

# Big data insights into the distribution and evolution of tropical diversity

Alexander Zizka

2018



UNIVERSITY OF  
GOTHENBURG

Faculty of Science  
Department of Biological and Environmental Sciences

Opponent: Prof. Jérôme Chave  
Examiner: Prof. Mari Källersjö  
Supervisors: Prof. Alexandre Antonelli & Assist. Prof. Christine D. Bacon

©Alexander Zizka

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without written permission.

Zizka, A (2018): Big data insights into the distribution and evolution of tropical diversity. PhD thesis. Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden.

**Cover image:** South American evergreen rainforest on the Guiana shield, with topology. View from an inselberg close to the Nouragues field station, French Guiana (4°05'N, 52°41'W).

Copyright ISBN print: 978-91-88509-13-0

ISBN digital: 978-91-88509-14-7

Digital version available at <http://hdl.handle.net/2077/55303>

Printed by BrandFactory AB

*To my family*



# Table of Contents

<b>Abstract</b>	
<b>Svensk sammanfattning</b>	
<b>Publications included</b>	<b>1</b>
<b>Introduction</b>	<b>3</b>
Tropical biodiversity . . . . .	3
Biomes and biome shifts . . . . .	9
Big data - a new era in biogeography . . . . .	11
<b>Objectives</b>	<b>17</b>
<b>Methods</b>	<b>19</b>
Study groups . . . . .	19
Data sources . . . . .	19
Data analysis . . . . .	21
<b>Results and Discussion</b>	<b>25</b>
<b>Conclusions</b>	<b>33</b>
<b>Paper contributions</b>	<b>35</b>
<b>References</b>	<b>37</b>
<b>Abbreviations</b>	<b>51</b>
<b>Acknowledgements</b>	<b>53</b>



# Abstract

Tropical America (the Neotropics) and tropical Africa have comparable climate and share a geological history as parts of Gondwana. Nevertheless, the Neotropics today harbour roughly three times more flowering plant species than tropical Africa. The role of evolutionary history in generating this pattern remains poorly understood, mostly because collecting biological specimens in the tropics is difficult. Species occurrence information from collections in museums and herbaria has the potential to overcome this gap and, for the first time, enable an understanding of tropical biodiversity on a global scale across the tree of life. However, uncertain data quality and methodological limitations to process large amounts of data often hamper the use of collection records in biogeographic analyses, especially in historical biogeography invoking phylogenetic trees. In this thesis I first (co-)develop three software tools to process large amounts of species occurrence data in biogeography: (1) *CoordinateCleaner* to test and insure data quality in large data sets of species and fossil occurrences, (2) *SpeciesGeoCoder* to include large-scale species distribution data in historical biogeography, and (3) *Infomap Bioregions* to delimit taxon-specific bioregions. I then apply these tools to identify processes underlying the evolution of tropical diversity across multiple taxonomic groups. The results suggest a significantly higher species turnover in the Neotropics compared to other tropical regions and identify this region, especially Amazonia, as a global species pump. Furthermore, shifts among different bioregions and biomes are more common than expected in evolutionary lineages of the Neotropics, and are potential drivers of diversification. The results show that biome shifts into seasonally dry biomes are particularly common, and increased during the last 20 million years, especially in the Bombacoideae (Malvaceae), a pantropical group with highest diversity in the Neotropics. The presented results shed further light on the evolutionary history of the differences in biodiversity across Earth's tropical regions, and provide a methodological route forward to integrate large-scale species occurrence data with information on species' evolutionary relationships to reveal general processes underlying the evolution of biodiversity across taxonomic borders.

## Keywords

Amazonia, automated data cleaning, biome shifts, Bombacoideae, data quality, GBIF, Neotropics, tropical plant diversity.





# Svensk sammanfattning

Tropiska Amerika (den Nya Världens tropiker, eller Neotropikerna) och tropiska Afrika har jämförbara klimat och en gemensam geologisk historia, eftersom båda var delar av Gondwanaland. Trots detta finns idag ungefär tre gånger fler blomväxter i Neotropikerna än i tropiska Afrika. Även om det råder enighet om att den nuvarande mångfalden är ett resultat av en kombination av flera processer både i nutid och i dåtid, så känner man dåligt till vilken betydelse evolutionen har för uppkomsten av denna mångfald; bristen på kunskap hör i hög grad samman med svårigheten med att samla in biologiska prover i tropikerna.

Den snabbt ökande tillgängligheten till information från museer och herbarier om arters förekomst har potential att överbrygga denna okunskap och, för första gången, möjliggöra en förståelse på global nivå för den tropiska biodiversiteten och för släktskapen mellan organismer (livets träd). Användningen av dessa data i biogeografiska analyser, särskilt i historisk biogeografi som utgår från fylogenetiska träd, försvåras emellertid ofta av att data är av låg kvalitet och av metodologiska begränsningar vid bearbetning av stora datamängder.

I denna avhandling (med)utvecklar jag initialt tre mjukvaror för att behandla och bearbeta stora mängder data om artförekomst i biogeografi: (1) *CoordinateCleaner* för att testa och säkerställa datakvalitet i stora dataset av art- och fossilförekomster, (2) *SpeciesGeoCoder* för att inkludera storskaliga artutbrednings data i historisk biogeografi, och (3) *Infomap Bio-regions* för att avgränsa taxonspecifika bioregioner. Jag använder sedan dessa verktyg för att identifiera processer som ligger till grund för utvecklingen av tropisk mångfald i flera taxonomiska grupper.

Resultaten tyder på en signifikant högre artomsättning i Neotropikerna jämfört med i andra tropiska områden och identifierar denna region, i synnerhet Amazonas, som en global "art-pump" – en plats där många arter bildas och därifrån sprids till hela världen. Dessutom är skiften mellan olika bioregioner och biom mer vanligt förekommande än förväntat i evolutionära utvecklingslinjer i Neotropikerna, och växlingarna utgör en potentiell drivkraft för diversifiering. Detta är särskilt tydligt inom Bombacoideae (Malvaceae), där upprepade skiften till torra biomer är vanligt förekommande.

Dessa resultat sprider ytterligare ljus över evolutionshistorien avseende de påtagliga skillnaderna i biodiversitet mellan jordens tropiska områden, och erbjuder en metodik för att integrera artförekomstdata med information om arters evolutionära släktskap så att de allmänna processer som ligger till grund för utveckling av biologisk mångfald över taxonomiska gränser kan urskiljas.



# Publications included

This thesis is based on the following manuscripts and papers, referred to in the text by their Roman numerals. Paper I is reprinted under a CC-BY-NC-ND and **Papers III-V & VII** under a CC-BY license.

\* indicates equal contribution/shared first authorship.

## New methods for big data in historical biogeography

**I** - Maldonado C, Molina CI, **Zizka A**, Persson C, Taylor CM, Albán J, Chilquillo E, Rønsted N & Antonelli A (2015), Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography* **24**(8): 973-984.

**II** - **Zizka A**, Silvestro D, Andermann T, Azevedo J, Duarte Ritter C, Farooq H, Herdean A, Ariza Salazar MdA, Scharn R, Svanteson S, Wengström N, Zizka V & Antonelli A, CoordinateCleaner: new tools for the automated cleaning of geographic occurrences from biological collections. *Manuscript, submitted to Methods in Ecology and Evolution*.

**III** - \*Töpel M, \***Zizka A**, Calió MF, Scharn R, Silvestro D & Antonelli A (2017), Species-GeoCoder: Fast categorisation of species occurrences for analyses of biodiversity, biogeography, ecology and evolution. *Systematic Biology* **66**(2): 145-151.

**IV** - Edler D, Guedes T, **Zizka A**, Rosvall M & Antonelli A (2017), Infomap Bioregions: Interactive mapping of biogeographical regions from species distributions. *Systematic Biology* **66**(2): 197-204.

## Big data insights into the distribution and evolution of tropical diversity

**V** - **Zizka A**, ter Steege H, de Céo M & Antonelli A (2018), Finding needles in the haystack: Where to look for rare species in the American Tropics. *Ecography* **40**: 321-330.

**VI** - \*Antonelli A, \***Zizka A**, \*Carvalho FA, Scharn R, Silvestro D, Bacon CD & Condamine F, Amazonia is the primary source of Neotropical biodiversity. *Manuscript, submitted to Proceedings of the National Academy of Sciences USA*.

**VII** - \*Antonelli A, \***Zizka A**, Silvestro D, Scharn R, Cascales-Miñana B & Bacon CD (2015), An engine for global plant diversity: highest evolutionary turnover and emigration in the American Tropics. *Frontiers in Genetics* **6**: 14pp.

**VIII - Zizka A**, Carvalho-Sobrinho JG, Pennington RT, Queiroz L, Alcantara A, Baum D, Bacon CD & Antonelli A, Biome shifts are common, directional, and increase through time in Bombacoideae (Malvaceae). *Manuscript*.

#### **Additional publications not included in this thesis**

**Zizka A** & Antonelli A (2018), Mountains of Biodiversity. *Nature* **555**: 173-174.

Guedes TB, Sawaya RJ, **Zizka A**, Laffan S, Faurby S, Pyron RA, Bérnils RS, Janssen M, Passos P, Prudente ALC, Cisneros-Heredia DF, Braz HB, Nogueira CdC & Antonelli A (2018), Patterns, biases, and prospects in the distribution and diversity of Neotropical snakes. *Global Ecology and Biogeography* **27**(1): 14-21.

\* Schmidt M, \* **Zizka A**, Traoré S, Ataholo M, Chatelain C, Daget P, Dressler S, Hahn K, Kirchmair I, Krohmer J, Mbayngone E, Müller JV, Nacolma BMI, Ouédraogo A, Ouédraogo O, Sambaré O, Schumann K, Wieringa J, Zizka G & Thiombiano A (2017), Diversity, distribution and preliminary conservation status of the flora of Burkina Faso. *Phytotaxa* **304**(1): 215 pp.

Silvestro D, **Zizka A**, Bacon CD, Cascales-Miñana B, Salamin N, & Antonelli A (2016), Fossil Biogeography: A new model to infer dispersal, extinction, and sampling from palaeontological data. *Philosophical Transactions of the Royal Society B* **371**: 1691.

**Zizka A**, Thiombiano A, Dressler S, Nacolma BMI, Ouédraogo A, Ouédraogo I, Zizka G, Hahn K & Schmidt M (2015), The Vascular Plant Diversity of Burkina Faso (West Africa) - A quantitative analysis and implications for conservation. *Candollea* **70**(1): 9-20.

**Zizka A**, Thiombiano A, Dressler S, Nacolma BMI, Ouédraogo A, Ouédraogo I, Zizka G, Hahn K & Schmidt M (2015), Traditional plant use in Burkina Faso (West Africa): a national-scale analysis with focus on traditional medicine. *Journal of Ethnobiology and Ethnomedicine* **11**(9): 10 pp.

**Zizka A**, Govender N & Higgins SI (2014), How to tell a shrub from a tree: a life history perspective from South African savanna. *Austral Ecology* **39**(7): 767-778.

# Introduction

The diversity of life on Earth—biodiversity—has fascinated humans for thousands of years. One question at the core of this fascination is: "What generates biodiversity?" or phrased differently: "Why are there more species in some areas than others, despite a similar environment?". A prime example are the world's tropical regions, where tropical America harbours more than three times as many plant species than tropical Africa, and probably more species than tropical Africa and Asia combined. To this date, the scientific understanding of this pattern, and the underlying evolutionary processes remain incomplete. This lack of knowledge is problematic, because protecting tropical biodiversity will be one of the major challenges for the current generation (Butchart *et al.*, 2010; Rockström *et al.*, 2009; Steffen *et al.*, 2015), and understanding the evolutionary history of this diversity can be critical to guide area and taxon prioritization for effective conservation.

## Tropical biodiversity

Biodiversity has multiple dimensions, including genetic diversity ("How diverse is the genetic material?"), organismic diversity ("How many species are there?") and ecological or functional diversity ("How diverse are forms and ecological roles?") (Gaston and Spicer, 2004). A large number of indices exist to measure biodiversity, each putting focus on slightly different aspects (Scheiner *et al.*, 2017; Tucker *et al.*, 2016). Species richness, the count of species in an area, is a simple and intuitive measure of organismic diversity and among the few with data available on a global scale; and thus often used as proxy for overall biodiversity.

There are an estimated 4–10 million species on Earth today (Costello *et al.*, 2013), and potentially even many more (Larsen *et al.*, 2017). This diversity is not equally distributed on the globe, but increases from the poles to the equator, peaking between the Tropics of Cancer and Capricorn (app. 23.4° North and 23.4° South). This latitudinal diversity gradient has often been explained by higher energy availability or long-term climatic stability in the Tropics (among others; see Mittelbach *et al.*, 2007 and Wiens and Donoghue, 2004 for reviews). Surprisingly however, species richness also largely differs among the tropical continents despite the similar environment (Fig. 1), suggesting a critical importance of continental evolutionary history on global biodiversity.

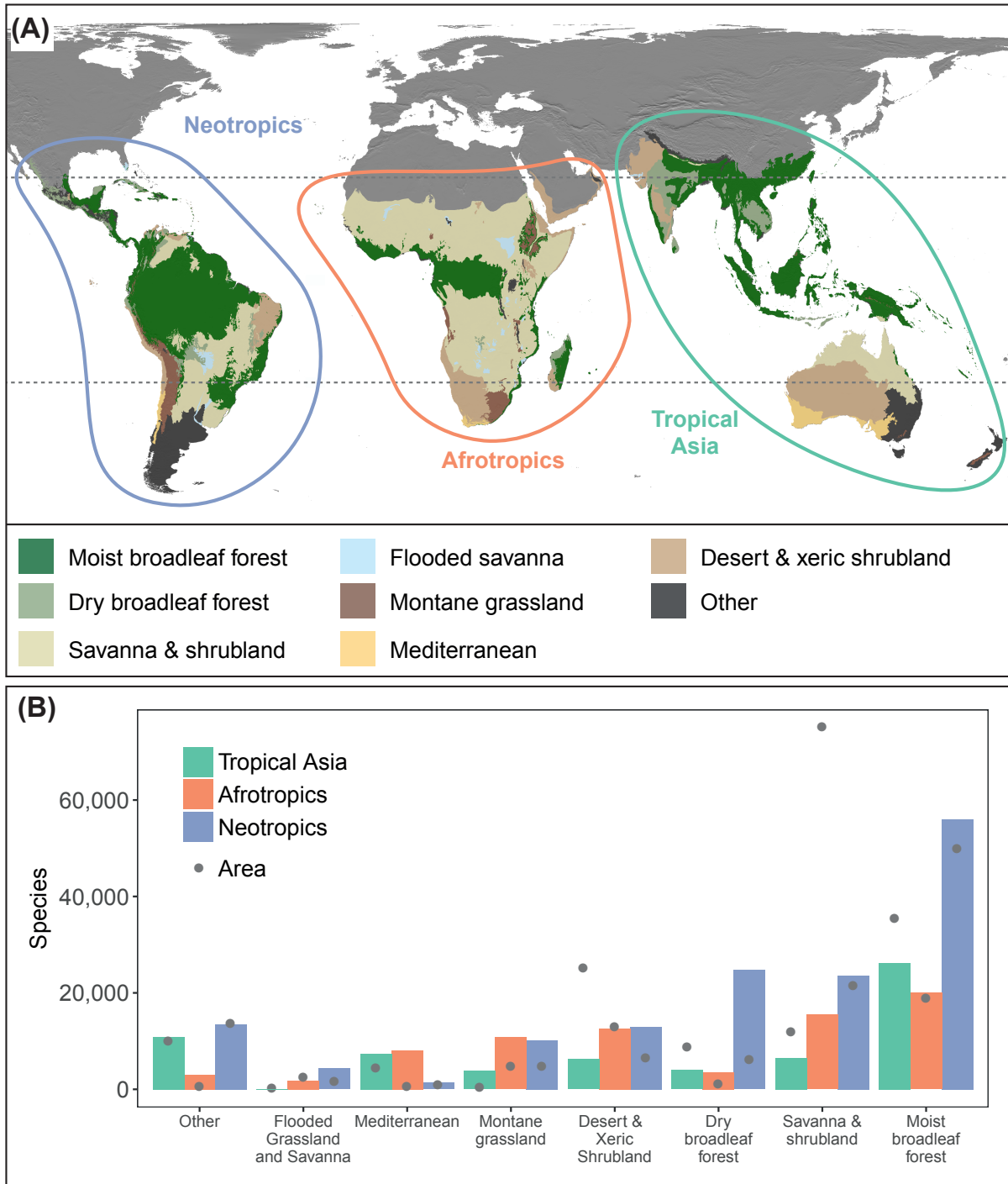
## Species richness among tropical realms

Tropical America, Africa and Asia vary strongly in their floristic composition. For instance, only four percent of tropical tree species are shared among these regions and only between 14%–24% of tree genera might be shared (Dexter *et al.*, 2015; Ricklefs and Renner, 2012; Slik *et al.*, 2015, but see Gentry, 1993). This divergence has inspired the classification of the world's Tropics into separate floristic realms (Good, 1953; Takhtajan, 1987) comparable to zoogeographic regions for animals (Holt *et al.*, 2013; Wallace, 1876). The exact definition

of these realms and the affinities among them are controversial (e.g. Slik *et al.*, 2017), but a widely used scheme divides the Tropics into four realms: Neotropics (tropical America), Afrotropics (tropical Africa), Australasia and Indomalaya (hereafter combined as tropical Asia, Fig. 1, Olson *et al.*, 2001). This definition is admittedly broad and includes habitats often not considered tropical (e.g. based on mean annual temperature, or annual minimum temperature). However, it has the advantage that area, latitudinal range, and biomes among the three realms are similar, and these are thus more comparable for the purpose of this thesis.

The Neotropics, Afrotropics and tropical Asia are generally comparable in size and environmental conditions, but differ drastically in species richness of flowering plants (angiosperms). There are an estimated 90,000 angiosperm species in the Neotropics, 27,000 species in the Afrotropics and 50,000 species in tropical Asia (Lebrun, 2001; Thomas, 1999; Whitemore, 1998). The outstanding species richness of the Neotropics, in particular in comparison with Africa, has long been noted from individual clades (Gentry, 1982; Raven, 1976; Richards, 1973) and was later confirmed by spatially explicit global analyses (Govaerts, 2001; Kreft and Jetz, 2007; Slik *et al.*, 2015). The difference in species richness seems to be consistent across biomes (Fig. 1) and prevails when taking into account area, climate or productivity (Fig. 2, Banin *et al.*, 2014; Couvreur, 2015; Parmentier *et al.*, 2007). Interestingly, the Neotropics are similarly outstanding in many animal groups, including amphibians (Wiens, 2007), birds (Grenyer *et al.*, 2006; Somveille *et al.*, 2013) and mammals (Ceballos and Ehrlich, 2006, Fig. 3), and some evidence suggests that functional diversity might follow a similar pattern as species richness (Chomicki and Renner, 2015; Reu *et al.*, 2011).

A rigorous comparison of species richness needs to standardize methodology and to account for sampling effort (Engemann *et al.*, 2015; Whittaker *et al.*, 2001). The few studies directly comparing tropical species richness among continents based on standardized area and methodology (mostly using vegetation plots), have almost exclusively focussed on trees with more than 10 cm diameter at breast height in evergreen and semi-evergreen rainforest (rainforest hereafter). It was based on data from rainforest that the pattern was first observed and the idea of Africa as "the odd man out" due to the unexpectedly low species counts was developed (Richards, 1973). Since then, particularly high species richness in the Neotropics (and possibly tropical Asia) as well as comparatively low species numbers in Africa have been reported from plots repeatedly (Cáceres *et al.*, 2012; Gentry, 1988; Ricklefs and Renner, 2012; Terborgh *et al.*, 2016) and suggested that the Amazon basin might be up to three times as species rich as central Africa (Parmentier *et al.*, 2007). It must however be noted that some areas of the Afrotropics are also very species rich on the local scale, reaching plot-level species-counts similar to Neotropical plots, but rarely so (Cáceres *et al.*, 2012; Couvreur, 2015; Parmentier *et al.*, 2007).



**Figure 1** The major biomes in the Neotropics, Afrotropics and tropical Asia, and their area and raw species counts of flowering plants. For the majority of biomes the species richness is highest in the Neotropics, a pattern especially prominent in forest biomes. Realms, biomes and biome area from Olson *et al.* (2001) (tropical Asia combines Indomalaya and Australasia, see text for details), species estimates from GBIF (2016). Species can occur in multiple areas. The grey dots show the relative area of each biome in each realm.

Virtually no information on inter-realm comparison of diversity exists from biomes other than rainforests. Few data are available only from single locations or from case studies on individual taxonomic groups. Realm-wide data compiled for this thesis suggest, that the Neotropics are particular species rich for most biomes (Fig. 1) also when accounting for area (Fig. 2), although not to the same extent as in rainforests. A well-known exception are the three tropical Mediterranean climate regions (cold, rainy winters and dry hot summers, Fig. 1), where the South African Cape is the most species rich, followed by South-Western Australia and Central Chile (Cowling and Rundel, 1996; Valente and Vargas, 2013). Some evidence suggests that diversification rates might not have been different in seasonal dry biomes (Lavin *et al.*, 2000), but studies explicitly comparing species richness in these biomes across tropical realms are virtually non-existent, potentially due to the difficulties to define these biomes across realm borders (Dexter *et al.*, 2015; Moncrieff *et al.*, 2016). All tropical realms include hot-spots with particularly high diversification rates, for instance the Páramo and the Cerrado in the Neotropics, the succulent Karoo and Cape region in the Afrotropics and South-Western Australia in tropical Asia (Madriñán *et al.*, 2013).

### **Potential drivers of differences in species richness**

The drivers of the diversity differences among the tropical realms of the world have puzzled biologist for decades (Gentry, 1982; Raven and Axelrod, 1974), but are surprisingly little studied (Couvreur, 2015). The high diversity of the Neotropics is particularly intriguing compared to the Afrotropics, because these realms represent comparable continental masses and shared a direct connection until c. 110 million years ago as parts of Gondwana. Tropical Asia in contrast largely consists of islands, making comparisons more complex. Accordingly, attempts to explain the drivers and mechanisms underlying the differences in diversity among the tropical realms have mostly focused on the Neotropics and Afrotropics, and rainforests in particular.

Generally, the biodiversity in any given area is shaped by three fundamental processes: speciation, extinction, and dispersal. All of these processes are critically influenced by evolutionary history, for instance via past competition and other biotic interactions, continental drift, and mountain uplift. In this context two major (not mutually exclusive) perspectives for the Tropics exist: to either consider the Afrotropics as relatively depauperate (e.g. Morley, 2000; Raven and Axelrod, 1974), or to consider the Neotropics as relatively enriched (e.g. Antonelli and Sanmartín, 2011; Gentry, 1982). The former usually invokes higher extinction rate in the Afrotropics through time, whereas the latter usually implies higher speciation rates in the Neotropics. A considerable number of hypotheses concerning the specific factors causing extinction and speciation in both realms exist (see Couvreur, 2015 and Antonelli and Sanmartín, 2011 for an overview).



Historically, current and past climatic conditions have been considered as major drivers of the differences in species richness. Following the observation that species richness correlates with precipitation and temperature (e.g. Barthlott *et al.*, 2005; Jaramillo *et al.*, 2010; Kreft and Jetz, 2007), and that African rainforests are more seasonal (Richards, 1973) and generally drier (Malhi and Wright, 2004), African rainforests would sustain only lower numbers of species. Additionally, a long term aridification throughout the Cenozoic (the last 66 million years) especially pronounced in Africa supposedly led to a continuous loss of evergreen forest area and lower productivity (Jetz and Fine, 2012; Morley, 2000), associated with range contractions and extinctions of rainforest lineages. This effect might have been exacerbated by strong dry phases during the Pleistocene (the last 2.6 million years, Gasse, 2000; Stager *et al.*, 2011), causing the retraction of rainforests into few refugia (Hamilton, 1981; Parmentier *et al.*, 2007; Raven and Axelrod, 1974; Ray and Adams, 2001). Indeed, past and current climate influence tree diversity in rainforests (Jaramillo *et al.*, 2010; Stropp *et al.*, 2009), and fossil pollen records from the recent past support a contraction of rainforests in Africa following aridification (e.g. Bouimetarhan *et al.*, 2015; Dupont *et al.*, 2000; Ivory *et al.*, 2012; Kirchmair, 2017; Morley, 2000; Vincens *et al.*, 2007). In contrast, climate models and fossils suggest more stable rainforests in the Neotropics throughout the Pleistocene (Bush and Oliveira, 2006; Leite *et al.*, 2016; Mayle *et al.*, 2004) and the entire Cenozoic (Jaramillo *et al.*, 2006), although potentially with high spatial and temporal variation (Carnaval and Moritz, 2008; Jaramillo *et al.*, 2010; Morley, 2000).

However, while aridification and linked rainforest retractions in Africa are well supported at least for the Pleistocene, their magnitude likely varied strongly across regions (Andrews and Bamford, 2008; Ivory and Russell, 2016) and there is fossil evidence for extensive rainforests in Africa throughout the Cenozoic (Bonnefille, 2010; Jacobs, 2004; Pan *et al.*, 2006; Utescher and Mosbrugger, 2007). Hence, the role of extinctions induced by rainforest retraction as driver of current day diversity, is unclear and has recently been challenged conceptually and empirically. The postulated past refugia of rainforests in the Afrotropics would still have been large compared to rainforests on some oceanic islands (Terborgh *et al.*, 2016); nevertheless the latter can be more diverse today, as for example shown in one hectare plots on the Solomon and Fiji islands (Keppel *et al.*, 2010). Furthermore, newly available molecular evidence from numerous plant clades does not support higher extinction rates in afrotropical lineages, for instance in palms (Baker and Couvreur, 2013; Kissling *et al.*, 2012), Annonaceae (Erkens *et al.*, 2012), and Chrysobalanaceae (Bardon *et al.*, 2013). Instead, these studies and others, for example on Campanulaceae (Lagomarsino *et al.*, 2016), Burseraceae (Fine *et al.*, 2014), Fabaceae (Richardson *et al.*, 2001), Sapotaceae (Armstrong *et al.*, 2014) or the Páramo vegetation (Madriñán *et al.*, 2013) suggest particularly high speciation rates in the Neotropics, supporting the idea of the Neotropics as particularly enriched compared to the Afrotropics. However, an integrative perspective on the issue, synthesizing information from multiple taxonomic groups is missing to date.

## The Neotropics - a global biodiversity hotspot

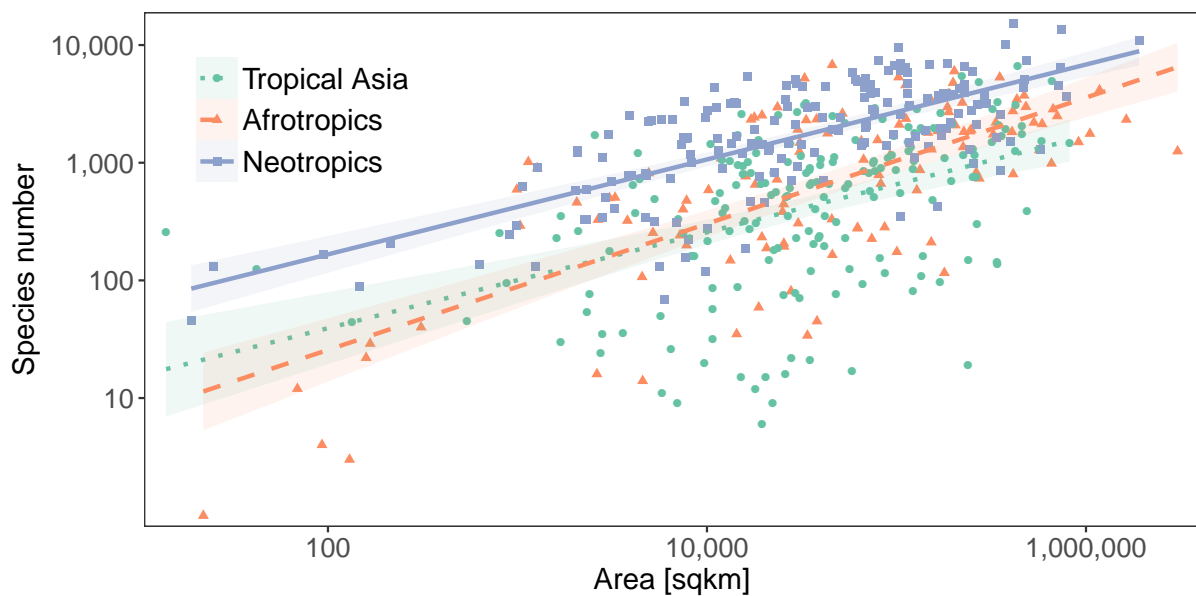
The Neotropical realm roughly stretches from northern Mexico to central Argentina (including the Caribbean, Fig. 1), with the precise definition differing slightly among authors (Holt *et al.*, 2013; Morrone, 2014; Morrone, 2017; Olson *et al.*, 2001; Schultz, 2005). The realm includes multiple biomes and habitats (Fig. 1) as well as a large number of different ecoregions, such as the Andean highlands, the Amazonian and Atlantic rainforests, the Cerrado, and large dry areas (Hughes *et al.*, 2013; Olson *et al.*, 2001). Climate varies significantly (the region includes the wettest and driest places on Earth), but in large areas it is generally moist and warm, although with varying seasonality.

Regions of particularly high diversity in the Neotropics are Central America, the northern Andes, western Amazonia, south-east Brazil and the Caribbean (Barthlott *et al.*, 2005). In general, Neotropical centres of high species richness are often linked to mountain chains (Knapp, 2002; Kruckeberg and Rabinowitz, 1985) and high topographic diversity (Kessler, 2002; Kreft and Jetz, 2007). A central pattern in Neotropical biogeography is the division of plant taxa into two groups, being predominantly species-rich in the Andes and relatively species-poor in Amazonia, or vice-versa (Gentry, 1982). Recent studies stress the importance of rare species, suggesting that they contribute a major proportion to the regional species richness, at least for Amazonian trees (ter Steege *et al.*, 2013; ter Steege *et al.*, 2016).

The outstanding biodiversity of the Neotropics has been linked to multiple factors, including: the uplift of the Andes (Luebert and Weigend, 2014; Pirie *et al.*, 2018; Särkinen *et al.*, 2007), past marine incursions (Hoorn *et al.*, 2010), habitat specialization (Fine *et al.*, 2014; Gentry, 1982; Higgins *et al.*, 2011), biotic interactions (Kursar *et al.*, 2009; Lagomarsino *et al.*, 2016), favourable current climate conditions (Kreft and Jetz, 2007), long-term climate stability or climatic fluctuations in the past (Haffer, 1969, see Antonelli and Sanmartín, 2011; Haffer, 2008 for reviews). Most likely a combination of factors are causing the high diversity observed today, but the mechanisms and their interaction remain elusive.

Large parts of the Neotropics are scarcely studied (Feeley and Silman, 2011; Kier *et al.*, 2005; Tobler *et al.*, 2007). For instance, Pimm and Joppa (2015) expect that 41% of all globally yet undiscovered plant species are to be found in Mexico, Panama, Colombia, Peru and Ecuador, whereas Feeley (2015) estimates that the median number of collected seed plant specimens in tropical America is as low as 0.01 per square kilometre and that 13% of the area are not represented by any collections. In addition to the sparse collection effort, collections and taxonomic knowledge are strongly biased towards few well-studied locations (Hopkins, 2007; Nelson *et al.*, 1990), often related to easy access and research stations (Hijmans *et al.*, 2000; Meyer *et al.*, 2016).

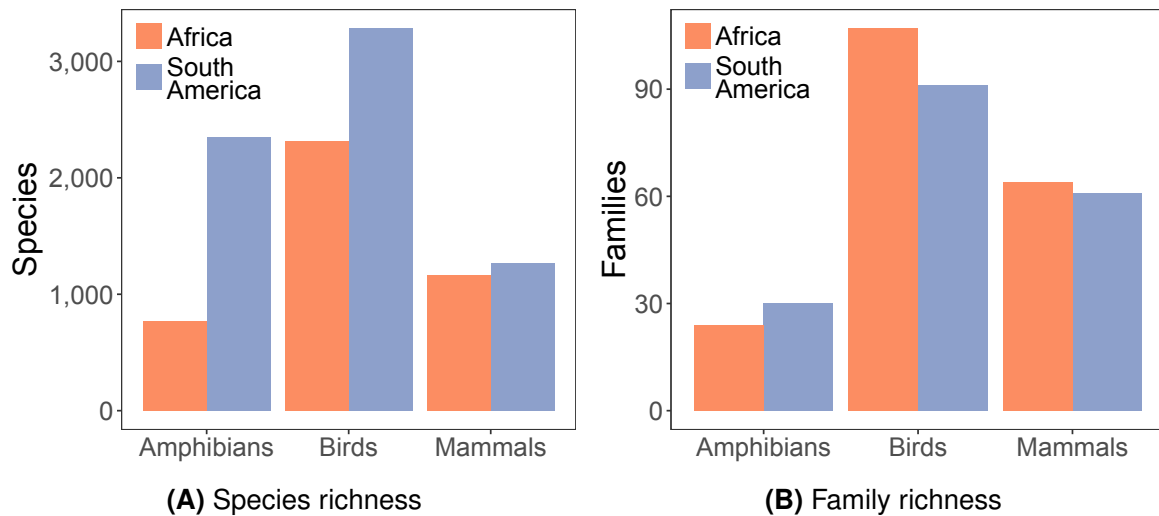
Unfortunately, large parts of the Neotropics are conservation hot-spots under strong human land-use pressure (Mittermeier *et al.*, 2011; Myers *et al.*, 2000). Increasing habitat destruction in combination with high biodiversity and the scarce knowledge of this diversity lead to severe problems for conservation. The situation is particularly critical for rare species with low numbers of individuals and potentially restricted range size. Thus, a better knowledge of Neotropical plant diversity and its spatial distribution, in particular for rare species is needed.



**Figure 2** Species richness of ecoregions in the Neotropics, Afrotropics and tropical Asia in relation to ecoregion size. Each point represents an ecoregion from Olson *et al.* (2001). The species richness is consistently higher in ecoregions from the Neotropics. The lines show individual linear regression smoothers for each realm. Estimates based on data from GBIF (2016). Note the logarithmic scale.

## Biomes and biome shifts

In contrast to ecoregions and realms, which are described by species or phylogenetic composition, biomes can be defined by functional similar plant groups and their environmental demands (Moncrieff *et al.*, 2015). For instance, the most widely used global definition of biomes, based on expert-generated vegetation units, includes among others "Tropical & Subtropical Moist Broadleaf Forests" or "Tropical & Subtropical Grasslands, Savannas & Shrublands" (Fig. 1, Olson *et al.*, 2001). The definition and delimitation of biomes is controversial and a set of alternative definitions exists, for example based on climate, phenology or vegetation height and productivity (e.g. Buitenwerf and Higgins, 2016; Higgins *et al.*, 2016; Moncrieff *et al.*, 2015; Pennington *et al.*, 2004; Särkinen *et al.*, 2011; Whittaker,



**Figure 3** Taxonomic richness of selected animal groups with distribution data available from mainland South America and Africa. The species richness is consistently higher in South America. Family richness is more balanced and does not show a clear pattern. Estimates based on species ranges from International Union for the Conservation of Nature (IUCN).

1962). The usefulness of the individual approaches depends on the scientific question, but despite important differences, the mentioned definitions generally agree concerning the major biomes in the southern hemisphere. Irrespective of the definition, biomes can be interpreted as large-scale representation of species ecological niches (= abiotic and biotic requirements, Hutchinson, 1957).

Ecological niche conservatism is the observed tendency of species to retain their ecological niche through time. Phylogenetic Niche Conservatism (PNC) is the resulting pattern of closely related species having similar ecological niches (Crisp and Cook, 2012; Harvey and Pagel, 1991; Losos, 2008; Wiens and Donoghue, 2004). PNC has been observed in many traits and many lineages, and is a key concept for understanding speciation as well as macro-evolutionary patterns and processes (Crisp and Cook, 2012; Wiens and Graham, 2005; Wiens *et al.*, 2010). The assessment of how conserved a niche is depends on temporal and organismic scale. Some evidence suggests that niches are little conserved over short times and on small scale (e.g. population level) but strongly conserved on a larger temporal and taxonomic scale (Peterson, 2011). This might be due to the fact that moving into a new environment might be easy, but evolving a stable, competitive phenotype might be difficult, leading to higher extinction rate in new environment (Edwards and Donoghue, 2013)

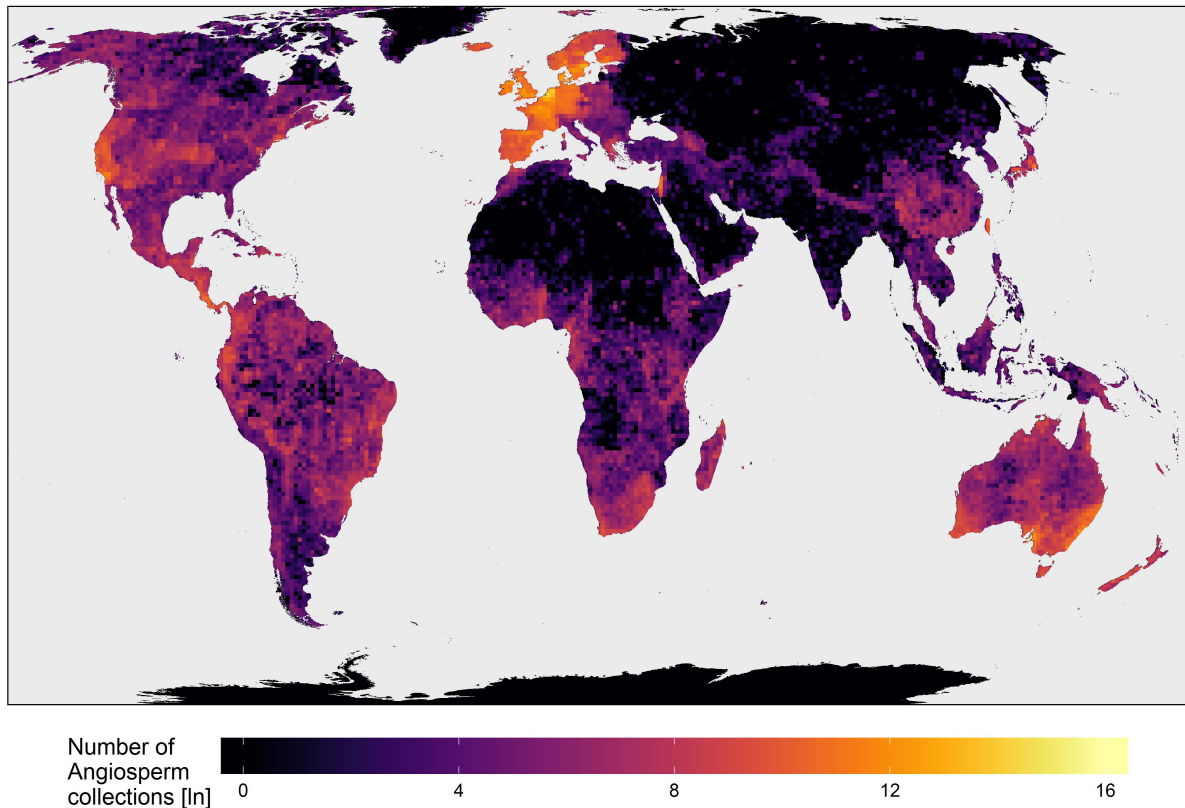
Biomes are large-scale representations of a species' ecological niche, and thus for the majority of plant lineages seem conserved over evolutionary time ("biome conservatism",

Crisp and Cook, 2012; Crisp *et al.*, 2009; Donoghue, 2008; Pennington *et al.*, 2009). However, the alternative scenario of lineages crossing biome borders through evolutionary time is also observed regularly ("biome shift", Dexter *et al.*, 2015; Donoghue and Edwards, 2014; Edwards and Donoghue, 2013; Fine *et al.*, 2014; Lohmann *et al.*, 2013; Simon *et al.*, 2009). There is evidence for shifts into and out of all major biomes, but the numbers of these shifts differ considerably, possibly dependent on factors such as the biome size, their spatial and ecological adjacency, perimeter length, biome age, their connectivity through time, biotic interactions and traits of individual evolutionary lineages (Crisp *et al.*, 2009; Donoghue and Edwards, 2014). Furthermore, dry and cold biomes seem less prone to receive lineages as compared to wetter and warmer biomes (Qian *et al.*, 2017). Biome shifts and biome conservatism can determine global biodiversity, but it is unclear how common they are among tropical plant lineages and how they impact diversification (Crisp *et al.*, 2009; Wiens and Donoghue, 2004). Few existing case studies indicate a possible relation between biome shifts and increased diversification in some systematic groups such as in Fabaceae and Proteaceae (Onstein *et al.*, 2016), or *Viburnum* (Koenen *et al.*, 2013; Spriggs *et al.*, 2015), but not in others, e.g. *Protea* (Valente *et al.*, 2009). Additionally, the effect of biome-shifts on diversification might differ among geographic realms (Sauquet *et al.*, 2009).

## **Big data - a new era in biogeography**

Biogeography is *"the science that attempts to document and understand spatial patterns of biodiversity. It is the study of distributions of organisms, both past and present, and of related patterns of variation over the Earth in numbers and kinds of living things"* (Lomolino *et al.*, 2010; page 3). Historical biogeography is concerned with reconstructing the origin and evolutionary history of taxa and geographic areas, by combining information of species distribution and ecology with information on their evolutionary history (phylogenies) (Wiens and Donoghue, 2004). The biogeography of angiosperms is of particular interest, because they are one of the most diverse systematic groups and because of their role as primary producers and ecosystem engineers, providing food and habitat for most other organisms.

Available data on tropical ecosystems and their biodiversity is limited. This is even the case for basic geographic and taxonomic information on the distribution, and evolutionary relatedness of species. For most part of the history of modern biogeographic research logistic difficulties have restricted data collection in the Tropics to certain geographic areas and taxonomic groups. While accessibility has generally improved lately, high logistic investment remains a challenge to the study of many tropical ecosystems. The aggregation of existing data from museums and herbaria offers a unique opportunity to, at least partly, leverage this problem and holds great potential to refine the picture of tropical diversity and the processes underlying its evolution.



**Figure 4** The number of angiosperm collection records available from Global Biodiversity Information Facility (GBIF) in March 2016. Only geographically unique records (no records from one species with identical coordinates) and records falling into the sea or on country centroids were excluded using *CoordinateCleaner*. More than 34 million records are available globally, but some regions remain without records and sampling effort is very uneven. Resolution 100 x 100 km, note the logarithmic scale.

Natural history collections and (vegetation-)survey archives are biological data repositories, holding information of species distributions from hundreds of years of scientific collection effort. Over the last decade the digitization of museum and herbarium collections made unprecedented amounts of species distribution information publicly available (Feeley, 2015). In particular, large platforms, for instance GBIF ([www.gbif.org](http://www.gbif.org)) and the IUCN ([www.iucn.org](http://www.iucn.org)) play a critical role by enabling free access to species distribution information from museums, herbaria, and regional observation initiatives all over the world. GBIF alone currently includes more than 140 million geo-referenced collection records from almost 200,000 angiosperm species (as of February 2018, see also Fig. 4), and for many animal groups even more data are available.

While there are still large gaps to fill (Fig. 4; Bebber *et al.*, 2010; Engemann *et al.*, 2015;

Feeley, 2015) and collecting biases to overcome (Daru *et al.*, 2018; Hopkins, 2007; Meyer *et al.*, 2016; ter Steege *et al.*, 2011), the increased availability of species distribution data, for the first time, enables global-scale, data-driven approaches to biogeography, that have increased our understanding of biodiversity and underlying patterns. A comparable revolution in data availability is currently ongoing for DNA-sequences, based on Next Generation Sequencing (NGS) techniques. NGS allows for the cost-effective analysis of hundreds of genes or even entire genomes for a large number of organisms (see Levy and Myers, 2016 for an overview), and can even be applied to historic specimens. This development promises an explosion in the availability of genetic data for non-model organisms, providing even bigger and better sampled phylogenies as basis for historical biogeography, and potentially rendering a unified phylogenetic tree of life for all (macroscopic) species possible (Eiserhardt *et al.*, 2018; Hudson, 2008). Indeed, whole genome sequences for the majority of all living bird species are currently being analysed (<http://avian.genomics.cn/en/>).

For historical biogeography, the increase in spatial and genetic data has enormous potential. For the first time, a better understanding of global biodiversity on evolutionary time-scales is possible. However, the use of this "treasure box" is still limited by the inability of current methods and work-flows to process large amounts of data in a quick and reproducible way, while ensuring high data quality (Boakes *et al.*, 2010; Dickinson *et al.*, 2010; Meyer *et al.*, 2016; Ruete, 2015).

### **Methods in historical biogeography**

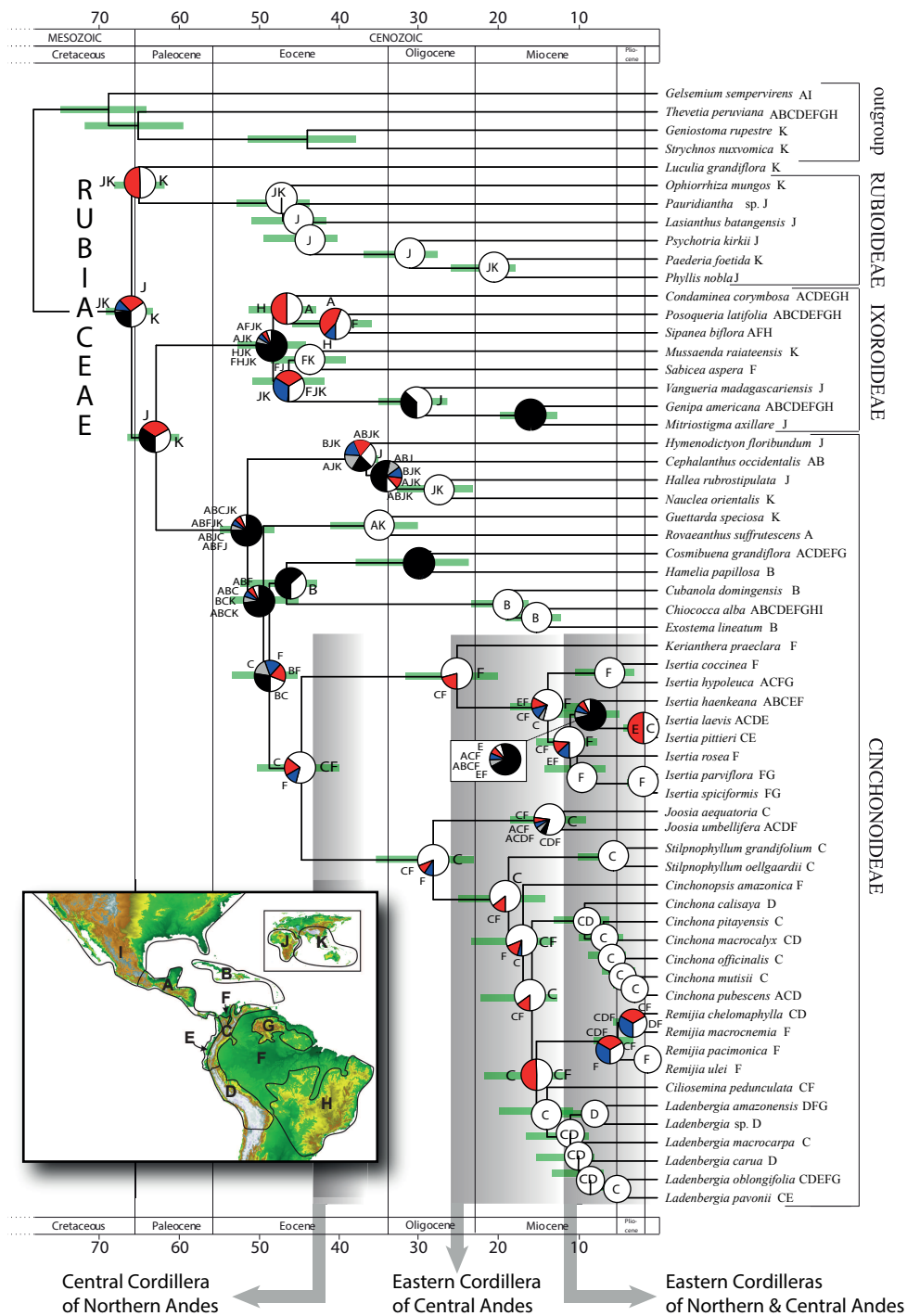
Typical analyses in historical biogeography combine information on the geographic distribution of recent taxa with dated, DNA-based phylogenies to infer the potential distributions of ancestral species (Fig. 5). The combination of geographic and phylogenetic data can link evolution to events in Earth history (either narratively or statistically, e.g. Morlon *et al.*, 2016) or can even inform on these events (Bacon *et al.*, 2015). In particular, area-specific speciation and extinction rates, area connectivity, niche evolution, and competition can be addressed. Most tools to reconstruct geographic ranges through time are confined to a classification of recent taxa into relatively few, discrete areas defined by large-scale geographic barriers, such as islands, continents or biomes (FitzJohn, 2012; Goldberg *et al.*, 2011; Matzke, 2014; Rabosky *et al.*, 2013; Ree and Smith, 2008; Yu *et al.*, 2015, but see Nylinder *et al.*, 2014; Quintero *et al.*, 2015 for alternative approaches that do not rely on discrete areas).

For studies including moderate numbers of taxa and a small geographic resolution (Fig. 5) the area classification of recent species is often based on expert knowledge or regional floras and check-lists. If occurrence records are used, they are often derived from individually curated databases and processed with Geographic Information Systems (GIS) based on a graphical user interface. These approaches are time-consuming, error-prone, difficult to

reproduce and slow, and cannot be scaled to process large amounts of distribution data or big phylogenies, especially if the number of areas analysed increases (Landis *et al.*, 2013). Large amounts of data call for automated and scalable data processing, accounting for data quality. Three practical challenges to the use of large-scale species distribution data in historical biogeography are:

1. **Area delimitation.** Higher spatial resolution is often desirable and is theoretically possible, but finding biologically meaningful delimitations on a regional scale can be difficult and is often arbitrary and impractical. Moreover, barriers are taxon specific.
2. **Area classification.** Classification of species to areas based on expert knowledge or small individually curated data sets is time-consuming, error-prone, difficult to reproduce and not scalable to large data sets.
3. **Data quality.** Geo-referencing errors, taxonomic misidentification and sampling biases are rarely explicitly taken into account. Data cleaning is often performed using graphical user-interface GIS (such as ArcGIS or QGIS), which is subjective, difficult to reproduce and not feasible for large data sets.





**Figure 5** An example for an analysis in historical biogeography: The distribution of extant taxa is classified into large discrete areas (capital letters from the inset map) from which ancestor distributions are reconstructed backwards through time (pie charts) based on a dated molecular phylogeny. Ancestral distributions and their changes can then be interpreted in the light of past geological events, in this case the Andean uplift. Modified from Antonelli *et al.* (2009).



# Objectives

The aim of this thesis is to provide a temporally and spatially large-scale perspective on the evolution of the globally outstanding Neotropical biodiversity. To this end I combine large-scale species occurrence data with fossils and phylogenetic information to evaluate the role of diversification rates, bioregion connectivity and biome shifts. I co-develop new software tools for the use of large amounts of species occurrence data in historical biogeography, with a special focus on data quality and reproducibility. Specifically, this thesis addresses three questions:

1. How can large-scale species distribution data best be used in historical biogeography while ensuring high data quality?
2. What is the role of diversification and dispersal for the differences in species richness among tropical regions?
3. Which role have bioregion connectivity and biome shifts played in the evolution of Neotropical diversity?



# Methods

## Study groups

Evidence from multiple distantly related lineages is necessary to identify and understand general biogeographic patterns and processes and to make generalizations beyond case studies. Therefore this thesis includes studies on multiple organism groups:

**Papers I-IV** used multiple systematic groups to test novel methodology and to exemplify the functioning of the presented tools. These studies did not primarily aim to explain biological patterns and therefore are only mentioned here briefly: **Paper I** used the plant tribe Cinchoneae from the Coffee family (Rubiaceae) to test the reliability of public databases; **Paper II** used angiosperms to exemplify automatic data cleaning of geographic occurrences and fossils; **Paper III** used Neotropical birds to demonstrate the handling of millions of occurrence records and **Paper IV** used global ranges of all amphibians and mammals to demonstrate taxon-specific bioregionalization.

**Papers V-VII** explored biogeographic patterns and processes in different systematic groups. **Paper V** investigated the distribution of rare Neotropical angiosperms, including more than 67,000 species. **Paper VI** crossed the plant-animal boundary and compared the history of bioregion shifts in amphibians, angiosperms, birds, ferns, mammals and squamates (4,450 species in total) across the Neotropics and throughout the Cenozoic. **Paper VII** analysed diversification rates and range shifts of c. 22,600 species of angiosperms globally.

Finally, **Paper VIII** was a case study on biogeography, dispersal and biome shifts in the Malvaceae subfamily Bombacoideae, one of the most iconic clades of the southern hemisphere. The Bombacoideae comprise 176 species of woody plants in 17 genera, including species of high economic importance, such as the African Baobab (*Adansonia digitata*) and South American Kapok (*Ceiba pentandra*) (Fig. 6, Yoon *et al.*, 2016; Zizka *et al.*, 2015). The subfamily is most likely monophyletic (Carvalho-Sobrinho *et al.*, 2016), and distributed throughout the Tropics, with c. 90% of the species endemic to the Neotropics (Fig. 7).

Bombacoideae occur in a variety of different environments and comprise specialized local endemics (*Catostemma* spp.) as well as widespread and abundant generalists (*A. digitata* and *C. pentandra*). Their pan-tropical distribution and the potential high variability in niche width, together with their economic value made the Bombacoideae a suited model to understand drivers of diversification.

## Data sources

This thesis was mostly based on three types of data:

## Geographic occurrence records

The biogeographic analyses were based on over 100 million geographic occurrence records of angiosperms, ferns, mammals, amphibians, and squamates from GBIF ([www.gbif.org](http://www.gbif.org)). Additionally, the analyses included information from various other sources. **Paper I** added a private, expert curated data set of occurrence records obtained from field work and herbaria. **Paper III** used occurrence records of more than 9,000 bird species from the ebird database ([www.ebird.org](http://www.ebird.org)), a public repository specific to birds. **Paper IV** used expert range maps from the IUCN ([www.iucn.org](http://www.iucn.org)) for 6,069 species of amphibians. **Paper V** compared GBIF data to independent data from the Amazon Tree Diversity Network (ATDN), a database of Amazonian vegetation plots (ter Steege *et al.*, 2013). **Paper VIII** added data from other public sources ([www.splink.cria.org.br](http://www.splink.cria.org.br), African Plants Database, 2017; Dauby *et al.*, 2016; Schmidt *et al.*, 2017; The Botanical Information and Ecology Network, 2015) as well as field collections and expert knowledge from the author and collaborators. For **Papers V-VII** records were geographically cleaned and processed using the tools developed in **Papers I-III** and taxonomically scrubbed using the Taxonomic Name Resolution Service (Boyle *et al.*, 2013) (**Papers I-VII**) or manually based on taxonomic literature (**Papers I, VIII**).

## Fossils

Fossils are a key resource for historical biogeography. They allow to age-calibrate phylogenetic trees, to improve and verify ancestral state reconstruction and ancestral range estimations, and to understand the past geographic distribution of bioregions and biomes. Furthermore, fossils can provide an independent line of evidence to estimate diversity and diversification rates through time (Silvestro *et al.*, 2014). **Papers II, VII and VIII** explicitly included fossil information. **Paper II** used c. 20,000 angiosperm fossils from the Paleobiology Database (PBDB) ([www.paleobiodb.org](http://www.paleobiodb.org)) to test fossil-specific cleaning tools. **Paper VI** used a data set of more than 9,500 fossils from the PBDB to test if fossils can be used for large-scale biogeographic inferences across the Tropics. Finally, **Paper VII** used two fossils from the literature to age-calibrate a phylogeny of the Bombacoideae (see below).

## Phylogenetic trees

Phylogenetic trees describe the evolutionary relationships among species, based on genetic information from DNA sequences and/or morphology. **Papers III, IV, VI-VIII** included analyses of different phylogenetic trees. **Paper III** used a phylogeny of birds including virtually all species of birds (Jetz *et al.*, 2012) to demonstrate stochastic mapping implemented in *SpeciesGeoCoder* and **Paper IV** used a phylogeny comprising all extant species of mammals (Faurby and Svenning, 2015) to plot biogeographic regions inferred with *InfoMap Bioregions*. **Papers VI and VII** used a comprehensive plant phylogeny including c. 32,000 plant species (Zanne *et al.*, 2014), to estimate diversification and dispersal rates

through time. **Paper VI** additionally used available phylogenies of amphibians (c. 3,300 species, Pyron and Wiens, 2011, 2013), birds (c. 9,900 species, Jetz *et al.*, 2012), ferns (c. 1,100 species, Lehtonen *et al.*, 2017), mammals (c. 5,000 species, Rolland *et al.*, 2014), and squamates (snakes, lizards, and amphisbaenians or "worm lizards", c. 4,200 species, Pyron *et al.*, 2013) to analyse bioregion shifts in the Neotropics. **Paper VIII** inferred a new phylogeny of the Bombacoideae based on genetic data from Carvalho-Sobrinho *et al.* (2016) to reconstruct geographic range evolution and the effect of biome shifts in the clade.

## Data analysis

### Bioinformatic framework

Data-handling, preparation and statistical analyses for this thesis were performed in R (R Core Team, 2018). R is an open-source programming language with a large user-base in ecology and evolutionary biology. Many of the analyses particularly used the following R-packages: *geosphere* (Hijmans, 2016), *rgdal* (Bivand *et al.*, 2017), *rgeos* (Bivand and Rundel, 2014), *raster* (Hijmans, 2016), *sp* (Bivand *et al.*, 2013; Pebesma and Bivand, 2005), *sf* (Pebesma, 2018) for spatial analyses; *tidyverse* (Wickham, 2017) and *rgbif* (Chamberlain *et al.*, 2015) for data handling and preparation; and *ape* (Paradis *et al.*, 2004), *diversitree* (FitzJohn, 2012), *geiger* (Harmon *et al.*, 2008) and *phytools* (Revell, 2012) for analyses involving phylogenetic trees. Furthermore, parts of *SpeciesGeoCoder* (**Paper III**) are written in python, an open-source language common in bioinformatic applications, and *Infomap Bioregions* (**Paper IV**) is written in JavaScript, an object-oriented programming language common for web-applications.

### Phylogenetic inference

**Paper VIII** used Bayesian Evolutionary Analyses Sampling Trees (BEAST) v. 1.8.1 (Drummond *et al.*, 2012) for phylogenetic inference and divergence time estimation of the Bombacoideae, based on five genetic markers (nuclear + plastid). The analysis used an uncorrelated log-normal molecular clock model, a Yule pure birth speciation model with no starting tree and the Generalised Time-Reversible (GTR) +  $\Gamma$  model of nucleotide substitution with four rate categories. The topology was constraint to the topology presented by Carvalho-Sobrinho *et al.* (2016), in that *Pseudobombax*, as well as the clade comprising *Pseudobombax* and *Pochota fendleri*, *Ceiba* and the clade comprising *Ceiba* and *Neobuchinia* were fixed to be monophyletic. The phylogeny was dated using two fossils. First, a macro-fossil of *Malvaciphyllum macondicus* from the Cerrejón Paleocene forests of Colombia (Wing *et al.*, 2009) which we assigned to the crown node of Bombacoideae + the Malvoideae outgroup species (the *Malvatheca* clade *sensu* Baum *et al.*, 2004) with an exponential prior with mean 0.7 and an offset of 58 million years following (Carvalho *et al.*, 2011). Second, a flower fossil of *Eriotheca prima* from Rodolfo de Lima and Salard-Chebouldaëff (1981)

assigned to the crown node of the clade comprising *Ceiba*, *Eriotheca*, *Spirotheca*, *Pochota*, and *Pseudobombax* with an exponential prior with a mean of 2.3 and an offset age of 41.26 million years.

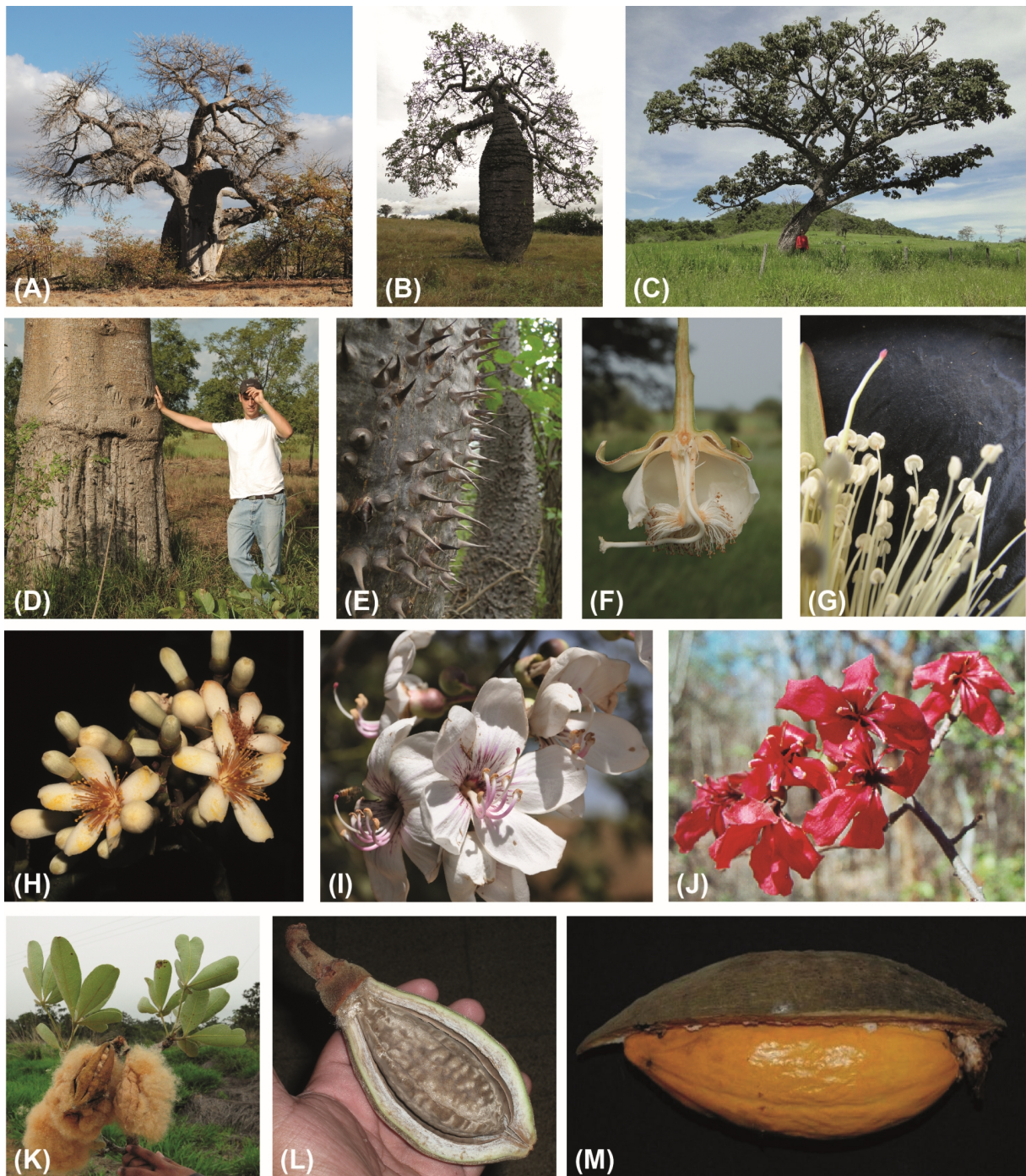
### **Geographic range evolution**

**Papers V** and **VIII** used the Dispersal-Extinction-Cladogenesis (DEC) model (Ree and Smith, 2008) to estimate the geographic range evolution of Neotropical clades of amphibians, angiosperms, birds, ferns, mammals, and squamates, and the Bombacoideae, respectively. The DEC model estimated geographic range evolution based on a phylogenetic tree and discrete present-day distribution ranges. Range evolution was reconstructed along the phylogenetic branches based on a transition rate matrix defining instantaneous transition rates among discrete range states (i.e. defined geographic areas). The DEC model is implemented in R as part of the BioGeoBEARS package (Matzke, 2014).

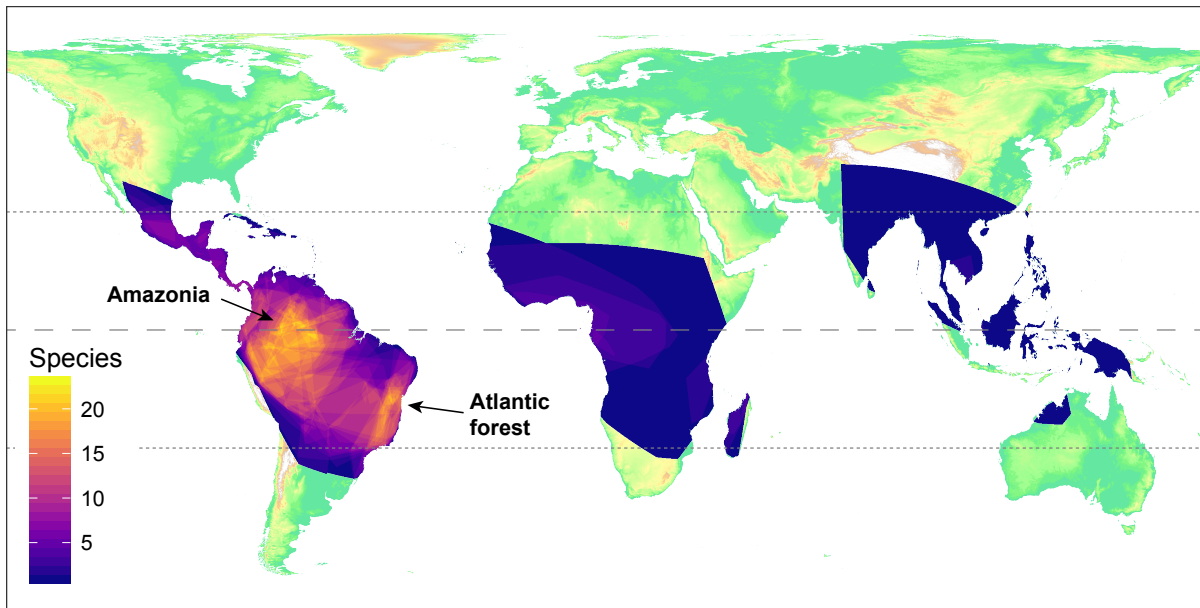
### **Diversification rates**

Several methods exist to estimate the rates at which new species come into existence (speciation) and go extinct (extinction) from phylogenetic trees. These rates, in particular their sum (diversification rate = speciation – extinction) can be compared among different clades, areas and trait states, to infer their impact on the standing diversity of any organism group. This thesis used a group of State-specific-Speciation-and-Extinction (SSE) models to estimate speciation and extinction rates for specific areas or biomes. **Paper VII** used the Multiple state Speciation and Extinction (MuSSE) (FitzJohn, 2012) model to estimate speciation and extinction rates in tropical vs. non-tropical as well as Neotropical vs. Afrotropical vs. tropical Asian angiosperms. **Paper VIII** used the Binary state Speciation and Extinction (BiSSE) (Maddison *et al.*, 2007) and Geographic state Speciation and Extinction (GeoSSE) (Goldberg *et al.*, 2011) models to estimate the impact of biome shifts on the diversification of Bombacoideae. BiSSE is a parametric model to calculate the probability that a group of taxa has evolved as observed given a trait effect. The model specifies two extinction and two speciation rates, one for each trait state, and rate of transitions between the states, thus it can be used to compare rates between trait states. The MuSSE and GeoSSE models are extensions of this model to allow for multiple trait states and widespread ancestors, respectively.





**Figure 6** Examples of Bombacoideae species. **(A)-(E)** habit, **(F)-(J)** flowers, **(K)-(M)** fruits. **(A)** *Adansonia digitata*, **(B)** *Cavanillesia umbellata*, **(C)** *Pseudobombax tomentosum*, **(D)** the succulent stem of *A. digitata* with the author as scale, **(E)** the prickled stem of *Spirotheca elegans*, **(F)** *A. digitata* flower, **(G)** monotheccate anthers of *Pochota fendleri*, **(H)** *Eriotheca* sp., **(I)** *Ceiba glaziovii*, **(J)** *Ceiba rubriflora*, **(K)** *Eriotheca pubescens* with kapok after fruit dehiscence, **(L)** *Pseudobombax tomentosum* with kapok before fruit dehiscence, **(M)** *Catostemma* sp. with fleshy endocarp. Photos by: A. Zizka **(A)**, M. Machado **(B, C, J, L)**, G. Zizka **(D, F)**, J. Carvalho-Sobrinho **(E, K)**, P. Kaminski **(G)**, L. Queiroz **(H)**, R. Machado **(I)**, and C. Zartman **(M)**.



**Figure 7** The global distribution of Bombacoideae. The group is widespread across all three tropical realms (Fig. 1), but the Neotropics are most diverse, in particular Amazonia and the Atlantic forest in Eastern Brazil. Warmer colours indicate more species, the lines show the equator and the Tropics of Cancer (23.4° N) and Capricorn (23.4° S).

# Results and Discussion

## Paper I

Species occurrence information from public databases, such as GBIF, have become widely used in biogeography, but taxonomic errors and geographical uncertainty put into question to what extent such data can be used to unveil patterns of biodiversity and distribution. This paper compares the precision and accuracy of continent-scale diversity patterns inferred based on species occurrence data from GBIF with those inferred from highly precise, but time-intensive manually compiled data, in the Rubiaceae clade Cinchoneae.

The two data sets overlap in the major diversity centres of the Cinchoneae, however the data from GBIF identifies additional false centres of high species richness detached from the core distribution of the clade. These erroneous records affect the inferred diversity patterns across different scales (grid, ecoregion, biome) and can lead to an overestimation of species richness in certain areas (false positives). Interestingly, the majority of these erroneous coordinates represent the geographical midpoints of countries and their political subdivisions. These coordinates supposedly have been geo-referenced long after the original specimens had been collected based on vague locality descriptions (e.g. "Venezuela"). The results demonstrate that the estimation of species distribution still requires occurrence data of good quality. In practice, this means ideally applying substantial amounts of taxonomic knowledge, time and funding on field work and verifying subsets of public databases. However, when this is not a viable option, automatically removing uncertain data together with automatic cleaning procedures can suffice to reveal general diversity patterns and identify main centres of diversity.

## Paper II

Species occurrence records from public databases have become an indispensable resource in ecological, biogeographic and palaeontological research, but issues with data quality can diminish the usefulness of these data. This paper presents *CoordinateCleaner*, a novel, open-source, user-friendly tool to automatically identify records with potentially problematic geographic information, based on errors common to biological and fossil collection databases (see **Paper I**). *CoordinateCleaner* is implemented in R (R Core Team, 2018) and, together with extensive documentation, is available through the major public repository for this language (<https://cran.r-project.org>). The ongoing development of *CoordinateCleaner* can be tracked via GitHub ([www.github.com/azizka](http://www.github.com/azizka)), an open-source software development platform. Detailed documentation including tutorials for data cleaning is also available at GitHub (<https://github.com/azizka/CoordinateCleaner/wiki>). Figure 8A shows the application of *CoordinateCleaner* in an exemplary work-flow for historical biogeography.

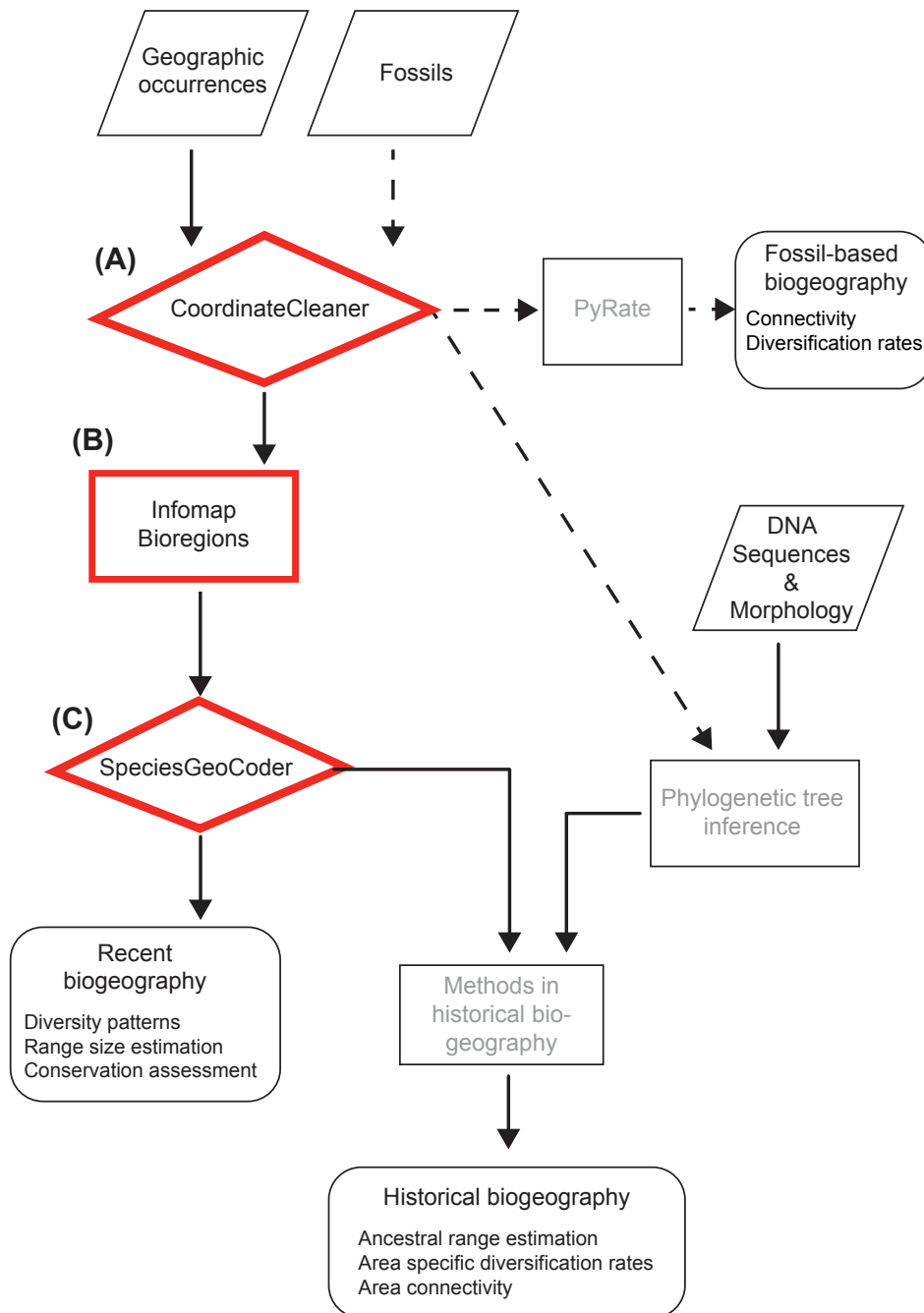
Empirical analyses of all angiosperm records available from GBIF and PBDB, the most commonly used providers for distribution information in ecology and biogeography for recent and fossil data, respectively, suggest that 3.6% (> 3.3 million) of the records available from the GBIF and 8% (1,521) of all angiosperm fossil records available from PBDB are potentially problematic. Furthermore, conversion errors and rasterized collection schemes might be relatively common in GBIF (potentially 18.5% of all contributing data sets).

### **Paper III**

The availability of species point-occurrence records has increased exponentially in the last decade, but there is a lack of bioinformatic tools to use these data in historical biogeography, where often discrete areas are needed. This paper presents *SpeciesGeoCoder*, an open-source tool to use large amounts of species occurrence records in biogeography, facilitating point-to-polygon classification, visualization of diversity patterns and automated conservation assessments (Fig. 8). *SpeciesGeoCoder* is implemented in python and R and together with extensive documentation is available through the major public repository for R ([www.cran.r-project.org](http://www.cran.r-project.org)). The ongoing development of *SpeciesGeoCoder* can be tracked via GitHub ([www.github.com/azizka/speciesgeocodeR](http://www.github.com/azizka/speciesgeocodeR) and [www.github.com/mtop/speciesgeocoder](http://www.github.com/mtop/speciesgeocoder)). Detailed documentation including tutorials is also available at GitHub ([www.github.com/azizka/speciesgeocoder/wiki](http://www.github.com/azizka/speciesgeocoder/wiki)). An empirical example based on 200 million occurrence records suggests that lowland species of birds crossed the Isthmus of Panama about twice as frequently as montane species with a marked increase in the number of dispersals during the last 10 million years. Figure 8C shows the application of *SpeciesGeoCoder* in an example work-flow for historical biogeography.

### **Paper IV**

Biogeographic regions defined by species composition and turnover are a core concept in biogeography, yet few methods for data-driven delimitation of such regions exist, and existing methods often are not accessible in a user-friendly way. This paper presents *Infomap Bioregions*, an extremely fast and user-friendly software tool to identify taxon-specific biogeographic areas from species occurrence information based on a recently developed method in Network and Information theory. *Infomap Bioregions* is implemented in JavaScript and available via a web-based graphical-user-interface at <http://bioregions.mapequation.org/>. *Infomap Bioregions* handles millions of occurrence records as well as global range maps within minutes, accounts for varying sampling intensity through adaptive resolution (i.e. resolution is higher where more data is available) and provides concise summary output for each bioregion, including indicator species. Empirical examples using global distributions of mammals and amphibians are largely concordant with an expert-based bioregionalization for both groups.



**Figure 8** Exemplary work-flow to use large-scale species occurrence data in historical biogeography. Tools presented in this thesis work are marked in red. **(A)** *CoordinateCleaner* to clean geo-referencing errors common in collection data, **(B)** *Infomap bioregions* to infer taxon specific bioregions, and **(C)** *SpeciesGeoCoder* to assign records to discrete areas and visualize species richness. Solid lines: recent occurrences and DNA, dashed lines: fossils. Light grey are tools and methods not further developed in this thesis, but used for data analysis.

## **Paper V**

Rare species are a critical component of the Neotropical diversity, but their spatial distribution remains poorly understood. This paper uses GBIF data to identify putatively rare angiosperm species within the Neotropics and the Amazonian rainforest, and to analyse their spatial distribution. The results indicate that rare species are relatively homogeneously distributed throughout most parts of the lowland Neotropics and lowland Amazonia, and the fraction of rare species collections is low throughout these areas, but significantly higher in highlands. Collections of rare species are mainly clustered in the montane areas of Central America, the northern Andes, the Guiana shield and the Atlantic forest in eastern Brazil; consistent with expected and known centres of endemism. The relatively lower and constant fraction of rare species collections in the lowlands suggests that factors influencing rarity might operate relatively uniformly throughout these areas, but might be different in montane regions, possibly related to higher topographic and habitat heterogeneity.

For 20% of those species with only two occurrence records in GBIF, these records are more than 200 km apart, and for 5% even more than 1,700 km, suggesting that a considerable proportion of rare plant species have surprisingly large distribution ranges. There is a common pattern of disjunct species distribution within the Andes, the Atlantic rainforest in eastern Brazil and between the Atlantic rainforest and Amazonia, but no clear pattern within other lowland areas. In lowlands, the second record of many rare species may be found virtually anywhere (Fig. 9A), urging the need for intensive and broad biological sampling. The observed common disjunctions within and among mountain ranges as well as between Amazonia and the coastal regions of eastern Brazil confirm known taxonomic affinities among these regions. The results are supported by nearly independent data from vegetation plots.

In summary, the results of this paper confirm long-held ideas in Neotropical ecology: that there are centres of rare plant species within the Neotropics, in particular in the montane regions. The findings shed further light on this issue by showing that the fraction of rare species is relatively constant in large parts of the lowland Neotropics and Amazonia, and that some species, despite having very few collections, have surprisingly large distribution ranges. Moreover, they show that disjunct distributions of rare species are mostly linked within mountain ranges and in many cases largely unpredictable in lowland areas.

## **Paper VI**

The Neotropics comprise many different biomes and habitats, and it is unclear how these are connected on evolutionary time-scale and what role biotic interchange among these habitats played for the evolution of the standing diversity in the region. This paper infers the timing and origin of the living biota in all major Neotropical biomes by performing a cross-

taxonomic biogeographic analysis based on 4,450 species from six major clades across the tree of life (angiosperms, ferns, birds, frogs, mammals, and squamates).

The results show that biotic interchange is common among Neotropical regions on an evolutionary time-scale, with a total of more than 4,500 dispersal events across all taxa and regions. All regions have served as source and recipient of lineages, and there was generally high congruence in the directionality of dispersal events across taxa. For instance, all taxa showed a substantial interchange between Amazonia and Mesoamerica, the Atlantic Forests, the Cerrado and Chaco, and the Andean Grasslands. Interchange with the West Indies and the Dry Western South America was low for all groups. Rates of dispersal seem to have fluctuated through time rather than being constant, with a particular increase in the last 10 million years. These results contrast the view that bioregion and biome shifts over evolutionary time are rare events, and imply that even very dissimilar regions—in terms of climatic and environmental variables and inherent biota—do not evolve in isolation, but are biologically interconnected over evolutionary time-scales. For instance the interchange of lineages between the rainforests of lowland Amazonia and the Andean Grasslands shows that although these shifts require substantial eco-physiological adaptations, they have played a substantial role in the assembly of alpine ecosystems.

The connectivity patterns among regions across all studied taxonomic groups indicate Amazonia as the primary source of diversity (Fig. 9B), supplying over 2,800 lineages to other regions (more than four times as many as the second most important source region). Notably, Amazonia provided more species to a landmass that has only been minimally connected to it (Mesoamerica), rather than to the surrounding dry areas in northern and central-east South America, with which Amazonia shares long borders. The results suggest Amazonia as the primary source of Neotropical biodiversity: not only did it generate enormous *in situ* diversity, it also supplied lineages to all other Neotropical regions, across all studied taxonomic groups, and throughout the Cenozoic.

## **Paper VII**

The processes underlying the latitudinal diversity gradient and the difference in diversity among tropical regions remain controversial. This paper estimates area-specific diversification rates, as well as the timing and direction of range shifts of extant angiosperms between tropical and non-tropical zones, as well as into and out-of the three major tropical realms of the world.

There are no significant differences between the speciation and extinction of tropical and non-tropical angiosperms, suggesting that, at least in plants, the latitudinal biodiversity gradient might primarily derive from other factors than differential rates of diversification. In contrast, plant lineages in the Neotropics show significantly higher speciation and extinction

rates, than those in the Afrotropics or tropical Asia, on average 2-2.5 times higher (Fig. 9C-D). This suggests an exceedingly rapid evolutionary turnover, i.e., Neotropical species being formed and replaced by one another at unparalleled rates, also reflecting that South American plant diversity is characterized by a relatively large number of recent, species-rich radiations. The causes underlying these differences remain elusive, but might be associated with the substantial landscape dynamics that have affected northern South America since the Miocene, among other continent-specific differences such as biome sizes, niche space, and climatic history.

The number of range shifts out of tropical regions was higher than the number of shifts into these areas in the first half of the Cenozoic, but reached equilibrium around 30 Million years ago (Ma). Among the three tropical realms the Neotropics stand out from other continents by having "pumped out" more species than it received through most of the last 66 million years. While tropical Africa and tropical Asia showed similar mean rates of immigration and emigration through time, range shifts out of tropical America were consistently more frequent than those entering it throughout most of the Cenozoic. Thus, the results suggest that the Neotropics might have functioned as a "species pump" for the rest of the world.

## **Paper VIII**

Some evidence suggests that lineages can shift between biogeographic regions and biomes more often than previously assumed on evolutionary time-scales. However only little is known on the effect of these shifts on speciation and extinction rates of these lineages. This paper is a case study elucidating the biogeographic history of the Bombacoideae and testing the effect of biome shifts on the evolution of plants.

The phylogenetic dating estimates the age of the group at around 53.5 - 59.3 Ma, and confirms the possibility of at least two recent radiations in *Ceiba* and *Eriotheca*. The bioregionalization identifies eight biogeographic regions in the group, most of them in the Americas. Ancestral range estimation based on these dates and bioregions suggests numerous inter-continental dispersal events: potentially ancient dispersals from South America to Africa, and subsequent, more recent dispersals from Africa to Asia. The relative young age of the group and the age of the inferred dispersal events contradict the hypothesis on the distribution of Bombacoideae as a relict of Gondwanan breakup.

Biome shifts seem to be common in Bombacoideae. The ancestral state reconstruction suggests numerous shifts between biomes across the phylogeny (Fig. 9E), the majority of shifts were from rainforest into seasonally dry biomes in the last 20 million years. The direction and timing of these shifts are remarkably congruent with the hypothesis of a raise to ecological importance of savannas and woodlands during this period (Edwards and Smith, 2010). Furthermore, the results reject a significant relation between seasonally dry

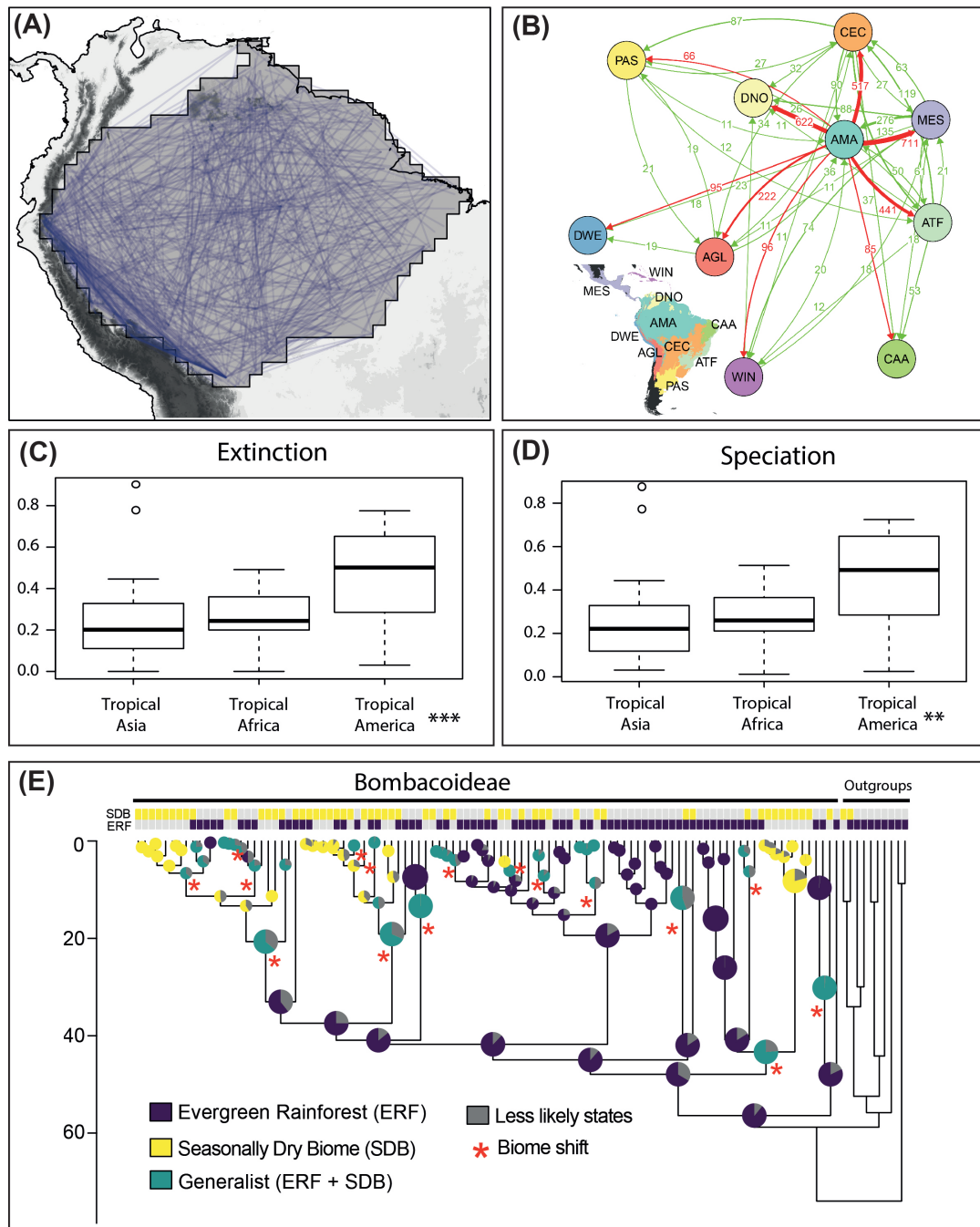


biomes and increased diversification rates, suggesting other factors than biome shifts as driver of diversification, at least in Bombacoideae.

### **On the importance of case studies and field work**

Ultimately, a detailed understanding of each plant species, population and individual from a genetic, physiological and evolutionary perspective would be desirable. Until this is feasible, alternative approaches for understanding the biosphere have to be employed. As outlined in the introduction and shown with the results, big data from collections have merits and potentials for understanding large-scale patterns and processes in biogeography. Building on hundreds of years of collection effort allows an unprecedented level of geographic and taxonomic generalization and hopefully will continue to generate new insights on key processes of the evolution of life. This is particularly promising in the Tropics, where data generation is difficult. The results of this thesis give some examples on how large-scale analyses can add pieces to the enormous task of understanding how tropical biodiversity evolves. The resulting large-scale understanding of tropical biodiversity is urgently needed and timely, as the next decade may decide on the success of conserving this diversity for future generations (e.g. Rockström *et al.*, 2009; Steffen *et al.*, 2015).

However, there are limitations of large-scale approaches in general and data from collections in particular. Even leaving obvious challenges such as low data quality, sampling biases and the vastly differing collection methodology aside, large-scale approaches should be a complement to, and ideally inform on, data collection and in-depth studies on smaller scales (taxonomically and geographically). Large scale cross taxonomic analyses in biogeography combine data and results from many different individuals, populations, species and evolutionary lineages, under the assumption that a common directional signal will emerge above random noise. By definition, biological characteristics of species or populations can hardly be accommodated. Thus, while these analyses are essential to unveil general patterns and the underlying processes they fail to inform on the specific ecological and physiological mechanisms. In contrast, studies focused on specific taxonomic groups can integrate ecology, phylogeny, and physiology to understand processes and mechanisms. Therefore, taxonomically and geographically focussed studies, and particularly data collection in the field, despite being time and resource consuming, remain essential and were also performed on several occasions during this PhD thesis (in French Guiana, South Africa, Panamá and Chile).



**Figure 9** Main results on the evolution of tropical diversity. **(A)** The connectivity between two collections of rare species in Amazonia. The hue of the connecting line indicates the amount of shared rare species between sites. **(B)** The total number of evolutionary range shifts among Neotropical bioregions for angiosperms, amphibians, birds, ferns, mammals, and squamates. The colours represent different bioregions shown on the inset map. **(C, D)** Extinction and speciation rate of 17 orders of angiosperms in the tropical realms. Rates are significantly higher in the Neotropics, indicating higher species turnover. **(E)** The evolution of biomes in the Bombacoideae. Changes in colour indicate a biome shift. The pie charts visualize the uncertainty in the ancestral state estimation.

# Conclusions

Taxonomic and geographic verification of species distribution data from public repositories, such as GBIF, remains essential for biogeographic analyses. In cases where a detailed verification is not possible, automatically removing uncertain data may be sufficient to represent general diversity patterns. Such an automated cleaning procedure dealing with common errors in collection data, as well as functions to use these data to identify taxon-specific bioregions and to use them in historical biogeography are implemented in the *Coordinate-Cleaner*, *SpeciesGeoCoder* and *Infomap Bioregions* presented in this thesis (Fig. 8).

The globally outstanding diversity of the Neotropics seems driven by diversification and dispersal (Fig. 9). The results of this thesis suggest that the species richness of angiosperms found today in the Neotropics as compared to tropical Africa and tropical Asia is associated with significantly higher speciation and extinction rates in the Neotropics—and thereby higher species turnover and shorter average species longevity. Extinction probability of a species, and therefore turnover rates of evolutionary lineages, are linked to species' population- and range-size. For the Neotropics there seem to be centres of rare and range restricted species, in particular in the Andes, stressing the importance of this mountain chain for Neotropical angiosperms diversity.

A substantial fraction of the rapidly evolving Neotropical diversity has been exported to tropical Africa and Asia via emigration, throughout the Cenozoic. Linage migration and the connectivity among different regions within the Neotropics was relatively high on evolutionary time-scales for plants and animals. In particular, biome shifts—the shift of evolutionary lineages among broad environmental settings—were prevalent on evolutionary time, and did, as in the case of the Bombacoideae, increase through time. Amazonia emerged as a central source of diversity, donating an unexpectedly high number of lineages to the surrounding regions.

In summary, this thesis demonstrates how big data driven analyses can contribute to understand the evolution of the outstanding biological diversity observed today in (Neo-)tropical ecosystems. The presented tools facilitate the use of large-scale species occurrence data in biogeography and the results points to the evolutionary connectivity among different bioregions and biomes as a potentially important, but under-appreciated, factor underlying the evolution of the globally outstanding biodiversity of the Neotropics.



## Paper contributions

**Paper I:** *Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases?*—Alexander Zizka (AZ) participated in the study design, compiled part of the data, analysed the data together with the co-authors and contributed to writing the manuscript.

**Paper II:** *CoordinateCleaner: new tools for the automated cleaning of geographic occurrences from biological collections.*—AZ conceived of this study, designed the algorithms with contributions of the co-authors, implemented the tool, analysed the data; and wrote the manuscript with contributions from the co-authors.

**Paper III:** *SpeciesGeoCoder: Fast categorisation of species occurrences for analyses of biodiversity, biogeography, ecology and evolution.*—AZ participated in the study design, implemented the R-Version of the software and contributed to the manuscript.

**PAPER IV:** *Infomap Bioregions: Interactive mapping of biogeographical regions from species distributions.*—AZ participated in the study design, provided the empirical data and contributed to writing the manuscript.

**PAPER V:** *Finding needles in the haystack: Where to look for rare species in the American Tropics.*—AZ designed the study with contributions from the co-authors, analysed the data and wrote the manuscript with contributions from the co-authors.

**PAPER VI:** *Amazonia is the primary source of Neotropical biodiversity.*—AZ participated in the study design, analysed the data with contribution from all authors and contributed to writing the manuscript.

**PAPER VII:** *An engine for global plant diversity: highest evolutionary turnover and emigration in the American Tropics.*—AZ participated in the study design, analysed the data with contributions from the co-authors, and contributed to writing the manuscript.

**PAPER VIII:** *Biome shifts are common, directional, and increase through time in Bombacoideae (Malvaceae).*—AZ conceived of this study together with the co-authors, AZ analysed the data and wrote the manuscript with contributions from all authors.



## References

- African Plants Database (2017) *Conservatoire et Jardin botaniques de la Ville de Genève South African National Biodiversity Institute Conservatoire et Jardin botaniques de la Ville de Genève and South African National Biodiversity Institute Pretoria*. [www.ville-ge.ch/musinfo/bd/cjb/afri](http://www.ville-ge.ch/musinfo/bd/cjb/afri).
- Andrews P & Bamford M (2008) Past and present vegetation ecology of Laetoli, Tanzania. *J Hum Evol* **54**(1): 78–98.
- Antonelli A & Sanmartín I (2011) Why are there so many plant species in the Neotropics? *Taxon* **60**: 403–414.
- Antonelli A, Nylander JA, Persson C & Sanmartín I (2009) Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proc Natl Acad Sci USA* **106**(24): 9749–9754.
- Armstrong KE, Stone GN, Nicholls JA, Valderrama Escallón E, Anderberg AA, Smedmark J, Gautier L, Naciri Y, Milne R & Richardson JE (2014) Patterns of diversification amongst tropical regions compared: A case study in Sapotaceae. *Front Genet* **5**: 1–13.
- Bacon CD, Silvestro D, Jaramillo C, Smith BT & Chakrabarty P (2015) Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proc Natl Acad Sci USA* **112**(24): 6110–6115.
- Baker WJ & Couvreur TL (2013) Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical Biogeography. *J Biogeogr* **40**(2): 286–298.
- Banin L *et al.* (2014) Tropical forest wood production: A cross-continental comparison. *J Ecol* **102**(4): 1025–1037.
- Bardon L, Chamagne J, Dexter KG, Sothers CA, Prance GT & Chave J (2013) Origin and evolution of Chrysobalanaceae: insights into the evolution of plants in the Neotropics. *Bot J Linn Soc* **171**(1): 19–37.
- Barthlott W, Mutke J, Rafiqpoor D, Kier G & Kreft H (2005) Global centers of vascular plant diversity. *Nov Acta Leopoldina* **92**(342): 61–83.
- Baum DA, Smith SDW, Yen A, Alverson WS, Nyffeler R, Whitlock BA & Oldham RL (2004) Phylogenetic relationships of Malvatheca (Bombacoideae and Malvoideae; Malvaceae sensu lato) as inferred from plastid DNA sequences. *Am J Bot* **91**(11): 1863–1871.
- Bebber DP, Carine MA, Wood JRI, Wortley AH, Harris DJ, Prance GT, Davidse G, Paige J, Pennington TD, Robson NKB & Scotland RW (2010) Herbaria are a major frontier for species discovery. *Proc Natl Acad Sci USA* **107**(51): 22169–22171.
- Bivand R & Rundel C (2014) *Rgeos: Interface to Geometry Engine - Open Source (GEOS)*.

- Bivand R, Keitt T & Rowlingson B (2017) *rgdal: Bindings for the Geospatial Data Abstraction Library*.
- Bivand RS, Pebesma E & Gomez-Rubio V (2013) *Applied spatial data analysis with R, Second edition*. Springer.
- Boakes EH, McGowan PJK, Fuller RA, Chang-Qing D, Clark NE, Connor KO & Mace GM (2010) Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biol* **8**(6): e1000385.
- Bonnefille R (2010) Cenozoic vegetation, climate changes and hominid evolution in tropical Africa. *Glob Planet Change* **72**(4): 390–411.
- Bouimetarhan I, Dupont L, Kuhlmann H, Pätzold J, Prange M, Schefuß E & Zonneveld K (2015) Northern Hemisphere control of deglacial vegetation changes in the Rufiji uplands (Tanzania). *Clim Past* **11**(5): 751–764.
- Boyle B, Hopkins N, Lu Z, Raygoza Garay JA, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, McKay SJ, Lowry S, Freeland C, Peet RK & Enquist BJ (2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics* **14**(1): 16pp.
- Buitenwerf R & Higgins SI (2016) Convergence among global biogeographical realms in the physiological niche of evergreen and deciduous vegetation. *Glob Ecol Biogeogr* **25**: 704–715.
- Bush MB & Oliveira PE de (2006) The rise and fall of the Refugial Hypothesis of Amazonian speciation: A paleoecological perspective. *Biota Neotrop* **6**(1): 1–20.
- Butchart SHM *et al.* (2010) Global biodiversity: Indicators of recent declines. *Science* **328**: 1164–1168.
- Cáceres M de *et al.* (2012) The variation of tree beta diversity across a global network of forest plots. *Glob Ecol Biogeogr* **21**(12): 1191–1202.
- Carnaval AC & Moritz C (2008) Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *J Biogeogr* **35**(7): 1187–1201.
- Carvalho MR, Herrera FA, Jaramillo CA, Wing SL & Callejas R (2011) Paleocene Malvaceae from northern South America and their biogeographical implications. *Am J Bot* **98**(8): 1337–1355.
- Carvalho-Sobrinho JG, Alverson WS, Alcantara S, Queiroz LP, Mota AC & Baum DA (2016) Molecular phylogenetics and evolution revisiting the phylogeny of Bombacoideae (Malvaceae): novel relationships, morphologically cohesive clades, and a new tribal classification based on multilocus phylogenetic analyses. *Mol Phylogenet Evol* **101**: 56–74.



- Ceballos G & Ehrlich PR (2006) Global mammal distributions, biodiversity hotspots, and conservation. *Proc Natl Acad Sci USA* **103**(51): 19374–19379.
- Chamberlain S, Ram K, Barve V & Mcglinn D (2015) Rgbif: Interface to the Global Biodiversity Information Facility 'API'.
- Chomicki G & Renner SS (2015) Phylogenetics and molecular clocks reveal the repeated evolution of ant-plants after the late Miocene in Africa and the early Miocene in Australasia and the Neotropics. *New Phytol* **207**(2): 411–424.
- Costello MJ, May RM & Stork NE (2013) Can we name Earth's species before they go extinct? *Science* **339**: 413–416.
- Couvreur TLP (2015) Odd man out: why are there fewer plant species in African rain forests? *Plant Syst Evol* **301**(5): 1299–1313.
- Cowling R & Rundel P (1996) Plant diversity in Mediterranean-climate regions. *Trends Ecol Evol* **11**: 362–366.
- Crisp MD & Cook LG (2012) Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? *New Phytol* **196**(3): 681–694.
- Crisp MD, Arroyo MTK, Cook LG, Gandolfo MA, Jordan GJ, McGlone MS, Weston PH, Westoby M, Wilf P & Linder HP (2009) Phylogenetic biome conservatism on a global scale. *Nature* **458**(7239): 754–756.
- Daru BH, Park DS, Primack RB, Willis CG, Barrington DS, Whitfield TJ, Seidler TG, Sweeney PW, Foster DR, Ellison AM & Davis CC (2018) Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytol* **217**(2): 939–955.
- Dauby G *et al.* (2016) RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys* **18**: 1–18.
- Dexter KG, Smart B, Baldauf C, Baker TR, Balinga MPB, Brienen RJW & Fauset S (2015) Floristics and biogeography of vegetation in seasonally dry tropical regions. *Int For Rev* **17**: 9–32.
- Dickinson JL, Zuckerberg B & Bonter DN (2010) Citizen science as an ecological research tool: challenges and benefits. *Annu Rev Ecol Syst* **41**: 149–172.
- Donoghue MJ (2008) A phylogenetic perspective on the distribution of plant diversity. *Proc Natl Acad Sci USA* **105**: 11549–11555.
- Donoghue MJ & Edwards EJ (2014) Biome shifts and niche evolution in plants. *Trends Ecol Evol* **45**: 547–572.
- Drummond AJ, Suchard Ma, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**(8): 1969–1973.

- Dupont LM, Jahns S, Marret F & Ning S (2000) Vegetation change in equatorial West Africa: time-slices for the last 150 ka. *Palaeogeogr Palaeoclimatol Palaeoecol* **155**(1-2): 95–122.
- Edwards EJ & Donoghue MJ (2013) Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. *J Exp Bot* **64**(13): 4047–4052.
- Edwards EJ & Smith SA (2010) Phylogenetic analyses reveal the shady history of C4 grasses. *Proc Natl Acad Sci USA* **107**(6): 2532–2537.
- Eiserhardt W *et al.* (2018) A roadmap for global synthesis of the plant tree of life. *American Journal of Botany*. *Am J Bot* **in press**.
- Engemann K, Enquist BJ, Sandel B, Boyle B, Jørgensen PM, Morueta-Holme N, Peet RK, Violle C & Svenning J-C (2015) Limited sampling hampers big data estimation of species richness in a tropical biodiversity hotspot. *Ecol Evol* **5**(3): 807–820.
- Erkens RH, Chatrou LW & Couvreur TL (2012) Radiations and key innovations in an early branching angiosperm lineage (Annonaceae; Magnoliales). *Bot J Linn Soc* **169**(1): 117–134.
- Faurby S & Svenning J-C (2015) Molecular Phylogenetics and evolution a species-level phylogeny of all extant and late Quaternary extinct mammals using a novel heuristic-hierarchical Bayesian approach. *Mol Phylogenet Evol* **84**: 14–26.
- Feeley K (2015) Are we filling the data void? An assessment of the amount and extent of plant collection records and census data available for tropical South America. *PlosOne* **10**(4): 17pp.
- Feeley KJ & Silman MR (2011) The data void in modeling current and future distributions of tropical species. *Glob Chang Biol* **17**(1): 626–630.
- Fine PV, Zapata F & Daly DC (2014) Investigating processes of neotropical rain forest tree diversification by examining the evolution and historical biogeography of the proteae (Burseraceae). *Evolution* **68**(7): 1988–2004.
- FitzJohn RG (2012) Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods Ecol Evol* **3**(6): 1084–1092.
- Gasse F (2000) Hydrological changes in the African tropics since the Last Glacial Maximum. *Quat Sci Rev* **19**(1-5): 189–211.
- Gaston KJ & Spicer JI (2004) *Biodiversity: an introduction*. Oxford: Blackwell Science Ltd: 208pp.
- GBIF (2016) *Magnoliophyta*, <https://doi.org/10.15468/dl.om11gi>.

- Gentry AH (1982) Neotropical floristic diversity: Phylogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Ann Missouri Bot Gard* **69**(3): 557–593.
- Gentry AH (1988) Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann Missouri Bot Gard* **75**(1): 1–34.
- Gentry AH (1993) “Diversity and floristic composition of lowland tropical forest in Africa and South America”. *Biol Relationships between Africa South Am.* Ed. by P Goldblatt. New Haven: Yale University Press: 500–548.
- Goldberg EE, Lancaster LT & Ree RH (2011) Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst Biol* **60**(4): 451–65.
- Good R (1953) *The geography of the flowering plants*. 2nd. London: Longmans Green and Co: 516pp.
- Govaerts R (2001) How many species of seed plants are there. *Taxon* **50**: 1085–1090.
- Grenyer R *et al.* (2006) Global distribution and conservation of rare and threatened vertebrates. *Nature* **444**: 93–96.
- Haffer J (2008) Hypotheses to explain the origin of species in Amazonia. *Brazilian J Biol* **68**: 917–947.
- Haffer J (1969) Speciation in Amazonian forest birds. *Science* **165**: 131–137.
- Hamilton AC (1981) The quaternary history of African forests: Its relevance to conservation. *Afr J Ecol* **19**: 1–6.
- Harmon LJ, Weir JT, Brock CD, Glor RE & Challenger W (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**: 129–131.
- Harvey PH & Pagel MD (1991) *The comparative method in evolutionary biology*. Oxford: Oxford University Press: 239pp.
- Higgins MA, Ruokolainen K, Tuomisto H, Llerena N, Cardenas G, Phillips O, Vásquez R & Räsänen M (2011) Geological control of floristic composition in Amazonian forests. *J Biogeogr* **38**: 2136–2149.
- Higgins SI, Buitenwerf R & Moncrieff GR (2016) Defining functional biomes and monitoring their change globally. *Glob Chang Biol* **22**(11): 3583–3593.
- Hijmans R, Garrett K, Huamán Z, Zhang D, Schreuder M & Bonierbale M (2000) Assessing the geographic representativeness of Genebank collections: The case of Bolivian wild potatoes. *Conserv Biol* **14**(6): 1755–1765.
- Hijmans RJ (2016) *raster: Geographic Data Analysis and Modeling*.

- Holt BG, Lessard J-P, Borregaard MK, Fritz Sa, Araújo MB, Dimitrov D, Fabre P-H, Graham CH, Graves GR, Jønsson KA, Nogués-Bravo D, Wang Z, Whittaker RJ, Fjeldså J & Rahbek C (2013) An update of Wallace's zoogeographic regions of the world. *Science* **339**(6115): 74–78.
- Hoorn C *et al.* (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**: 927–931.
- Hopkins MJG (2007) Modelling the known and unknown plant biodiversity of the Amazon basin. *J Biogeogr* **34**: 1400–1411.
- Hudson M (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Resour* **8**: 3–17.
- Hughes CE, Pennington RT & Antonelli A (2013) Neotropical plant evolution: assembling the big picture. *Bot J Linn Soc* **171**(1): 1–18.
- Hutchinson GE (1957) Concluding remarks. *Cold Spring Harb Symp Quant Biol* **22**: 415–427.
- Ivory SJ & Russell J (2016) Climate, herbivory, and fire controls on tropical African forest for the last 60ka. *Quat Sci Rev* **148**: 101–114.
- Ivory SJ, Lézine AM, Vincens A & Cohen AS (2012) Effect of aridity and rainfall seasonality on vegetation in the southern tropics of East Africa during the Pleistocene/Holocene transition. *Quat Res* **77**(1): 77–86.
- Jacobs BF (2004) Palaeobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savannah biomes. *Philos Trans R Soc B* **359**(1450): 1573–83.
- Jaramillo C, Rueda MJ & Mora GG (2006) Cenozoic Plant Diversity in the Neotropics. *Science* **311**(March): 1893–1896.
- Jaramillo C *et al.* (2010) Effects of rapid global warming at the Paleocene-Eocene boundary on Neotropical vegetation. *Science* **311**(May): 518–522.
- Jetz W, Thomas GH, Joy JB, Hartmann K & Mooers AO (2012) The global diversity of birds in space and time. *Nature* **491**(7424): 444–448.
- Jetz W & Fine PV (2012) Global gradients in vertebrate diversity predicted by historical area-productivity dynamics and contemporary environment. *PLoS Biol* **10**(3): e1001292.
- Keppel G, Buckley YM & Possingham HP (2010) Drivers of lowland rain forest community assembly, species diversity and forest structure on islands in the tropical South Pacific. *J Ecol* **98**(1): 87–95.

- Kessler M (2002) The elevational gradient of Andean plant endemism: Varying influences of taxon-specific traits and topography at different taxonomic levels. *J Biogeogr* **29**(9): 1159–1165.
- Kier G, Mutke J, Dinerstein E, Ricketts TH, Küper W, Kreft H & Barthlott W (2005) Global patterns of plant diversity and floristic knowledge. *J Biogeogr* **32**(7): 1107–1116.
- Kirchmair I (2017) “Biogeography of West African gallery forests”. PhD thesis. Goethe University Frankfurt: 199pp.
- Kissling WD, Eiserhardt WL, Baker WJ, Borchsenius F, Couvreur TLP, Balslev H & Svenning J-C (2012) Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide. *Proc Natl Acad Sci USA* **109**(19): 7379–7384.
- Knapp S (2002) Assessing patterns of plant endemism in Neotropical uplands. *Bot Rev* **68**(1): 22–37.
- Koenen EJM, Vos JMD, Atchison GW, Simon MF, Schrire BD, Souza ERD, Queiroz LPD & Hughes CE (2013) Exploring the tempo of species diversification in legumes. *South African J Bot* **89**: 19–30.
- Kreft H & Jetz W (2007) Global patterns and determinants of vascular plant diversity. *Proc Natl Acad Sci USA* **104**(14): 5925–5930.
- Kruckeberg AR & Rabinowitz D (1985) Biological aspects of endemism in higher plants. *Annu Rev Ecol Syst* **16**: 447–479.
- Kursar TA, Dexter KG, Lokvam J, Pennington RT, Richardson JE, Weber MG, Murakami ET, Drake C, McGregor R & Coley PD (2009) The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proc Natl Acad Sci USA* **106**(43): 18073–18078.
- Lagomarsino LP, Condamine FL, Antonelli A, Mulch A & Davis CC (2016) The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytol* **210**: 1430–1442.
- Landis MJ, Matzke NJ, Moore BR & Huelsenbeck JP (2013) Bayesian analysis of biogeography when the number of areas is large. *Syst Biol* **62**(6): 789–804.
- Larsen BB, Miller EC, Rhodes MK & Wiens JJ (2017) Inordinate fondness multiplied and redistributed: the number of species on Earth and the new pie of life. *Quarterly Rev Biol* **92**(3): 229–262.
- Lavin M, Thulin M, Labat J & Pennington RT (2000) Africa, the odd man out: Molecular biogeography of dalbergioid legumes (Fabaceae) suggests otherwise. *Syst Bot* **25**(3): 449–467.
- Lebrun J-P (2001) *Introduction à la flore d’Afrique*. Paris: CIRAD, Ibis Press: 145pp.

- Lehtonen S, Silvestro D, Karger DN, Scotese C, Tuomisto H, Kessler M, Peña C, Wahlberg N & Antonelli A (2017) Environmentally driven extinction and opportunistic origination explain fern diversification patterns. *Sci Rep* **7**: 4831.
- Leite YLR, Costa LP, Loss AC, Rocha RG, Batalha-Filho H, Bastos AC, Quaresma VS, Fagundes V, Paresque R, Passamani M & Pardini R (2016) Neotropical forest expansion during the last glacial period challenges refuge hypothesis. *Proc Natl Acad Sci* **113**(4): 1008–1013.
- Levy SE & Myers RM (2016) Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet* **17**(1): 95–115.
- Lohmann LG, Bell CD, Calió MF & Winkworth RC (2013) Pattern and timing of biogeographical history in the Neotropical tribe Bignonieae (Bignoniaceae). *Bot J Linn Soc* **171**(1): 154–170.
- Lomolino MV, Riddle BR, Whittaker RJ & Brown JH (2010) *Biogeography*. 4th. Sunderland: Sinauer Associates: 764pp.
- Losos JB (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol Lett* **11**(10): 995–1003.
- Luebert F & Weigend M (2014) Phylogenetic insights into Andean plant diversification. *Front Ecol Evol* **2**(June): 1–17.
- Maddison WP, Midford PE & Otto SP (2007) Estimating a binary character's effect on speciation and extinction. *Syst Biol* **56**(5): 701–710.
- Madriñán S, Cortés AJ & Richardson JE (2013) Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Front Genet* **4**(192): 1–7.
- Malhi Y & Wright J (2004) Spatial patterns and recent trends in the climate of tropical rainforest regions. *Philos Trans R Soc B Biol Sci* **359**(1443): 311–329.
- Matzke NJ (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst Biol* **63**(6): 951–970.
- Mayle FE, Beerling DJ, Gosling WD & Bush MB (2004) Responses of Amazonian ecosystems to climatic and atmospheric carbon dioxide changes since the last glacial maximum. *Philos Trans R Soc B Biol Sci* **359**(1443): 499–514.
- Meyer C, Weigelt P & Kreft H (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol Lett* **19**(8): 992–1006.
- Mittelbach GG *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol Lett* **10**(4): 315–331.

- Mittermeier RA, Turner WR, Larsen FW, Brooks TM & Gascon C (2011) “Global biodiversity conservation: The critical role of hotspots”. *Biodivers Hotspots*. Ed. by F Zachos & J Habel. Berlin Heidelberg: Springer. Chap. 1: 22pp.
- Moncrieff GR, Hickler T & Higgins SI (2015) Intercontinental divergence in the climate envelope of major plant biomes. *Glob Ecol Biogeogr* **24**: 324–334.
- Moncrieff GR, Bond WJ & Higgins SI (2016) Revising the biome concept for understanding and predicting global change impacts. *J Biogeogr* **43**: 863–873.
- Morley RJ (2000) *Origin and evolution of tropical rain forests*. Chichester, England: Wiley: 362pp.
- Morlon H, Lewitus E, Condamine FL, Manceau M, Clavel J & Drury J (2016) RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol Evol* **7**: 589–597.
- Morrone JJ (2014) Biogeographical regionalisation of the Neotropical region. *Zootaxa* **3782**(1): 1–110.
- Morrone JJ (2017) *Neotropical Biogeography. Regionalization and Evolution*. Boca Raton, FL: CRC Press: 282pp.
- Myers N, Mittermeier RA, Mittermeier CG, Fonseca GaB da, Kent J, da Fonseca GAB & Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**(6772): 853–858.
- Nelson BBW, Ferreira CAC, da Silva MF & Kawasaki ML (1990) Endemism centres, refugia and botanical collection density in Brazilian Amazonia. *Nature* **345**(6277): 714–716.
- Nylinder S, Lemey P, De Bruyn M, Suchard MA, Pfeil BE, Walsh N & Anderberg AA (2014) On the biogeography of centipeda: A species-tree diffusion approach. *Syst Biol* **63**(2): 178–191.
- Olson DM *et al.* (2001) Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience* **51**(11): 933–938.
- Onstein RE, Jordan GJ, Sauquet H, Weston PH, Bouchenak-Khelladi Y, Carpenter RJ & Linder HP (2016) Evolutionary radiations of Proteaceae are triggered by the interaction between traits and climates in open habitats. *Glob Ecol Biogeogr* **25**(10): 1239–1251.
- Pan AD, Jacobs BF, Dransfield J & Baker WJ (2006) The fossil history of palms (Arecaceae) in Africa and new records from the Late Oligocene (28–27 Mya) of north-western Ethiopia. *Bot J Linn Soc* **151**(1): 69–81.
- Paradis E, Claude J & Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.

- Parmentier I *et al.* (2007) The odd man out? Might climate explain the lower tree  $\alpha$  diversity of African rain forests relative to Amazonian rain forests? *J Biogeogr* **95**(5): 1058–1071.
- Pebesma E (2018) *sf: Simple Features for R*.
- Pebesma E & Bivand R (2005) Classes and methods for spatial data in R. *R News* **5**(2): 21pp.
- Pennington RT, Cronk QCB & Richardson JA (2004) Introduction and synthesis: Plant phylogeny and the origin of major biomes. *Philos Trans R Soc B* **359**(1450): 1455–1464.
- Pennington RT, Lavin M & Oliveira-Filho A (2009) Woody plant diversity, evolution, and ecology in the tropics: Perspectives from seasonally dry tropical forests. *Annu Rev Ecol Evol Syst* **40**(1): 437–457.
- Peterson AT (2011) Ecological niche conservatism: A time-structured review of evidence. *J Biogeogr* **38**: 817–827.
- Pimm SL & Joppa LN (2015) How many plant species are there, where are they and at what rate are they going extinct. *Ann Missouri Bot Gard* **100**: 170–176.
- Pirie MD, Maas PJM, Wilschut R, Melchers-Sharrott H & Chatrou L (2018) Parallel diversifications of *Crematosperma* and *Mosannonna* (Annonaceae), tropical rainforest trees tracking Neogene upheaval of the South American continent. *R Soc Open Sci* **5**: 171561.
- Pyron RA & Wiens JJ (2011) A large-scale phylogeny of Amphibia including over 2,800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Mol Phylogenet Evol* **61**(2): 543–583.
- Pyron RA & Wiens JJ (2013) Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proc R Soc B* **280**: 20131622.
- Pyron RA, Burbrink FT & Wiens JJ (2013) A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol Biol* **13**(93): 53pp.
- Qian H, Jin Y & Ricklefs RE (2017) Patterns of phylogenetic relatedness of angiosperm woody plants across biomes and life-history stages. *J Biogeogr* **44**(6): 1383–1392.
- Quintero I, Keil P, Jetz W & Crawford FW (2015) Historical biogeography using species geographical ranges. *Syst Biol* **64**(6): 1059–1073.
- R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J & Alfaro ME (2013) Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat Commun* **4**: 1–8.



- Raven PH (1976) "Ethics and attitudes". *Conserv Threat Plants*. Ed. by J Simmons. New York and London: Springer: 155–179.
- Raven PH & Axelrod D (1974) Angiosperm biogeography and past continental movements. *Ann Missouri Bot Gard* **61**(3): 539–673.
- Ray N & Adams J (2001) A GIS-based vegetation map of the world at the Last Glacial Maximum (25,000-15,000 BP). *Internet Archaeol* **11**(11): 0–15.
- Ree RH & Smith SA (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst Biol* **57**(1): 4–14.
- Reu B, Proulx R, Bohn K, Dyke JG, Kleidon A, Pavlick R & Schmidtlein S (2011) The role of climate and plant functional trade-offs in shaping global biome and biodiversity patterns. *Glob Ecol Biogeogr* **20**(4): 570–581.
- Revell LJ (2012) phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* **3**(2): 217–223.
- Richards P (1973) "Africa the odd man out". *Trop For Ecosyst Africa South Am*. Ed. by M BJ. Smithsonian Insitute Press: 21–66.
- Richardson J, Pennington RT, Pennington TD & Hollingsworth PM (2001) Rapid diversification of a species-rich genus of Neotropical rain forest trees. *Science* **293**: 2242–2245.
- Ricklefs RE & Renner SS (2012) Global correlations in tropical tree species richness and abundance reject neutrality. *Science* **335**(6067): 464–7.
- Rockström J *et al.* (2009) A safe operating space for humanity. *Nature* **461**(7263): 472–475.
- Rodolfo de Lima M & Salard-Cheboldaef M (1981) Palynologie des bassins de Gandarela et Fonseca. *Bol IG* **12**: 33–54.
- Rolland J, Condamine FL & Jiguet F (2014) Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *PLOS Biol* **12**(1): e1001775.
- Ruete A (2015) Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance maps. *Biodivers Data J* **3**: e5361.
- Särkinen T, Iganci JR, Linares-Palomino R, Simon MF & Prado DE (2011) Forgotten forests - issues and prospects in biome mapping using Seasonally Dry Tropical Forests as a case study. *BMC Ecol* **11**: 27.
- Särkinen TE, Newman MF, Maas PJ, Maas H, Poulsen AD, Harris DJ, Richardson JE, Clark A, Hollingsworth M & Pennington RT (2007) Recent oceanic long-distance dispersal and divergence in the amphiatlantic rain forest genus *Renealmia* L.f. (Zingiberaceae). *Mol Phylogenet Evol* **44**(3): 968–980.

- Sauquet H, Weston PH, Anderson CL, Barker NP, Cantrill DJ, Mast AR & Savolainen V (2009) Contrasted patterns of hyperdiversification in Mediterranean hotspots. *Proc Natl Acad Sci USA* **106**(1): 221–225.
- Scheiner SM, Kosman E, Presley SJ & Willig MR (2017) The components of biodiversity, with a particular focus on phylogenetic information. *Ecol Evol* **7**(16): 6444–6454.
- Schmidt M *et al.* (2017) Diversity, distribution and preliminary conservation status of the flora of Burkina Faso. *Phytotaxa Monogr* **304**(1): 1–215.
- Schultz J (2005) *The ecozones of the world: The ecological division of the ecosphere*. Berlin Heidelberg: Springer: 252pp.
- Silvestro D, Salamin N & Schnitzler J (2014) PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods Ecol Evol* **5**: 1126–1131.
- Simon MF, Grether R, Queiroz LP de, Skema C, Pennington RT, Hughes CE, Queiroz LPD, Skema C, Pennington RT & Hughes CE (2009) Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proc Natl Acad Sci USA* **106**(48): 20359–20364.
- Slik J *et al.* (2015) An estimate of the number of tropical tree species. *Proc Natl Acad Sci USA* **112**(24): 7472–7477.
- Slik J *et al.* (2017) Phylogenetic classification of the world's tropical forests. *Proc Natl Acad Sci USA*: 2–7.
- Somveille M, Manica A, Butchart SHM & Rodrigues ASL (2013) Mapping global diversity patterns for migratory birds. *PLoSone* **8**(8): e70907.
- Spriggs EL, Clement WL, Sweeney PW, Madriñán S, Edwards EJ & Donoghue MJ (2015) Temperate radiations and dying embers of a tropical past: The diversification of *Viburnum*. *New Phytol* **207**(2): 340–354.
- Stager JC, Ryves DB, Chase BM & Pausata FSR (2011) Catastrophic drought in the Afro-Asian monsoon region during Heinrich event 1. *Science* **331**(6022): 1299–1302.
- Steffen W *et al.* (2015) Planetary boundaries: Guiding human development on a changing planet. *Science* **347**: 1259855.
- Stropp J, ter Steege H & Malhi Y (2009) Disentangling regional and local tree diversity in the Amazon. *Ecography (Cop)* **32**(1): 46–54.
- Takhtajan A (1987) *Floristic Regions of the World*. Ed. by A Conquist. Berkley: University of California Press: 522pp.
- ter Steege H *et al.* (2013) Hyperdominance in the Amazonian tree flora. *Science* **342**(6156): 325–334.

- ter Steege H, Haripersaud PP, Bánki OS & Schieving F (2011) A model of botanical collectors' behavior in the field: Never the same species twice. *Am J Bot* **98**(1): 31–37.
- ter Steege H, Vaessen RW, Cardena-Lopez D, Sabatier D, Antonelli A, Mota de Oliveira S, Pitman NC, Møller Jørgensen P & Salomão RP (2016) The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Sci Rep* **6**: 29549.
- Terborgh J, Davenport LC, Niangadouma R, Dimoto E, Mouandza JC, Schultz O & Jaen MR (2016) The African rainforest: Odd man out or megafaunal landscape? African and Amazonian forests compared. *Ecography (Cop)* **39**(2): 187–193.
- The Botanical Information and Ecology Network (2015) *BIEN3*, <http://bien.nceas.ucsb.edu/bien/>.
- Thomas WW (1999) Conservation and monographic research on the flora of Tropical America. *Biodivers Conserv* **8**(8): 1007–1015.
- Tobler M, Honorio E, Janovec J & Reynel C (2007) Implications of collection patterns of botanical specimens on their usefulness for conservation planning: An example of two neotropical plant families (Moraceae and Myristicaceae) in Peru. *Biodivers Conserv* **16**: 659–677.
- Tucker CM *et al.* (2016) A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol Rev* **92**: 698–715.
- Utescher T & Mosbrugger V (2007) Eocene vegetation patterns reconstructed from plant diversity - A global perspective. *Palaeogeogr Palaeoclimatol Palaeoecol* **247**: 243–271.
- Valente LM & Vargas P (2013) Contrasting evolutionary hypotheses between two mediterranean-climate floristic hotspots: The cape of southern Africa and the Mediterranean basin. *J Biogeogr* **40**(11): 2032–2046.
- Valente LM, Reeves G, Schnitzler J, Mason IP, Fay MF, Rebelo TG, Chase MW & Barraclough TG (2009) Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution* **64**: 745–760.
- Vincens A, Garcin Y & Buchet G (2007) Influence of rainfall seasonality on African lowland vegetation during the Late Quaternary: Pollen evidence from Lake Masoko, Tanzania. *J Biogeogr* **34**(7): 1274–1288.
- Wallace A (1876) *The geographic distribution of animals*. Cambridge: Cambridge University Press: 503pp.
- Whitmore T (1998) *An introduction to tropical rain forests*. 2nd. Oxford: Oxford University Press: 296pp.
- Whittaker RH (1962) Classification of natural communities. *Bot Rev* **28**(1): 239pp.

- Whittaker RJ, Willis KJ & Field R (2001) Scale and species richness: Towards a general, theory of species diversity hierarchical. *J Biogeogr* **28**(4): 453–470.
- Wickham H (2017) *tidyverse: Easily install and load the 'Tidyverse'*.
- Wiens JJ (2007) Global patterns of diversification and species richness in amphibians. *Am Nat* **170**: 86–106.
- Wiens JJ & Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends Ecol Evol* **19**(12): 639–643.
- Wiens JJ & Graham CH (2005) Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annu Rev Ecol Evol Syst* **36**: 519–539.
- Wiens JJ, Ackerly DD, Allen AP, Anacker BL, Buckley LB, Damschen EI, Davies TJ, Grytnes J-A, Harrison SP, Hawkins BA, Holt RD, McCain CM & Stephens PR (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol Lett* **13**: 1310–1324.
- Wing SL, Herrera F, Jaramillo CA, Gómez-Navarro C & Labandeira CC (2009) Late Paleocene fossils from the Cerrejón Formation, Colombia, are the earliest record of Neotropical rainforest. *Proc Natl Acad Sci USA* **106**(44): 18627–18632.
- Yoon SY, Kim D-J, Sung YJ, Han S-H, Aggangan NS & Shin S-J (2016) Enhancement of enzymatic hydrolysis of kapok *Ceiba pentandra* (L.) Gaertn seed fibers with potassium hydroxide pretreatment. *Asia Life Sci* **25**(1): 17–29.
- Yu Y, Harris AJ, Blair C & He X (2015) RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. *Mol Phylogenet Evol* **87**: 46–49.
- Zanne AE *et al.* (2014) Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**(7486): 89–92.
- Zizka A, Thiombiano A, Dressler S, Nacoulma BMI, Ouédraogo A, Ouédraogo I, Ouédraogo O, Hahn K, Schmidt M & Zizka G (2015) Traditional plant use in Burkina Faso (West Africa): A national-scale analysis with focus on traditional medicine. *J Ethnobiol Ethnomed* **11**(9): 10pp.

# Abbreviations

- ATDN** Amazon Tree Diversity Network. 18
- BEAST** Bayesian Evolutionary Analyses Sampling Trees. 19
- BiSSE** Binary state Speciation and Extinction. 20
- DEC** Dispersal-Extinction-Cladogenesis. 20
- GBIF** Global Biodiversity Information Facility. 12, 18, 23–25, 32
- GIS** Geographic Information Systems. 13, 14
- GTR** Generalised Time-Reversible. 19
- IUCN** International Union for the Conservation of Nature. 10, 12, 18
- Ma** Million years ago. 28
- MuSSE** Multiple state Speciation and Extinction. 20
- NGS** Next Generation Sequencing. 13
- PBDB** Paleobiology Database. 18, 24
- PNC** Phylogenetic Niche Conservatism. 10
- SSE** State-specific-Speciation-and-Extinction. 20



# Acknowledgements

I am thankful to:

First and foremost my supervisor Alexandre Antonelli for unconditioned support, excellent advice, ample opportunities to learn, the possibility to participate in so many exciting research projects and great field and conference trips together. Alex, thank you for putting your trust in me.

Christine Bacon, my co-supervisor for advice and help at many occasions, and good times during field work in French Guiana and beyond.

Daniele Silvestro and Søren Faurby, for help and advice on countless questions, stimulating discussions at work and in the pub.

All members of the systematics group at BioEnv, in particular Bengt Oxelman, Bernard Pfeil and Claes Persson for advice, instructive Journal Club discussions and an unforgettable excursion to South Africa.

Mari Källersjö for examining this thesis, and helping to keep it on track.

All co-authors, for the productive and stimulating collaborations.

Camila Duarte Ritter for being a good friend and a constant source of joy, inspiration and entertainment.

Somnath and Trirantha for being great friends and neighbours and making Thursdays, the best day of the week.

My friends at the department Josué, Tobi, Harith & Yann for making the time at BioEnv about more than work, lots of fun and shared beers.

Ylva Heed for being so fabulous and the most helpful and generous colleague and friend.

All the current and former colleagues at BioEnv and the team of innebandiers that made the time here even more enjoyable: Sven T., Daniela A., Lisa A., Brigitte M. Andrei H., Erik T., Allison P., Climbié F., Najara A., Maria do Ceo, Gustavo H., Filipe S., Thomas H., Jonna E, Angela C., Ruud S., Daniel E., Patrik C, Dom B., Pavel M., Juan-David C., Beatriz N. & Anna A.

The department of biological and environmental sciences at the University of Gothenburg for providing a great infrastructure and excellently designed third-cycle education system.

The ForBio Research School in Systematics for offering a large number of high-quality PhD courses around Scandinavia for free.

ForBio and the Adlerberska foundation for financial support.

This PhD was funded by a Starting Grant from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013, ERC Grant Agreement n. 331024)