to be composed of over 10^{14} cells, of which about 10% are animal cells. »

Savage, 1977

1. Contexte

Les microbiotes intestinaux sont des communautés bactériennes ectosymbiotiques vivant dans l'intestin de certains organismes et notamment ceux des mammifères (Savage 1977). Ces communautés forment des systèmes complexes qui se révèlent d'une importance capitale pour certaines fonctions physiologiques fondamentales de l'hôte, comme sa nutrition ou sa réponse immunitaire, et participe donc à sa fitness (Hooper & Gordon 2001; Bäckhed *et al.* 2005; Ley *et al.* 2008; Hooper *et al.* 2012). Par exemple, les (cellules animales des) herbivores n'ont pas les capacités biochimiques de digérer la cellulose qui est pourtant le composant majeur de leur régime alimentaire. Par contre, les bactéries de leur microbiote possèdent quant à elles les enzymes capables de métaboliser ce sucre. Ainsi, l'herbivore dépend directement de la fermentation assurée par son microbiote pour acquérir les ressources nécessaires à sa survie.

Il m'apparaît donc intéressant de comprendre l'évolution et la structure des microbiotes intestinaux de mammifères tout autant que l'évolution et la structure des assemblages de mammifères. Par ailleurs, nous allons voir qu'au-delà de cet aspect intégratif, la biogéographie et l'étude des microbiotes intestinaux posent en fait des questions proches et utilisent des méthodes quasi similaires.

2. Les parallèles conceptuels et théoriques entre l'étude des relations hôtes-symbiontes et la biogéographie

La biogéographie, qu'elle soit historique ou écologique, entretient de nombreux liens conceptuels avec l'étude des relations hôtes-symbiontes (ou parasites).

2.1. Biogéographie historique, relations hôtes-parasites et évolution des génomes

« Evolutionary associations among genes, organisms and geographical areas have traditionally been studied by biologists from different disciplines, with little interaction between them. »

Page & Charleston, 1998

La description et la compréhension de la dynamique évolutive d'un système à deux « partenaires » avec l'un des deux partenaires « emboîtés dans » ou « suivant » (« tracking ») l'autre est un problème général qui trouve des applications dans différentes disciplines. Ainsi, l'étude de l'évolution du couple parasite-hôte présente des similitudes importantes avec l'étude de la biogéographie historique (p. ex. Brooks 1981), tout comme l'étude de l'évolution du couple gène/espèce (p. ex. Doyle 1992). En fait, les trois disciplines (biogéographie historique, étude des relations hôtes-symbiontes (ou parasites) et évolution des génomes) utilisent les mêmes objets conceptuels avec des méthodes très proches pour répondre à des questions similaires (Page & Charleston, 1998). En effet, on peut s'apercevoir, en s'appuyant l'encadré 4, que les trois disciplines manipulent toutes deux objets et quatre concepts qui prennent des formes et des noms différents mais qui sont similaires. Simplement, le système d'étude est radicalement différent ce qui provoque l'émergence de disciplines différentes et relativement cloisonnées, nuisant clairement au développement de chacune d'entre elles (Page & Charleston, 1998). Heureusement, la reconnaissance de la similarité conceptuelle entre les trois disciplines semble aujourd'hui mieux établie (Page & Charleston 1998; Felsenstein 2004).

2.2. Biogéographie écologique et microbiologie

Par ailleurs, il apparaît aussi de plus en plus que la microbiologie en général (et l'étude des relations hôtes-symbiotes ou hôtes-parasites en particulier) peut utiliser une approche de biogéographie écologique (ou d'écologie des communautés). Le cadre conceptuel et théorique proposé par Vellend (2010) pour l'écologie des communautés est en fait fort utile en microbiologie (Nemergut *et al.* 2013) et en particulier dans l'étude des microbiotes (Costello *et al.* 2012). Par exemple, dans le cas du microbiote humain, Costello *et al.* (2012) rapprochent des concepts venant des deux disciplines comme le couple perturbation/prise d'antibiotiques, ou utilise les concepts écologiques de méta-communauté ou d'invasion pour tenter de mieux comprendre l'assemblage du microbiote humain. Les parallèles entre les deux disciplines sont donc forts, mais il existe tout de même des zones encore mal explorées, notamment l'influence de l'histoire ancienne (*i.e.* la co-évolution) sur l'assemblage des microbiotes actuels.

3. L'apport des diversités à plusieurs facettes en microbiologie

Dans cette thèse, je me suis demandé comment les indices de diversité phylogénétique et fonctionnelle pouvaient participer au rapprochement de la biogéographie écologique et de la biogéographie historique, notamment pour parvenir à une approche unifiée et une compréhension élargie des patrons de diversités (Wiens & Donoghue 2004; Cavender-Bares *et al.* 2009, 2012). Je pense qu'un tel rapprochement est aussi nécessaire dans l'étude des microbiotes, et que les indices de diversité à plusieurs facettes peuvent jouer un rôle clef. En fait, on peut s'apercevoir :

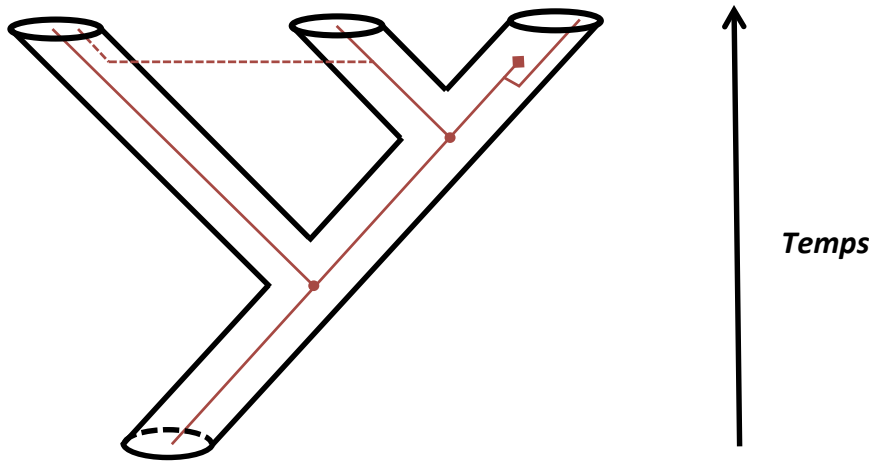
- (1) que de tels indices sont très utilisés en microbiologie. En effet, comme l'atteste clairement la figure 4 de l'introduction générale, près de 30% des articles qui utilisent le terme « diversité phylogénétique » sont classés dans la catégorie « microbiologie »
- (2) qu'une part importante des indices de diversité phylogénétique ont en fait été créés par des microbiologistes (voir le matériel supplémentaire 1 de l'article 1)

Ainsi, les conclusions de la partie I sont aisément transposables à l'étude microbiologique et des microbiotes. Dans cette optique, je propose dans le chapitre VI d'étudier la composition du microbiote de 33 espèces de mammifères.

4. Stratégie de travail

Dans cette sous-partie III.1, je vais me focaliser sur la compréhension de deux systèmes emboîtés : la répartition géographique des mammifères sur la terre et la répartition des bactéries au sein des intestins de différentes espèces de mammifères. Bien que très différents en terme d'échelle spatiale, ces systèmes sont tous les deux complexes, car ils comprennent un grand nombre de lignées. En ce sens, je pense que le cadre méthodologique et conceptuel du chapitre II, qui propose de décomposer le long d'une échelle phylogénétique la structure des assemblages, est adapté aux deux systèmes. En particulier je vais tester dans les deux cas si les processus proposés dans la littérature, et parfois considérés comme mutuellement exclusifs, n'agissent en fait pas à différentes échelles phylogénétiques.

A. Schéma général



B. Détails des objets et processus en fonction des disciplines

Domaines de recherche		Biogéographie historique	Interactions hôtes parasites ou symbiontes	Evolution des génomes	Légende
Objets d'étude		Aire - Espèce	Espèce – Symbionte	Espèce - Gène	
Processus	Transfert horizontal	Dispersion	Transfert d'hôte	Transfert de gène	
	Duplication	Spéciation sympatrique	Spéciation dans l'hôte	Duplication génique	
	Co-séparation	Spéciation par vicariance	Co-spéciation	Spéciation	
	Extinction	Extinction	Extinction	Perte de gène	

Encadré 4. Parallèles entre biogéographie, parasitologie et génomique.

Le document A présente le schéma général d'évolution dans le temps des objets d'études (Aire, Espèce, Gènes) accompagné de l'illustration des quatre processus fondamentaux (transfert horizontal, duplication, co-séparation et extinction). Le document B précise les termes employés pour décrire les objets et les processus dans trois champs disciplinaires différents, mettant ainsi en évidence leurs similarités conceptuelles. La légende du document A est associée au document B pour plus de clarté.

CHAPITRE V

L'ENVIRONNEMENT ET L'ESPACE MODULENT LA RÉPARTITION MONDIALE DES OISEAUX ET DES MAMMIFÈRES À DIFFÉRENTES ÉCHELLES PHYLOGÉNÉTIQUES

Résumé

La compréhension des processus responsables de la distribution des espèces à large échelle spatiale est un enjeu majeur en écologie. Classiquement, on distingue les processus de type historique (dispersion et diversification) et des processus de type écologique (essentiellement des processus de niche). Dans cette analyse, nous développons une nouvelle méthode, appelée BDTT (Beta-Diversity through Time) qui nous permet de montrer que les deux processus agissent en fait à des échelles phylogénétiques différentes. L'histoire semble moduler la distribution géographique des branches profondes alors que l'environnement paraît être plutôt important à fine échelle phylogénétique. De façon intéressante, nous montrons que de tels résultats sont vrais pour les mammifères, mais non pour les oiseaux, ce qui peut s'expliquer par la plus grande capacité de dispersion de ces derniers. En résumé, nous montrons à quel point il est important de faire varier l'échelle phylogénétique en biogéographie afin de séparer les effets de différents processus.

Decomposing the global beta diversity of mammals and birds along a phylogenetic scale reveal the differential effects of space and contemporary climate

F. Mazel, R. O. Wüest, J. Renaud, S. Lavergne and W. Thuiller

Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France.
CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France

Introduction.

The complex structure of past and present-day geographical distribution of species and lineages has long fascinated naturalists (Humbolt and Bonpland 1805, Darwin 1859, Wallace 1876). Generally, species distributions are considered to be the outcome of three interrelated processes: historical events (such as dispersal or vicariance, e.g. Simpson 1980; Sanmartín & Ronquist 2004), environmental filtering (e.g. Root 1988a; b) and biotic interactions (e.g. Pedersen *et al.* 2014; Silvestro *et al.* 2015). Nevertheless, jointly studying these processes has been difficult, notably because their importance vary among spatial and/or phylogenetic scale (Rosenzweig 1995, Lomolino *et al.* 2010). In particular, different processes might shape lineage distribution at different phylogenetic scales but we lack understanding on how deep time events and contemporary climate relate to present day distributions (Webb *et al.* 2002, Cavender-Bares *et al.* 2006).

We might expect deep biogeographical events such as dispersal or vicariance to act on the distribution of higher taxonomic levels at a deep phylogenetic scale (e.g. orders, families, etc.), while contemporary environment shape the distribution of genera and species. This view is particularly well supported on island systems where adaptive radiations have been observed. In a island radiation, populations of a given ancestor species reaches a newly created archipelago from the mainland and subsequently diversifies in several lineages. This is the case for example in Hawaï with the tetragnatha spiders (Gillespie 2004) and silverword alliance (Baldwin and Sanderson 1998) but also in less remote islands such as the caribbean with the anoles lizard (Losos 2009) and in south America fossils records with the so-called notoungulates (Simpson 1980). This means that between mainland and island assemblages, spatial turnover of high taxonomic levels is the product of dispersal whereas that of lower taxonomic levels results from adaption to different environments. For mammals and birds, we

might expect deep biogeographical events to have been shaped by historical connectivity between landmasses that depends on climate configurations and plate tectonics.

On the contrary, one could expect biogeographical events (such as long distance dispersal) to act at a recent phylogenetic scale and environment or habitat type to shape the distribution of higher lineages. This would imply that an ancient niche divergence promoted the emergence of two lineages adapted to different environment. Then, the diversification within each of the climatic regimes could have been driven by allopatric diversification coupled with strong niche conservatism (Crisp and Cook 2012). This has been shown for example, for swallowtail butterflies where the two main clades inhabit different climatic regimes (temperate Vs. tropical, Condamine *et al.* 2012).

Understanding spatial patterns of biodiversity have usually relied on the magnitude and shape of the relationship between single β diversity measure such as species or phylogenetic turnover (compositional and phylogenetic β diversity, CBD and PBD, Whittaker 1960; Graham & Fine 2008; Anderson *et al.* 2011) and environmental and spatial distances. While relatively straightforward and easy to understand, the use of such synthetic indices may in fact limit our understanding of the processes shaping lineages distribution. Indeed, if different processes act at different phylogenetic scales, the use of a single index of turnover between assemblages will mix different independent signals and possibly blur the overall pattern. Moreover, while some studies have examined the beta-diversity through a relatively crude taxonomic scale (i.e. analysing beta-diversity of species, genus and families, see for example Qian 2009; Kreft & Jetz 2010), very few have used the phylogenetic information to decompose the spatial variation of lineages composition.

To overcome these limits, we propose to decompose the classical measure of phylogenetic β diversity along the phylogenetic scale (Beta Diversity through time,

BDTT). Instead of a single value of turnover, we thus obtain a profile of turnover through evolutionary time (BDTT profile, see Fig. 1 and Rosauer *et al.*, 2014 for a related approach). This BDTT profile thus starts with the CBD (i.e. turnover of the tips of the phylogenetic tree) and progressively goes back up in the hierarchic phylogenetic classification (see Fig. 1). PBD then relates to the mean of the BDTT profile (Fig 1.C), i.e. it represents a synthetic measure of the profile, as it is averaged over the whole phylogenetic time scale. This new approach may reveal the differential effect of space and environment.

More specifically, we decomposed the Phylogenetic version of the classical Simpson index (Simpson 1943, Leprieur *et al.* 2012, Holt *et al.* 2013) to better describe and understand the relative importance of space and contemporary environment in present day structure of birds and mammals assemblages worldwide. We used available species distributions coupled with phylogenetic trees for birds and mammals to ask three main questions.

First, are geographical and environmental distances shaping bird and mammal distributions at the same phylogenetic time scale? Second, are these relative influences variable within each of the two groups, i.e. is there a good congruence among orders or not? Third, do past geographical distances explain the turnover of deep lineages better than present-day isolation? We hypothesise that if vicariance events (or isolation because of geographic distance) have been the major drivers of range evolution, deep lineages turnover should be more related to past geographical distances. On the contrary if long-distance dispersal has been predominant in the past, independently of past geographic distances, we should not see any effect of past geographic information on BDTT profile.

Results and Discussion

Relative effects of space and environment

At the scale of this analysis and for both birds and mammals species, true turnover was significantly positively related to broad environmental and spatial distances (Fig. 2). The relationship between species turnover and spatial distance –the so called distance decay relationship– is a well studied pattern in (macro)ecology (Rosenzweig 1995) and has been already reported at this spatial scale (e.g. Qian & Ricklefs 2012). This significant relationship explains why species assemblages can be grouped in biogeographic realm that harbour clear spatial organisation (Wallace 1876, Kreft and Jetz 2010, Holt

et al. 2013, Vilhena and Antonelli 2015) and have been suggested to be –at least partly– the product of historical contingencies, i.e. vicariance events or dispersal. The positive significant relationship between species turnover and environmental distances has also been found at this scale (e.g. Buckley & Jetz 2008) and may correspond to the signature of environmental filtering. Overall, the fact that space better explained species turnover than environment for the two groups (Fig. 2) could reflect the important effect of historical events Vs. species differentiation across climatic gradients. In principle, we can't rule out the possibility that our spatial distance hide other environmental distances (e.g. Anderson *et al.* 2011).

We then analysed turnover of higher lineages using the branching pattern of the phylogenetic tree of each group (see principle in Fig. 1). This time-calibrated branching pattern is somewhat linked to the classical taxonomic rank assignment in genus, families and orders (see distributions of taxonomic rank ages on the phylogenetic scale in Fig. 2 and for example Qian 2009; Kreft & Jetz 2010). Nevertheless the congruence is relatively low and we believe our approach is thus more transparent and less arbitrary. In fact, decomposing lineage diversity through evolutionary time reveals the differential impacts of space and environment on lineage distributions. Across the two groups, we found that the pure spatial effect was generally hump-shaped, although much more for mammals than for birds, meaning that geographical distances relatively better predict the turnover of deep lineages than the turnover of species (Fig. 2). These profile do not only derive from the observed species beta diversity and the phylogenetic tree shape but rather represent a non-random spatial distribution of higher lineages in the phylogeny (compare observed profile and null 95% CI envelope in blue on Fig. 2). On the contrary, environmental effects appeared to be monotonically decreasing through time and becoming rapidly non-significant, meaning that environmental distance better predict the turnover of species than deeper lineages (Fig 2). Nevertheless, this decreasing profile is not as steep as one could expect from the observed species beta diversity and the structure of the phylogenetic tree only, meaning that the turnover of certain relatively deep lineages is linked to broad environment gradient as well (compare observed profile and null 95% CI envelope in orange on Fig. 2). All these results were robust to tree uncertainties (compare multiple observed profiles in Fig. 2), to the relaxation of the linear hypothesis

between faunal turnover and space or environment and to our stratified sample design (Supp. Mat. 1).

These relative trends could be the result of old dispersal events followed by diversification across major climatic gradients. There are many examples of such sequences of processes, for example lemur primates only occur in Madagascar due to an ancient dispersal event and diversified within the different climatic regime of the island (Ganzhorn et al. 2006). Overall our findings demonstrate that environment and space do not influence lineage distributions at the same phylogenetic scale and thus shed new lights on the spatial structure of biodiversity. Because this structure is a highly complex system, biogeographers and macro-ecologists have to describe it with some synthetic, tractable measures (e.g. beta-diversity indices). This simplification then allows to handle a relatively few number of variables against which hypotheses can be tested. Our approach lies somewhat between studies that focus on one single parameter (i.e. that studies one single beta diversity values) and those that study all lineages separately (e.g. Borregaard *et al.* 2014). Nevertheless we assumed that spatial and environmental effects could act differently along the phylogenetic scale but not among different parts of the tree, thus neglecting heterogeneity within the tree.

Comparing BDTT between mammals and birds and among orders revealed interesting trends. The fact that environmental profiles were similar for birds and mammals is consistent with the deterministic nature of niche processes and similar niche conservatism strength. On the contrary, deep-time faunal similarities were much better explain by geographical distances (than species similarities) in mammals than in birds. This is consistent with the higher dispersal abilities of birds that possibly erased, over evolutionary time, the potential spatial signal of in-situ diversification.

We found that these overall trends hide a diversity of response across the different parts of each of the tree (Supp. Mat. 2). Decomposing the turnover within orders showed variable trends, notably that orders with a clear disjunction patterns show clear BDTT patterns (Supp. Mat. 2). For example, the strength of the relationship between primate turnover and space show a clear increase from species turnover to higher clades, a pattern that matches the clear geographical disjunction of some lineages of primates (*e.g.* new vs. old world monkeys). Also, it appears that the spatial effect was more variable through time than the environmental one, a pattern that could be explained by the somewhat contingent nature of dispersal.

Overall we showed that the pure spatial effect had an important imprint on the turnover of lineages at all phylogenetic scales up to 65 Ma for the two clades. These geographic patterns can be explained by historical dispersal events between the closest landmasses, such as the arrivals of rodents (family Caviomorpha) in South America via long distance dispersal from Africa around 35 Ma (Sallam et al. 2009). But the important question is rather to which extent these dispersal events are predictable or not. Dispersal is generally thought to be linked to (1) species dispersal abilities that are mediated by functional traits (Sanmartín and Ronquist 2004, Whitmee and Orme 2013) and to some measure of distance between two regions (de Queiroz 2005, Gillespie et al. 2012). As a consequence, we hypothesise that ancient dispersal probability between two landmasses should be linked to (1) past connectivity between the landmasses and (2) dispersal abilities of the studied taxa.

The imprint of tectonics

We tested this hypothesis by reconstructing the paleo-coordinates of the studied grid cells and testing their ability to explain faunal turnover at deep phylogenetic scales. We used past crow-fly geographic distance as a proxy of past historical connectivity between regions. Our hypothesis is that relatively close grid cells in the past may have been more prone to dispersal events than distant grid cells. In this case, we expect higher deep-time similarities between grid cells relatively more connected in the past. On the contrary, grid cells that have been more isolated from each other in the past have probably less experienced dispersal events and thus may harbor lower deep-time faunal similarities.

Historical proximity was significantly related to deep faunal turnover in mammals but not in birds, supporting the hypothesis that historical proximity is more important for taxa with lower dispersal abilities, because the high dispersal would erase the signal of in-situ diversification (Fig 3 and Supp Mat. 3). It is important to note that while our results could be theoretically explained by repeated vicariance events, it seems in fact highly unlikely. Indeed, the breakup of pangea started well before the cenozoic era so as vicariance events could only occur before the K/T boundary, i.e. in the Mesozoic, while the data we analyse here is restricted to the Cenozoic. As a consequence, we suggest that the relationship we found is rather the product of relatively long-distant dispersal constrained by historical proximity between

landmasses. It is surprising to find such a clear relationship with a relatively naïve proxy of historical connectivity. In fact, we found that the overall pattern was mostly driven by the particular case of Australia. Indeed, when this continent was removed from the analysis, the congruence between past distance and deep faunal turnover vanished (Supp. Mat. 4). Australia represents an important zoogeographic realm because it contains a very unique set of species and thus show high turnover with the rest of the world (Kreft & Jetz 2007; Holt *et al.* 2013 and see Supp. Mat. 5.1). In absolute terms, the turnover of higher lineages is necessarily lower than the turnover of species due to the hierarchical structure of the phylogenetic tree. Nevertheless, it appears that Australia shows a relatively higher faunal isolation of higher lineages (Supp. Mat. 5.2-6). Because Australia was geographically more isolated –relatively to other landmasses in the world– in the past, it explains why Australia drives the overall pattern of congruence between BDDT and past distances we found. Except from the particular case of Australia, we did not find a signature of links between past geographical proximity and turnover of deep lineages. This could be explained by the real uncoupling of the two processes or by two main methodological limits of our approach when it is linked to past connectivity. First, our approach can't directly disentangle the signal of recent and ancient dispersal events. For example, in the case of mammals, the Great American Biotic Interchange brought only recently new orders in South America (i.e. in the last 15 Ma, Webb 2006; Woodburne 2010 but see Bacon *et al.* 2015), for example cetartiodactyls or carnivoras. Thus, because of these recent events, turnover of deep lineages (such as the stem branches of these orders) is reduced between North and South America and does not represent the turnover of lineages that were only restricted to South America at this given time slice. Also, while being isolated for most of its history (Simpson 1980, but see Bacon *et al.* 2015), South America surely harboured in the past a rather unique set of high taxonomic categories (Simpson 1980), but the joint effect of recent migration (i.e. the great American biotic interchange) coupled with high extinction rates of native lineages (Simpson 1980) removed the signal of important deep lineages turnover. On the contrary, Australia did not undergo secondary connections with other continents followed by massive migration events, thus retaining its unique deep lineages.

Second, our proxy for historical connectivity is in fact a raw measure of past geographical proximity. It is

probably oversimplified because dispersal do not necessarily occurs in a straight line but rather depends on the geographical structure of habitat (i.e. connectivity). For example, two pairs of grid cells with the same geographical distances may in fact harbour very different connectivity if for example, one pair is separated by ocean while the other pairs is separated by suitable habitat. Future studies may try to test alternative measures of connectivity, such as least cost distances considering the different barrier effects of seas or landmasses with unfavourable environmental conditions (e.g. deserts, mountain chains). An interesting avenue would also be to take into account the past distribution of climates to track back the potential 'corridor' of suitable climate for lineages trough time.

Conclusion

The importance of phylogenetic scale decomposition in the studies of diversity patterns have long been emphasized in the context of lineage richness (e.g. O'Brien *et al.* 1998, 2000; Qian & Ricklefs 2007; Qian *et al.* 2009; Terlizzi *et al.* 2009; Kreft & Jetz 2010) but have not been carried out in the context of a calibrated time-scale and faunal composition (i.e. beta diversity patterns). Here we show that such decomposition reveal the differential effect of geographical and environmental distance on faunal turnovers. We found that while environment better explain turnovers of recent lineages (i.e. species), geography better predict deep faunal turnover (e.g. genus or families). This could suggest that biogeography impact the distribution of broad lineages while environment select for finer differences. We further show that the impact of geography on the distribution of broad lineages is shaped by species dispersal abilities and historical proximity between regions.

Acknowledgements

The research leading to these results had received funding from the European Research Council under the European Community's Seven Framework Programme FP7/2007-2013 Grant Agreement no. 281422 (TEEMBIO). All authors belong to the Laboratoire d'Écologie Alpine, which is part of Labex OSUG@2020 (ANR10 LABX56).

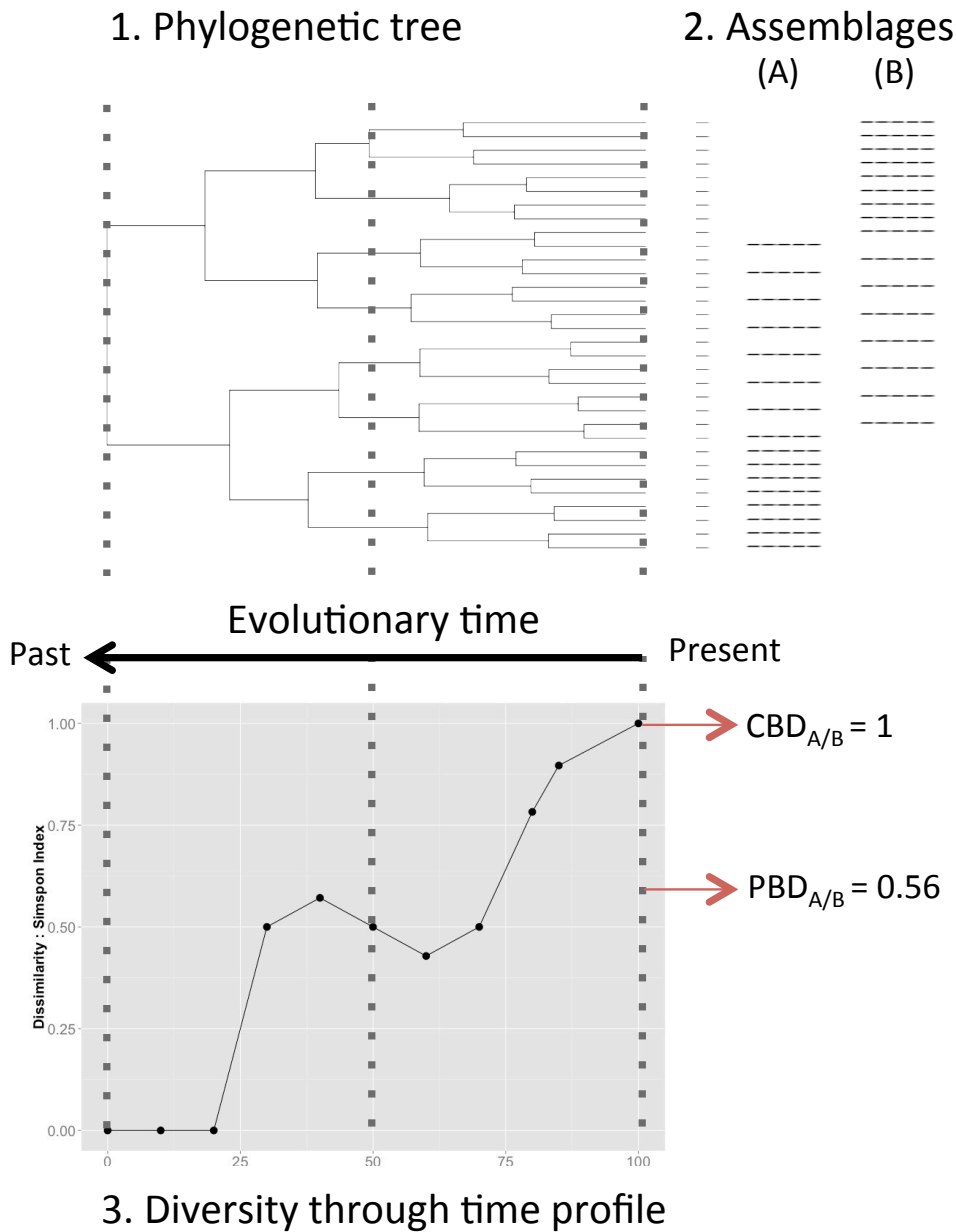


Figure 1. Theoretical example of the β diversity through time (BDTT) approach. The figure presents a (1) phylogenetic tree associated with (2) two assemblages (A and B) and (3) the corresponding compositional β -diversity (CBD), phylogenetic β -diversity (PBD) and BDTT measures between A and B. The phylogenetic tree (1) contains 16 species whose presence in assemblage A and/or B are depicted by a line in front of each tip (2). Below the tree is depicted the BDTT profile along the same time scale with CBD and PBD given for information.

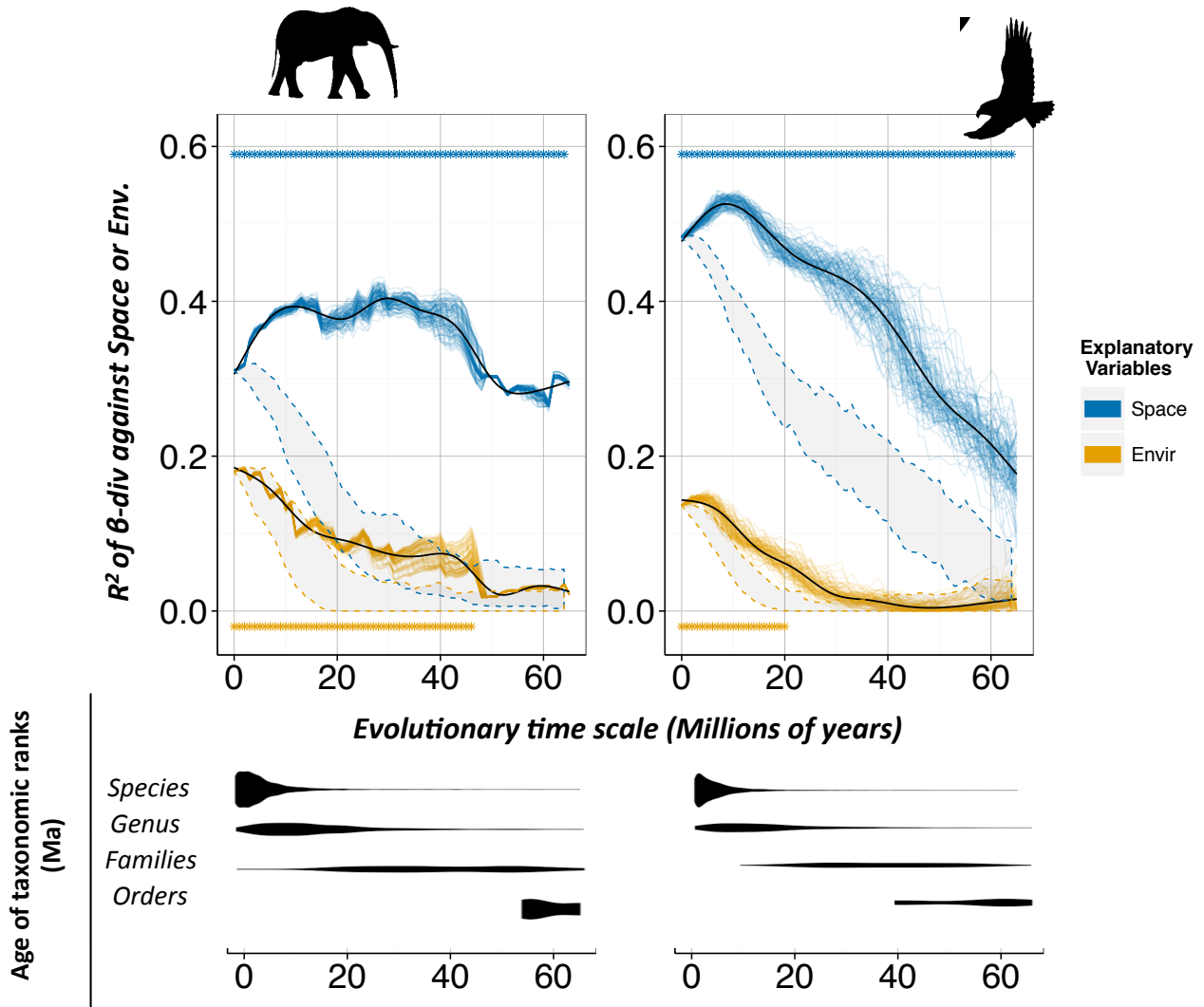


Figure 2. The relative effect of spatial and geographical distances on the turnover of lineages through time. The figure presents, for birds and mammals separately, the variation of the strength of the Pearson correlation (Y-axis) between faunal turnover and geographic or environmental distances along the phylogenetic scale at which faunal turnover is defined (X-axis) based on the stratified sample. Multiple profiles of correlation refer to alternative phylogenetic trees. Black line corresponds to a smoothed fit. The shaded area corresponds to the 95% confidence interval of a null model that randomizes phylogenetic relationships but keeps species composition constant. The distribution of stem ages of four classical taxonomic ranks (species, genus, families and orders) is given along the time-scale used to define faunal turnover. *:significant correlations, i.e. for a given slice, more than 90% of the observed phylogenetic decompositions support a significant relationship at the 5% level.

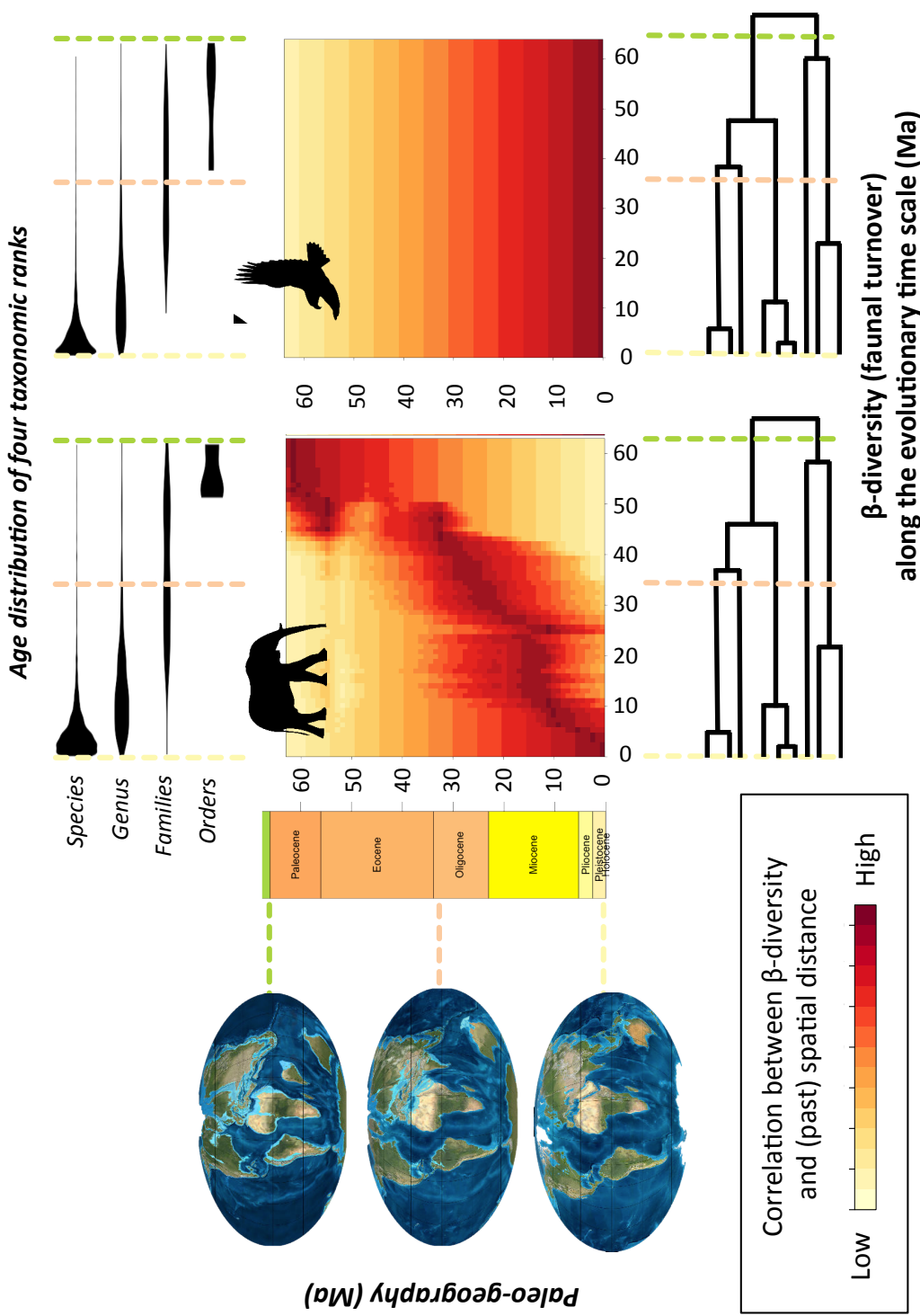


Figure 3. The historical imprints of past continental position on BDTT. The figure presents, for birds and mammals separately, the variation of the strength of the correlation between faunal turnover and geographic distances (see colours in the legend) along the phylogenetic scale at which faunal turnover is defined (X-axis) and the date at which geographical distance is computed (Y-axis). More specifically, for each evolutionary time slice where turnover is computed (X-axis), we rank past geographical distances according to their variance explanation. Ranks presented are the median over phylogenetic 100 trees with red colour indicating a relative better fit. Paleogeographic positions are depicted near the Y-axis (based on Scotese maps). The distribution of stem ages of four classical taxonomic ranks (species, genus, families and orders) is given along the time-scale used to define faunal turnover. An illustration of the method used to compute BDTT is given below the X-axis. Phylogenetic trees depicted below the plots, as well as the three coloured lines, are given here for illustration purpose and do not correspond to the real mammal and bird trees, nor any specific time slice used for computations, respectively.

Material and Methods

Distribution data

We derived occurrence data for birds and mammals on worldwide grid cells of approximately 200*200 km. For mammals, we used the distribution maps provided by the Mammal Red List Assessment (<http://www.iucnredlist.org/>) for 4616 mammals. For birds, winter ranges distribution maps were extracted from BirdLife (<http://www.birdlife.org/>) for 9993 species. The best resolution to use these maps is still under discussion in the literature (Storch et al. 2012, Jenkins et al. 2013) so we used a resolution commonly used at global scale (e.g. Safi *et al.* 2011; Holt *et al.* 2013). Domestic, aquatic and semi aquatic mammals were excluded from the analysis.

Stratified sample

Because of computational constraints, we simplified some part of our analysis by subsampling 200 grid cells out of 3600. In order to correctly sample the environmental (summarized by the two first axis of a PCA on the global worldclim variables) and the geographical space we used the cube method (Deville 2004) based on PCA axis scores and geographical position of each grid cells (see Supp. Mat. 6).

Phylogenies

For mammals, we used the 100 calibrated and dated ultrametric phylogenetic tree updated by Fritz et al. (2009) from Bininda-Emonds et al. (2007). For Birds, we used the first 100 phylogeny proposed by Jetz *et al.* (2012).

β diversity metrics

Metrics

The species beta-diversity or dissimilarity between two assemblages is generally decomposed into two distinctive components (Baselga 2010). The first component quantifies the degree of replacement of species between two sites and is often called the 'true turnover' component or 'spatial turnover' (Gaston and Blackburn 2008, Baselga 2010). This component is independent of species richness difference between assemblages and is the component of choice to delineate zoogeographic regions (Kreft and Jetz 2010, Holt et al. 2013, Mouillot et al. 2013). Here we choose to use to decompose the classical Sorensen metric (Sørensen 1948) using the framework proposed by Baselga (2010).

The general Sørensen dissimilarity metric can be written as:

$$\text{Eq.1} \quad \beta_{sor} = \frac{b+c}{a+b+c}$$

and is decomposed in a 'true turnover' component - initially proposed by Simpson (1943)-:

$$\text{Eq.2} \quad \beta_{sim} = \frac{\min(b,c)}{a+\min(b,c)}$$

where a = the number of species shared by the two grid cells and b and c represent the number of species unique to each grid cell.

The Sørensen metric is tightly close to the Jaccard index (Jaccard 1901) and both are commonly used in macro-ecology (e.g. Buckley & Jetz 2008; Qian & Ricklefs 2012; Mouillot *et al.* 2013). The decomposition of beta-diversity metrics into their two distinct components have been some matters of debates (Baselga 2010, 2012, Podani et al. 2013, Mouillot et al. 2013, Leprieur and Oikonomou 2014, Legendre 2014, Baselga and Leprieur 2015). We choose to use the decomposition proposed by Baselga (2010) because its 'true turnover' component is not mathematically constrained by difference in species richness (Baselga and Leprieur 2015), a key property not fulfilled by other proposed frameworks.

Decomposition through time

For a given metric (e.g. β_{sor} or β_{sim} see Eq.1-2), classical measures typically rely on a single number to describe the turnover between two assemblages. This number may reflect species dissimilarities (Baselga 2010) or branch length dissimilarities if a phylogeny is integrated (Leprieur *et al.* 2012). Nevertheless, relying on a single number to describe assemblage dissimilarities may be a problem if different processes act at different phylogenetic scale because the use of a single index will mix different independent signals and possibly blur the overall pattern. We thus propose to study a profile of dissimilarity through time (see Fig. 1). More specifically, we trace back the lineages composition similarities between grid cells for each time slice along the phylogenetic tree. This decomposition thus separates the turnover of recent (e.g. species) vs. deep (e.g. genus or families) lineages.

Environmental and geographical data

For environment, we used the Euclidian distance on the two first axis of a PCA of worldclim variables. These axes mostly represent effect of temperature, precipitation and seasonality. For geographical distance we simply used the length of the shortest strait line between grid cells. To derive past spatial distance between our grid cells, we used the free software GPLATE (Boyden et al. 2011, Williams et al. 2012). Basically, it uses the movement of static polygons (parts of continents) around its corresponding Eulerian pole to reconstruct the past geographic position of the continents and oceans. Details concerning the original data source and synthesis can be found in Seton *et al.* (2012). We obtained for each millions year, the past coordinates of the centres of our grid cells and derive corresponding distance matrices.

Link between β diversity profile and geography and environment

To link the turnover profiles to environmental and geographical distance between grid cell we used multiple regression on distance matrices (MRM, Lichstein 2006) with randomisation test to assess significance. We then used variance partitioning (Legendre 2007) to extract the pure and shared effects of environment and space. We used the pure space effect and the total effect for environment (pure environment + shared affect).

Null model

To understand the impact of species turnover on higher lineage turnover, we shuffled species identity on the whole phylogenetic trees and recomputed the BDTT profile and its correlation to environment and space. This null profile of correlation enables us to evaluate the impact of the tree structure itself (independent of the geography) on our interpretations. Notably, if species turnover appears significantly linked to environmental or geographical distances, we expect that turnover of deeper lineages may also be significant because of the hierarchical structure of the phylogenetic tree. As a consequence, the comparison of the observed and null profile of correlations enables us to disentangle biological from statistical effects. The null model we choose is commonly used to compare CBD and PBD (e.g. Graham & Fine 2008; Graham *et al.* 2009; Leprieur *et al.* 2012; Weinstein *et al.* 2014).

References

- Anderson, M. J. et al. 2011. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. - *Ecol. Lett.* 14: 19–28.
- Bacon, C. D. et al. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. - *Proc. Natl. Acad. Sci. U. S. A.* 112: 6110–5.
- Baldwin, B. G. and Sanderson, M. J. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). - *Proc. Natl. Acad. Sci.* 95: 9402–9406.
- Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. - *Glob. Ecol. Biogeogr.* 19: 134–143.
- Baselga, A. 2012. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. - *Glob. Ecol. Biogeogr.* 21: 1223–1232.
- Baselga, A. and Leprieur, F. 2015. Comparing methods to separate components of beta diversity (M. Spencer, Ed.). - *Methods Ecol. Evol.*: n/a–n/a.
- Bininda-Emonds, O. R. P. et al. 2007. The delayed rise of present-day mammals. - *Nature* 446: 507–12.
- Borregaard, M. K. et al. 2014. Node-based analysis of species distributions (RB O'Hara, Ed.). - *Methods Ecol. Evol.* 5: 1225–1235.
- Boyden, J. A. et al. 2011. *Geoinformatics* (GR Keller and C Baru, Eds.). - Cambridge University Press.
- Buckley, L. B. and Jetz, W. 2008. Linking global turnover of species and environments. - *Proc. Natl. Acad. Sci. U. S. A.* 105: 17836–17841.
- Cavender-Bares, J. et al. 2006. Phylogenetic structure of floridian plant communities depends on taxonomic and spatial scale. - *Ecology* in press.
- Condamine, F. L. et al. 2012. What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. - *Ecol. Lett.* 15: 267–277.
- Crisp, M. D. and Cook, L. G. 2012. Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? - *New Phytol.* 196: 681–94.
- Darwin, C. 1859. The origin of species by means of natural selection, or, The preservation of favored races in the struggle for life. (J Murray, Ed.). - The Werner Company.
- De Queiroz, A. 2005. The resurrection of oceanic dispersal in historical biogeography. - *Trends Ecol. Evol.* 20: 68–73.
- Deville, J.-C. 2004. Efficient balanced sampling: The cube method. - *Biometrika* 91: 893–912.
- Fritz, S. A. et al. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. - *Ecol. Lett.* 12: 538–49.
- Ganzhorn, J. U. et al. 2006. Lemur Biogeography. - In: *Primate biogeography. Developments in Primatology: Progress and Prospects.* Springer US, pp. 229–254.
- Gaston, K. and Blackburn, T. 2008. *Pattern and Process in Macroecology.*
- Gillespie, R. 2004. Community assembly through adaptive radiation in Hawaiian spiders. - *Science* 303: 356–9.
- Gillespie, R. G. et al. 2012. Long-distance dispersal: a framework for hypothesis testing. - *Trends Ecol. Evol.* 27: 47–56.
- Graham, C. H. and Fine, P. V. A. 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. - *Ecol. Lett.* 11: 1265–77.
- Graham, C. H. et al. 2009. Phylogenetic structure in tropical hummingbird communities. - *Proc. Natl. Acad. Sci. U. S. A.* 106 Suppl: 19673–19678.
- Holt, B. G. et al. 2013. An Update of Wallace's Zoogeographic Regions of the World. - *Science* (80-.). 339: 74–78.
- Humboldt, A. von and Bonpland, A. 1805. *Essai sur la géographie des plantes: accompagné d'un tableau physique des régions équinoxiales, fondé sur des mesures exécutées, depuis le dixième degré de latitude boréale jusqu'au dixième degré de latitude australe,*

Partie III.1 – Chapitre V

- pendant les années 1799, 1800, 1801., - Levrault Schoell, Paris.
- Jaccard, P. 1901. Etude comparative de la distribution florale dans une portion des Alpes et du Jura. in press.
- Jenkins, C. N. et al. 2013. Global patterns of terrestrial vertebrate diversity and conservation. - *Proc. Natl. Acad. Sci. U. S. A.* 110: E2602–2610.
- Jetz, W. et al. 2012. The global diversity of birds in space and time. - *Nature* 491: 444–448.
- Kreft, H. and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. - *Proc. Natl. Acad. Sci. U. S. A.* 104: 5925–5930.
- Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species distributions. - *J. Biogeogr.* 37: 2029–2053.
- Legendre, P. 2007. Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. - *J. Plant Ecol.* 1: 3–8.
- Legendre, P. 2014. Interpreting the replacement and richness difference components of beta diversity. - *Glob. Ecol. Biogeogr.* 23: 1324–1334.
- Leprieur, F. and Oikonomou, A. 2014. The need for richness-independent measures of turnover when delineating biogeographical regions (R Ladle, Ed.). - *J. Biogeogr.* 41: 417–420.
- Leprieur, F. et al. 2012. Quantifying phylogenetic beta diversity: distinguishing between “true” turnover of lineages and phylogenetic diversity gradients. - *PLoS One* 7: e42760.
- Lichstein, J. W. 2006. Multiple regression on distance matrices: a multivariate spatial analysis tool. - *Plant Ecol.* 188: 117–131.
- Lomolino, M. V. et al. 2010. *Biogeography*. - Sinauer Associates.
- Losos, J. B. 2009. Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles.
- Mouillot, D. et al. 2013. The challenge of delineating biogeographical regions: nestedness matters for Indo-Pacific coral reef fishes (W Jetz, Ed.). - *J. Biogeogr.* 40: 2228–2237.
- O'Brien, E. M. et al. 1998. Climate and woody plant diversity in southern Africa: relationships at species, genus and family levels. - *Ecography (Cop.)* 21: 495–509.
- O'Brien, E. M. et al. 2000. Climatic gradients in woody plant (tree and shrub) diversity: water-energy dynamics, residual variation, and topography. - *Oikos* 89: 588–600.
- Pedersen, R. Ø. et al. 2014. Macroecological evidence for competitive regional-scale interactions between the two major clades of mammal carnivores (Feliformia and Caniformia). (D Russo, Ed.). - *PLoS One* 9: e100553.
- Podani, J. et al. 2013. A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data. - *Ecol. Complex.* 15: 52–61.
- Qian, H. 2009. Global comparisons of beta diversity among mammals, birds, reptiles, and amphibians across spatial scales and taxonomic ranks. - *J. Syst. Evol.* 47: 509–514.
- Qian, H. and Ricklefs, R. E. 2007. A latitudinal gradient in large-scale beta diversity for vascular plants in North America. - *Ecol. Lett.* 10: 737–44.
- Qian, H. and Ricklefs, R. E. 2012. Disentangling the effects of geographic distance and environmental dissimilarity on global patterns of species turnover. - *Glob. Ecol. Biogeogr.* 21: 341–351.
- Qian, H. et al. 2009. The latitudinal gradient of beta diversity in relation to climate and topography for mammals in North America. - *Glob. Ecol. Biogeogr.* 18: 111–122.
- Root, T. 1988a. Environmental factors associated with avian distributional boundaries. - *J. Biogeogr.* 15: 489–505.
- Root, T. 1988b. Energy constraints on avian distributions and abundances. - *Ecology* 69: 330–339.
- Rosauer, D. F. et al. 2014. Phylogenetic generalised dissimilarity modelling: a new approach to analysing and predicting spatial turnover in the phylogenetic composition of communities. - *Ecography (Cop.)* 37: 21–32.
- Rosenzweig, M. L. 1995. *Species Diversity in Space and Time*. - Cambridge University Press.
- Safi, K. et al. 2011. Understanding global patterns of mammalian functional and phylogenetic diversity. - *Philos. Trans. R. Soc. London, Ser. B* 366: 2536–2544.
- Sallam, H. M. et al. 2009. Fossil and molecular evidence constrain scenarios for the early evolutionary and biogeographic history of hystricognathous rodents. - *Proc. Natl. Acad. Sci. U. S. A.* 106: 16722–7.
- Sanmartín, I. and Ronquist, F. 2004. Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. - *Syst. Biol.* 53: 216–43.
- Seton, M. et al. 2012. Global continental and ocean basin reconstructions since 200Ma. - *Earth-Science Rev.* 113: 212–270.
- Silvestro, D. et al. 2015. The role of clade competition in the diversification of North American canids. - *Proc. Natl. Acad. Sci. U. S. A.* 112: 8684–8689.
- Simpson, G. G. 1943. Mammals and the nature of continents. - *Am. J. Sci.* 241: 1–31.
- Simpson, G. G. 1980. *Splendid isolation : the curious history of South American mammals*. in press.
- Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. - *I kommission hos E. Munksgaard*.
- Storch, D. et al. 2012. Universal species-area and endemics-area relationships at continental scales. - *Nature* 488: 78–81.
- Terlizzi, A. et al. 2009. Beta diversity and taxonomic sufficiency: Do higher-level taxa reflect heterogeneity in species composition? - *Divers. Distrib.* 15: 450–458.
- Vilhena, D. A. and Antonelli, A. 2015. A network approach for identifying and delimiting biogeographical regions. - *Nat. Commun.* 6: 6848.
- Wallace, A. 1876. *The geographical distribution of animals*. - Cambridge Univ. Press.
- Webb, S. 2006. The great American biotic interchange: patterns and processes. - *Ann. Missouri Bot. Gard.* 245–257.
- Webb, C. O. et al. 2002. Phylogenies and Community Ecology. - *Annu. Rev. Ecol. Evol. Syst.* 33: 475–505.
- Weinstein, B. G. et al. 2014. Taxonomic, phylogenetic, and trait Beta diversity in South American hummingbirds. - *Am. Nat.* 184: 211–24.
- Whitmee, S. and Orme, C. D. L. 2013. Predicting dispersal distance in mammals: a trait-based approach. - *J. Anim. Ecol.* 82: 211–21.
- Whittaker, R. H. 1960. *Vegetation of the Siskiyou Mountains, Oregon and California*. - *Ecol. Monogr.* 30: 279.
- Williams, S. et al. 2012. An open-source software environment for visualizing and refining plate tectonic reconstructions using high-resolution geological and geophysical data sets. - *GSA Today* 22: 4–9.
- Woodburne, M. O. 2010. The Great American Biotic Interchange: Dispersals, Tectonics, Climate, Sea Level and Holding Pens. - *J. Mamm. Evol.* 17: 245–264.

CHAPITRE VI

LE RÉGIME ALIMENTAIRE ET LA PHYLOGÉNIE DE L'HÔTE MODULENT LA COMPOSITION DES MICROBIOTES À DIFFÉRENTES ÉCHELLES PHYLOGÉNÉTIQUES

Résumé

Tous les mammifères vivent en symbiose avec des communautés bactériennes (microbiotes) présentes dans leur intestin. Cependant, notre compréhension des processus responsables de l'évolution du système symbiotique hôte : communauté reste limitée. En particulier, l'importance relative de la descendance verticale (c'est-à-dire le long de la phylogénie de l'hôte) et horizontale (c'est-à-dire entre hôtes éloignés ou depuis l'environnement) est mal connue. Dans cette analyse, nous montrons que la phylogénie et le régime alimentaire de l'hôte influencent les microbiotes à différentes échelles phylogénétiques bactériennes et sont en fait largement indépendants. La phylogénie de l'hôte explique la répartition des bactéries à fine échelle phylogénétique (par exemple les « espèces » bactériennes) alors que le régime alimentaire de l'hôte explique la répartition à plus large échelle phylogénétique (par exemple les « ordres » de bactéries), ce qui est en accord avec un conservatisme de niche trophique à haut niveau taxonomique des bactéries. Par ailleurs, nous montrons que la structure emboîtée des régimes alimentaires de l'hôte (les omnivores sont des carnivores et des herbivores) résulte en une structure emboîtée des communautés de bactéries et pourrait partiellement expliquer les dynamiques macroévolutives du régime alimentaire des mammifères. Finalement, le signal phylogénétique (dans l'arbre des hôtes) des compositions bactériennes observées (aussi appelé « phyllosymbiose ») peut s'expliquer par un nombre important de co-spéciations entre hôte et bactéries dénotant une descendance verticale d'une part importante de la communauté. En définitive, nos résultats permettent de réconcilier l'effet de la phylogénie et du régime alimentaire de l'hôte en les décomposant le long d'une échelle phylogénétique bactérienne.

Note sur la participation

Cet article est le fruit d'un travail commun avec mon ami et collègue Mathieu Groussin, du MIT (Boston, USA). Cette collaboration a été initiée lors de mon voyage en Amérique du Nord en 2014 et mon implication est très forte (co-leadership de l'article).

Phylogenetic decomposition of mammalian gut microbiota reveals the evolution of community symbiosis

Groussin. M.^{1*}, Mazel F.^{2,3*}, Sanders J.⁴, Smillie C.¹, Lavergne S.^{2,3}, Thuiller W.^{2,3} & Alm E.¹

¹ Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

² Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France.

³ CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France

⁴ Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue room 3085, Cambridge, Massachusetts 02138, USA

All mammals live in mutualistic symbiosis with complex bacterial communities (microbiota) present in their gut. Yet, our understanding of the processes responsible for the evolution of host:community symbiotic systems is limited, obscuring our knowledge on gut microbiota assembly. Notably, how vertical descent along host phylogeny and horizontal descent due to host dietary shifts relatively shape gut microbiota composition is unknown. Here, we show that host phylogeny and host diet drive community assembly at different temporal scales of bacterial evolution and are independent processes. Host phylogeny discriminates bacterial lineages at recent time scales, consistent with intimate host-microbe interactions, while host diet separates out more ancient lineages owing to ecological conservatism of high ranking bacterial clades. We show that the nested structure of diets results in nested gut microbiota compositions, possibly related to macroecological observations of dietary shifts in mammals. We further highlight that at the time scale at which host phylogeny exerts its strongest influence, extant mammalian gut microbiota compositions retained an ancient signal of host diversification. Finally, we reveal a strong and global pattern of co-speciation between hosts and gut bacteria, indicative of a community-scale vertical inheritance of this complex community symbiosis. Together, our results shed light on the underlying complexity of processes driving the evolution of mammal:microbiota symbiotic systems and reconcile previous, seemingly inconsistent empirical observations on the relative contribution of host phylogeny and diet. Also, our results open new interesting avenues to study and further understand the functional drivers of mammalian evolution.

The bacterial consortium (microbiota) living in mammalian guts form complex and malleable ectosymbiotic communities with their host (1, 2). They are engaged in mutualistic interactions, benefiting from resources provided by the host, while having positive impacts on the physiological state and fitness of the host (3, 4). The evolution of 1:1 symbiotic relationships between a host and a single symbiont has been extensively studied (5-7). However, little is known about the mechanisms involved in the long-term evolution of host:community symbiotic relationships. In particular, we are unaware of the long-term relative influence of vertical vs. horizontal descent in driving the composition of gut microbiota.

To ultimately understand the nature and the magnitude of the evolutionary forces shaping these communities that symbiotically evolve along the tree of Mammals, it is essential to analyze them in a phylogenetic context. If vertical transmission of symbionts occurs, it should result in microbiota compositions that recapitulate host phylogeny (*i.e.* 'phylosymbiosis' (8, 9)). Neutral

inheritance (drift) can entail these vertical patterns due to passive mother-to-infant transmission and allopatric host speciations. Alternatively, selection can also lead to phylosymbiosis if a host trait that is conserved along the host phylogeny (*e.g.* immune genes) selects for a particular microbiota composition (10, 11). Horizontal transmission, however, may occur if a selective trait decoupled from host phylogeny (*e.g.* diet in Mammals) promotes the acquisition of particular bacterial symbionts from the environment or from distantly-related hosts (12, 13). Depending on the magnitude of this horizontal selective force and the amount of bacteria involved, the phylosymbiosis signal may be erased rapidly, making gut microbiota compositional similarities related to differences in trait values rather than phylogenetic distances.

Here, we consider both host evolutionary history (summarized by host phylogeny) and host diet (as captured by food items) as the main factors shaping guts microbiota vertically and horizontally, respectively (1). Several attempts have been made to determine which of these two

factors exerts a stronger influence, often concluding in favor of diet (1, 12, 14), although the topic remains debated (8). Furthermore, both confirming (1, 15) and disproving (12, 16) results about phylosymbiosis have been reported at the scale of mammals, questioning the long-term evolutionary importance of host phylogeny in driving gut microbiota composition and obscuring our understanding of the evolution of host:community symbiotic systems.

We argue that the problem over the relative effects of host phylogeny and diet has been misidentified and that part of the confusion arises from the fact the selective pressures can operate at different temporal scales of bacterial evolution. For example, sister bacterial lineages might be functionally redundant and distribute randomly with respect to a given diet (e.g. they possess the same essential enzymes needed for herbivores to digest cellulose) but not with respect to host immune system (e.g. they harbor different antigens). Consequently, understanding the determinants of microbiota assembly along host evolution first requires the use of appropriate metrics and analytical procedures that are able to partition the effects of factors possibly acting at different phylogenetic scales.

Here, we fill this gap by developing a new method called BDTT (for Beta-Diversity Through Time, beta-diversity measuring the amount of lineage compositional dissimilarities between two communities). BDTT seeks to explicitly capture the relationship between temporal diversification of bacteria and the evolution of their hosts by performing a time decomposition of microbiota compositions. From the leaves to the root, the phylogenetic community tree of all sequences (e.g. 16S rDNA reads) is continuously sliced, either by time or evolutionary distance. For all slices, the microbiota compositions are determined and correlated to host phylogenetic or diet distances. We analyzed a dataset of 33 mammalian gut microbiotas (12), composed of 44,444 dereplicated and chimera-free amplicons of the 16S rRNA gene (V2 region) (see Methods). To apply BDTT, we reconstructed and time calibrated the phylogenetic community tree of all sequences (see Methods). Note that, here, times given in Millions of years should not be considered as accurate divergence times between bacterial lineages, but, as evolutionary distance, should be considered as proxies of time (see Methods).

While both host phylogeny and diet strongly correlate with community composition,

we observe a striking difference of bacterial temporal/phylogenetic scale at which host phylogeny and diet shape gut microbiota. Host phylogeny governs community assembly at recent bacterial phylogenetic scales, while host diet mostly determines the composition of more ancient lineages (Fig. 1). The recent-scale discrimination of bacterial lineages by host phylogeny is consistent in time with a model of co-evolution between host-specific and bacteria-specific markers. It promoted the emergence of intimate interactions, such as the interplay with components of the innate and adaptive immune system (17) or the physical interactions with host mucopolysaccharides permitted by the bacterial CCF complex to mediate gut colonization (18). We cannot reject, though, the potential selection of recent bacterial lineages by small differences in dietary compounds that are not captured by our dietary distance matrix. Diet, on the other hand, is probably a less stringent factor and discriminates between more ancient bacterial lineages. This is consistent with the ecological conservatism of ancient bacterial taxa (19), with all lineages belonging to these ancient taxa sharing sets of functional traits that are involved in the digestion of dietary compounds (12). Consequently, neutral assembly is more important with respect to diet at recent time scales of bacterial evolution, owing to higher degrees of functional equivalence within community members. A large number of control experiments suggest that our analysis uncovers a genuine and robust signal of scale disparity between the two factors and is not affected by systematic biases (see Supplementary Results). In particular, our results are robust to topological uncertainties of the bacterial phylogenetic community tree. Importantly, our results demonstrate that claiming domination of one process over another is inaccurate when gut microbiota is reduced to a single dimension. As the gut microbiota is a complex system, correlation outcomes with any factor depend on the (phylogenetic) scale to which the complex system is defined and observed.

The time scale disparity of host phylogeny and diet may emerge from two different patterns: recent bacterial lineages correlated to host phylogeny may be nested in or independent from more ancient diet-related bacterial lineages. To distinguish between these two hypotheses, we removed the diet-related ancient lineages (i.e. those significantly correlated to diet at the time slice where the diet correlation is maximal (Supp. Fig. 17) from the gut communities and re-run BDTT. The correlation with host phylogeny remains very strong at recent slices, close to its

previous level, demonstrating that the vast majority of bacterial lineages targeted by host phylogeny and diet are independent. The decrease in correlation, even though it is slight, is nevertheless higher than expected by chance (p -value < 0.01) and we estimate that only 8% of host phylogeny-related recent lineages are nested within diet-related ancient lineages (see Supp. Fig. 18). Taken together these results strongly suggest that the vast majority of bacterial lineages targeted by host phylogeny and diet are independent and emphasize the progress made in the description and understanding of gut microbiota assembly when decomposing microbiota compositions through time (BDTT).

After disentangling the relative effect of host phylogeny and diet, we can now describe the mechanistic features responsible for the evolution of the host:community symbiosis between gut microbiota and mammalian hosts for each factor separately.

We first investigated whether macroevolutionary patterns of dietary shift in mammals are linked to microevolutionary patterns in the gut microbiota. Transition rates between trophic strategies are biased, with transitions from herbivory or carnivory into omnivory and vice versa occurring at much higher rates than transitions between herbivory and carnivory, which are very rare (20). To ask whether the nested structure of diets (*i.e.* omnivory regimes include herbivory and carnivory regimes) results in a gut microbiota with nested compositions we partitioned our slice-specific beta-diversities into true turnover (resulting from lineage replacement) and nestedness (resulting from non-random lineage losses) components (see Methods) (21). At the ancient time slices where beta-diversity highly correlates with diet, a significant signal of bacterial community nestedness is observed (Supp. Fig. 12-13), indicating that some gut microbiota are subsets of others. We suggest that omnivorous guts are responsible for this community nestedness. To confirm this hypothesis, we used a subset of the whole bacterial community containing only ancient lineages that correlate with diet and not host phylogeny to represent both hosts and bacterial lineages on the same ordination space. This technique (OMI, see methods) separates bacterial lineages along a niche gradient, here defined by the host diet (Figure 2). It confirmed the nestedness component of mammalian gut communities, with the first axis (representing most of the variance (77,6%)) separating carnivores from herbivores, with omnivores in-

between. It further reveals ecologically differentiated clades adapted to specific dietary niches. Some bacterial lineages exhibit niche breadth encompassing carnivorous hosts only, while others have niches restricted to herbivorous guts. However, no niche differentiation is associated with the omnivorous diet, whatever the phylogenetic resolution of the microbiota. Taken together, these results suggest that herbivorous and carnivorous guts contain specialist diet-related bacterial lineages, contrarily to omnivores which filter lineages from these more circumscribed herbivorous and carnivorous collections. We can also reasonably speculate that the high rates of transition into omnivory from herbivory and carnivory and vice versa are in part facilitated by the composite and flexible composition of omnivorous gut microbiota.

We next investigated in more details the processes associated to host phylogeny that drive the evolution of microbiota symbiosis. We found that the correlation with host phylogeny at all recent bacterial scales is almost entirely driven by true lineage turnovers, contrarily to what we found with diet. This pattern is likely the consequence of niche sorting by co-adapted host-related factors. It is consistent with previous observations that gut microbiota are highly malleable systems that can undergo rapid compositional turnovers (2, 22). But it also raises the question of whether the phylosymbiosis signal is conserved over long evolutionary time. Thus, a probabilistic and phylogenetic approach to model and quantify the dynamics of lineage turnover (gain and loss of bacterial lineages) along the mammalian phylogeny is required to capture the phylosymbiosis signal. We reconstructed bacterial community sizes for all mammalian ancestors and compared these estimations with a null model where host phylogeny is disrupted (see Methods). Our approach departs from previous ad-hoc, parsimony and distance-based methods that provided conflicted results (8, 9, 12, 23). Ancestors with a significantly higher number of gut bacterial lineages than randomly expected denote clades containing hosts that have conserved a signal for ancestral bacterial lineage content and so, for host speciations. We also measured the magnitude of this phylosymbiosis signal by computing Standard Effect Sizes (SES) (see Methods) and observed that most of the clades harbor a significant and strong signal for phylosymbiosis. Unexpectedly, the gut microbiota of some clades older than 80Myr still harbor imprints of ancient host speciation, a pattern only detectable through our phylogenetic and

probabilistic approach. The degree of compositional conservation decreases linearly with time, consistent with a progressive erasure of the phyllosymbiosis signal owing to multiple turnovers over millions of years of host evolution.

Are gut bacterial symbionts vertically inherited over millions of years of mammalian evolution? Actually, the presence/absence profiles can be generated by two different processes: strict vertical inheritance through co-speciations or host-shifts among hosts. For instance, three closely-related hosts can possess a bacterial symbiont in their gut to the exception of other hosts. This lineage can evolve vertically or being horizontally inherited several times in time, leaving imprints in the symbiont

phylogenetic tree (see Supp. Fig. 21). Both processes result in the same patterns of presence/absence. But they can be differentiated with probabilistic models of host tree/symbiont tree reconciliation, modeling co-speciation and host-shifts events (see Methods and Supp. Fig. 21). We highlight that 65% of OTUs evolved with more co-speciation events than host-shift events (Figure 4), this proportion being by far higher than randomly expected under a null model controlling for the over-fitting of the host tree during the search for the most likely scenario of evolutionary events. These results strongly suggest that vertical inheritance is the main evolutionary process experienced by mammalian gut symbionts, explaining the strong correlation with host phylogeny at recent time scales (Figure 1).

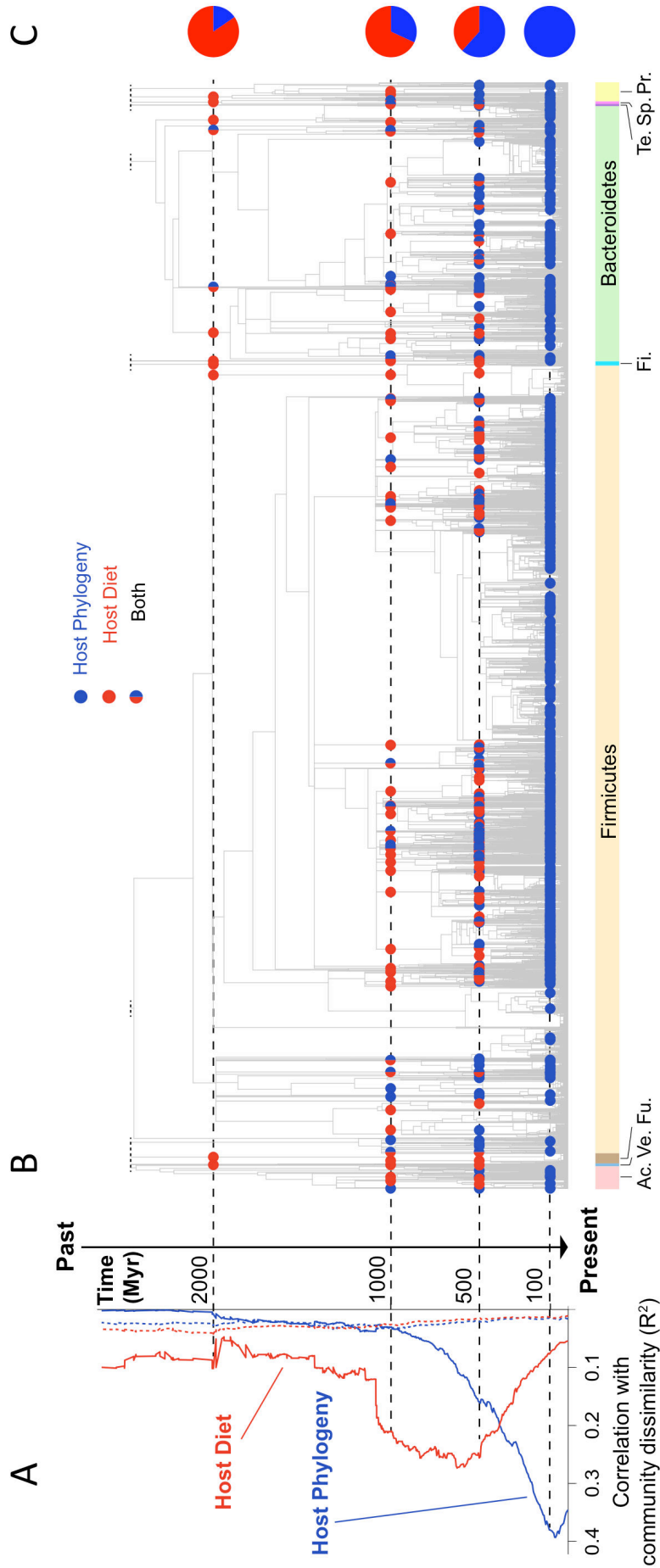


Figure 1. Host phylogeny and diet influence microbiota composition at different phylogenetic scale. A) Variation of the strength of the mantel correlation (X-axis) between microbiota dissimilarities (β diversity) and host phylogenetic (blue) or diet (red) distances along the bacterial phylogenetic scale at which dissimilarities are defined (Y-axis). The colored dashed lines correspond to the upper part of the 95% null envelope (see methods). B) Bacterial correlations with host phylogeny and diet at the individual lineage level according to PERMANOVA tests (see methods). Significant correlations are represented with colored dots on the bacterial phylogenetic community tree at four time slices, from recent to ancient phylogenetic scales. Colors are as in A). C) Pie charts presenting the % of lineages correlated to either host phylogeny or diet. Note that the phylogenetic scale is common for the three plots, which are connected with the dashed black lines corresponding to the four time slices. Main bacterial lineages are depicted below the tree (Ac.: Actinobacteria; Ve.: Verrucomicrobia; Fu.: Fusobacteria; Fi.: Fibrobacteres; Te.: Tenericutes; Sp.: Spirochaetes; Pr.: Proteobacteria).

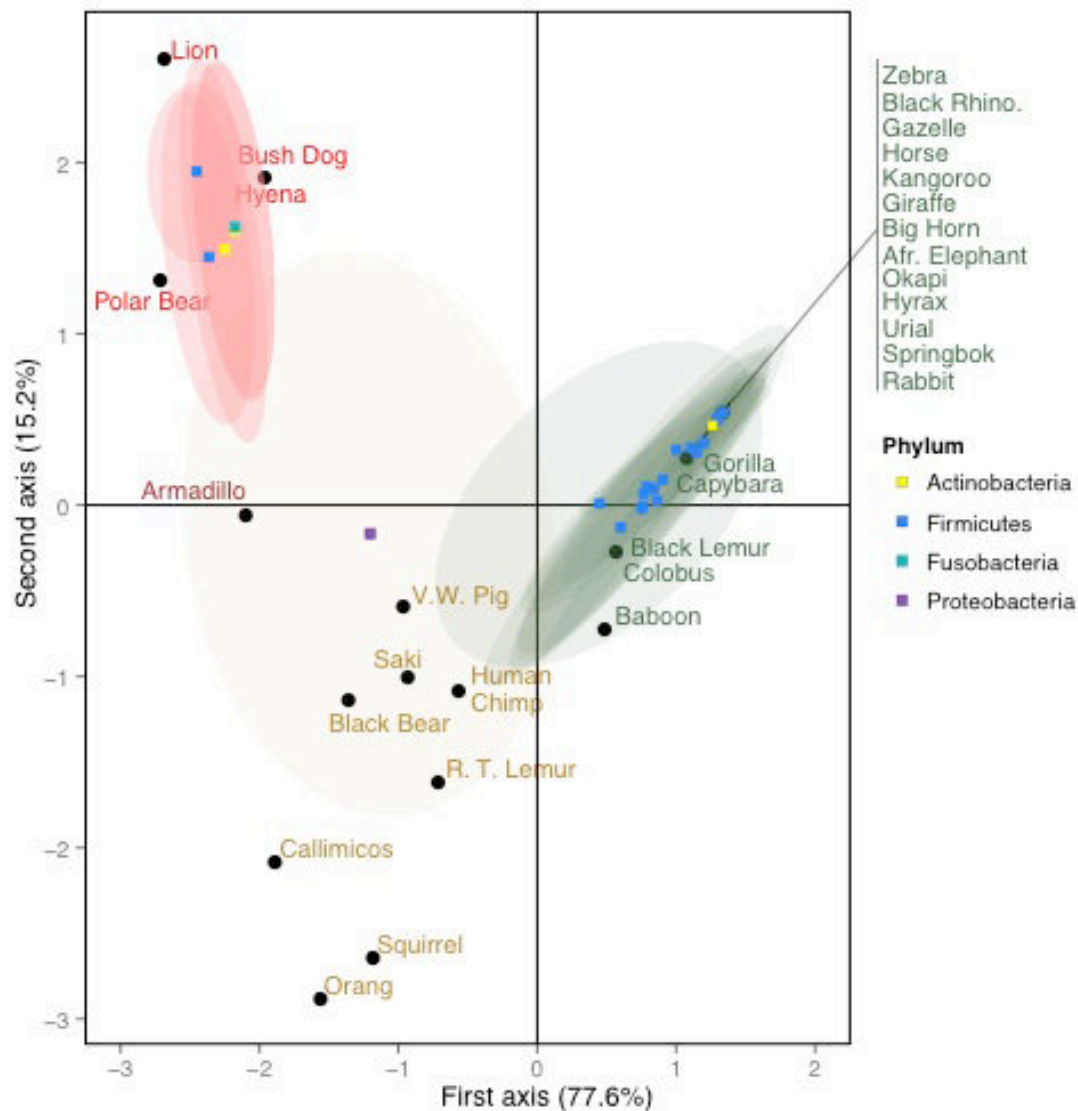


Figure 2. Ecological niche differentiation of bacterial lineages according to host diet. Mammalian hosts and single bacterial lineages are represented in the same ordination space. Black dots correspond to mammals and colored squares denote bacterial lineages. Only bacterial lineages with a distribution across hosts that is significantly correlated to diet and not to host phylogeny are represented. The bacterial lineages has been defined at the time slice where the correlation between host diet and microbiota composition is high. The breadth of a bacterial lineage niche is represented with an ellipse while niche centers are colored according to bacterial phyla (see legend). Host names are color-coded according to broad dietary categories: carnivores (red), herbivores (green) or omnivore (brown). Bacterial niche breadths that are restricted to a particular diet, highlighting ecological differentiations, are colored accordingly.

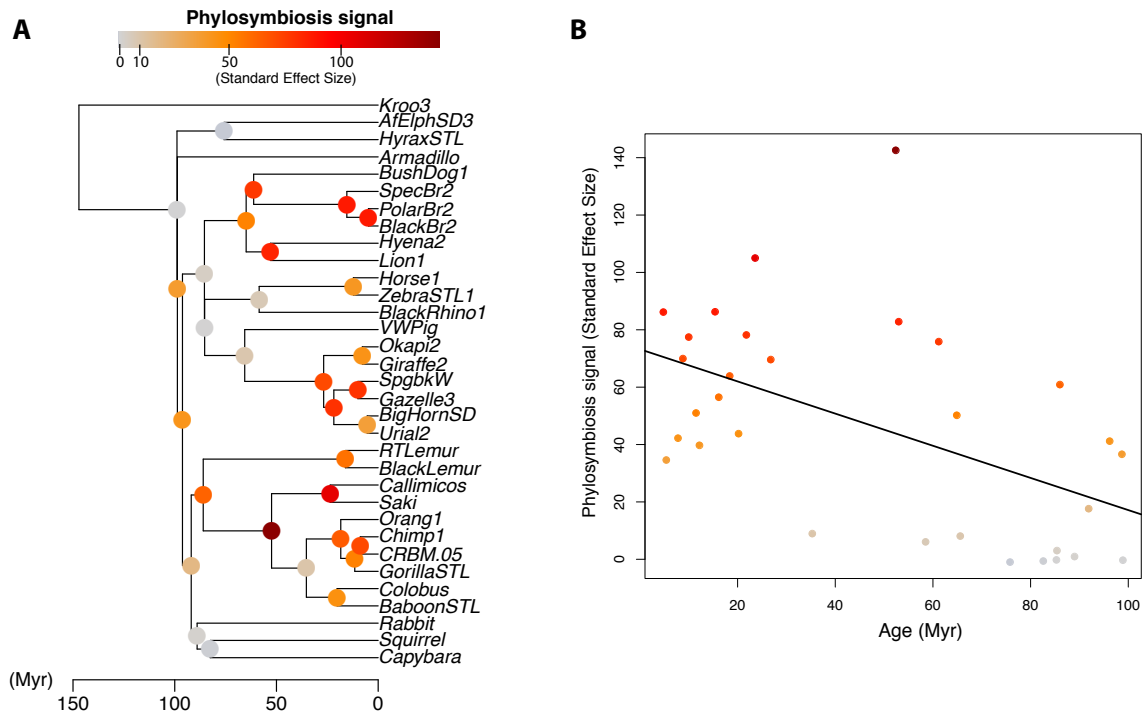


Figure 3. The decrease of phylosymbiotic signal with time. A) Mammalian host calibrated phylogeny. Internal nodes are colored according to the degree of the phylogenetic conservatism of community composition (strength of the phylosymbiotic signal, see legend). For each node, the phylosymbiotic signal is measured as the deviation between the inferred ancestral bacterial community size (*i.e.* number of OTUs) and a distribution of null expectations (see methods). A large value indicates that the inferred community size is much larger than expected by chance. B) Linear decay of the conservation of ancient community compositions in extant mammalian microbiota. Each dot represents an internal node of the phylogenetic tree presented in A). The phylosymbiotic signal at a given ancestor (Y-axis) is plotted against its age in Myr (X-axis). A linear regression is fitted to the data (black line) but a *p-value* is not provided owing to the non-independence of the points.

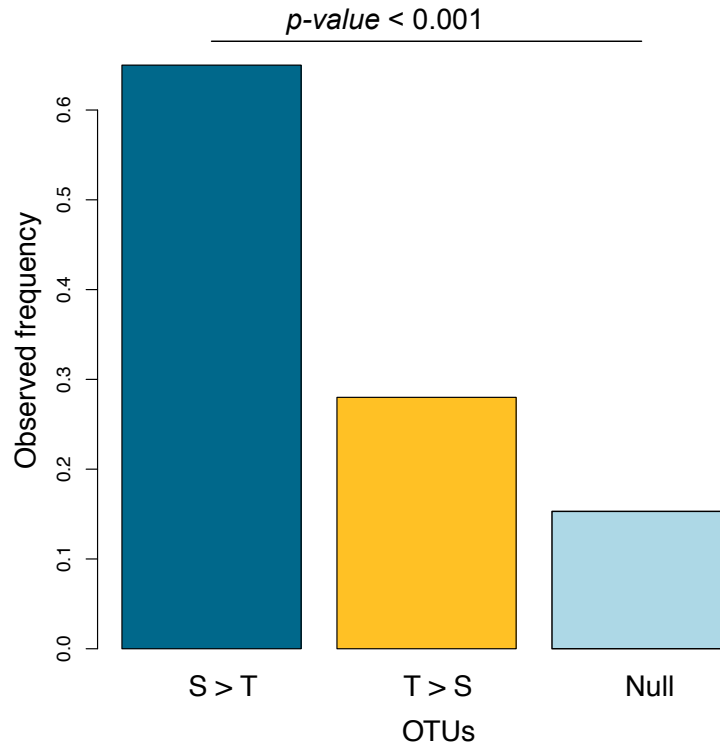


Figure 4: The vertical inheritance of gut bacterial symbionts in mammals. The figure presents the observed frequencies of OTUs that harbor more co-speciation (S) than host shifts (T) events (dark blue bar) or more host shifts than co-speciation events (yellow bar). These observed frequencies are derived from an explicit model of co-evolution between host phylogeny and each individual OTU phylogeny, accounting for co-speciation, host-shift, intra-host speciation and extinction events. The light blue bar corresponds to the mean frequency of OTUS harboring more co-speciation than host shifts events (S>T) under a null model (see methods). The observed frequency of OTUs that harbor more co-speciation than host shifts events is much higher than expected under this null model ($p\text{-val}<0.001$).

References

1. R. E. Ley *et al.*, Evolution of mammals and their gut microbes. *Science*. **320**, 1647–1651 (2008).
2. L. A. David *et al.*, Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. **505**, 559–563 (2014).
3. L. V. Hooper, J. I. Gordon, Commensal host-bacterial relationships in the gut. *Science*. **292**, 1115–1118 (2001).
4. C. A. Lozupone, J. I. Stombaugh, J. I. Gordon, J. K. Jansson, R. Knight, Diversity, stability and resilience of the human gut microbiota. *Nature*. **489**, 220–230 (2012).
5. N. A. Moran, J. P. McCutcheon, A. Nakabachi, Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet*. **42**, 165–190 (2008).
6. G. M. Bennett, N. A. Moran, Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci U S A*. **112**, 10169–10176 (2015).
7. J. L. Sachs, R. G. Skophammer, J. U. Regus, Evolutionary transitions in bacterial symbiosis. *Proc Natl Acad Sci U S A*. **108 Suppl 2**, 10800–10807 (2011).
8. H. Ochman *et al.*, Evolutionary Relationships of Wild Hominids Recapitulated by Gut Microbial Communities. *Plos Biol*. **8**, e1000546–8 (2010).
9. R. M. Brucker, S. R. Bordenstein, The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. *Science*. **341**, 667–669 (2013).
10. J. Schluter, K. R. Foster, The evolution of mutualism in gut microbiota via host epithelial selection. *Plos Biol*. **10**, e1001424 (2012).
11. T. Hosokawa, Y. Kikuchi, N. Nikoh, M. Shimada, T. Fukatsu, Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *Plos Biol*. **4**, e337 (2006).
12. B. D. Muegge *et al.*, Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. *Science*. **332**, 970–974 (2011).
13. F. Delsuc *et al.*, Convergence of gut microbiomes in myrmecophagous mammals. *Mol Ecol*. **23**, 1301–1317 (2013).
14. R. N. Carmody *et al.*, Diet Dominates Host Genotype in Shaping the Murine Gut Microbiota. *Cell Host and Microbe*. **17**, 72–84 (2015).
15. P. L. Oh *et al.*, Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. *The ISME Journal*. **4**, 377–387 (2010).
16. P. A. Vaishampayan *et al.*, Comparative metagenomics and population dynamics of the gut microbiota in mother and infant. *Genome Biology and Evolution*. **2**, 53–66 (2010).
17. H.-J. Wu, E. Wu, The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes*. **3**, 4–14 (2012).
18. S. M. Lee *et al.*, Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature*. **501**, 426–429 (2013).
19. L. Philippot *et al.*, The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Micro*. **8**, 523–529 (2010).
20. S. A. Price, S. S. B. Hopkins, K. K. Smith, V. L. Roth, Tempo of trophic evolution and its impact on mammalian diversification. *Proc Natl Acad Sci U S A*. **109**, 7008–7012 (2012).
21. A. Baselga, Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*. **19**, 134–143 (2010).
22. L. A. David *et al.*, Host lifestyle affects human microbiota on daily timescales. *Genome Biol*. **15**, R89 (2014).
23. R. M. Brucker, S. R. Bordenstein, The roles of host evolutionary relationships (genus: *Nasonia*) and development in structuring microbial communities. *Evolution*. **66**, 349–362 (2012).

**SYNTHÈSE & DISCUSSION
DE LA SOUS-PARTIE III.1**

Dans cette sous-partie, nous avons montré que dans des systèmes d'études radicalement différents, la décomposition des similarités compositionnelles le long d'une échelle phylogénétique peut permettre de distinguer les effets relatifs de différents processus qui rassemblent les communautés de mammifères ou bien les communautés de microbes dans leur système digestif. Dans les deux cas, ces études tentent de reconstruire l'histoire des assemblages, mais à des échelles radicalement différentes.

1. Similarités entre les deux études

Nous montrons que l'influence de l'espace et du climat sur la répartition géographique des mammifères et des oiseaux à l'échelle du globe varie le long de l'échelle phylogénétique. En particulier, l'espace agit de façon prépondérante à une échelle phylogénétique intermédiaire, notamment chez les mammifères, alors que l'environnement a un effet majoritaire à fine échelle phylogénétique, par exemple au niveau des espèces. On peut interpréter ces résultats en proposant que l'histoire (dispersion intercontinentale, vicariance) influence plutôt la distribution des clades profonds, alors que l'environnement sélectionne différentes espèces parmi ces clades profonds. La même approche méthodologique permet de montrer que les microbiotes présents dans chacun des 33 mammifères étudiés gagnent à être décrits – comme leurs hôtes – à différentes échelles évolutives. Le régime alimentaire de l'hôte semble impacter des clades bactériens profonds alors que la phylogénie de l'hôte joue plutôt sur des échelles évolutives plus fines. De façon intrigante, si l'on considère la phylogénie des hôtes comme une mesure de l'histoire et le régime alimentaire comme une mesure de l'environnement, les résultats obtenus dans les deux études sont opposés.

En fait, il apparaît que la phylogénie de l'hôte n'est pas forcément une mesure de l'histoire, mais plutôt d'un environnement dont nous n'avons pas de mesure directe, mais qui est conservé le long de la phylogénie de l'hôte, par exemple le système immunitaire (Hooper *et al.* 2012). Dans le cas de la biogéographie des mammifères, l'espace pourrait aussi « cacher » des variables environnementales non mesurées. Cependant, il est bien établi que les événements historiques de dispersion et de

vicariance ont joué un rôle fondamental dans la biogéographie des mammifères (Simpson 1980), ce qui conforte nos résultats dans ce cas. De plus, nous montrons que la géographie passée permet de mieux expliquer les compositions à large échelle évolutive, ce qui conforte l'idée que l'histoire (ancienne) agit à des échelles phylogénétiques profondes. On peut cependant noter que l'histoire récente (par exemple le quaternaire, et notamment ses oscillations climatiques) est aussi un élément important de la répartition des lignées, mais à une échelle phylogénétique plus fine (Hewitt 2000, 2004; Lister 2004).

En définitive, nous montrons que l'analyse des patrons à différentes échelles phylogénétique, par exemple via l'utilisation de notre méthode, est prometteuse. Nous le faisons pour la première fois dans le cadre de la diversité β , mais nous avons montré dans le chapitre II son utilité dans le cadre de l'analyse de la diversité α (voir aussi O'Brien *et al.* 1998, 2000; Qian & Ricklefs 2007; Qian *et al.* 2009; Terlizzi *et al.* 2009; Kreft & Jetz 2010).

2. Limites et améliorations possibles de la méthode BDTT

2.1. Limites de la méthode

La méthode proposée comprend cependant certaines limites. En particulier, il ne faut pas confondre les similarités de composition établies par exemple à 20 millions d'années sur l'arbre phylogénétique et la similarité des faunes à cette époque (que ce soit la faune fossile ou les descendants actuels de ces faunes-là). Par exemple, les carnivores et les artiodactyles sont aujourd'hui présents en Amérique du Sud, mais ils n'y sont arrivés que récemment (3-15 Ma), lors du grand échange biotique entre les deux continents américains (GABI, Webb 2006; Woodburne 2010, mais voir Bacon *et al.* 2015). Cependant, lorsque l'on calculera, avec la méthode BDTT, la similarité faunistique à 30 Ma entre les deux continents, les branches menant aux carnivores seront présentes dans les deux régions, et la similarité sera totale, bien qu'à cette époque, les carnivores n'étaient pas présents au Sud. En définitive, la méthode proposée ne reconstruit en aucun cas la présence passée des lignées dans l'assemblage. Ceci contraste directement avec les méthodes de biogéographie historique classiques (Ronquist & Sanmartín 2011).

Cependant, notre approche possède certains avantages, comme celle de ne pas avoir à définir des aires d'endémisme et de pouvoir prendre en compte explicitement l'influence du climat actuel et passé. Cette approche pourrait donc constituer un premier pas vers le rapprochement de la biogéographie historique et de la biogéographie écologique.

2.2. Améliorer les mesures de connectivité passée

Nous utilisons la méthode BDTT pour évaluer l'impact potentiel de la configuration passée des continents sur les distributions actuelles des lignées en mesurant les distances euclidiennes entre assemblages tout au long du cénozoïque (Chapitre V). Notre mesure de proximité historique tente de quantifier un coût à la dispersion et pourrait être améliorée, notamment en prenant en compte la part relative de croûte continentale et océanique dans le calcul des distances entre assemblages. En effet, ces deux types de croûtes terrestres possèdent des altitudes très contrastées : alors que la croûte continentale (+300 m d'altitude en moyenne) peut être émergée, surtout pendant les glaciations (p. ex. la Manche), la croûte océanique (-2400 m d'altitude en moyenne) est toujours immergée et représente donc une bonne mesure pour la limitation de la dispersion. Par exemple, la ligne de Wallace correspond à un plancher océanique et non à un plancher continental (Lomolino *et al.* 2010). Par ailleurs, il serait intéressant d'ajouter une dimension climatique à notre analyse de la connectivité passée. Le climat tropical et la forêt tropicale dominaient largement la surface terrestre au début du cénozoïque (voir par exemple Burnham & Johnson (2004) pour une perspective sud-américaine) pour finalement se contracter petit à petit jusqu'à leur position actuelle (Zachos *et al.* 2001). On peut imaginer que la niche de certains clades a évolué vers des climats plus froids et secs, permettant la colonisation et la radiation dans les zones tempérées à un temps *t*. Si ces événements sont assez rares et non réversibles, c'est-à-dire si l'on observe un fort conservatisme de niche, on peut imaginer que les patrons actuels de composition auront gardé la trace d'une telle histoire et que l'on pourra la révéler en utilisant la méthode BDTT. Par exemple, si l'on imagine qu'un certain nombre de clades commencent à radier dans les zones tempérées à partir du miocène, on observera aujourd'hui une faible similarité d'espèces entre les régions tempérées et les régions tropicales (donc un bon pouvoir explicatif du climat

actuel au temps 0 du BDTT). Si l'on remonte le long de la phylogénie, la similarité au niveau du miocène (et au-delà) sera alors mal expliquée par le climat actuel (les deux régions ont des climats différents, mais des compositions similaires), mais probablement mieux par le climat passé (les deux régions avaient alors des climats similaires et des compositions similaires).

3. La biogéographie des hôtes peut-elle expliquer la composition du microbiote ?

Finalement, j'ai appliqué l'approche aux deux systèmes de façon relativement indépendante. J'entrevois cependant d'intéressantes perspectives en croisant ces systèmes d'études. Il serait par exemple possible de tester si la biogéographie des hôtes a une influence sur la composition des microbiotes. On peut en effet supposer que des hôtes en sympatrie* puissent partager davantage de lignées bactériennes que des hôtes en allopatrie* (à environnement égal) simplement par transfert horizontal ou via transfert depuis l'environnement. Pour l'instant, de telles influences ont été montrées à petite échelle phylogénétique de l'hôte, notamment parmi les grands singes du bassin du Congo (Moeller *et al.* 2013) ou même au sein de groupes d'humains (Yatsunenko *et al.* 2012). Il serait alors intéressant de comparer les similarités de microbiotes entre hôtes à leur proximité géographique, pour notamment tester si l'influence de la géographie tient à des échelles phylogénétiques (des hôtes) larges. La proximité géographique est un prérequis pour le transfert horizontal de bactérie entre hôtes. Il apparaît donc logique qu'une plus longue proximité géographique augmente la probabilité de transfert. On pourrait tester cette hypothèse en comparant le pouvoir prédictif de la proximité géographique actuelle avec celle intégrée sur des temps évolutifs, en utilisant les méthodes classiques de la biogéographie historique pour reconstruire les aires ancestrales des hôtes et reconstruire un profil de proximité géographique dans le temps. Cependant, il faut rester conscient que pour tester de telles hypothèses, un jeu de données conséquent est nécessaire pour distinguer les effets de la phylogénie, du régime alimentaire et de la géographie. En fait, il faut des hôtes qui représentent toutes les combinaisons de cas possibles (par exemple, des hôtes proches phylogénétiquement et en terme de régime alimentaire, mais éloignés géographiquement) afin de disposer d'une puissance statistique adéquate pour détecter les effets relatifs de chacun des facteurs.

SOUS-PARTIE III.2
LA GÉOGRAPHIE DE LA CONVERGENCE

INTRODUCTION DE LA SOUS-PARTIE III.2





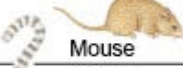




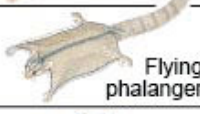




Niche	Placental Mammals	Australian Marsupials
Burrower	Mole 	Marsupial mole 
Anteater	Anteater 	Numbat (anteater) 
Mouse	Mouse 	Marsupial mouse 
Climber	Lemur 	Spotted cuscus 
Glider	Flying squirrel 	Flying phalanger 
Cat	Bobcat 	Tasmanian "tiger cat" 
Wolf	Wolf 	Tasmanian wolf 

Figure 9. La convergence évolutive entre marsupiaux et placentaires.

Copyright © 2015 The Pennsylvania State University

Dans le chapitre précédent, nous montrons qu'il existe une relation statistique forte entre le taux de similarité faunistique de deux assemblages et leur proximité environnementale. Cette observation suggère que des espèces différentes sont adaptées à des niches climatiques différentes et qu'elles ne cohabitent donc pas. Elles sont filtrées par l'environnement de sorte que l'espèce A n'est pas retrouvée dans le même espace géographique et environnemental que l'espèce B. Pour confirmer cette hypothèse, il apparaît important d'incorporer une dimension fonctionnelle à notre étude. En d'autres termes, il serait d'autant plus convaincant de montrer que des espèces vivantes dans des environnements différents possèdent des caractéristiques fonctionnelles différentes adaptées à chacun de ces environnements, ou au contraire, que des espèces différentes vivantes dans des environnements similaires possèdent des caractéristiques fonctionnelles similaires. Dans une perspective évolutive, il est intéressant de se demander à quel point la sélection a pu, de façon indépendante dans plusieurs lignées, façonner les caractéristiques fonctionnelles des espèces en fonction de l'environnement. Ce phénomène que l'on appelle la convergence évolutive (ou parallélisme*) avait déjà été noté par Darwin :

« Des animaux appartenant à deux lignées d'ancêtres très distincts peuvent, en effet, s'être adaptés à des conditions semblables, et avoir ainsi acquis une grande ressemblance extérieure ; mais ces ressemblances, loin de révéler leurs relations de parentés tendent plutôt à les dissimuler. [...] Ainsi, la forme du corps et les membres en forme de nageoires sont des caractères purement analogues lorsque l'on compare la baleine aux poissons, parce qu'ils constituent dans les deux classes une adaptation spéciale en vue d'un mode de locomotion aquatique »

Darwin, 1859

Dans cette sous-partie, je présente deux études (chapitre VII et annexe 3.2.1) qui visent à détecter l'effet de la convergence évolutive à différentes échelles.

Question générale : La convergence est-elle un processus majeur au cours de l'évolution?

Les convergences morphologiques, physiologiques, biochimiques, comportementales et même « microbiotiques » (Delsuc *et al.* 2014) qui permettent à des espèces relativement éloignées dans la phylogénie d'occuper des niches écologiques similaires représentent un argument majeur dans la démonstration de l'importance de la sélection dans l'évolution des espèces. Il faut noter que ce champ de recherche a été révolutionné par l'avènement de la phylogénie moléculaire (Felsenstein 2004) qui a modifié de façon importante les classifications taxonomiques classiques. Dans le cas des mammifères par exemple, la phylogénie moléculaire a révélé des liens de parenté insoupçonnés entre lignées et a placé des taxons morphologiquement similaires dans des groupes totalement différents. C'est par exemple le cas de l'ordre - longtemps instable - des « insectivores » qui regroupaient des familles holarctiques, par exemple les solénodontidés, et des familles africaines, par exemple les tenrécidés. Ces deux familles sont maintenant placées dans des ordres distincts (afrosoricidae et soricomorpha, respectivement). En fait, la détection de la convergence évolutive va se baser sur la comparaison des caractères fonctionnels dans un cadre phylogénétique. Dans cette sous-partie, on va distinguer deux échelles de convergence (Lomolino *et al.* 2010).

Annexe 3.2.1 et Chapitre VII : La convergence au niveau des espèces

La convergence au niveau des taxons est traditionnellement étudiée et reconnue comme importante (Lomolino *et al.* 2010). Elle correspond notamment à de nombreuses « règles écologiques » comme celle de Bergmann (augmentation de la taille des espèces avec la latitude) ou encore celle d'Allen (diminution de la taille des extrémités avec la latitude). En pratique, il s'agit, pour un ensemble d'espèces, de corrélérer un trait fonctionnel à une variable environnementale en prenant en compte la similarité phylogénétique entre les espèces. Dans ce chapitre, je présente un résumé de l'article 9 (texte entier en annexe 3.2.1) qui améliore les techniques actuelles de détection de convergence en prenant en compte les variations de taux d'évolution au sein d'une phylogénie.

Chapitre VII : La convergence au niveau des assemblages entiers d'espèces

La convergence au niveau des assemblages d'espèces correspond à l'évolution dans des lignées phylogénétiques distinctes de « types morphologiques » similaires adaptés à des niches particulières. En somme, cette convergence correspond à une multitude de convergences au niveau d'espèces qui cooccurrent dans une région donnée (Lomolino *et al.* 2010). Il existe des exemples emblématiques de convergence au niveau des assemblages, comme ceux des cyclidés des lacs d'Afrique de l'Est. Dans ce chapitre, nous tentons de montrer à quel point ce phénomène est répandu chez les mammifères à l'échelle du globe.

CHAPITRE VII

LA CONVERGENCE ÉVOLUTIVE DES ESPÈCES ET DES ASSEMBLAGES DE MAMMIFÈRES

Résumé de l'annexe 3.2.1 (article 9, accepté dans *Ecology*)

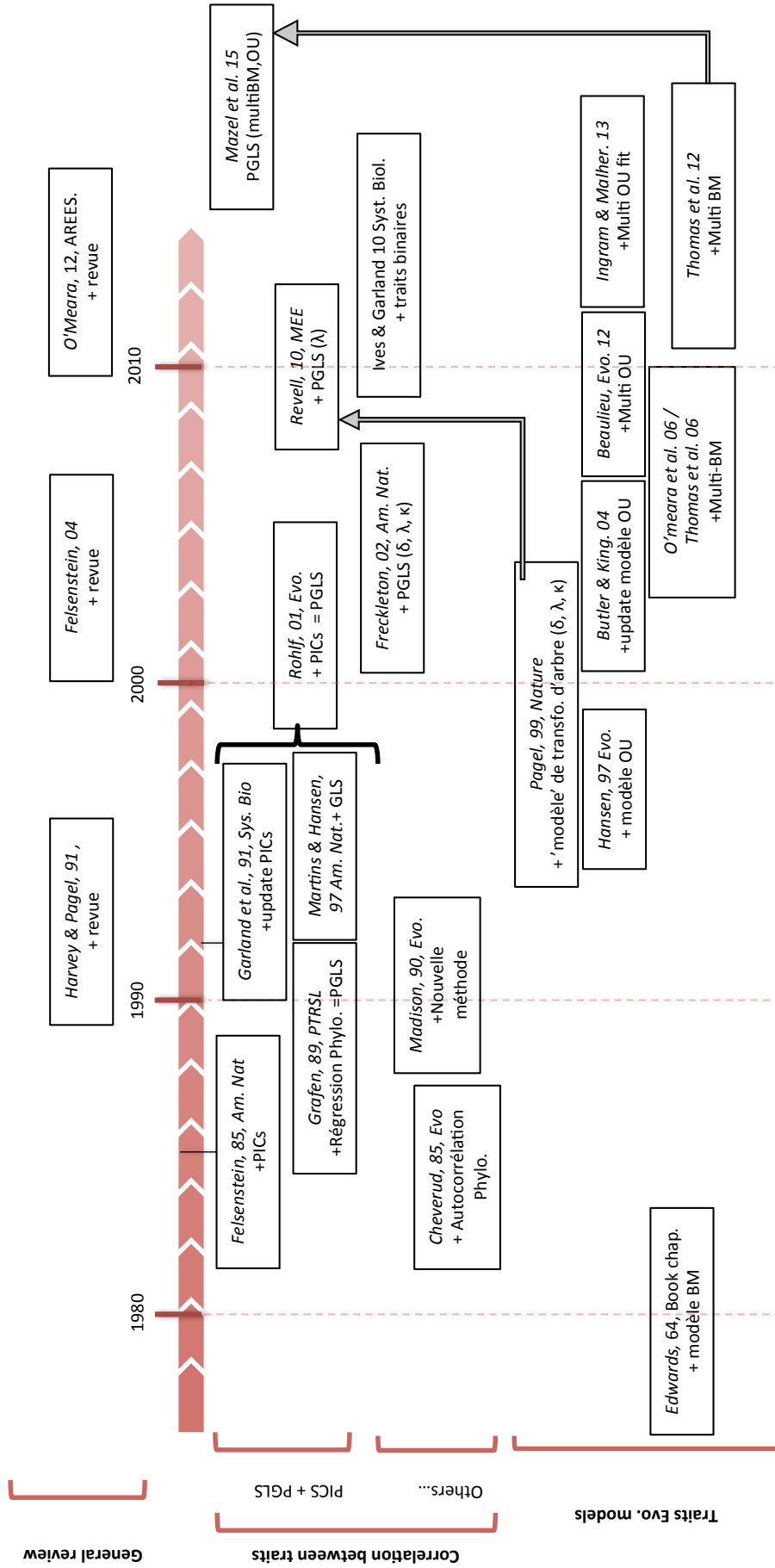
*« Differences in rates of evolution,
and not only divergent evolution at comparable rates,
are among the reasons for the great diversity of organisms on Earth. »*

Simpson, 1953

Les méthodes comparatives représentent un outil clef dans la compréhension des processus écologiques et évolutifs et sont utilisées pour tester des hypothèses sur l'évolution corrélée des traits (Felsenstein 1985; Harvey & Pagel 1991). De nombreuses méthodes ont été proposées par le passé pour permettre de relier des traits fonctionnels d'espèces tout en prenant en compte leur proximité phylogénétique (Felsenstein 1985, Grafen 1989, Maddison 1990, Lynch 1991, Garland et al. 1992, Martins and Hansen 1997, Diniz-Filho et al. 1998, Freckleton et al. 2002, Paradis and Claude 2002, Ives and Garland 2010; voir **l'encadré 5** pour un aperçu de la chronologie des articles fondateurs de la discipline). Aujourd'hui, la méthode des moindres carrés phylogénétique (« PGLS ») semble être particulièrement utilisée. Dans sa forme originale, elle assume en fait que les traits évoluent de façon brownienne le long de la phylogénie, mais peut théoriquement incorporer des modèles plus complexes. Nous montrons dans cette analyse que le PGLS, dans sa forme actuelle, peut faire de graves erreurs si l'hétérogénéité des taux d'évolution n'est pas prise en compte dans le modèle. Nous proposons une solution simple à ce problème.

Note sur la participation

Cet article est le fruit d'une collaboration avec Jonathan T. Davies et Pedro R Peres-Neto, tout deux basés à Montréal (Canada). Cette collaboration a été initiée lors de mes voyages en Amérique du Nord durant les étés 2013 et 2014.



Encadré 5. Chronologie des articles importants dans le développement des outils statistiques pour l'analyse des données interspécifiques. Les articles sont groupés par sous-disciplines (en ligne). La frise ne prétend pas être une bibliographie exhaustive mais permet simplement de donner un aperçu global de l'histoire du champ disciplinaire.

Bibliographie de l'encadré 5

- Beaulieu, J. and Jhwieng, D. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. - *Evolution* (N. Y). 66: 2369–2383.
- Butler, M. A. and Kings, A. A. 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. - *Am. Nat.* 164: 683–695.
- Cheverud, J. M. et al. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. - *Evolution* (N. Y). 39: 1335–1351.
- Edwards, A. W. F. and Cavalli-Sforza, L. L. 1964. Reconstruction of evolutionary trees. - In: Heywood, V. H. and McNeill, J. (eds), *Phenetic and phylogenetic classification*. pp. 67–76.
- Felsenstein, J. 1985. Phylogenies and the comparative method. - *Am. Nat.*: 1–15.
- Felsenstein, J. 2004. *Inferring Phylogenies*.
- Freckleton, R. P. et al. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. - *Am. Nat.* 160: 712–726.
- Garland, T. et al. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. - *Syst. Biol.* 41: 18–32.
- Grafen, A. 1989. The phylogenetic regression. - *Philos. Trans. R. Soc. London B*: 119–157.
- Hansen, T. F. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation. - *Evolution* (N. Y). 51: 1341.
- Harvey, P. and Pagel, M. 1991. *the Comparative Method in Evolutionary Biology*. - Oxford University Press.
- Ingram, T. and Mahler, D. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. - *Methods Ecol. Evol.* 4: 416–425.
- Ives, A. and Garland, T. 2010. Phylogenetic logistic regression for binary dependent variables. - *Syst. Biol.* 59: 9–26.
- Maddison, W. P. 1990. A Method for Testing the Correlated Evolution of Two Binary Characters: Are Gains or Losses Concentrated on Certain Branches of a Phylogenetic Tree? - *Evolution* (N. Y). 44: 223–539–557.
- O'Meara, B. 2012. Evolutionary inferences from phylogenies: a review of methods. - *Annu. Rev. Ecol. Evol. Syst.* 43: 267–285.
- O'Meara, B. C. et al. 2006. Testing for different rates of continuous trait evolution using likelihood. - *Evolution* (N. Y). 60: 922–933.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. - *Nature* 401: 877–884.
- Revell, L. 2010. Phylogenetic signal and linear regression on species data. - *Methods Ecol. Evol.* 1: 319–329.
- Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. - *Evolution* (N. Y). 55: 2143–2160.
- Thomas, G. H. and Freckleton, R. P. 2012. MOTMOT: models of trait macroevolution on trees. - *Methods Ecol. Evol.* 3: 145–151.
- Thomas, G. H. et al. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. - *Proc. Biol. Sci.* 273: 1619–24.

Résumé de l'article 10 (en préparation)

Alors que l'extraordinaire diversité de la vie sur Terre a longtemps fasciné les biogéographes, l'importance relative de la contingence historique et du déterminisme pour expliquer son évolution reste débattus. Le déterminisme peut apparaître lorsque des lignées indépendantes subissant les mêmes contraintes environnementales montrent des tendances évolutives similaires et prédictibles, comme c'est le cas avec les convergences. À l'échelle des mammifères, les convergences entre placentaires et marsupiaux servent souvent d'exemple emblématique mais n'ont jamais été clairement quantifiées. Dans cette analyse, nous comparons les similarités fonctionnelles entre espèces et entre assemblages d'espèces à des attendus neutres, c'est-à-dire produit par la dérive génétique uniquement. Nous montrons qu'il existe de nombreuses convergences au niveau des espèces, notamment entre marsupiaux et placentaires. Par ailleurs, nous prouvons que ces convergences sont transférées au niveau des assemblages entiers d'espèces, notamment entre l'Australie et l'hémisphère nord, et particulièrement entre régions qui ont des conditions climatiques semblables. Notre étude permet donc de confirmer quantitativement des intuitions longtemps restées invérifiées.

The global biogeography of non-neutral trait evolution in mammals

F. Mazel, R. Wüest, J. Renaud, S. Lavergne and W. Thuiller

Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France.
CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France.**Introduction**

While the extraordinary diversity of life on Earth has long fascinated biogeographers (Humbolt and Bonpland 1805, Wallace 1876), the relative importance of historical contingency and determinism in explaining its evolution remains unresolved (Simpson 1944, Gould 2002, Vermeij 2006, Burbrink et al. 2012, Gillespie 2013). Determinism may arise when independent lineages experiencing the same environmental constraint show similar and predictable evolutionary trends. In particular, evolutionary theory predicts that natural selection will drive the convergence of morphological, physiological and/or behavioural traits of two unrelated species experiencing similar environmental conditions (Arendt and Reznick 2008). This prediction has been successfully tested in recently isolated systems (either lakes or true islands), with emblematic groups of species such as cyprinid fishes (Burbrink et al. 2012), tetragnatha spiders (Gillespie 2004), anole lizards (Mahler et al. 2013) or sticklebacks (Rundle et al. 2000). Species-level evolutionary convergence has also been documented at large spatial scales, for example for fishes in the Pacific (Ingram and Kai 2014) or in the Antarctic (Rutschmann et al. 2011), salamander in North America (Kozak et al. 2009) or desert rodents at a global scale (Mares 1993). Nevertheless, the extent at which these evolutionary trends are common in nature and contribute to the building of species assemblages remain discussed (Moen et al. 2009, Lomolino et al. 2010, Gillespie 2013). Unexpectedly, the existence of species-level convergence for entire and large clades such as all mammals of the world has never been evaluated in a quantitative way. Instead, it is usually accepted that, for example, placental and marsupial mammals harbour extreme cases of species convergence (Lomolino et al. 2010) but this has never been evaluated on entire species assemblages (e.g. all species from distinct regions). Indeed, species-level convergence events are predicted to scale up to convergence of entire assemblages if distinct species converge to similar ecological functions in distinct regions (Ingram and Kai 2014). Some empirical work suggests that assemblage-level convergence exist but the amount of evidences remain scarce. For example, some desert lizards (Melville et al. 2006) and stream fishes (Ibañez et al. 2009) communities from distantly related regions appear to show

assemblage scale evolutionary convergence. In the case of mammals, it has been common to use species level convergence events to suggest that placental and marsupial convergences scale up to entire assemblages, for example between Australia and North America (Lomolino et al. 2010). This expectation has never been properly tested. Instead, arguments for assemblage level convergence between Australia and North America have been mostly graphical (see Figure 9 in the introduction of this chapter), emphasising remarkable pairs of highly convergent species (discussed in Lomolino et al. 2010). However, those examples does not represent the whole mammalian fauna of the two continents but rather show the most striking examples of species convergence, and published studies have neither used quantitative data of functional trait of species, nor neutral expectations of evolution.

In this paper, we build on the most up-to date data on functional traits and phylogeny for all mammals of the world (~4600 mammal species) to test the extent of convergence at the assemblage level. More specifically, we first simulate neutral trait evolution on the worldwide species-level phylogeny and ask whether observed functional distance between species appear lower than expected under the neutral model, possibly indicating deterministic convergence. To scale up from species to assemblage level, we use the IUCN range maps distribution of nearly all mammals to derive assemblage composition worldwide at coarse grain size (200*200km). We then project species assemblages in a multi-dimensional functional trait space (Petchey and Gaston 2006, Villéger et al. 2008) and estimate the functional overlap between assemblages using functional β diversity indices (Siefert et al. 2013). We then take advantage of the previously estimated neutral expectations of species traits to compute for each pair of assemblages, a distribution of functional β diversity expected under a neutral model of trait evolution. We predict that if evolutionary convergence occurs at the scale of species assemblage, some pairs of assemblage should show very low functional β diversity (i.e. high similarity) compared to neutral phylogenetic expectations and (2) will experience similar environmental conditions driving convergent functional assemblage of species.

Results

We derived evolutionary neutral expectations for mammal functional trait distances and showed that a large number of pairs of species harbour lower (or, alternatively higher) dissimilarity than expected under neutral functional divergence. This suggests important functional convergences between unrelated species (or, alternatively functional divergences, see Fig. 1). In particular, marsupial and placental showed recurrent cases of convergence toward some guilds such as predators, large herbivores or small insectivores (see Fig. 1). Other cases of convergences occur within orders, for example within Chiroptera (Supp. Mat. 2).

We then up-scaled these neutral and observed species similarities to assemblages by deriving observed and neutral functional β diversity. Lower (/higher) than expected functional β diversity is characterized by negative (/positive) SES. Overall, $SES.\beta_{sim}$ values were biased toward negative values (median of the all SES distribution is -0.63) and significantly convergent assemblages (at the 1% threshold) were more than six times more frequent than significantly divergent assemblages in respect to a model of pure neutral divergence. Importantly, we found that $SES.\beta_{sim}$ was positively and significantly related to environmental distances (slope mean estimate 0.16; 95% CI: 0.15-0.17) meaning that pairs of assemblages experiencing similar environments tended to be functionally more similar than expected under a neutral model of trait evolution. This trend appears to be even more marked when we only consider functionally similar assemblages (see lower quantile regression lines in Fig. 1) rather than functionally dissimilar ones (see higher quantile regression line in Fig. 1). We further show that these convergent and divergent pairs of assemblages were highly structured in space (Fig. 2). In particular, we found that convergent assemblages disproportionately concerned those assemblages with very high phylogenetic beta diversity (Fig 2.C), notably Australian versus northern temperate regions (Fig 2.D). Divergent assemblages appeared also geographically structured and associated temperate and tropical assemblages, in particular temperate North American and tropical South Asian assemblages (Fig 2.A).

Discussion

In this paper, we described for the first time the evolutionary convergence and divergence of mammalian species and assemblages at a global scale. We first assessed the degree of species convergence using the most up-to-date data on functional and phylogenetic similarities between species. We observed important case of convergence between placental and marsupial, as it is commonly admitted (Springer et al. 1997, Losos

2008, Lomolino et al. 2010). Indeed, we found many examples of phenotypes that have independently evolved in placentals and marsupials such as large herbivores (Cetartiodactylae Vs. Macropodidae), small flying, tree-dwelling species (Pteromyini Vs. Petaurus) or 'rodent-like' (Mus Vs. Antechinus). While our results are not surprising, they represent the first attempt to qualitatively test convergences at the scale of mammals with functional and phylogenetic data and an explicit neutral model of trait evolution.

We expected these species-level evolutionary convergences to scale-up to entire assemblages if repeated species convergence occurs in two distinct assemblages that experience similar environmental conditions. In this case, both assemblages would be independently built by *in-situ diversification* and multiple evolutionary convergence events (ISE, Moen et al. 2009, 2013). Here, we found that, for mammals on a global scale, the extreme cases of functional similarities between assemblages often included one Australian assemblage (Fig 2D). As Australian mammals are mainly composed of marsupials (besides Chiroptera), we suggest that evolutionary convergence indeed occurred at the assemblage level in mammals. Such convergences have been already discovered for other taxa such as desert lizards (Melville *et al.* 2006) or stream fishes (Ibañez *et al.* 2009). For example, Melville *et al.* (2006) showed repeated convergent evolution between one North American and one Australian assemblage of lizards. In this case, the two continents are inhabited by two unrelated clades such as migration between continents is negligible. As a consequence, lower functional dissimilarities than expected under neutral evolution are likely the product of repeated evolutionary convergence towards similar ecological functions, which shaped functionally similar species assemblages from a distinct evolutionary context.

Since dispersion of lineages between mammalian assemblages is not negligible, the functional similarities between assemblages can result from two non-exclusive processes: an *in-situ diversification* process we just described earlier (ISE), but also a process of ecologically conservative dispersal (ECD, Moen *et al.* 2009, 2013). The ecologically conservative dispersal hypothesis postulates that (1) some lineages show less similarity than expected under a neutral model of evolution because of niche conservatism (Crisp and Cook 2012) and (2) important dispersal events allow it to invade new assemblages. In this case, we also expect that the functional similarity between two invaded assemblages will be lower than expected under a BM model of evolution. As a consequence, ECD processes may also explain the pattern we found. For example, at large spatial and phylogenetic scales, Cetartiodactyls and Carnivoran

invaded South America (Simpson 1980, Flynn 1998) from North America, well after acquiring their major functional role, potentially leading to ECD. Even at smaller spatial scales, recent allopatric speciation coupled with strong niche conservatism probably prevent species from occurring in secondary sympatry due to biotic interactions limiting local coexistence of functionally similar species (Pigot and Tobias 2013) and lead to lower functional turnover that one would expect from a BM. As a consequence, it is likely that the observed relationship between ecological convergence and environmental similarity is the product of both ECD and ISE. Nevertheless we showed that for extreme functional similarities, *in-situ* diversification was likely to be important as in concerned the well-known example of marsupial and placental. Nevertheless it is true that our analysis is not designed to directly disentangle the relative role of ISE and ECD because it does not explicitly compare the timing of colonization/diversification of species and the timing of functional evolution (Moen *et al.* 2009, 2013). Such analysis is beyond the scope of this study but may deserve future investigation.

Another important result of our study is that functional similarities between assemblages were significantly related to environmental similarities. Ecological theory predict that niche-based processes should lead to functional similarity between assemblages experiencing similar environmental conditions while neutral theory (Hubbell 2001) predict that such similarities occurs at random (Siefert *et al.* 2013). As a consequence, our results suggest that niche based process may play an important role in structuring the functional composition of mammals assemblage at a global scale, either by directly triggering convergent trait evolution or via environmental filtering of lineages experiencing niche conservatism. Note that in this analysis, we built a neutral model of functional composition of assemblage worldwide by simulating traits evolution under a neutral model (BM). This contrast with classical studies on functional β diversity (e.g. Siefert *et al.* 2013) that compared observed β diversity with null and not neutral expectations (Gotelli and McGill 2006). The main difference is that neutral expectations derive from analytic model of trait evolution while null expectations derive from the shuffling of one particular attribute of the dataset (for example, the species identity in the functional dendrogram). The use of a neutral model as we did here allows to specifically emphasize the geography of (non-neutral) trait evolution by merging classical approaches of trait macroevolution with a macro-ecological perspective on species assemblages. Former studies of species assemblages have mostly used species shuffling as a null model, thus hindering to highlight and contrast alternative

evolutionary scenarios shaping species assemblages. Finally we also found cases of higher than expected functional dissimilarities between assemblages, although to a much lesser extent than the reverse. These pairs of assemblages appeared slightly dissimilar in term of climates. Spatially, they mostly concerned tropical versus temperate regions, possibly indicating divergent selection due to different climate.

In summary we show in this paper that combining a macro-evolutionary and a macro-ecological approach lead to important insights into the evolutionary and ecological assembly of mammalian faunas at a global scale. Some future studies may want to explore potential limits of our approach. For example, we simulated here neutral evolution by using a Brownian model of evolution with a single parameter of evolutionary rate. As it is likely that heterogeneous rate of trait evolution occurs in nature (Simpson 1944), especially in large datasets (O'Meara 2012), it would be interesting to compare this single-rate neutral model with a multi-rate neutral model (Mazel *et al.* 2015).

Acknowledgements

The research leading to these results had received funding from the European Research Council under the European Community's Seven Framework Programme FP7/2007-2013 Grant Agreement no. 281422 (TEEMBIO). The authors of this study belong to the Laboratoire d'Écologie Alpine, which is part of Labex OSUG@2020 (ANR10 LABX56).

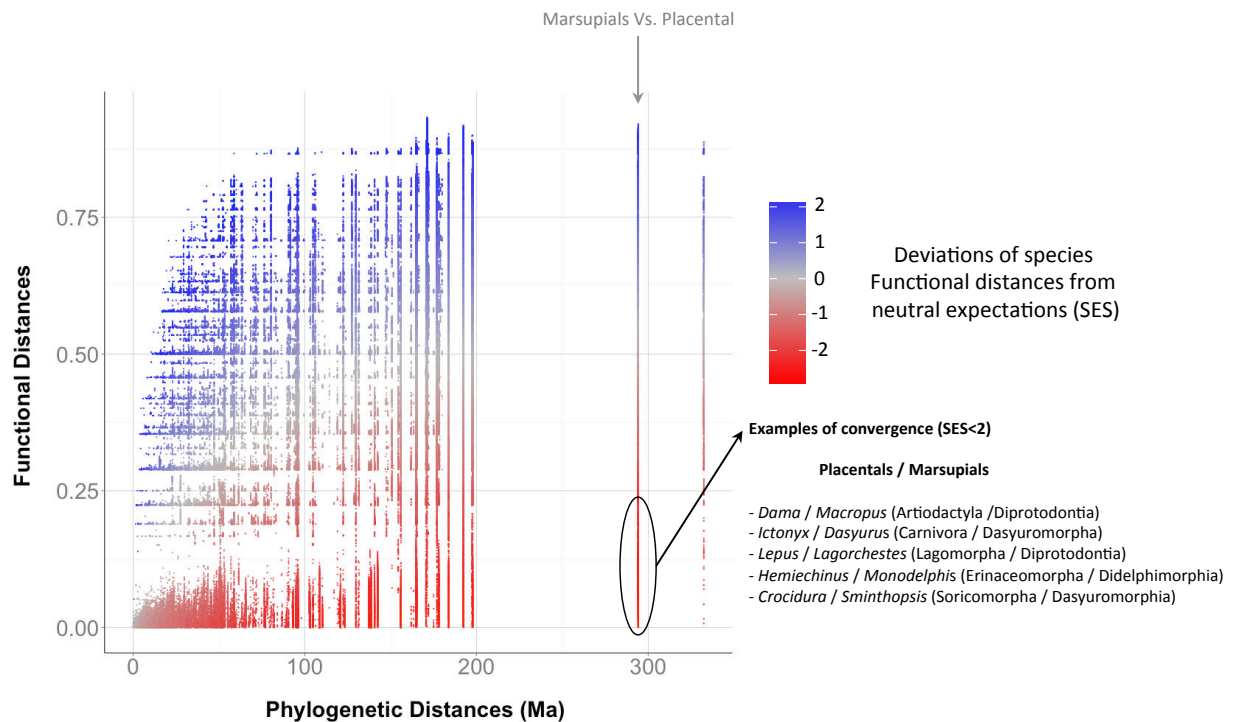


Figure 1. Relationship between phylogenetic and functional distances at the scale of mammalian species. The figure compares the phylogenetic distances (X-axis, in Ma) and the functional distance (Y-axis) between all species of mammals used in this study. Each pair of species is coloured according to the deviation (Standard Effect Size) of functional distance to a neutral model of evolution (Brownian motion). For placental vs. marsupial comparison, we give example of some convergent pairs of species (by giving their genus and order).

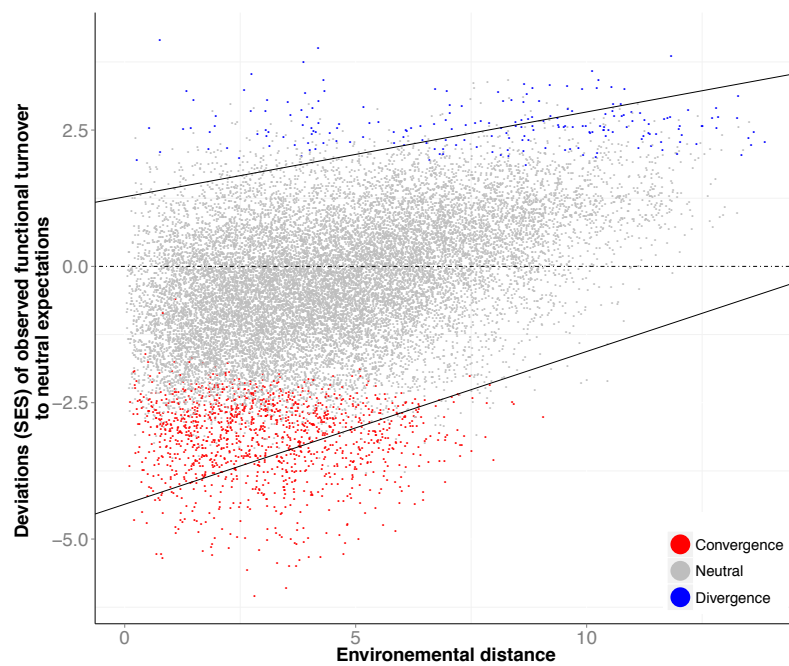


Figure 2. Relationship between the functional and environmental dissimilarities of assemblages. Deviations of observed functional β diversity to a neutral model of evolution (Standard Effect Sizes) is plotted against environmental dissimilarities between assemblages worldwide. Environmental dissimilarities are Euclidian distance between assemblages in an environmental space of the two first axes of a PCA of all Worldclim variables. Black lines correspond to lower (1%) and higher (99%) quantile regression fits. If the observed functional β diversity between two assemblages is significantly lower (higher) than neutral expectations, points are coloured in red (blue).

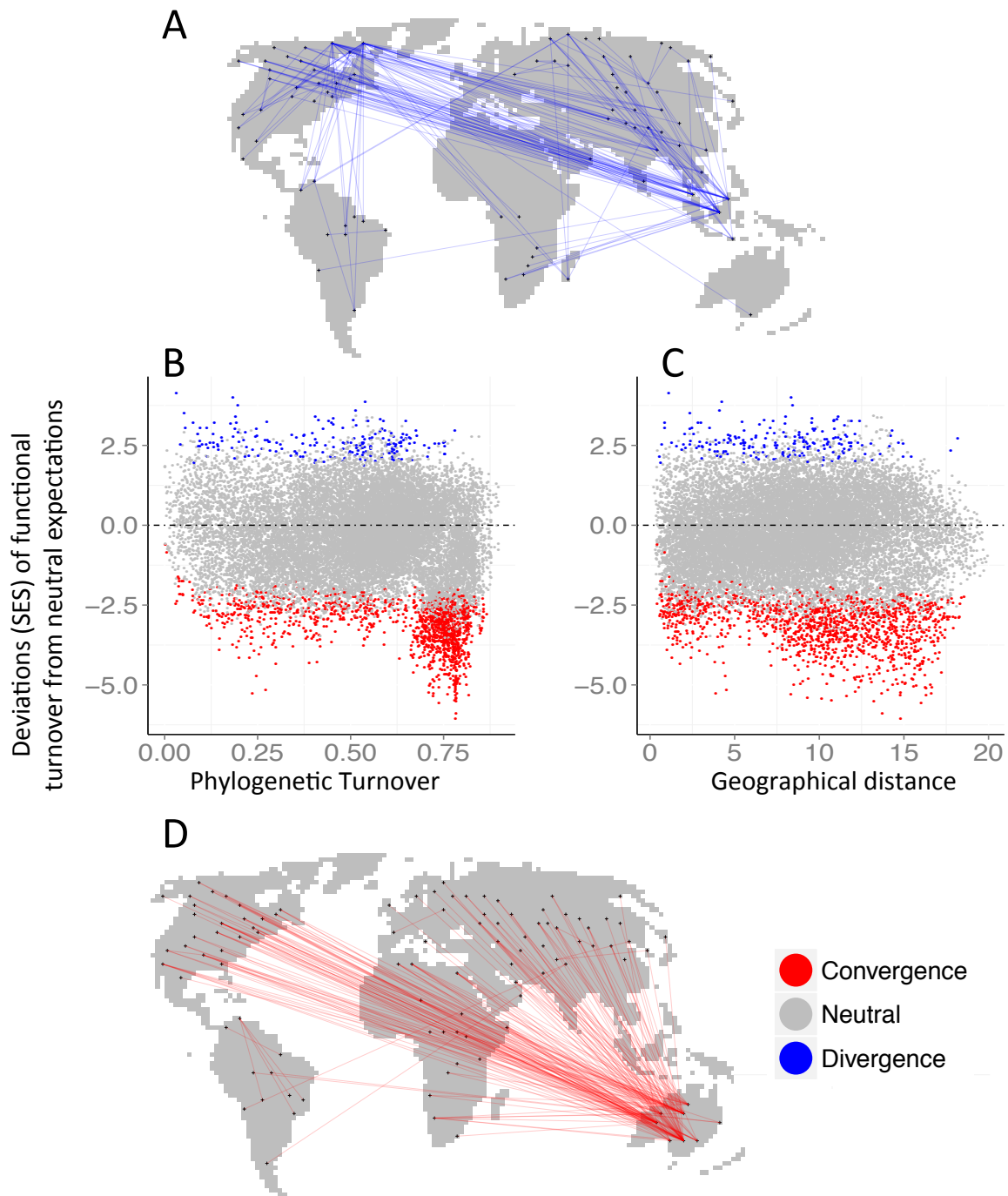


Figure 3. Spatial patterns of non-neutral functional dissimilarities between assemblages. Deviations of observed functional β diversity to a neutral model of evolution (Standard Effect Sizes) is plotted against phylogenetic β diversity (B) or geographical distances (C) between assemblages. The figure also depicts the top 1% highest and lowest deviations of observed functional β diversity to neutral expectations in subplot A (blue lines) and D (red lines) respectively.

Material and Method

Data

Phylogenetic and functional distances between species

To derive the phylogenetic distances between species, we used the calibrated and dated ultrametric phylogenetic tree updated by Fritz et al. (2009) based on Bininda-Emonds et al. (2007). We then updated the tree by replacing the Carnivora clade in this phylogeny with a highly resolved supertree published more recently (Nyakatura and Bininda-Emonds 2012).

To derive the functional distances between species, we used four different types of traits (diet, body mass, activity cycle and foraging height) extracted from the online database provided by Wilman et al. (2014) which represents appropriate niche dimension within assemblages (Belmaker and Jetz 2015). Because they ignore intraspecific variations, we may overestimate niche overlap between sympatric species but we assume that, at the scale of our study, intraspecific variations remains negligible compared to interspecific variations. Since we used both continuous and discrete variables we used the Gower metric to compute distances between each pairs of species using the function *dist.ktab* of the *ade4* package (Pavoine et al. 2009). Each four category was given equal weight. For continuous variable (body mass), Euclidian distance was used while for multi-choice nominal variables, we used the coefficient of Gower and Legendre (1986).

Distributions maps

We derived occurrence data for 4616 mammals worldwide for 3616 grid cells of approximately 200*200 km using the distribution maps provided by the Mammal Red List Assessment (<http://www.iucnredlist.org/>). The best resolution to use these maps is still under discussion in the literature (Storch et al. 2012, Jenkins et al. 2013). We here used a resolution commonly used at global scale (Safi et al. 2011, Holt et al. 2013). Domestic, aquatic and semi aquatic mammals were excluded from the analysis.

Stratified sample

Because of computational constrains, we simplified our analyses by subsampling 200 pixels out of 3616. We choose these pixels in order to maximise the representativeness of the environmental and the spatial space. The environmental space was summarized by the first two axes of a PCA on the

global Worldclim variables. We used the cube method (Deville 2004) based on the PCA axes scores and the geographic position of each pixels (see Supp. Mat. 1) to select 200 representative pixels.

Species-level Analysis

To derive neutral expectations of functional distances between species, we simulated 100 sets of traits values under a neutral model of evolution, namely the Brownian Motion model (BM; Edwards and Cavalli-Sforza 1964). The BM model explicitly simulates drift along the mammals phylogenetic tree and is thus adequate to test for convergent evolution (Melville et al. 2006, Muschick et al. 2012, Ingram and Kai 2014). To calibrate the BM model before the simulation, we fitted it to each trait independently. We then used these fitted parameters to simulate neutral traits values and to compute, for each pair of species, a distribution of neutral distances. We then assessed to which extent observed functional distances (see below) departed from the neutral distribution by computing the standard effect size functional distance, as traditionally used in the case of null models in ecology (Eq. 1):

$$ses. distance = \frac{Distance_{obs} - mean(Distance_{null})}{sd(Distance_{null})} \quad \text{Eq. 1}$$

Assemblage-level Analysis

β Diversity Metric.

To quantify the functional (and phylogenetic) similarities between assemblages we used a functional and phylogenetic version of the classical Simpson index. The Simpson index was initially defined in term of species similarities between two assemblages (Eq 2):

$$\beta_{sim} = \frac{\min(b,c)}{a + \min(b,c)} \quad \text{Eq.2}$$

where *a* = the number of species shared by the two grid cells and *b* and *c* represent the number of species unique to each grid cell. It can be easily extended to phylogenetic similarities by considering branch length instead of species (Bryant et al. 2008, Leprieur et al. 2012). In this case, *a* becomes the length of branches shared by the two grid cells and *b* and *c* represent the length of branches unique to each grid cell. In a functional context, shared functional characteristics may be viewed in terms of shared functional space (Villéger et al. 2013) or shared functional dendrogram branch lengths (Petchey and Gaston 2007, Leprieur et al. 2012). Here we use a functional dendrogram to depict

distance as commonly done in the literature (Safi et al. 2011, Thuiller et al. 2014, Belmaker and Jetz 2015) and because it facilitate comparison with phylogenetic similarities. To convert the functional distance into an ultrametric dendrogram, we applied a hierarchical cluster algorithm using the UPGMA method (function *hclust* in R). Note that initial functional distances and (transformed) dendrogram functional distances were highly correlated (pearson mantel $r=0.88$, $p=0.001$ based on 999 permutations).

We choose to use the β_{sim} metric because of two reasons. First, it quantifies the degree of replacement of functional branch length (Gaston and Blackburn 2008, Baselga 2010) without being affected by the difference of richness between assemblages (Kreft and Jetz 2010, Holt et al. 2013, Mouillot et al. 2013). Indeed it represents the true-turnover component of the widely used PhyloSor metric, the later being a functional (phylogenetic) version of the Sørensen similarity metric (Sørensen 1948, Bryant et al. 2008, Baselga 2010, Leprieur et al. 2012). Second, it mainly quantifies functional (resp. phylogenetic) turnover between functionally (resp. phylogenetically) close species in both assemblages ('terminal' metric sensu Swenson 2011), for example between large herbivores (or between close relatives in its phylogenetic version). This is very important here since we aim at comparing similar functional types between the two assemblages, and not comparing all species of one assemblage with all species of the other assemblage, which will be the case if we would have used the COMDIST (Webb et al. 2008) or Pst metric (Hardy and Senterre 2007).

Observed and neutral β Diversities.

We used β_{sim} to compute observed functional and phylogenetic turnover between assemblages of the world. We used our neutral trait expectations to derive a distribution of neutral functional β_{sim} values for each pair of assemblages. Deviations from these expectations were computed using standard effect size ($SES.\beta_{sim}$, see Eq. 1) and significance assessed using the relative rank of observed value compared to neutral values as we will do under a null model.

Statistical relationships between assemblage convergence and environment

To relate $SES.\beta_{sim}$ to environmental dissimilarities we used two approaches. First, we fitted quantile regressions using the *rq* function of the *quantreg* R package. Because the observations are not statistically independent from each other (they corresponds to pairs of assemblages), we also used generalized linear mixed models with pairs of sites

set as a random effect, fitted within a Bayesian framework (MCMCglmm, Hadfield 2010). Each MCMC chain was ran for a total of 50,000 generations, retaining estimated parameters every 25 steps (we verified this yielded a low autocorrelation between estimates), and removing the first 25% sampled parameters as a burnin period.

References

- Arendt, J. and Reznick, D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? - *Trends Ecol. Evol.* 23: 26–32.
- Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. - *Glob. Ecol. Biogeogr.* 19: 134–143.
- Belmaker, J. and Jetz, W. 2015. Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients (H Arita, Ed.). - *Ecol. Lett.* 18: 563–571.
- Bininda-Emonds, O. R. P. et al. 2007. The delayed rise of present-day mammals. - *Nature* 446: 507–12.
- Bryant, J. A. et al. 2008. Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. - *Proc. Natl. Acad. Sci. U. S. A.* 105: 11505–11.
- Burbrink, F. T. et al. 2012. Evidence for determinism in species diversification and contingency in phenotypic evolution during adaptive radiation. - *Proc. Biol. Sci.* 279: 4817–26.
- Crisp, M. D. and Cook, L. G. 2012. Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? - *New Phytol.* 196: 681–94.
- Deville, J.-C. 2004. Efficient balanced sampling: The cube method. - *Biometrika* 91: 893–912.
- Edwards, A. W. F. and Cavalli-Sforza, L. L. 1964. Reconstruction of evolutionary trees. - In: Heywood, V. H. and McNeill, J. (eds), *Phenetic and phylogenetic classification*. pp. 67–76.
- Flynn, J. 1998. Recent advances in South American mammalian paleontology. - *Trends Ecol. Evol.* 13: 449–454.
- Fritz, S. A. et al. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. - *Ecol. Lett.* 12: 538–49.
- Gaston, K. and Blackburn, T. 2008. *Pattern and Process in Macroecology*. - Wiley & Sons.
- Gillespie, R. 2004. Community assembly through adaptive radiation in Hawaiian spiders. - *Science* 303: 356–9.
- Gillespie, R. G. 2013. Adaptive radiation: convergence and non-equilibrium. - *Curr. Biol.* 23: R71–4.
- Gotelli, N. and McGill, B. 2006. Null versus neutral models: what's the difference? - *Ecography (Cop.)* 29: 793–800.
- Gould, S. J. 2002. *The Structure of Evolutionary Theory*. - Harvard University Press.
- Gower, J. C. and Legendre, P. 1986. Metric and Euclidean properties of dissimilarity coefficients. - *J. Classif.* 3: 5–48.
- Hadfield, J. D. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. - *J. Stat. Softw.* 33: 1–22.
- Hardy, O. J. and Senterre, B. 2007. Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. - *J. Ecol.* 95: 493–506.
- Holt, B. G. et al. 2013. An Update of Wallace's Zoogeographic Regions of the World. - *Science (80-.)* 339: 74–78.
- Hubbell, S. 2001. *The unified neutral theory of biodiversity and biogeography*. - Princeton University Press.
- Humboldt, A. von and Bonpland, A. 1805. *Essai sur la géographie des plantes : accompagné d'un tableau physique des régions équinoxiales, fondé sur des mesures exécutées, depuis le dixième degré de latitude boréale jusqu'au dixième degré de latitude australe, pendant les années 1799, 1800, 1801.* - Levrault Schoell, Paris.
- Ibañez, C. et al. 2009. Convergence of temperate and tropical stream fish assemblages. - *Ecography (Cop.)* 32: 658–670.

Partie III.2 – Chapitre VII

- Ingram, T. and Kai, Y. 2014. The geography of morphological convergence in the radiations of Pacific *Sebastes* rockfishes. - *Am. Nat.* 184: E115–31.
- Jenkins, C. N. et al. 2013. Global patterns of terrestrial vertebrate diversity and conservation. - *Proc. Natl. Acad. Sci. U. S. A.* 110: E2602–2610.
- Kozak, K. H. et al. 2009. Can parallel diversification occur in sympatry? Repeated patterns of body-size evolution in coexisting clades of north american salamanders. - *Evolution* (N. Y.) 63: 1769–1784.
- Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species distributions. - *J. Biogeogr.* 37: 2029–2053.
- Leprieux, F. et al. 2012. Quantifying phylogenetic beta diversity: distinguishing between “true” turnover of lineages and phylogenetic diversity gradients. - *PLoS One* 7: e42760.
- Lomolino, M. V. et al. 2010. *Biogeography*. - Sinauer Associates.
- Losos, J. B. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. - *Ecol. Lett.* 11: 995–1003.
- Mahler, D. et al. 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. - *Science* (80-.). 341: 292–295.
- Mares, M. A. 1993. Desert rodents, seed consumption, and convergence: evolutionary shuffling of adaptations. - *Bioscience* 43: 372–379.
- Mazel, F. et al. 2015. Improving phylogenetic regression under complex models of evolution. *Ecology*. Accepted
- Melville, J. et al. 2006. Intercontinental community convergence of ecology and morphology in desert lizards. - *Proc. Biol. Sci.* 273: 557–63.
- Moen, D. S. et al. 2009. Community assembly through evolutionary diversification and dispersal in Middle American treefrogs. - *Evolution* 63: 3228–47.
- Moen, D. S. et al. 2013. Evolutionary conservatism and convergence both lead to striking similarity in ecology, morphology and performance across continents in frogs. - *Proc. Biol. Sci.* 280: 20132156.
- Mouillot, D. et al. 2013. The challenge of delineating biogeographical regions: nestedness matters for Indo-Pacific coral reef fishes (W Jetz, Ed.). - *J. Biogeogr.* 40: 2228–2237.
- Muschick, M. et al. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. - *Curr. Biol.* 22: 2362–8.
- Nyakatura, K. and Bininda-Emonds, O. R. P. 2012. Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. - *BMC Biol.* 10: 12.
- O'Meara, B. 2012. Evolutionary inferences from phylogenies: a review of methods. - *Annu. Rev. Ecol. Evol. Syst.* 43: 267–285.
- Pavoine, S. et al. 2009. On the challenge of treating various types of variables: application for improving the measurement of functional diversity. - *Oikos* 118: 391–402.
- Petchey, O. L. and Gaston, K. J. 2006. Functional diversity: back to basics and looking forward. - *Ecol. Lett.* 9: 741–758.
- Petchey, O. L. and Gaston, K. J. 2007. Dendrograms and measuring functional diversity. - *Oikos* 116: 1422–1426.
- Pigot, A. L. and Tobias, J. A. 2013. Species interactions constrain geographic range expansion over evolutionary time. - *Ecol. Lett.* 16: 330–8.
- Rundle, H. D. et al. 2000. Natural selection and parallel speciation in sympatric sticklebacks. - *Science* 287: 306–8.
- Rutschmann, S. et al. 2011. Parallel ecological diversification in Antarctic notothenioid fishes as evidence for adaptive radiation. - *Mol. Ecol.* 20: 4707–21.
- Safi, K. et al. 2011. Understanding global patterns of mammalian functional and phylogenetic diversity. - *Philos. Trans. R. Soc. London, Ser. B* 366: 2536–2544.
- Siefert, A. et al. 2013. Functional beta-diversity patterns reveal deterministic community assembly processes in eastern North American trees. - *Glob. Ecol. Biogeogr.* 22: 682–691.
- Simpson, G. G. 1944. *Tempo and Mode in Evolution*. - Columbia University Press.
- Simpson, G. G. 1980. *Splendid isolation: the curious history of South American mammals*. - New Haven: Yale University Press.
- Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. - *I kommission hos E. Munksgaard*.
- Springer, M. S. et al. 1997. The chronicle of marsupial evolution. - In: Givnish, T. and Sytsma, K. (eds), *Molecular Evolution and Adaptive Radiations*. Cambridge University Press., pp. 129–161.
- Storch, D. et al. 2012. Universal species-area and endemics-area relationships at continental scales. - *Nature* 488: 78–81.
- Swenson, N. G. 2011. Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. - *PLoS One* 6: e21264.
- Thuiller, W. et al. 2014. The European functional tree of bird life in the face of global change. - *Nat. Commun.* 5: 3118.
- Vermeij, G. J. 2006. Historical contingency and the purported uniqueness of evolutionary innovations. - *Proc. Natl. Acad. Sci. U. S. A.* 103: 1804–9.
- Villéger, S. et al. 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. - *Ecology* 89: 2290–2301.
- Villéger, S. et al. 2013. Decomposing functional β -diversity reveals that low functional β -diversity is driven by low functional turnover in European fish assemblages. - *Glob. Ecol. Biogeogr.* 22: 671–681.
- Wallace, A. 1876. *The geographical distribution of animals*. - Cambridge Univ. Press.
- Webb, C. O. et al. 2008. *Phylocom: software for the analysis of phylogenetic community structure and trait evolution*. - *Bioinformatics* 24: 2098–2100.
- Wilman, H. et al. 2014. *EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals*. - *Ecology* 95: 2027–2027.

SYNTHÈSE & DISCUSSION DE LA SOUS-PARTIE III.2

Dans cette dernière partie, nous proposons de nouvelles solutions méthodologiques pour détecter la convergence au niveau des espèces (annexe 4). Nous étudions ensuite la convergence au niveau des assemblages pour les mammifères à travers le monde.

Il est important de souligner que l'analyse du chapitre VII ne suit pas les recommandations de l'annexe 3.2.1 et qu'il faudra donc veiller à vérifier nos résultats. En effet, nous montrons dans l'annexe 3.2.1 que l'hétérogénéité des taux d'évolution browniens est un processus important à prendre en compte pour modéliser de façon fidèle une évolution neutre. Dans l'état actuel de l'analyse des convergences de mammifères, nous ne prenons pas en compte cette possibilité ce qui impacte probablement certaines de nos conclusions. Je pense cependant que la convergence (au niveau des espèces et au niveau des assemblages) entre marsupiaux et placentaires ne reflète pas cette variabilité des taux d'évolution. Il faudra cependant développer des attentes neutres relaxant l'hypothèse d'homogénéité des taux d'évolution pour en avoir le cœur net.

SYNTHÈSE & DISCUSSION DE LA PARTIE III

La Partie III de ma thèse visait à mieux comprendre la distribution spatiale des mammifères et des oiseaux à l'échelle du globe en tirant profit de l'utilisation des indices de diversité phylogénétique et fonctionnelle. Cette discussion s'attache à rappeler les principaux résultats de cette partie en les reliant à quelques perspectives de travail. En particulier, je discute l'intérêt de l'importance de l'échelle spatiale, de la comparaison de groupes différents, et de nouvelles perspectives de comparaison entre PD et FD.

1. Travailler à différentes échelles spatiales

Dans cette partie III, nous nous sommes contentés de travailler à une seule échelle spatiale, mais il apparaît que l'importance relative des processus historiques et écologiques peut varier le long de l'échelle spatiale (Levin 1992; Chave 2013). Ainsi, on pourrait réanalyser les profils de BDTT à différentes résolutions spatiales, par exemple, en utilisant les grains d'études utilisés dans le chapitre III (article 4, de 100*100km à 1000*1000km). On pourrait par exemple s'attendre à avoir un effet prédominant de l'histoire à vaste échelle spatiale et un effet prédominant de l'environnement à échelle intermédiaire.

2. Comparer la BDTT de groupes taxonomiques distincts

2.1. En biogéographie à large échelle

La sous-partie III.1 montre comment une description de la composition des assemblages à plusieurs échelles phylogénétiques permet de décortiquer l'influence relative de différents processus majeurs, notamment les processus de type historique (dispersion à large échelle, vicariance) et les processus écologiques liés à la niche des espèces. En particulier, nous montrons comment ceux-ci peuvent avoir des effets emboîtés ou indépendants le long de l'échelle phylogénétique : la biogéographie historique conditionne la distribution des lignées à large échelle phylogénétique alors que les processus de niche la conditionnent à plus fines échelles. En définitive, cette

interprétation est en accord avec le fait que l'histoire conditionne la formation des royaumes (Kreft & Jetz 2010; Holt *et al.* 2013), et qu'à l'intérieur de ces royaumes l'environnement définit des biomes différents qui présentent des faunes distinctes, adaptées à leurs conditions environnementales. En fait, nos résultats soutiennent cette interprétation pour les mammifères, mais pas pour les oiseaux. Pour ces derniers, il apparaît en effet que les processus de type historique sont moins importants à large échelle phylogénétique. On peut tenter de relier ces résultats avec les capacités relatives de dispersion des mammifères et des oiseaux. Il est en effet possible que des capacités supérieures de dispersion conduisent à un effacement du signal de biogéographie historique dans les branches profondes de l'arbre. En effet, des évènements de colonisation multiples vont potentiellement homogénéiser la distribution des lignées profondes et brouiller le signal de l'histoire ancienne sur les patrons actuels. De manière générale, les oiseaux semblent en effet dotés de meilleures capacités de dispersion que les mammifères, ce qui permet d'expliquer l'absence de signal historique pour les oiseaux, mais pas pour les mammifères. Par ailleurs, l'effet de la proximité historique des continents a laissé une empreinte dans les distributions actuelles de mammifères, mais pas sur celles des oiseaux, comme attendu. En définitive, il est donc intéressant de contraster les importances relatives de différents processus entre groupes aux caractéristiques écologiques différentes. C'est par exemple ce que proposent Sanmartín & Ronquist (2004) en comparant l'histoire biogéographique des plantes et des animaux dans l'hémisphère sud et en montrant que la distribution des animaux est le produit d'une vicariance ancienne alors que celle des végétaux est davantage conditionnée par les multiples évènements de dispersion récents. De telles comparaisons représentent des pistes de travail intéressantes pour le futur. Au sein des mammifères et des oiseaux, l'analyse de la variabilité entre les ordres (Figure 2 du matériel supplémentaire 3.1.1) révèle d'ores et déjà des résultats intéressants qui mériteront d'être creusés. En particulier, il apparaît intéressant de contraster la variabilité de l'effet de l'espace et de l'environnement sur les profils de BDTT. On peut prédire que les effets géographiques (interprétés ici comme historiques) seront davantage variables que les effets environnementaux, car ils dépendent d'évènements rares, difficiles à prévoir (« contingent ») alors que les processus de niches devraient être plus constants, car ils sont déterministes. Cependant, on pourrait aussi supposer que les effets historiques entre les ordres dépendent des caractéristiques de dispersion de ces ordres. Au-delà des

mammifères et des oiseaux, il existe quelques groupes pour lesquels des données géographiques, phylogénétiques et parfois fonctionnelles existent à l'échelle du globe. On pourrait par exemple ajouter à notre analyse des groupes taxonomiques tels que les poissons coralliens (Mouillot *et al.* 2013) ou les champignons (Tedersoo *et al.* 2014).

2.2. En microbiologie, pour l'étude des microbiotes

Nous montrons dans le chapitre VI comment une analyse BDTT permet de préciser l'importance relative de la phylogénie et du régime alimentaire de l'hôte sur la composition du microbiote intestinal. Nous montrons même que ces conclusions restent valides dans les deux sous clades majoritaires de notre analyse, c'est-à-dire les Firmicutes et les Bacteroidetes (voir les figures supplémentaires 12 et 13 en annexe 3.1.2). Cependant, il serait intéressant de se demander si certains clades bactériens ont plutôt tendance à être liés au régime alimentaire de l'hôte, ou suivent plutôt la phylogénie. Dans ce dernier cas par exemple, il serait intéressant de regarder la distribution des OTUs à forte transmission verticale (ou horizontale) au sein de la phylogénie des bactéries, dans un premier en temps en caractérisant le signal phylogénétique du type de transmission et ensuite en tentant de relier ces taux aux caractéristiques fonctionnelles des bactéries (capacité de dispersion par exemple). Par ailleurs on pourrait tester si ces lignées qui co-spécient beaucoup avec l'hôte (i.e. celles liées à la phylogénie de l'hôte) sont aussi celles pour lesquelles il a été montré une forte héritabilité en intraspécifique (voir Goodrich *et al.*, 2014, pour le cas humain). En d'autres termes, observe-t-on un lien entre micro et macroévolution dans ce cas ?

Par ailleurs, nous avons limité notre étude au seul microbiote intestinal. Cependant, il en existe d'autres chez les animaux, comme ceux de la peau ou de la bouche. Il serait donc intéressant de les inclure dans les futures analyses afin de proposer une vision plus complète de l'évolution de l'ensemble de notre microbiote.

3. Combiner les diversités phylogénétique et fonctionnelle

L'utilisation combinée des diversités phylogénétique et fonctionnelle est prometteuse. Dans cette partie, je discute dans un premier temps l'avènement des données fonctionnelles en microbiologie et dans un second temps je propose des perspectives sur la façon d'utiliser les deux facettes conjointement.

3.1. Combiner les informations fonctionnelles et phylogénétiques dans l'étude des microbiotes

Notre analyse des microbiotes intestinaux de mammifères s'est basée sur la composition phylogénétique des assemblages (en utilisant une phylogénie des séquences d'ARN ribosomal 16S) et leur lien avec les caractéristiques de l'hôte. Cependant, tout comme l'analyse biogéographique de la même sous-partie (chapitre VI), nous n'avons pas utilisé de données fonctionnelles qui pourraient pourtant préciser la nature des liens entre hôtes et symbiontes. Étant donné qu'il reste difficile de cultiver en laboratoire la plupart des bactéries du tractus intestinal des mammifères, on ne peut pas mesurer directement les caractéristiques fonctionnelles des OTUS bactériens. Cependant, il est possible d'estimer le contenu génique des microbiotes (c'est-à-dire le microbiome*), notamment via le séquençage métagénomique de nouvelle génération (shotgun). Même si la présence d'un gène ne garantit en aucun cas qu'il soit transcrit en ARN, traduit en protéines, et donc qu'il participe au phénotype de la bactérie, il s'agit d'une première approche intéressante (Lozupone *et al.* 2012). Par exemple, Muegge *et al.* (2011) montrent que les microbiotes des mammifères herbivores et des carnivores contiennent des gènes très différents. En particulier, ils montrent que les herbivores possèdent davantage d'enzymes anaboliques des acides aminés alors que les carnivores contiennent davantage d'enzymes cataboliques. En d'autres termes, les herbivores possèdent un microbiote qui les aide à construire des acides aminés à partir de composés carbonés plus simples alors que les carnivores ont un microbiote qui leur permet plutôt de cataboliser les acides aminés pour en retirer de l'énergie. Ceci correspond bien au fait que le régime alimentaire des carnivores est riche en acide aminé, au contraire des herbivores. Ceci représente donc un formidable exemple de coévolution entre les hôtes et leur symbiontes, qui n'aurait pas pu être révélé par

l'analyse des seules séquences 16S. De façon générale, il semble que les fonctions soient généralement conservées alors que les lignées bactériennes ne le sont pas (Lozupone *et al.* 2012). Il serait alors intéressant de se demander (1) à quel point les fonctions sont généralement conservées ? (L'équivalent du débat sur la prévalence du conservatisme de niche en macroévolution), (2) sont-elles conservées à différentes échelles phylogénétiques ? (3) À quel point ce conservatisme varie-t-il en fonction de l'échelle phylogénétique et de la fonction considérée (par exemple, la fonction ribosomale est forcément très largement conservée) ?

3.2. Expliciter les modèles macro-évolutifs en biogéographie

La sous-partie III.2 tente de croiser l'information phylogénétique et l'information fonctionnelle pour mieux comprendre l'évolution des traits et leur répercussion sur les patrons biogéographiques. En général, l'utilisation conjointe de la PD et de la FD propose de contraster leurs valeurs, mais sans faire directement appel à des modèles d'évolution explicites (Safi *et al.* 2011; Weinstein *et al.* 2014). Par exemple, Safi *et al.* (2011) utilisent un modèle additif généralisé (GAM) pour modéliser la relation entre la PD (Faith 1992) et la FD (Petchey & Gaston 2007) et en déduire des résidus de FD indépendants de la PD. Il apparaît ainsi que les zones tropicales possèdent moins de diversité fonctionnelle « qu'attendu » selon leur diversité phylogénétique, ce qui selon les auteurs pourrait être le résultat d'une extinction plus importante ou d'une évolution plus rapide dans les régions tempérées et d'une évolution plus lente dans les régions tropicales (« niche conservatism »). Ceci représente un premier pas important dans l'analyse conjointe des diversités phylogénétiques et fonctionnelles à large échelle spatiale. Cependant, je pense que l'on pourrait tirer profit de l'utilisation de modèles d'évolution développés en macroévolution. En effet, de tels modèles permettent de tester différents scénarii d'évolution avec des hypothèses claires et précises et représentent donc le lien direct entre l'espace fonctionnel et l'arbre phylogénétique. Par exemple le modèle brownien (Edwards & Cavalli-Sforza 1964) et ses diverses extensions (Pagel 1999; Eastman *et al.* 2011) permettent de modéliser l'effet de dérive seule et représentent donc des modèles de choix pour étudier les déviations à la neutralité. Le modèle d'*Ornstein-Uhlenbeck* (Hansen 1997) permet lui de modéliser la

sélection stabilisante, mais aussi les convergences évolutives (Ingram & Kai 2014). En utilisant de tels outils, je pense que l'on pourrait relier de façon plus claire la PD et la FD, notamment en explicitant une FD attendue en fonction d'une phylogénie, d'une répartition spatiale et d'un modèle d'évolution. Il me semble qu'avec cette approche, on peut aller plus loin dans la compréhension des processus sous-jacents, puisque l'on ne s'intéresse pas uniquement à l'éradication de « l'effet de la phylogénie », mais bien à décrire explicitement les processus en jeu et les hypothèses que l'on assume. En macroévolution, cette dichotomie d'approche se reflète notamment dans les méthodes utilisées, qui peuvent être explicites sur les processus modélisés (voir les différents modèles présentés plus haut) ou non (voir par exemple les techniques de vecteur propres phylogénétiques, Diniz-Filho *et al.* 1998). Il apparaît que lorsque des modèles explicites sont disponibles, leur utilisation devrait être préférée (voir Freckleton *et al.* 2011 pour une critique des modèles non explicites en macroévolution). C'est notamment ce que j'ai tenté de faire dans le chapitre VII de cette thèse en comparant un modèle neutre explicite aux données observées (mais voir la synthèse du chapitre VII pour une discussion plus détaillée sur le type de modèle neutre à utiliser).

Dans certains cas, les modèles explicites n'existent tout simplement pas et on se cantonne généralement à « gommer » l'effet de ces processus, qui sont alors souvent perçus comme « indésirables ». Deux exemples indépendants sur la considération de la dispersion en biogéographie appuient cette proposition.

(1) Le premier provient du traitement de la dimension spatiale en biogéographie écologique. Dans ce domaine, de nombreuses études tentent de relier des variables écologiques (présence ou abondance d'espèces, richesse spécifique) aux variables environnementales, notamment afin de confirmer l'importance de la niche écologique. La dimension spatiale de l'analyse est souvent réduite à un « danger » que l'on appelle l'autocorrélation spatiale* (Dormann *et al.* 2007) qui peut systématiquement biaiser les tests de significativité. Ainsi, il s'agit souvent de retirer cet effet spatial, qui peut être le résultat de la dispersion limitée ou des interactions biotiques, mais pas de le modéliser explicitement (Peres-Neto & Legendre 2010; Dray *et al.* 2012).

(2) Le deuxième provient du traitement de la dispersion en biogéographie historique. Dans ce domaine, on retrouve deux écoles de pensées dont les approches (Sanmartín

2012) considèrent la dispersion soit comme un bruit de fond sans intérêt (modèles « process-based », Morrone 2009), soit comme un processus à part entière (modèles « event-based », Ronquist & Sanmartín 2011).

En définitive, l'utilisation de modèles explicites n'est pas toujours possible, car il n'existe pas pour tous les systèmes d'études, et notamment pour les plus complexes. Ceci est typiquement le cas en écologie des communautés, où il n'existe pas de modèle mathématique explicite à comparer aux observations et les modèles nuls (ou un des paramètres des données est randomisé) sont davantage utilisés (voir Gotelli & McGill 2006 pour une discussion sur la différence entre modèles nuls et modèles neutres). Cependant, des modèles explicitement neutres commencent à émerger, notamment le modèle neutre de Hubbel (Hubbell 2001; Rosindell *et al.* 2011; mais voir aussi Pigot & Etienne 2014; O'Dwyer *et al.* 2015) et semblent prometteurs.

BIBLIOGRAPHIE GÉNÉRALE
DE LA PARTIE III
(HORS ARTICLES)

Partie III – Bibliographie

- Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A. & Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. *Science (New York, N.Y.)*, **307**, 1915–20.
- Bacon, C.D., Silvestro, D., Jaramillo, C., Smith, B.T., Chakrabarty, P. & Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 6110–5.
- Belmaker, J. & Jetz, W. (2015). Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients (H. Arita, Ed.). *Ecology Letters*, **18**, 563–571.
- Brooks, D.R. (1981). Hennig's Parasitological Method: A Proposed Solution. *Systematic Zoology*, **30**, 229–249.
- Burnham, R.J. & Johnson, K.R. (2004). South American palaeobotany and the origins of neotropical rainforests. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 1595–610.
- Cavender-Bares, J., Ackerly, D.D. & Kozak, K.H. (2012). Integrating ecology and phylogenetics: the footprint of history in modern-day communities 1. *Ecology*, **93**, 1–3.
- Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecology letters*, **12**, 693–715.
- Chave, J. (2013). The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecology letters*, **16 Suppl 1**, 4–16.
- Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J.M. & Relman, D.A. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science (New York, N.Y.)*, **336**, 1255–62.
- Darwin, C. (1859). *The origin of species by means of natural selection, or, The preservation of favored races in the struggle for life*. (J. Murray, Ed.). The Werner Company.
- Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., González, A. & Knight, R. (2014). Convergence of gut microbiomes in myrmecophagous mammals. *Molecular ecology*, **23**, 1301–17.
- Diniz-Filho, J.A.F., de Sant'Ana, C.E.R. & Bini, L.M. (1998). An Eigenvector Method for Estimating Phylogenetic Inertia. *Evolution*, **52**, 1247–1262.
- Doyle, J.J. (1992). Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic botany*, **17**, 144–163.
- Dray, S., Péliissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P.R., Bellier, E., Bivand, R., Blanchet, F.G., De Cáceres, M., Dufour, A.-B., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J. & Wagner, H.H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, **82**, 257–275.
- Eastman, J.M., Alfaro, M.E., Joyce, P., Hipp, A.L. & Harmon, L.J. (2011). A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution; international journal of organic evolution*, **65**, 3578–89.
- Edwards, A.W.F. & Cavalli-Sforza, L.L. (1964). Reconstruction of evolutionary trees. *Phenetic and phylogenetic classification* (eds V.H. Heywood & J. McNeill), pp. 67–76. London.
- F Dormann, C., M McPherson, J., B Araújo, M., Bivand, R., Bolliger, J., Carl, G., G Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W. & Others. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, **61**, 1–10.
- Felsenstein, J. (2004). *Inferring Phylogenies*, Sinauer. Sunderland MA.
- Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 1–15.
- Freckleton, R.P., Cooper, N. & Jetz, W. (2011). Comparative methods as a statistical fix: the dangers of ignoring an evolutionary model. *The American naturalist*, **178**, E10–E17.
- Freckleton, R.P., Harvey, P.H. & Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist*, **160**, 712–726.
- Garland, T., Harvey, P. & Ives, A. (1992). Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology*, **41**, 18–32.
- Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., Spector, T.D., Clark, A.G. & Ley, R.E. (2014). Human Genetics Shape the Gut Microbiome. *Cell*, **159**, 789–799.
- Gotelli, N. & McGill, B. (2006). Null versus neutral models: what's the difference? *Ecography*, **29**, 793–800.
- Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London*, **B**, 119–157.
- Hansen, T.F. (1997). Stabilizing Selection and the Comparative Analysis of Adaptation. *Evolution*, **51**, 1341.
- Harvey, P. & Pagel, M. (1991). *the Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford, UK.

Partie III – Bibliographie

- Hewitt, G.M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 183–95; discussion 195.
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–13.
- Holt, B.G., Lessard, J.P., Borregaard, M.K., Fritz, S.A., Araújo, M.B., Dimitrov, D., Fabre, P.H., Graham, C.H., Graves, G.R., Jønsson, K.A. & Others. (2013). An Update of Wallace's Zoogeographic Regions of the World. *Science*, **339**, 74–78.
- Hooper, L. V & Gordon, J.I. (2001). Commensal host-bacterial relationships in the gut. *Science (New York, N.Y.)*, **292**, 1115–1118.
- Hooper, L. V, Littman, D.R. & Macpherson, A.J. (2012). Interactions between the microbiota and the immune system. *Science (New York, N.Y.)*, **336**, 1268–73.
- Hubbell, S. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.
- Ingram, T. & Kai, Y. (2014). The geography of morphological convergence in the radiations of Pacific Sebastes rockfishes. *The American naturalist*, **184**, E115–31.
- Ives, A. & Garland, T. (2010). Phylogenetic logistic regression for binary dependent variables. *Systematic Biology*, **59**, 9–26.
- Kreft, H. & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, **37**, 2029–2053.
- Levin, S. (1992). The problem of pattern and scale in ecology. *Ecology*, **73**, 1943–1967.
- Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R. & Gordon, J.I. (2008). Evolution of mammals and their gut microbes. *Science (New York, N.Y.)*, **320**, 1647–51.
- Lister, A.M. (2004). The impact of Quaternary Ice Ages on mammalian evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 221–41.
- Lomolino, M. V., Riddle, B.R., Whittaker, R.J. & Brown, J.H. (2010). *Biogeography*. Sinauer Associates.
- Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K. & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, **489**, 220–30.
- Lynch, M. (1991). Methods for the analysis of comparative data in evolutionary biology. *Evolution*, **45**, 1065–1080.
- Maddison, W.P. (1990). A Method for Testing the Correlated Evolution of Two Binary Characters: Are Gains or Losses Concentrated on Certain Branches of a Phylogenetic Tree? *Evolution*, **44**, 223–539–557.
- Martins, E.P. & Hansen, T.F. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American naturalist*, **149**, 646–667.
- Moeller, A.H., Peeters, M., Ndjango, J.-B., Li, Y., Hahn, B.H. & Ochman, H. (2013). Sympatric chimpanzees and gorillas harbor convergent gut microbial communities. *Genome research*, **23**, 1715–20.
- Morrone, J.J. (2009). *Evolutionary Biogeography: An Integrative Approach with Case Studies*.
- Mouillot, D., De Bortoli, J., Leprieur, F., Parravicini, V., Kulbicki, M. & Bellwood, D.R. (2013). The challenge of delineating biogeographical regions: nestedness matters for Indo-Pacific coral reef fishes (W. Jetz, Ed.). *Journal of Biogeography*, **40**, 2228–2237.
- Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R. & Gordon, J.I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science (New York, N.Y.)*, **332**, 970–4.
- 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. & Ferrenberg, S. (2013). Patterns and processes of microbial community assembly. *Microbiology and molecular biology reviews : MMBR*, **77**, 342–56.
- O'Brien, E.M., Field, R. & Whittaker, R.J. (2000). Climatic gradients in woody plant (tree and shrub) diversity: water-energy dynamics, residual variation, and topography. *Oikos*, **89**, 588–600.
- O'Brien, E.M., Whittaker, R.J. & Field, R. (1998). Climate and woody plant diversity in southern Africa: relationships at species, genus and family levels. *Ecography*, **21**, 495–509.
- O'Dwyer, J.P., Kembel, S.W. & Sharpton, T.J. (2015). Backbones of evolutionary history test biodiversity theory for microbes. *Proceedings of the National Academy of Sciences*, **112**, 201419341.
- Page, R.D.M. & Charleston, M.A. (1998). Trees within trees: phylogeny and historical associations. *Trends in Ecology & Evolution*, **13**, 356–359.
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.
- Paradis, E. & Claude, J. (2002). Analysis of Comparative Data Using Generalized Estimating Equations. *Journal of Theoretical Biology*, **218**, 175–185.
- Peres-Neto, P.R. & Legendre, P. (2010). Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography*, **19**, 174–184.

Partie III – Bibliographie

- Petchey, O.L. & Gaston, K.J. (2007). Dendrograms and measuring functional diversity. *Oikos*, **116**, 1422–1426.
- Pigot, A.L. & Etienne, R.S. (2014). A new dynamic null model for phylogenetic community structure (J. Wiens, Ed.). *Ecology Letters*, **18**, n/a–n/a.
- Qian, H., Badgley, C. & Fox, D.L. (2009). The latitudinal gradient of beta diversity in relation to climate and topography for mammals in North America. *Global Ecology and Biogeography*, **18**, 111–122.
- Qian, H. & Ricklefs, R.E. (2007). A latitudinal gradient in large-scale beta diversity for vascular plants in North America. *Ecology letters*, **10**, 737–44.
- Ronquist, F. & Sanmartín, I. (2011). Phylogenetic Methods in Biogeography. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 441–464.
- 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**, 340–8.
- 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, Series B*, **366**, 2536–2544.
- Sanmartín, I. (2012). Historical Biogeography: Evolution in Time and Space. *Evolution: Education and Outreach*, **5**, 555–568.
- Sanmartín, I. & Ronquist, F. (2004). Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic biology*, **53**, 216–43.
- Savage, D.C. (1977). Microbial Ecology of the Gastrointestinal Tract. *Annual Reviews in Microbiology*, **31**, 107–133.
- Simpson, G.G. (1980). *Splendid isolation : the curious history of South American mammals*. New Haven : Yale University Press.
- Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Villarreal Ruiz, L., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., Smith, M.E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., Piepenbring, M., Phosri, C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A.L., Nilsson, R.H., Morgado, L.N., Mayor, J., May, T.W., Majuakim, L., Lodge, D.J., Lee, S.S., Larsson, K.-H., Kohout, P., Hosaka, K., Hiiesalu, I., Henkel, T.W., Harend, H., Guo, L., Greslebin, A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C., Drenkhan, R., Dearnaley, J., De Kesel, A., Dang, T., Chen, X., Buegger, F., Brearley, F.Q., Bonito, G., Anslan, S., Abell, S. & Abarenkov, K. (2014). Fungal biogeography. Global diversity and geography of soil fungi. *Science (New York, N.Y.)*, **346**, 1256688.
- Terlizzi, A., Anderson, M.J., Bevilacqua, S., Frascchetti, S., Włodarska-Kowalczyk, M. & Ellingsen, K.E. (2009). Beta diversity and taxonomic sufficiency: Do higher-level taxa reflect heterogeneity in species composition? *Diversity and Distributions*, **15**, 450–458.
- Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly review of biology*, **85**, 183–206.
- Webb, S. (2006). The great American biotic interchange: patterns and processes. *Annals of the Missouri Botanical Garden*, 245–257.
- Weinstein, B.G., Tinoco, B., Parra, J.L., Brown, L.M., McGuire, J.A., Stiles, F.G. & Graham, C.H. (2014). Taxonomic, phylogenetic, and trait Beta diversity in South American hummingbirds. *The American naturalist*, **184**, 211–24.
- Wiens, J. & Donoghue, M. (2004). Historical biogeography, ecology and species richness. *TRENDS in Ecology and Evolution*, **19**, 639–644.
- Woodburne, M.O. (2010). The Great American Biotic Interchange: Dispersals, Tectonics, Climate, Sea Level and Holding Pens. *Journal of mammalian evolution*, **17**, 245–264.
- Yatsunenkov, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R. & Gordon, J.I. (2012). Human gut microbiome viewed across age and geography. *Nature*, **486**, 222–7.
- Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science (New York, N.Y.)*, **292**, 686–93.

CONCLUSION GÉNÉRALE

1. Rappel des questions initiales

Dans cette thèse, nous proposons d'étudier les grands patrons biogéographiques via la description des différentes facettes de la diversité biologique (taxonomique, fonctionnelle et phylogénétique). En particulier :

(1) les différentes facettes de diversité des mammifères sont-elles congruentes à l'échelle du globe ? En d'autres termes, protéger le nombre d'espèces suffit-il à protéger la diversité fonctionnelle et évolutive ?

(2) Ces différentes facettes sont-elles également vulnérables aux changements globaux et en particulier à la perte d'habitat ?

(3) Ces différentes facettes sont-elles également couvertes par les zones de protection ?

(4) Que peuvent nous apporter ces différentes facettes dans la compréhension des grands patrons de répartition du vivant ?

En particulier, peut-on décomposer la distribution des lignées sur une échelle phylogénétique pour mieux comprendre les effets relatifs de l'histoire et de la niche écologique ?

2. Synthèse

La première partie de ma thèse, de nature méthodologique, visait à faire le tri des différentes mesures de diversité phylogénétique et fonctionnelle tout en les reliant à des questions écologiques et évolutives particulières. Je pense que nous avons proposé un premier tri conséquent de ces indices, mais il faut admettre plusieurs limites à nos travaux. Tout d'abord, plusieurs indices ont été publiés pendant la rédaction de l'étude, et d'autres le seront très probablement à (très) court terme, et il faudra donc rapidement refaire une synthèse. Ensuite, je pense que la « jungle » des indices n'est en fait que la partie émergée d'un problème plus profond. Cet « enfer vert » ne fait-il pas que refléter des racines théoriques encore incertaines et mal ancrées ? Ce manque de

« repères » théoriques, couplé à l'explosion des techniques de séquençage de l'ADN permettant d'obtenir des phylogénies rapidement et à bas coût, provoque, entre autres, une prolifération des indices de diversité phylogénétique. Il apparaît que le défi majeur de l'écologie des communautés soit la production d'un cadre conceptuel et théorique solide, à partir duquel on puisse d'abord unifier les multiples théories écologiques afin de produire des attendus qui pourront être infirmés - ou non - par les données empiriques, notamment la valeur des indices phylogénétiques et fonctionnels (Vellend, 2010).

La deuxième partie de ma thèse utilisait certains de ces indices pour montrer qu'à large échelle spatiale, les différentes facettes de la biodiversité étaient spatialement découplées et inégalement protégées et vulnérables. En ce sens, mon travail montre qu'une approche de conservation intégrative doit considérer les trois facettes indépendamment. Cet exercice était en fait largement descriptif et théorique et ne peut donc pas être directement appliqué en conservation. Cependant, je considère qu'il pourrait participer à une prise de recul permettant de percevoir la conservation de la biodiversité sous plusieurs facettes, et non plus uniquement avec le seul prisme de l'espèce.

La troisième partie de ma thèse, relativement découplée de la seconde sur le plan conceptuel, illustre quelques intérêts fondamentaux de la décomposition phylogénétique et du couplage entre phylogénie et fonction. Ces illustrations s'appuyaient sur des cas concrets, en l'occurrence des systèmes d'études emboîtés : l'ensemble des mammifères peuplant le globe et les bactéries peuplant l'intestin de certains de ces mammifères. Au premier abord, il peut être surprenant d'utiliser des objets biologiques si différents dans une même thèse. Je pense justement que cela constitue sa force et sa facette la plus intéressante. D'un point de vue de la connaissance pure, l'étude des microbiomes peut se nourrir – entre autres - des concepts biogéographiques et écologiques. Ces disciplines me semblent encore trop cloisonnées, alors qu'elles peuvent apprendre l'une de l'autre. Ce parallèle est aussi vrai dans le domaine appliqué des deux disciplines. Comme le fait justement remarquer Lozupone *et al.* (2012), les microbiomes, comme les macro-écosystèmes, font face à des menaces qui mettent en péril leur fonctionnement. La compréhension de ces deux systèmes pourra

permettre à l'avenir de mieux prédire les conséquences des menaces et surtout de proposer des plans d'action pour sauvegarder la santé humaine et celle de la biosphère.

3. Perspectives

Dans un avenir proche, plusieurs pistes de travail me semblent intéressantes. En terme conceptuels, j'envisage d'écrire (1) une revue synthétique sur les indices d'originalité phylogénétique et fonctionnelle des espèces (voir Partie I –Synthèse) et (2) un article d'opinion sur les liens entre l'étude des microbiotes et la biogéographie « classique ». En terme d'étude de cas, je vais continuer à travailler sur les deux systèmes d'études présentés dans cette thèse. D'une part, je vais poursuivre mes travaux sur les microbiotes d'animaux en collaboration avec M. Groussin, en analysant notamment un nouveau jeu de données. J'ai d'ailleurs participé au choix des espèces de mammifères à échantillonner au Zoo du Bronx, en essayant de sélectionner un ensemble d'espèces représentatif de la diversité des régimes alimentaires et de la phylogénie des mammifères. Par ailleurs, je compte continuer mes travaux en biogéographie « classique », notamment en tentant de combiner l'analyse BDTT et les données fonctionnelles. Je n'ai pas encore choisi de système d'étude en particulier car je crois que ce qui importe le plus, c'est justement de comparer les systèmes.

Bibliographie de la conclusion

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K. & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, **489**, 220–30.

Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly review of biology*, **85**, 183–206.

GLOSSAIRE

Allopatrie. Se dit de deux clades se distribuant dans des zones géographiques séparées.

Autocorrélation spatiale. L'autocorrélation spatiale reflète simplement le fait que des localités proches spatialement se ressemblent plus que des localités éloignées, que ce soit en terme de variables écologiques ou de variables environnementales.

Convergence évolutive. Les termes de parallélisme et de convergence ont été tous les deux employés dans la littérature pour décrire le phénomène de ressemblance fonctionnelle non héritée d'une même origine ancestrale et produite par la sélection naturelle. En fait, la distinction entre les deux termes apparaît historiquement floue et semble correspondre à une différence d'échelle phylogénétique (Arendt & Reznick 2008). Dans cette thèse, nous utiliserons le mot convergence dans tous les cas comme le proposent Arendt & Reznick (2008).

Diversité phylogénétique. Dans cette thèse, on utilisera le terme de 'diversité phylogénétique' pour n'importe quel indice quantifiant la structure phylogénétique d'une communauté, et on ne limitera pas son usage au seul indice du même nom (Faith 1992).

Diversité fonctionnelle. Dans cette thèse, on utilisera le terme de 'diversité fonctionnelle' pour n'importe quel indice quantifiant la structure fonctionnelle d'une communauté, et on ne limitera pas son usage au seul indice du même nom (Petchey & Gaston 2006).

Microbiome. L'ensemble des génomes du microbiote.

Microbiote. La communauté de lignées commensales, symbiotiques et pathogéniques qui peuplent le corps des métazoaires et métaphytes.

Niche écologique. Ensemble des conditions biotiques et abiotiques permettant la survie et la reproduction d'une espèce (voir les concepts de niche fondamentale et de niche réalisée)

Parallélisme. Voir 'Convergence évolutive'.

Phénotype. Ensemble des caractères observables d'un individu.

Phylogénie. Étude des relations de parenté entre êtres vivants, souvent (mais pas toujours) représentées par un arbre phylogénétique.

Sympatrie. Se dit de deux clades se distribuant dans les mêmes zones géographiques.

Trait fonctionnel. Caractéristique morphologique, physiologique ou phénologique d'un individu qui a un impact indirect sur sa performance (sa fitness).

Les mammifères sont présents sur la surface terrestre depuis au moins le crétacé et ont colonisé l'ensemble des continents et des océans. Mon travail s'attache à comprendre la nature et l'importance relative des processus qui ont pu conduire à la répartition géographique actuelle des mammifères et notamment aux similarités faunistiques entre régions du globe. En décrivant les facettes phylogénétiques et fonctionnelles de la biodiversité des mammifères, j'adopte une approche résolument intégrative, à l'interface entre la biogéographie historique et la biogéographie fonctionnelle.

À partir d'un travail de revue et de synthèse, la première partie de la thèse me permet de dégager un nombre limité de grandes lignes structurelles décrivant les facettes phylogénétiques et fonctionnelles de la biodiversité. En particulier, je mets en évidence comment la variation de l'échelle phylogénétique de travail peut permettre de mettre en lumière différents processus.

La seconde partie de ma thèse s'attache à utiliser cette approche pour expliquer la répartition des mammifères sur le globe. Nous montrons ainsi que les compositions faunistiques à large échelle phylogénétique (p. ex. pour les familles) semblent expliquées par l'isolement géographique des masses continentales au cours du cénozoïque alors que les compositions à fine échelle phylogénétique (p. ex. pour les espèces) sont davantage conditionnées par les affinités climatiques. Nous montrons ensuite que les régions ayant développées des faunes mammaliennes indépendantes présentent en fait des similarités fonctionnelles exceptionnelles, résultat d'une évolution convergente.

Comprendre les déterminants historiques et écologiques d'une telle répartition ne représente pourtant qu'un premier pas vers leur sauvegarde face aux multiples menaces anthropiques. La dernière partie de ma thèse entreprend de montrer l'importance d'une vision intégrative combinant les facettes phylogénétiques, fonctionnelles et taxonomiques pour la conservation de l'ensemble des dimensions de la diversité mammalienne.

Mammals inhabit the Earth since at least the Cretaceous and have colonised all continents and oceans. My work aims at understanding the nature and importance of the different processes leading to current geographical distributions of mammals and in particular faunal similarities between regions of the globe. By describing the phylogenetic and functional facets of mammalian diversity, I adopt an integrative approach, at the interface between historical and functional biogeography.

First, I review and synthesize most of the published multifaceted diversity metrics to propose a limited number of structural dimensions that parsimoniously describe the phylogenetic and functional structure of species assemblages. In particular, I point out how a multi-phylogenetic scale approach might improve our understanding of the mechanisms shaping diversity patterns.

Second, I use this approach to understand the geographical distribution of mammals worldwide. I show that at large phylogenetic scales (*e.g.* families) distributions seem to be driven by geographic isolation of the regions over the course of the Cenozoic Era, while at small phylogenetic scales (*e.g.* species), distributions are related to climatic affinities. In addition, I show that those regions having developed independent mammalian faunas show exceptional functional similarities due to evolutionary convergences.

Understanding the historical and ecological drivers of mammal biodiversity only represents the first step toward their conservation in the face of anthropic threats. The last part of my thesis shows the importance of an integrative approach using phylogenetic, functional and taxonomic facets to protect all dimensions of mammal diversity.

